GESPIN 2019

11 - 13 September

Book of Abstracts

6th Gesture and Speech in Interaction Conference

PADERBORN UNIVERSITY
The University for the Information Society

Contents

Foreword

The uniqueness of previous GESPIN meetings arises from examining gestures in tight coordination with speech (including its lexical, syntactical as well as prosodical properties). The 6th edition of the Gesture and Speech in Interaction (GESPIN) was held in Paderborn, Germany. For this meeting, we focused on the heterogeneity of this coordination. Topics and related questions were:

- Development of gesture-speech coordination: Can general principles of development be identified? Are there sensitive periods and developmental stages?
- Individual differences in coordinating speech and gestures: Are there developmental differences beyond infancy/childhood? How do various population groups (elder people, people with autism spectrum disorders, people with cognitive or language impairments) coordinate gesture and speech?
- Benefits of multimodal coordination for learning in individuals and in a variety of settings
- Computational models dealing with heterogenous data and/or generating behavior that differs across, for example, situation or addressees
- Cross-cultural differences on gesture-speech coordination: Is development following a universal path that is culturally shaped? How do cultural groups differ in how they coordinate gestures and speech?
- Heterogeneity across situations and contexts: Do situations differ due to familiarity with the environment or interlocutors? Do communicative genres require specific types of coordination?

The four keynote speakers addressed the focus of this 6th meeting from different perspectives:

Prof. Dr. Seyda Özçalışkan (Georgia State University) who focusses her research on children's earliest linguistic abilities and on the question whether gesture constitutes a robust aspect of the language learning process. In her talk, she pursues the idea of differences or delays in speech becoming first evident in gesture across different learners.

Prof. Dr. Petra Wagner (Bielefeld University) working on the relationship between prosodic expression in speech and gesture, and currently studying the impact of information structure and visibility between interlocutors on the cross-modal link in prosodic expression. By extending her focus to co-speech movements that are not considered as gestures in the traditional understanding, her talk contributes to the concept of embodiment of communication.

Dr. Alexis Heloir's (Université Polytechnique des Hauts de France) research centers around the question of how virtual agents contribute to understanding human behavior. By pointing to his transdisciplinary collaborations, he will address leading design principles of an agent creation and control framework called YALLAH.

Prof. Dr. Pilar Prieto (ICREA-Universitat Pompeu Fabra, Barcelona, Catalunya) with a strong research interest on the benefits of gesture in the second language classroom, especially with respect to embodied rhythmic movements that might have an effect on L2 pronunciation. In her talk, she proposes a multimodal approach in general and embodied prosodic trainings in specific that are essential to understanding L2 speech learning.

We were pleased that the talks and posters at this year's GESPIN meeting, just as at previous meetings, were presented by international contributors from a variety of countries, and based on them, we were able to offer different formats (workshops, data session) for sharing our research, experiences and ideas.

We also thank all the reviewers who engaged in the selection process this year, and we hope that in the future, GESPIN will remain a community of scholars devoted to the coordination of gestures and speech. Finally, we would like to acknowledge the work of the people at Paderborn University who helped a lot with organizing this event: Sabine Hendriks, and the student volunteers, Camilla Crawshaw, Lisa Enns, Monique Koke, Eileen Sygalla, and Jennifer Truhn.

Katharina J. Rohlfing, Angela Grimminger & Ulrich Mertens

Committees

Local Organizing Committee

Katharina J. Rohlfing (email address)
Angela Grimminger (email address)
Ulrich Mertens (email address)

Reviewers

Olga Abramov (Bielefeld University, Germany)
Manuel Bohn (Stanford University, USA)
Silvia Bonacchi (Warsaw University, Poland)
Jana Bressem (Technische Universität Chemnitz, Germany)
Hendrik Buschmeier (Bielefeld University, Germany)
Nina Capone Singleton (Seton Hall University, USA)
Alan Cienki (Vrije Universiteit Amsterdam, The Netherlands)
Jean-Marc Colletta (University of Grenoble, France)
Nevena Dimitrova (University of Applied Sciences and Arts of Western Switzerland)
Gaëlle Ferré (University of Nantes, France)
Tilbe Göksun (Koç University, Turkey)
Angela Grimminger (Paderborn University, Germany)
Michelle Guidetti (CNRS-CLLE-University Toulouse 2, France)
Marianne Gullberg (Lund University, Sweden)
Silva H. Ladewig (Europa-Universität Viadrina, Germany)
Vivien Heller (University of Wuppertal)
Konrad Juszczyk (Adam Mickiewicz University, Poland)
Maciej Karpiński (Adam Mickiewicz University, Poland)
Friederike Kern (Bielefeld University, Germany)
Sotaro Kita (University of Warwick, UK)
Carina Lüke (Paderborn University, Germany)
Zofia Malisz (Royal Institute of Technology in Stockholm, Sweden)
Iris Nomikou (University of Plymouth)
Seyda Özçalışkan (Georgia State University, USA)
Asli Özyürek (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands)
Karola Pitsch (University of Duisburg-Essen, Germany)
Wim Pouw (University of Connecticut, USA)
Katharina Rohlfing (Paderborn University, Germany)
Gale Stam (National Louis University, USA)
Lauren Stites (Georgia State University, USA)
Susanne Vogt (Hochschule Fresenius, Germany)
Paul Vogt (Tilburg University, The Netherlands)
Sławomir Wacewicz (Nicolaus Copernicus University, Poland)
Petra Wagner (Bielefeld University, Germany)
Przemysław Żywiczyński (Nicolaus Copernicus University, Poland)

Plenary Speakers

Pointing to words: How gesture provides a helping hand to language development across different learners

Şeyda Özçalışkan
Department of Psychology, Georgia State University, USA; seyda@gsu.edu

Children communicate using gestures before they speak, and continue to use gesture along with speech even after they begin to produce their first words. Does gesturing merely precede talking, or is it itself relevant to the language-learning process? If gesturing not only precedes language, but also reflects knowledge relevant to the developmental process responsible for language, then the differences and/or delays in speech should become first evident in gesture across different learners. I approach this question by examining early gesture and speech production of children with different developmental profiles—including children with autism, Down syndrome and typical development, who show unique strengths or weaknesses in their early gesture production. I ask whether early gesture predicts later speech across different learners, and if so, what underlies the link between early gesture and later spoken language development.

Prosody: Cross-modal Interactions of Form and Function

Petra Wagner
Bielefeld University, Germany; petra.wagner@uni-bielefeld.de

The strong link between prosodic expression in speech and gesture has been established through ample empirical evidence that prosodic prominence, prosodic phrasing as well as floor management is expressed and processed in a multi-modal fashion. However, the exact functional and formal relationship between the different modalities is still not well understood:
First, we lack knowledge about which aspects of prosodic expression, namely signal features such as pitch movements or rather structural features such as information structure, are actually reproduced across modalities. To shed light on this issue, I will present a recent series of analyses (Wagner et al., in press), where this question was tackled with a newly developed methodological approach: Listeners were asked to "reproduce" acoustically presented utterances in a drumming task. The results revealed that the patterns of drumming intensities closely resemble patterns of perceptual prominence gathered with established annotation schemes, and involving both prosodic experts and naive listeners. However, as unveiled by a Random Forest Analysis, the gestural reproductions of prosody were driven comparatively stronger by signal cues than by linguistic structure. Also, we found different strategies for the gestural interpretations of prosodic patterns: a largely signal-driven, and a more integrative strategy.
Second, we do not know much about the flexibility or stability of the cross-modal link in prosodic expression. While speech economy models predict a flexible relationship depending on communicative demands and cross-modal compensation, a strong view of cross-modal congruence predicts a stable cross-modal co-ordination. To explore this issue, I will present a series of recent studies of semi-spontaneous, task-oriented interaction (Wagner and Bryhadyr, 2017; Wagner et al., 2019a; 2019b; in prep.) aiming at a better understanding of the impact of (1) information structure and (2) visibility between interlocutors on the cross-modal link in prosodic expression. Our results once again confirm a strong cross-modal temporal co-ordination. Furthermore, we detected a systematic modulation of this co-ordination as a function of communicative demands: In important or unpredictable contexts, co-speech movements occur later and align tightly with corresponding pitch peaks if interlocutors can see each others' hands. Also, a lack of facial visibility between interlocutors leads to a significantly earlier production of corresponding co-speech movements. In summary, our results show that co-speech movements in general can express a rich set of signal and structural cues inherent in speech prosody, and that the degree of temporal co-ordination between speech and co-speech movements is a function of communicative needs. As a side result, we found that cross-modal prosodic link also extends to co-speech movements such as drumming or manual moves on a game board, which are not gestures in the traditional understanding of the term.

References

Wagner, P. and N. Bryhadyr (2017). Mutual Visibility and Information Structure Enhance Synchrony between Speech and Co-Speech Movements. *Journal of Multimodal Communication Studies 4(1-2):* 69-74.

Wagner, P., Cwiek, A., and B. Samlowski (in press). Exploiting the speech-gesture link to capture fine-grained prominence impressions and listening strategies. *Journal of Phonetics*

Wagner P., Bryhadyr, N., Schröer, M., and B. Ludusan (2019a). Does information-structural acoustic prosody change under different visibility conditions? *In: Proceedings of the International Congress of Phonetic Sciences 2019, Melbourne, Australia*.

Wagner P., Bryhadyr N., and M. Schröer (2019). Pitch Accent Trajectories across Different Conditions of Visibility and Information Structure - Evidence from Spontaneous Dyadic Interaction. *Proceedings of Interspeech 2019, Graz, Austria.*

Wagner, P. et al. (in prep.). The temporal coordination between speech prosody and co-speech movements as a function of communicative needs.

Understanding human behavior using virtual humans: lessons learned and upcoming challenges

Alexis Heloir
Université Polytechnique des Hauts de France, France; Alexis.Heloir@univ-valenciennes.fr

Animated Virtual Characters exhibit many desirable aspects for who wants to understand human behavior and language. Like their human counterpart, they can display a broad palette of multimodal stimuli, these stimuli can, however, be very precisely altered, fired at exact timestamps, or triggered by specific reactions of a subject taking part in an experiment.
Unfortunately, the integration of virtual characters into a full-fledged experiment setup requires a concentration of many diverse and specific skills which are often out of the reach of the team crafting the experiment. A lightweight, modular, well documented, and easy to deploy agent toolkit is still needed.
This talk starts depicting a series of trans-disciplinary collaborations which lead to the design and implementation of interactive virtual humans in experimental setups involving human counterparts. Each experiment was able to shed an original light on specific aspects of human language or behavior. This talk later focuses on the lessons learned during these trans-disciplinary collaborations and how we could infer from them the leading design principles of a new agent creation and control framework called YALLAH.
YALLAH stands for Yet Another Low-Level Avatar Handler. It is a framework supporting the creation of real-time interactive virtual humans by non-experts. After a quick overview of YALLAH's features, documentation, and ongoing projects using YALLAH, the talk will conclude by a discussion on how YALLAH could help the community understanding the coordination of gesture with speech.

Enacting prosody in the classroom: How the prosody in our hands helps us learn pronunciation in a second language

Pilar Prieto
ICREA-Universitat Pompeu Fabra, Barcelona, Catalunya; pilar.prieto@upf.edu

When we speak, we use rhythmic hand gestures which are coordinated with prominent parts of speech (e.g., beat gestures). In this talk I will discuss several experiments carried out in our research group that deal with how beat gestures and other embodied rhythmic movements facilitate the learning of second language pronunciation. Even though most of the research on the benefits of gesture in the second language classroom has focused on the effects of representational gestures (e.g., for the acquisition of vocabulary), little is known about the potential beneficial effects of embodied rhythmic movements on the learning of L2 pronunciation. A set of experiments will be presented. Experiments 1 and 2 will assess the potential benefits of observing and performing beat gestures on L2 pronunciation learning with intermediate Catalan learners of English. Experiments 3 and 4 will assess the benefits of hand-clapping on L2 pronunciation learning at initial stages of L2 acquisition of French by Catalan and Chinese native speakers. Widening the scope of this investigation, Experiments 5 and 6 will focus on the positive effects of using melodic and singing trainings for pronunciation learning. Based on the positive findings from these experiments, I will conclude that a multimodal approach is essential to understanding L2 speech learning. I will suggest that not only rhythmic trainings with beat gestures or hand-clapping procedures can act as scaffolding mechanisms for L2 speech production but also melodic trainings based on pitch mimicry and singing. Importantly, both types of embodied prosodic trainings could be successfully applied to language teaching and language treatment contexts.

Talks

Paper Session 1: Timing

Does gestural hierarchy align in time with prosodic hierarchy? Another modality to consider: Information structure

Olcay Turk
Victoria University of Wellington, School of Linguistics and Applied Language Studies, New Zealand

Abstract
This study investigates the coordination of gesture with prosody and information structure in Turkish. It has long been known that gesture has a hierarchical structure like prosody. It is also known that gesture is coordinated with prosody on a prominence-related micro level, but less is known about whether this coordination persists at higher levels in the hierarchies. Even less is known about a possible timing relationship to a modality that is also signalled by prosody - information structure. 3 hours of natural speech data was acquired from the narrations of four participants. The study tests the temporal coordination of gesture phrases with multiple levels of phrases within the prosodic hierarchy as well as with information structural units (e.g., topic/focus) that informs the prosodic phrasing. The results show that the hierarchy of alignment is preserved and gesture phrases align with the corresponding prosodic phrases. Information structure units and gesture phrases do not show perfect alignment, but there was a systematic overlap where complete gesture phrases contained the information structure units. Gesture phrase medial stroke + post-hold combinations provided a better anchor for alignment. Overall, the findings confirm multiple levels of alignment between hierarchical structures of gesture and prosody as well as providing empirical evidence for the claim that gesture is informed by information structure in addition to traditional semantic, pragmatic and phonological modalities.

The timing of pointing-speech combinations in typically developing and language-delayed toddlers

Angela Grimminger
Paderborn University, Germany

Abstract
Research on the development of the gesture–speech integrated system suggests that the temporal alignment becomes closer with progression in linguistic skills. In this study, the multimodal communicative combinations of pointing gestures with speech (vocalizations and first words) in two groups of 18-month-old children with different developmental trajectories in their linguistic development were analyzed: a group of typically developed children and a group of children delayed in language acquisition—as attested retrospectively by a standardized test. Using the reliable paradigm of the decorated room to elicit pointing behavior in children, the analyses focussed on the timing between the two modalities and the temporal distances between gesture and speech onsets. Similar patterns of gesture–speech integration were found for both groups.

Paper Session 2: Genre

Embodied reciprocity in conversational argumentation: Soliciting and giving reasons with Palm Up Open Hand gestures

Nora Schönfelder, Vivien Heller
University of Wuppertal, Germany

Abstract
Based on video-recorded peer interactions, this paper discusses the interactive functions of Palm Up Open Hand gestures in conversational argumentation. Drawing on conversation analysis, we demonstrate that PUOH gestures occur in sequential positions where new perspectives come up for discussion and divergent positions are established; they are thus resources for soliciting or giving reasons. It is argued that as publicly visible resources reciprocal PUOH gestures facilitate the orderly production of contiguous responses and ensure coherence between turns.

Encouraging Gesture Use in a Narration Task Increases Speakers' Gesture Rate, Gesture Salience and the Production of Representational Gestures

Alice Cravotta[1], Pilar Prieto[2,3], M. Grazia Busà[1]
[1]*Università degli Studi di Padova, Dipartimento di Studi Linguistici e Letterari (DiSLL), Italy*
[2]*Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Catalunya, Spain*
[3]*Universitat Pompeu Fabra, Departament de Traducció i Ciències del Llenguatge, Barcelona, Catalunya, Spain*

Abstract
Previous work has shown the positive effect of encouraging gestures in performing various tasks; in these studies, the participants generally appeared to gesture more when explicitly asked to do it. However, little attention has been paid to whether encouraging gestures also affects other gesture features, i.e., gesture type and salience. In this paper we explore this issue. Twenty native Italian speakers described the content of short comic strips to a listener in 2 conditions: Non-Encouraging gestures (N); Encouraging gestures (E). Co-speech gestures were manually coded and classified according to gesture type (Representational vs. Non-Representational) and gesture salience (Salient vs Non-Salient). The results show that instructing speakers to gesture led to an increase in gesture rate, in gesture salience, and in the number of representational gestures. By contrast, in the non-encouraging condition the rate of Non-Salient gestures was significantly higher, but no difference was found for Non-Representational gestures.

Children's viewpoint: Iconic co-speech gestures and their relation to linguistic structure across two communicative genres

Ulrich Mertens[1], Friederike Kern[2], Stefan Kopp[3], Olga Abramov[3], Anne Németh[2], Katharina J. Rohlfing[1]
[1]*Paderborn University, Germany;* [2]*Bielefeld University, Germany;* [3]*CITEC, Bielefeld University, Germany*

Abstract
In this study, two different communicative genres (explanation vs. report) were elicited in 38 German preschool children at the age of 4 years. In one part of the study, explanations of a game were elicited from the child. The game involved spatial movements and figures with various geometrical shapes. In a subsequent part, children reported about a puppet and its odd behaviour to their caregiver. We examined children's viewpoint in iconic co-speech gestures and related it to the children's event structures and linguistic structures that differed in terms of transitivity. Our findings suggest that children do not use viewpoints in a unified way—which had been reported from studies with adults. In contrast, our results indicate a great variability in the ways children use viewpoint in iconic co-speech gesture. We found that different communicative genres (explanation vs. report) evoke different viewpoints in gesture, due to their different event structure and linguistic structure. During the genre "explanation", O-VPT gestures occurred more frequently with intransitive utterances, whereas during the genre "report", C-VPT gestures occurred more frequently with transitive utterances. Moreover, neither of the events within the communicative genres exclusively evoked one specific viewpoint.

Paper Session 3: Methods

Quantifying Gesture-Speech Synchrony

Wim Pouw[1, 2], James A. Dixon[1]
[1]*Center for the Ecological Study of Perception and Action, University of Connecticut*
[2]*Department of Psychology, Education, & Child Studies, Erasmus University Rotterdam*

Abstract
Spontaneously occurring speech is often seamlessly accompanied by hand gestures. Detailed observations of video data suggest that speech and gesture are tightly synchronized in time, consistent with a dynamic interplay between body and mind. However, spontaneous gesture-speech synchrony has rarely been objectively quantified beyond analyses of video data, which do not allow for identification of kinematic properties of gestures. Consequently, the point in gesture which is held to couple with speech, the so-called moment of "maximum effort", has been variably equated with the peak velocity, peak acceleration, peak deceleration, or the onset of the gesture. In the current exploratory report, we provide novel evidence from motion-tracking and acoustic data that peak velocity is closely aligned, and shortly leads, the peak pitch (F0) of speech.

An integrative platform to capture the orchestration of gesture and speech

Christelle Dodane[1], Dominique Boutet[2], Ivana Didirkova[3], Fabrice Hirsch[1], Slim Ouni[4], Aliyah Morgenstern[5]
[1]*Université Paul Valéry, France;* [2]*Université de Rouen, France;* [3]*Université Paris 8, France;* [4]*Université de Lorraine, France;* [5]*Sorbonne Nouvelle-Paris 3, France*

Abstract
A number of studies have highlighted the coordination of gesture and intonation (Bolinger, 1983; Darwin, 1872; Cruttenden, 1997; Balog & Brentari, 2008; Roustan & Dohen, 2010) but the technological set-ups have been insufficient to couple the acoustic and gestural data with sufficient detail. In this paper, we present the MODALISA platform which enables language specialists to integrate gesture, intonation, speech production and content. The methods of data acquisition, annotation and analysis are detailed. The preliminary results of our pilot study illustrate strong correlations between gestures and intonation when they are simultaneously performed by the speaker. The correlations are particularly strong for proximal segments. Our aim is to expand those results and analyse typical and atypical populations across the lifespan.

Paper Session 4: Management of change

Smiling for negotiating topic transitions in French conversation

Mary Amoyal, Béatrice Priego-Valverde
Laboratoire Parole et Langage, CNRS, Aix-Marseille Université 5, France

Abstract
This study focuses on participants' smiling behavior as a resource for negotiating topic transitions in French conversations. The smile will be analyzed as a resource during topic transitions: through its intensities and its development. This study will show that the speaker's smiling dynamic contributes to initiating a transition and that the hearer tends to synchronize his/her smile with the speaker to ratify it. Index terms: smile, topic transition, conversation, alignment.

Hand gestures and pitch contours and their distribution at possible speaker change locations: a first investigation

Margaret Zellers[1], Jan Gorisch[2], David House[3], Benno Peters[1]
[1]*Institute for Scandinavian Studies, Frisian Studies, and General Linguistics, University of Kiel, Germany;*
[2]*Leibniz-Institute for the German Language, Mannheim, Germany;*
[3]*KTH Speech, Music & Hearing, Stockholm, Sweden*

Abstract
Smooth turn-taking in conversation depends in part on speakers being able to communicate their intention to hold or cede the floor. Both prosodic and gestural cues have been shown to be used in this context. We investigate the interplay of pitch movements and hand gestures at locations at which speaker change becomes relevant, comparing their use in German and Swedish. We find that there are some shared functions of prosody and gesture with regard to turn-taking in the two languages, but that these shared functions appear to be mediated by the different phonolog- ical demands on pitch in the two languages.

Paper Session 5: Coordination

Gesture-Speech Coordination in Expression of Motion: How Far to Zoom In to Observe Semantic Synchrony?

Katerina Fibigerova, Michèle Guidetti
CNRS-CLLE-University Toulouse 2, France

Abstract
The present paper/talk contributes to the discussion about coordination between gesture and speech from the semantic and morpho-syntactical perspective. What information is conveyed in co-speech gesture and how that information relates to the content of the co-occurring segment of speech? Does temporal synchronicity imply semantic synchronicity?
We tackled these questions in the context of description of motion events, in terms of combinations of a specific path (e.g. upward, downward, crossing) and a specific manner (e.g. walking, running, flying). We asked whether gesture depicts the same element(s) of motion that speech does and to ensure variability of verbal content we adopted the comparative method involving French and Czech speakers, two languages offering different patterns for expression of motion path and manner.
This paper/talk presents our most recent results that extend our previous studies in this field. After having observed gesture-speech semantic synchrony at the level of 'gesture-proposition' and 'gesture-word', it was time to zoom into individual words and explore the 'gesture-morpheme' level.

Gestural training benefits L2 phoneme acquisition: Findings from a production and perception perspective

Marieke Hoetjes, Lieke van Maastricht, Lisette van der Heijden
Radboud University, Centre for Language Studies, Nijmegen, The Netherlands

Abstract
This paper aims to study whether training with gestures benefits L2 phoneme acquisition from both a production and perception perspective. In the production study, Dutch learners of Spanish received pronunciation training for the phonemes /u/ and /θ/ in one of four conditions: audio-only, audio-visual, audio-visual with pointing gestures, or audio-visual with iconic gestures. Results show that in general, gestural training benefits L2 phoneme acquisition, but different gestures benefit the acquisition of different phonemes, possibly depending on their complexity. The perception study, in which L1 speakers of Spanish judged the L2 Spanish material on accentedness and comprehensibility, corroborate the findings from the production study: Including visual information in training generally lowered the perceived accentedness and increased the perceived comprehensibility of speech, but the type of phoneme matters. Together, these studies suggest that gestural training can benefit L2 phoneme acquisition, yet certain gestures work better for certain phonemes than others.

Acoustic Specification of Upper Limb Movement in Voicing

Wim Pouw[1,–2], Alexandra Paxton[1,3], Steven J. Harrison[1], James A. Dixon[1]
[1]*Center for the Ecological Study of Perception and Action, University of Connecticut*
[2]*Department of Psychology, Education, & Child Studies, Erasmus University Rotterdam*
[3]*Department of Psychological Sciences, University of Connecticut*

Abstract
Hand gestures communicate through the visual information created by movement. Recently, we found that there are also direct biomechanical effects of high-impetus upper limb movement on voice acoustics. Here we explored whether listeners could detect information about movement in the voicing of another person. In this exploratory study, participants listened to a recorded vocalizer who was simultaneously producing low- (wrist movement) or high- (arm movement) impetus movements at three different tempos. Listeners were asked to synchronize their own movement (wrist or arm movement) with the vocalizer. Listeners coupled with the frequency of the vocalizer arm (but not wrist) movements, and showed phase-coupling with vocalizer arm (but not wrist) movements. However, we found that this synchronization occurred regardless of whether the listener was moving their wrist or arm. This study shows that, in principle, there is acoustic specification of arm movements in voicing, but not wrist movements. These results, if replicated, provide novel insight into the possible interpersonal functions of gesture acoustics, which may lie in communicating bodily states.

Synchronization of (Dis)fluent Speech and Gesture: A Multimodal Approach to (Dis)fluency

Loulou Kosmala, Maria Candea, Aliyah Morgenstern
Sorbonne Nouvelle University, Paris, France

Abstract
Disfluency is verbally expressed by several markers (filled, unfilled pauses, repetitions, self- repairs, etc). This study is grounded in the functionally ambivalent view of *(Dis)fluency* following Crible, (2017) and Götz (2013), but with a multimodal and interactional approach. Previous research has shown a coordination between speech and gesture suspension (Gullberg, 2013, 2018; Seyfedinnipur 2006). The aim of our paper is thus to examine how (dis)fluent speech and gestures can be synchronized, and how visual-gestural features can provide a finer understanding of (dis)fluency. Our analyses are conducted on 3 pairs of French and American speakers interacting both in their L1 and their L2. (Dis)fluency markers were annotated according to their multimodal features. Qualitative analyses revealed how the notions of time suspension and planning associated with (dis)fluency were also found in gesture. This strongly supports the idea that (dis)fluency is to be considered a multimodal phenomenon, and its visual cues are essential for a closer examination of its pragmatic functions.

Posters

1 An exploration of verbal and non-verbal projectability and entrainment in choral productions in English interaction

Marina Noelia Cantarutti
Centre for Advanced Studies in Language and Communication, University of York, marina.cantarutti@york.ac.uk

Abstract
This paper adopts a conversation-analytic/interactional linguistic (CA-IL) approach (Sacks et al. 1974; Couper-Kuhlen and Selting 2001; Couper-Kuhlen and Selting, 2017) to the study of spontaneous choral productions. The focus of this exploration is placed on how clusters of verbal and non-verbal cues aid projectability, creating opportunity spaces (Lerner 2004) for a co-participant to come in, and how, in turn, these features and the mutual monitoring during talk contribute to the hearability of these choral productions as highly synchronised and entrained. Figure 1 at the end of this document illustrates the phenomenon.
A choral production here is defined as a co-production in overlap of the same or synonymous wording that acts as a form of pre-emptive completion (Lerner 2004; Lerner 1996; Lerner 1991) of a terminal-item or compound turn-constructional unit (TCU) of A's ongoing TCU. Following Walker (2016), cases are still considered choral even if the start is not strictly simultaneous if they are seen to happen within the 200-millisecond window experimentally established for processing (Levinson, 2014), known in interactional terms as a "blind spot" (Jefferson, 1986), where participants are still in speakership orientation.
This exploratory study is based on 23 cases of unelicited choral productions from 16 hours of two video-recorded spontaneous speech corpora (RCE, 2011; MCY, 2017) collected via direct recruitment. Each case is transcribed using GAT2 (Selting et al, 2011) in both its Basic and Fine conventions using FOLKER (Schmidt & Schütte, 2010) and Praat (Boersma & Weenink, 2016). In keeping with CA-IL methodology, phonetic detail is first annotated following a parametric impressionistic and auditory analysis (Local & Walker, 2005) and observations are subject to instrumental validation. Textgrids are exported onto ELAN (Brugman, Russel, & Nijmegen, 2004), where relevant shifts during the TCU preceding and including the choral production are noted on speaker-identifier tiers, including changes in head, eyebrow, torso and posture, mouth, and hand movement, as well as gaze direction, with due annotation of gestural phases (Kendon, 2004; McNeill, 2000). Each of the cases is analysed in terms of their social action, sequential position, and linguistic and embodied design and details are noted in transcripts and spreadsheets in order to identify potentially systematic and normative use of resources. As CA-IL practice establishes, the foreground-ing of certain verbal and non-verbal resources over others is established *locally* (Couper-Kuhlen and Selting, 1996) based on participant orientation. Therefore, the present study does not aim to assign a form-function mapping, but rather, analyse the role of the simultaneous deployment of resources in each particular interaction to describe common mutual orienta-tion processes that participants employ to format-tie (Goodwin, 1990) and entrain their pro-ductions.
The study is first concerned with the relevant lexico-grammatical and pragmatic design fea-tures that create projectability, their prosodic instantiation and their embodied accompani-ment prior to the achievement of the choral production. Preliminary results indicate that projection (Auer, 2005; Auer, 2009; Stukenbrock, 2018; Deppermann and Günthner, 2015) of the upcoming content is, for the present collection, mostly built syntactically and seman-tically, with the presence of two-part TCUs, such as [Cause + Consequence] or [Quotative + Quoted material], but generally through the establishment of clear dichotomies (see "Scoon") or repetitions that make the completion predictable for co-participants. However, projectability of possible opportunity spaces for B to come in are signalled prosodically and gesturally, with extended segmental duration accompanied with held gestures, or in other cases, the release of gestures aligned with the completion of a pitch contour (generally at the speaker's mid-range) at the end of the first part of compound TCUs. Rhythm and accompanying beat gestures are also found to be important in creating a predictable temporal template for B-incomings (Couper-Kuhlen 1993; Ogden and Hawkins 2015; Szczepek Reed 2006), and

as Figure 1 below shows, B can be visibly seen to be orienting to the rhythmic template established by A.

As to the choral production itself, it is not only the timing of the starting point, but also the possibility of a synchronized equal duration and concerted finish to the TCU that participants seem to be orienting to, even when the wording is not identical. In the collection, lexical and prosodic compression is found, which matches findings of experimental research on "forced" synchronizations, with an inter-speaker lag of no more than 40 ms (Cummins, 2003). Participants are seen to adapt aspects of the duration, and also are found to deploy hand beat gestures making up pikes of multimodal activity (Loehr, 2012; Ogden and Haw-kins, 2015) that contribute to the synchronisation in the ongoing choral production.

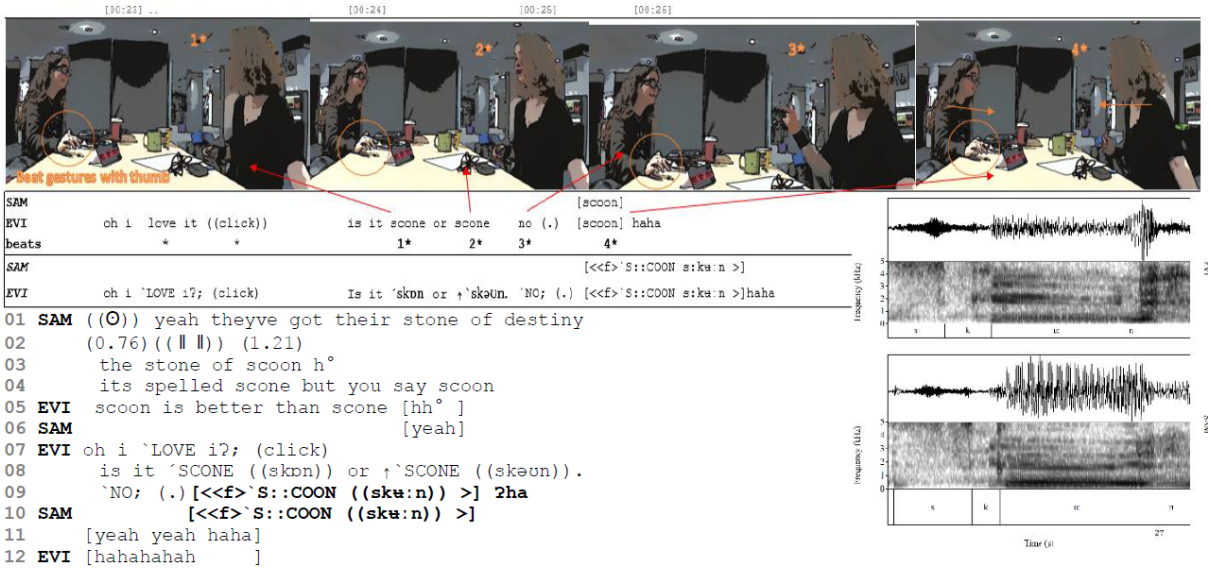

Figure 1: MCY20MUG "Scoon". Transcription (GAT2 Fine level in relevant areas) and partiture of a choral production with a time-aligned waveform + spectrogram.

References

Auer, P. (2005). Projection in Interaction and Projection in Grammar. *Text 25(1)*, 477.

Lerner, G. H. (2002). Turn-sharing: The Choral Co-Production of Talk-in-Interaction. In B. F. Cecilia Ford (Ed.), *The Language of Turn and Sequence* (pp. 225–256).

Walker, G. (2016). Phonetic Variation and Interactional Contingencies in Simultaneous Responses. *Discourse Processes, 53(4)*, 298–324.

2 Gestural portray of the public service interpreter: strategies of coping with source messages' nonverbal cues

Monika Chwalczuk
Université Paris Diderot, Laboratoire CLILLAC-ARP, monikachwalczuk@gmail.com

Abstract
While a growing body of researchers investigate gesture-speech correlations in language acquisition or second language teaching, few show interest in examining multimodality of interpreter-mediated interactions. My research goal is to create an authentic corpus of community interpreting events in public service settings, essential for answering the following question: what strategies do consecutive interpreters apply in order to render the nonverbal cues of the source messages?
Index terms: *community interpreting, interpreting strategies, multimodal corpora, gestural repertory, iconic gestures*

1. Introduction

Owing to the continuous development of computer-assisted speech analysis tools, multimodal corpora progressively become the prevalent way of investigating discursive patterns, especially those taking into account the role of gestures (Bigi & Saubesty, 2015). As it was widely discussed by gesture studies specialists (Cosnier, 1996) (Kendon, 2004), face-to-face communication does not limit to linear verbal exchanges, but follows the triangular structure of overlapping multimodal signals simultaneously combining *language, paralanguage* and *kinesis* (Poyatos, 1997). Nonetheless, no explicit code has so far been formulated in terms of coping with the nonverbal layer of the interaction. My research aims at determining how community interpreters deal with speakers' gestures. Do they opt for reproducing source gestures, explaining them verbally or rather omit them in their interpretation? That research project being in progress, the scope of this paper is to present the findings of the pilot study only, analyzing video recordings from interpreter-assisted pedopsychiatrical counselling sessions.

2. Methodology: data triangulation

As a first step, I conducted a survey covering the interpreters' perception of the role of gestures in their daily practice (Chwalczuk, 2019). The sample counted 60 professional interpreters working in legal, healthcare and social settings in 42 different languages. Semi-structured interviews with 12 of them were an opportunity to collect authentic examples of the specific use of gestures in their work all along their career and convince them to collaborate during filmed interactions.

2.1. Study sample and context

As a second step, I collected video recordings of an authentic interpreter-mediated interaction with a 6-year old Soudanese boy, accompanied by his parents to a pedopsychiatrical counseling session in Paris. The participants of the event are: the leading psychiatrist (f), the psychologist (f), the interpreter (m), the boy, the mother and the researcher (f). The interpreter is summoned to assist the boy's mother who only speaks French at a very basic level. Moreover, he plays the role of an intercultural mediator (Angelelli, 2004) (Benayoun & Navarro, 2015) pointing out the importance of cultural references and traditions with which the therapists are not familiar.

2.2. Corpus and its annotation

Two video sequences last respectively 32 and 66 minutes and have been manually annotated in ELAN using the gesture typology strongly inspired by the one of McNeill (1992), that is including: iconic, deictic, metaphoric gestures and emblems. Each participant is assigned a set of different tiers used to encode: transcription line, words line and gesture type.

3. Preliminary results

The analysis of abundant gestural productions in pedopsychiatrical sessions led to distinguishing two major tendencies: systematic reproduction of certain gesture types and gesture ritualization (Andrén, 2010) resulting in creation of a new shared repertory.

3.1. Mirroring of iconic gestures

Several iconic gestures were introduced to illustrate the notions which were being explained in the interaction. They were used both by the psychiatrist to demonstrate the meaning of difficult French

words (G + VE1; ex. *bandelette salivaire*), as by the mother to mimic actions that she did not manage to describe in words (G + NVE; ex. *se disperser*). The interesting phenomenon is that in such cases the interpreter tended to reproduce the key gesture in his contributions (G + VE) and the same gesture was subsequently mirrored by the recipient, reassuring of their understanding. An example of such triangular mirroring was for instance the explanation of the term *handball* brought up by the therapist (Figures 1 and 2).

Figure 1. Mirroring of iconic gestures [Handball]

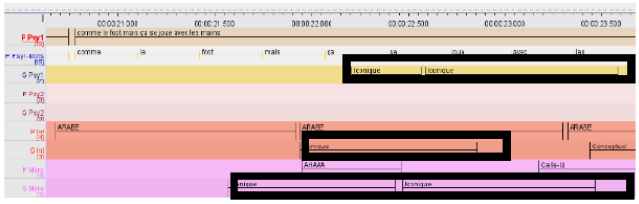

Figure 2. Mirroring in ELAN

3.2. Shared gestural repertory

The repeated reproduction of new gestural units led to establishing a strong semantic dependency between gestures and their verbal equivalents. Therefore, when one of the participants wanted to refer to such a term, they would evoke it both verbally and nonverbally, using the same gesture all along the interaction, or even during further sessions. In consequence, a shared gestural repertory was created.

4. Discussion

We could argue that due to their transparency (Andrén, 2010), the use of iconic gestures grants immediate access to the semantic content of language-coded messages. This convergence between the visually obtained information and their verbal interpretation reassures the patients that the interpreter repeats exactly what has been said, which suggests that skillful use of nonverbal cues may increase their role of trust builders in community interpreting. Therefore, the strategies of coping with nonverbal signal should become part of the interpreters' training.

References

Andrén, M. (2010). *Children's Gestures between 18 and 30 months.* Phd Thesis. Lund: Lund University.

Angelelli, C. (2004). *Medical Interpreting and Cross-cultural Communication.* Cambridge: Cambridge University Press.

Benayoun, J.-M., & Navarro, E. (2015). *Interprétation et médiation: Midgations, représentations et enjeux socioréférentiels.* Paris: Michel Houdiard.

Bigi, B., & Saubesty, J. (2015). Searching and retrieving multi-levels annotated data. *Gesture and Speech in Interaction - 4th Edition, (pp. 31-36)*. Nantes.

Chwalczuk, M. (2019). When gestures speak louder than words: The role of gestures in community interpreting events. In: I. Cobos López, *Estudios sobre traducción e interpretación. Especialización, didactica y nuevas lineas de investigación* (pp. 471-490). Valencia: Tirant humanidades.

Cosnier, J. (1996). Les gestes du dialogue. La communication non verbale. *Psychologie de la motivation,* pp. 129-138.

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

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

Poyatos, F. (1997). *Nonverbal communication and translation: New perspectives and challenges in literature, interpretation and translation.* Amsterdam/Philadelphia: John Benjamins.

3 Adaptation of multimodal communication strategies to noise and failure: evidence from a dyadic interaction task

James Trujillo, Linda Drijvers, Judith Holler and Asli Özyürek
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Abstract
In many natural face-to-face interactions, we are challenged with communicating in non-ideal settings, such as noisy environments. Typically, we are able to successfully communicate despite interference from noise. This partially due to our ability to "filter out" non-relevant noise, but also due to communicative adaptations made by the speaker, which help the signal to be more salient. The classic example of such adaptation is the Lombard Effect (Davis et al. 2006), which refers to involuntary changes in speech, such as in intensity and pitch, that make the speech more easily understood by the addressee. While much of this research has focused on adaptations in the auditory signal and in visible speech (ie. lip movements), co-speech hand gestures are an important part of communication which have not been well studied in noise from the production side. Previous research has shown links between speech and gesture production (Goldin-Meadow 1999), as well as the intentional shaping of gesture kinematics in response to communicative need (Holler and Wilkin 2011; Trujillo et al. 2018). Until now however, there is no research on how speakers respond multimodally to communication in noise. This is particularly relevant because co-speech gestures enhance a listener's understanding of speech in noise (Drijvers and Özyürek 2017). Here, I present results from a dyadic communication task carried out at the Lowlands music festival. In the task, participants wore headphones with varying (per round) levels of noise. One participant, called the Producer, communicated action verbs (one per round) to the Addressee. For our analyses, we look at the first two attempts at communicating each individual word, using qualitative methods to describe the communicative strategy and quantitative motion capture methods to assess kinematic features. Preliminary results show that increasing levels of noise do not affect the strategy used, nor is multimodality specifically associated with communicating in more or less noise. However, the kinematics of gestures are exaggerated in response to more noise. Similarly, when the first attempt at communication fails, participants typically persisted with the same communicative strategy and modality, but further modulated their kinematics in the second attempt. I discuss these results in relation to social shaping of gestures, speech-gesture trade-off, and the effects of noise and communicative failure on communicative strategy.

References

Davis C, Kim J, Grauwinkel K, Mixdorff H. (2006). Lombard speech: Auditory (A), Visual (V) and AV effects, *Proceedings of Speech prosody.*

Drijvers L, Özyürek A. (2017). Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension. *Journal of Speech, Language, and Hearing Research, 60*:212.

Goldin-Meadow S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Science, 3*:419–429.

Holler J, Wilkin K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics, 43*:3522–3536.

Trujillo JP, Simanova I, Bekkering H, Özyürek A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition, 180*:38–51.

4 The Visual Communication Heuristic: The Effect of Context on Gesture Production

Jacob Barker and Sotaro Kita
The University of Warwick, UK

Abstract

Speakers naturally produce co-speech gestures to help communicate their message (Kendon, 1994). When speakers produce gestures alongside speech, the speaker is presumably helping their listener to understand the message (Hostetter, 2011). If some gestures are motivated by the desire to communicate, they should be produced less frequently when they cannot be used to communicate; for example, when the speaker and the listener cannot see each other. Alibali, Heath and Myers (2001), found that when the speaker and listener could not see each other, the speaker produced a lower frequency of co-speech gestures than when the speaker and listener could see each other. This poster concerns this visibility effect on gesture production.

The traditional explanation for the visibility effect is that when the speaker and listener cannot see each other, the speaker's gestures cannot be used to communicate, thus the speaker has a weaker motivation to gesture (Alibali et al., 2001). This suggests the speaker's gesture rate is sensitive to whether his/her gestures are visible to the listener or not.

A more recent explanation for the visibility effect is that when the speaker and listener cannot see each other, the listener can no longer provide visual feedback to the speaker, weakening the speaker's motivation to gesture (Bavelas & Healing, 2013). If a speaker and listener cannot see each other, they can no longer make eye contact, see each other's facial expressions, or pick up on any other visual communication, and thus they can no longer provide each other with constant feedback. This feedback is part of the grounding process that takes place during conversations (Clark & Schaefer, 1989). Listeners provide constant "back channel" responses to let the speaker know if they are communicating effectively, or if they need to re-evaluate (Yngve, 1970). Bavelas and Healing argue that when listeners cannot provide back-channel responses, the grounding process breaks down and thus the speaker produced fewer gestures. When listeners are free to provide back channel responses verbally, the visibility effect is not observed (e.g. Bavelas, Gerwing, Sutton, & Prevost, 2008); according to Bavelas and Healing (2013), this is because when the speaker cannot see the listener, the listener provides more verbal feedback to compensate for the lack of visual feedback. This suggests the speaker's gesture rate is sensitive to whether the listener provides feedback through visual or verbal responses or not. This explanation for the visibility effect has not been directly tested through manipulating the verbal behaviour of the listener.

The current study tests different explanations for the visibility effect. Experiment 1 teased apart gesture visibility, facial visibility and the listener's verbal responsiveness. What was visible between the speaker and the listener was manipulated using three levels. The first level had no screen between the speaker and the listener. In this condition, the speaker and listener could see each other entirely. The second level had a large screen placed between the speaker and the listener. In this condition, the speaker and listener could not see each other at all. The final level featured a shoulder height screen between the speaker and listener. In this condition the speaker and listener could see each other's faces above the screen, but could not see each other from the shoulders down, thus they could not see each other's gestures. Listener verbal responsiveness was manipulated by having the experimenter acting as the listener, being either verbally responsive (i.e. providing verbal back channel responses) or non-responsive (i.e., being silent).

Participants produced more gestures in the no screen and shoulder height screen conditions than in the full screen condition. This suggests speakers produced more gestures when the speaker and listener could see each other's faces. Whether the listener could see the speaker's gestures did not influence the gesture rate. Listener responsiveness did not influence the speaker's gesture rate, either. The findings contradict the explanation for the visibility effect that speakers produce fewer gestures when their gestures cannot be seen by the listener (E.g. Alibali et al., 2001). Speakers produced comparable amounts of gestures in the full screen condition and the shoulder height screen condition, even though in the latter condition, the listener could not see gestures. This suggests the visibility effect is not caused by whether the listener can see the speaker's gestures or not.

The findings do not support the claim that speakers produce more gestures when the listener produces feedback than when they do not. (Bavelas & Healing, 2013). When the listener did not produce any verbal responses, speakers produced just as many gestures as when the listener did produce verbal responses. This suggests the visibility effect is caused by what the listener and speaker can see, not the listener's use of the conversational grounding process.

The results of experiment 1 support a novel hypothesis about the mechanism responsible for the visibility effect. A visual communication heuristic may trigger greater gesture rates from speakers when speakers can make eye contact with their listener, than when speakers cannot. In most everyday situations, the visibility of face co-varies with the visibility of gesture, thus such a heuristic leads to a desirable outcome, in which speakers produce more gestures when the listener can see the gestures (and the face) of the speaker. Furthermore, the heuristic would save the speaker the effort of determining if their gestures can be seen or not, which requires costly perspective taking processes.

The findings of Experiment 1, however, are compatible with two other alternative versions of a heuristic. The first alternative is that the heuristic determines if the speaker should produce a greater frequency of gestures whenever the speaker and listener can communicate to each other through any visual modality, including gesture and facial expressions. According to this view, the heuristic should trigger a greater number of gestures whenever the speaker's gestures (but not the face) can be seen. The second alternative is that the heuristic will determine the speaker should produce a greater frequency of gestures when the speaker has any visual indicator that their listener is present. According to this view the speaker should produce more gestures when they can see any part of the listener's body, not just those use to communicate visually.

Experiment 2 is currently being run to establish how the visual communication heuristic determines the rate at which speakers produce gestures. This experiment manipulated how the speaker and listener could see each other using four levels. In the first condition the speaker and listener could see each other entirely. In the second condition the speaker and listener could not see each other at all. In the third condition the speaker and listener could not see each other's faces, but could see each other from the chin down (thus could see gestures). In the fourth condition the speaker and listener could not see each other from the knees up (thus could not see the face nor gestures), but could see each other's shins and feet.

We hypothesise that if eye-contact needs to be made, we should observe the speaker producing more gestures when the speaker and listener can see each other's faces. If any part of the body used to communicate needs to be seen, we should observe the speaker producing more gestures when the listener can see the speaker's gestures, even in the situation where the listener cannot see the speaker's face. If any visible indicator of the listener's presence is sufficient, we should observe the speaker producing more gestures whenever the listener and speaker can see any part of each other.

The visual communication heuristic explains a range of visibility effects. Rather than determining whether gestures can be seen by the listener, the heuristic determines how frequently gestures should be produced, based on easy-to-process features of the communicative context.

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(2)*, 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(2)*, 495–520.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture, 13(1)*, 63–92.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13(2),* 259–294.

Hostetter, A. B. (2011). When Do Gestures Communicate? A Meta-Analysis. *Psychological Bulletin, 137(2)*, 297–315.

Kendon, A. (1994). Do Gestures Communicate?: A Review. *Research on Language and Social Interaction, 27(3)*, 175–200.

Yngve, V. H. (1970). On getting a word in edgewise. In: *Papers from the sixth regional meeting of the Chicago Linguistic Society* (pp. 567–578).

5 Aging, Working Memory, and Mental Imagery: Understanding gestural communication in younger and older adults

Burcu Arslan, Buse Nur Caba, and Tilbe Göksun
Koç University, Turkey

Abstract
Effects of aging are observed in language and in gestural communication skills. Previous studies showed that although gesture rate may be similar between younger and older adults, the use of representational gestures decreases with age. In this study, we focused on whether this change is observed across different contexts and it can be explained by working memory and/or mental imagery skills. We used three tasks to obtain spontaneous co-speech gestures from younger and older individuals. The results showed that older adults used fewer representational gestures only in a spatial context. However, in such a context, mental imagery skills but not working memory was associated with the use of representational gestures, regardless of age. Additionally, two age groups did not differ in terms of spatial word use in the spatial task. We concluded that gesture and speech production might have separate timelines in terms of being affected by aging process.

1 Introduction

Cognitive capacity peaks at around age 20 and then it starts to decline as people age (Salthouse, 2009). Language is an important domain in which we can observe the prominent manifestations of the cognitive decline (Kemper, 2015). Gestural communication is part of this language process. Although gesture rate does not differ between younger and older adults, these groups can produce different types of gestures (i.e., representational vs. nonrepresentational). In particular, elderly people use fewer representational gestures compared to younger adults (Cohen & Borsoi, 1996).

The change in the type of gestures has been attributed to different factors such as decrease in working memory and/or mental imagery abilities (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999). Since working memory capacity declines with age, simultaneous and effective organization of gesture and speech would become more difficult. As a result, elderly people can invest their limited working memory capacity to speech (Theocharopoulou, Cocks, Pring, & Dipper, 2015). Therefore, less energy would be spent for the organization of meaningfully rich co-speech gestures. In the case of mental imagery, on the other hand, Feyereisen and Havard (1999) asked their participants questions that would evoke visual, motor, and mental images and observed their gestures in such a context that mental imagery abilities would be required. However, to our knowledge, mental imagery has not been assessed directly in the earlier studies.

This study investigates the use of different types of gestures by younger and older groups in different tasks and examines the relationship between working memory capacity, mental imagery, and gesture use in these age groups. We hypothesize that the use of representational gestures will decrease by age and this pattern can be related to working memory and mental imagery, particularly for the older age group.

2 Methods of data acquisition, annotation and analysis

2.1 Participants

We recruited 16 younger adults (14 women) between 20 and 25 years old (*M*age = 21.75, *SD* = 1.61) and 17 older adults (9 women) between 60 and 75 years old (*M*age = 67, *SD* = 4.66). All of them had normal or corrected-to-normal vision and no diagnosis of a neurological disorder.

2.2 Materials and Procedure

After taking informed consents, we collected demographic information from the participants. For speech/gesture production, we presented three different tasks. First, we asked “What do you do in a normal day?” We asked this question to warm up the participants and observe baseline gesture rates in their daily speech. Second, we read three unfinished stories and asked participants to come up with an ending to these stories. Last, we asked the question of “How do you go to the closest market from your home?” to observe gesture frequency and types in a spatial context. To assess mental imagery, we used the Mental Imagery Task developed by Di Nuovo, Castellano, and Guarnera (2014). We also used the Corsi-span task to measure working memory performance (Corsi, 1972). All sessions were

carried out with standardized instructions by a single experimenter and videotaped for later transcription and coding.

After transcribing and coding co-speech gestures produced in each task, we classified them into iconic, deictic, metaphoric, beat gestures, and emblems (McNeill, 1992). We categorized iconic, metaphoric, deictic gestures as representational and beat and emblem gestures as nonrepresentational. Two independent coders identified and coded gestures for each with an interrater agreement of 91.9% for categorizing gestures (Cohen's kappa=.95)

2.3 Results

Preliminary results showed that the total number of words produced did not differ between two age groups neither in the daily question task nor in the story completion task, *t*s< .298, *p*s>.05. However, younger adults (*M* = 61.81, *SD* = 33.71) produced more words than the older adults (*M* = 29.59, *SD* = 33.37) in the address description task, *t*(31)= 2.756, *p*< .05. Gesture rates did not significantly differ between younger and older adults in any of these tasks, *t*s< -.493, *p*s >.05.

We then carried out a repeated measures ANOVA as age group being the between-subject factor and the use of gestures (the total number of gestures/the total number of words) in three different tasks as the within-subject factor. Results indicated only a main effect of gesture use in the tasks, *F*(2,30) = 3.16, *p*=.50. In particular, regardless of age, participants produced more gestures in the address description task. The same result was revealed for the representational gesture use, *F*(2,30) = 35.69, *p* < .001. In other words, both age groups used more representational gestures in the address description task (*M* = .85, *SD* = .20) than the daily question task (*M* = .34, *SD* = .32) and the story completion task (*M* = .27, *SD* = .22). Then, for the representational gestures when we ran the same analyses with two different ANCOVAs, controlling for the mental imagery test (MIT) scores or the Corsi total scores, the main effects were disappeared only for the MIT scores, *F*(2,28) = 3.19, *p*=.056 and *F*(2,28) = .09, *p*=.91, respectively for the Corsi and the MIT scores. Bivariate correlations among these variables (controlling for age) also indicated a correlation between the representational gesture use in this task and individuals' MIT scores, *r*= .64, *p*=0.008.

3 Discussion

Although we could not find a difference between two age groups in terms of the representational gesture use, overall, the participants produced more gestures and used more representational gestures in a spatial context. Our preliminary results suggest that mental imagery might play a bigger role for gesture production than working memory while talking about space, regardless of the age of the participant. We are still collecting data and these results are preliminary. For future analysis, we will code speech for the types of words participants use to examine whether spoken messages are also similar across age groups and whether speech can be affected by the MIT and Corsi scores.

References

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Ph. D. Thesis, McGill University).

Di Nuovo, S., Castellano S., & Guarnera M.A. (2014). *Mental Imagery Test.* Florence: Hogrefe

Cohen, R. L., & Borsoi, D. (1996). The role of gestures in description-communication: A cross-sectional study of aging. *Journal of Nonverbal Behavior, 20(1)*, 45-63.

Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior, 23(2)*, 153-171.

Kemper, S. (2015). Language production in late life. In: A. Gerstenberg, A. Voeste (Eds.). *Language development: The lifespan perspective.* Amsterdam: John Benjamins. 59-75.

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

Salthouse, T. A. (2009). When does age-related cognitive decline begin?. *Neurobiology of aging, 30(4)*, 507-514.

Theocharopoulou, F., Cocks, N., Pring, T., & Dipper, L. T. (2015). TOT phenomena: Gesture production in younger and older adults. *Psychology and Aging, 30(2)*, 245–25.

6 Degrees of explicitness in children's iconic gestures

Kristin Weiser-Zurmühlen[1], Friederike Kern[1], Ulrich Mertens[2], Olga Abramov[3], Anne Németh[1], Stefan Kopp[3], and Katharina J. Rohlfing[2]

[1]Bielefeld University, Germany; [2]Paderborn University, Germany; [3]CITEC, Bielefeld University, Germany

Abstract

While there is a rich body on research in iconicity in gesture, and their cognitive and pragmatic functions in discourse, less is known on how iconicity is actually achieved as a meaningful device. According to the Interface Hypothesis (Kita & Özyürek 2003, 17), iconic gestures do not only encode certain properties of the referent; they also add to the linguistic structure of the information about the referent. This implies that speakers have a choice (within the bounds of linguistic options provided by the respective language) when selecting features for gestural representation. Using an annotation scheme of practices of iconic gesturing that builds on Kendon (2004) and Streeck (2009), the paper's focus is on children's vague iconic gestures in which no clear referent can be identified. It provides insight into the sequential environments of such gestures from the perspective of speech-gesture interaction in language acquisition and through multimodal interaction analysis.

1 Introduction

Co-speech iconic gestures are produced regularly during interaction; according to GSA (Hostetter & Alibali, 2008), they occur when speakers simulate actions for the purpose of speaking. They are said to function not only as a window into cognitive processes on the side of the speaker (Goldin-Meadow et al. 1992), or the hearer (Alibali et al., 2012) but also as devices to establish and provide meaning and understanding in an ongoing, collaborative communication process (Goodwin 2003; 2011; Streeck 2009). In addition, iconicity in gesture needs to be established by speakers through a set of gestural practices, thus providing an analysis of objects or actions (Streeck 2009). While studies with adults propose that iconic gestures may be evolving in an ongoing interaction, with the co-participant co-constructing their meaning (Goodwin, 2003 and 2011; Streeck 2009), little is known about how children make use of gestural reference. Children are said to display a "gesture explosion" between three and four years of age (McNeill, 2005: 180), even though their gestures might not be adult-like at this point, i.e. with regard to frequency per clause, and gestural space (Levy & McNeill, 2013). To shed light onto children's gestural reference practices, we look at cases of vague iconic gesturing occurring in conversations between caregivers and their 4-year-old children from a perspective of language and gesture development. In these cases, iconicity cannot be derived easily from the gesture. Through describing the sequential environments of such gestures, we hope to gain knowledge about challenges of accomplishing gestural iconicity in children's talk.

2 Methods of data acquisition, annotation and analysis

2.1 Data and annotation

Data stem from a project on development of iconic gesturing in pre-school children (EcoGest; cf. Abramov et al., 2018). Video recordings from pre-school children (*n*=40) in different communicative tasks (explaining, re-telling, correcting, reporting) were obtained. The data were annotated in ELAN according to a categorisation system of gestural practices that builds on previously developed systems (Abramov et al., 2018; Kendon, 2004; Streeck 2009). They were divided into three categories: Mostly static *object-related practices* describe the shape of an object, i.e. by producing its shapes in the air, by sizing, or by modelling it (in this case, the hand itself represents something). Dynamic *action-related practices* depict objects or events by modeling actions or movements. Each of the two major categories comprises several sub-categories. Finally, indexing, i.e. by pointing gesture, is part of the third category.

2.2 Methods and procedure

The study follows methodological and theoretical principles of conversation analytically inspired interactive multimodal analysis (Goodwin, 2011; 2013). After the annotation process, problematic cases were collected that did not allow for a straightforward classification in the schema's sub-category system because no clear referent could be identified. These cases were transcribed in GAT 2 (Selting et al, 2009) and analysed in their sequential context. Focus was on the degree of (iconic) explicitness

in gesture, and in the lexical items the gestures co-occur with. With explicitness we refer to distinctness of the meaning conveyed by the gesture, or, in other words, how precise it can be understood to depict a certain semantic feature of the object or action it relates to.

2.3 Preliminary Results

First analyses indicate that there is a relation between low degree of explication in gesture and their relation to object: While not all object-related gestures have a low degree of explicitness, all gestures with a low degree of explicitness are object-related. Action-related gestures, on the contrary, seem to exhibit high explicitness throughout. Additionally, vague/unclear gestures occur in the context of troubles concerning the interactive understanding process. Thus, vague gestures seem to appear more likely in verbal repair activities and/or in co-occurrences with lexical items either unfitting or denoting vagueness as well. In other instances, the semantic focus signalled by the gesture differs from the one established by the co-participant, thus leading to sequences where meaning and shared knowledge are negotiated.

3 Discussion

The finding that object-related rather than action-related gestures exhibit vagueness can be interpreted as in accordance with the Gesture as Simulated Action [GSA] framework as well as phenomenological approaches to gesture production (Streeck, 2009). Both essentially argue that gesture has its source in real-live action, and that memories of action activate motor areas of the brain (Hostetter & Alibali, 2010). This might explain why gestures depicting actions seem to be more refined – at least at this point of development – than gestures depicting objects by relating to their form. Furthermore, the surrounding interactive context seems to be important: As long as all interlocutors display mutual understanding, gestures are more likely to provide relevant contextual information, even though the accompanying verbal information might be less precise (Streeck 2009). However, if repair activities are required, object-related gestures might turn out to be insufficient for ensuring understanding.

References

Abramov et al. (2018). *Annotation manual – EcoGest.* Universities of Bielefeld and Paderborn. Unpublished Manuscript.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds*. Cognition and Instruction, 9*, 201–219.

Goodwin, C. (2003). Pointing as Situated Practice. In S. Kita (Ed.). *Pointing: Where Language, Culture and Cognition Meet*, NJ: Lawrence Erlbaum, 217–241.

Goodwin, C. (2011). Contextures of action. In J. Streeck, C. Goodwin & C. LeBaron (Eds.). *Embodied interaction. Language and body in the material world.* Cambridge: Cambridge University Press, 182-193.

Goodwin, C. (2013). The Co-operative, Transformative Organization of Human Action and Knowledge. *Journal of Pragmatics, 46(1)*, 8–23.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review, 15(3)*, 495-514.

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

Kita, S. & A. Özyürek (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. (2005). *Gestures and Thought.* Chicago: University of Chicago Press.

McNeill, D. (2014). Gesture–speech–unity: Phylogenesis, ontogenesis, and microgenensis. *Language, Interaction and Acquisition, 5(2*), 137-185.

Streeck, J. (2009). *Gesturecraft. The manu-factering of meaning.* Amsterdam: John Benjamins.

7 Rhythmic movements with objects at 9 months are related to proximal deictic gestures at 12 months

Eva Murillo, Ignacio Montero, and Marta Casla
Universidad Autónoma de Madrid, Facultad de Psicología, Spain;
eva.murillo@uam.es, nacho.montero@uam.es, marta.casla@uam.es

Abstract
The aim of this study is to explore the production of rhythmic movements immersed in multimodal patterns at 9 and 12 months of age, and its relationship with proximal deictic gestures production. We observed 18 infants when they were 9 and 12months-old in a free play situation with a primary caregiver. Results showed that rhythmic movements are frequently produced, especially with objects in the hands. The production of rhythmic movements with objects in the hand at 9 months is related to the production of proximal deictic gestures three months later.

1 Introduction
Gestures and speech are parts of a single communication system in adults. As Iverson and Thelen (1999) proposed, the association between speech and motor actions can be traced developmentally, from the very early links between hand and mouth to the early development of the gesture-vocal system. Iverson & Thelen (1999) described four phases on the development of the coupled speech-gesture system. The first phase is denominated “Initial linkages”. It starts from birth, and includes the spontaneous hand-mouth coordination and the Babkin reflex, as evidence of this initial hand-mouth linkage. The second phase, comprising from six to eight months of life, is named “emerging control”. Rhythmic vocal and manual movements coincide in this phase where the canonical babbling emerges. In the third phase, communicative gesture production precedes and predicts the emergence of the first words. This phase, comprising from 9 to 14 months of life, is denominated “emergence of gestures and words”. The last phase is named “emergence of synchronous speech and gesture”. During this phase, the onset of meaningful and synchronous gesture - word combinations is proposed.
Gestures production in coordination with vocal elements have a key role during language learning. The development of multimodal communicative patterns results predictive of subsequent linguistic acquisitions at different points of the language development (Goldin-Meadow & Butcher, 2003; Rowe & Goldin-Meadow, 2009). Following Iverson and Thelen’s model, it is not until the third phase that we can observe the emergence of gestures and its progressive coordination with words. In this period, rhythmic movements of hands and arms tend to decrease and canonical babbling gives way to word-like productions. Rhythmic repetition is replaced by more controlled communicative resources. However, little is known about how rhythmic movements are progressively replaced by movements that have a clear communicative intention, such as gestures. Our hypothesis is that rhythmic movements are frequently produced with objects in the hands. Before the transition to the use of gestures, some rhythmic movements are immersed in multimodal patterns; that is, rhythmic movements are frequently produced with other elements such as vocalizations or with a social use of gaze. The performance of rhythmic movements with objects accompanied by social gaze or vocalizations can serve as an opportunity to learn how to bring other’s attention towards a reference. Specifically, we hypothesize that behaviours including rhythmic movements, objects and vocalization or social gaze are frequently produced in the period of transition between the phase of “emerging control” to the “emergence of gestures and words” proposed by Iverson and Thelen (1999). Moreover, we also hypothesize that these behaviours can facilitate the transition to the proximal deictic use. Our prediction is that the production of rhythmic behaviours at 9 months of age will be a frequent phenomenon, although there is a tendency to decrease through the last trimester of the first year of age. We also expect to find that the production of multimodal rhythmic behaviours with objects at 9 months of age will be related to proximal deictic productions at 12 months of age.

2 Methods of data acquisition, annotation and analysis

2.1 Participants

We observed 18 infants in a free play situation with a primary caregiver in natural settings when they were 9 and 12 months-old. The sessions lasted for 12 minutes when children were 9 months-old and 15 minutes when children were 12 months-old. All the sessions were coded using ELAN 5.2 software and analysed using Matlab R2018b and SPSS v24.

2.2 Coding system

We coded the rhythmic behaviours produced with the arms or hands. We considered a motor behaviour as rhythmic when it was repeated at least two consecutive times. We coded a rhythmic behaviour as "with object" when the child produces this movement with an object in the hand, or bangs an object with the hand or with another object. We considered a rhythmic movement as "without object" when the movement is produced with no object in the hands, i.e., the child shakes her hand(s) or arm(s) or bangs a surface with the hand(s). We also coded gestures, vocalizations and the social use of gaze according to categories described in previous studies.

3 Results

Rythmic movements production is a frequent behaviour in the repertoire of 9 – 12-month-olds. We conducted a repeated-measures ANOVA taking the frequency of rhythmic movents as the dependent variable. The use of object (with object vs. withouth object), the social use of gaze (social gaze vs. no social gaze), the vocalization (with vocalization vs. without vocalization) and the age (9 months vs. 12 months) were the factors. Preliminary results showed an effect of object use, with more rhythmic movements produced with an object in the hand than without it ($F(1,18)= 4.80$; $p=.042$), withouth social gaze ($F(1,18)= 19.30$; $p>.001$) and withouth vocalization ($F(1,18)= 39.03$; $p>.001$), and there is a tendency of rhythmic movement to decrease with age (($F(1,18)= 4.01$; $p=.061$). However, we found some interesting interactions: when rhytmic movements are produced with social gaze, they are more frequently produced withouth vocalizations at 9 months of age, but this difference is not so clearly found when children are 12 months-old. We also found a higher frequency of rhythmic movements produced without vocalization and with an objetc in the hand at 9 months of age than at 12 months of age. This reflects a tendency with age to produce rhythmic movements immersed in multimodal patterns, specially when the movement is produced with an object in the hand. Finally, we grouped the rhythmic movements produced with gaze and/or vocalizations as "multimodal rhythmic movements" and tested their relationship with deictic gestures (distal and proximal) produced at 12 months of age. We conducted a Spearman's correlation between multimodal and unimodal rhythmic movements (with and whithout object) and proximal and distal decitic gestures (produced in multimodal or unimodal patterns). We found a significant and positive correlation between the frequency of multimodal rhythmic movements produced with an object at 9 months of age and the frequency of proximal multimodal ($\sigma = .530$; $p=.020$) and unimodal ($\sigma = .57$; $p=.010$) deictic gestures produced at 12 months of age. We did not find this relationship for rhythmic movements withouth object or for distal deictic gestures.

4 Discussion

Preliminary results suggest that rhythmic movements produced with objects in a social context can provide children the opportunities to explore how other's attention to a shared referent works. In this sense, this kind of movements could give way to proximal deictic gestures, in which the child directs other's attention to an object that remains in contact with the hand. Further research is needed to analyse the effect of these rhythmic movements on adult's contingent response.

References

Goldin-Meadow, S. & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.). *Pointing: Where language, culture, and cognition meet* (85-107). New Jersey: LEA.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture*. Journal of Consciousness Studies, 6(11-12)*, 19-40.

Rowe, M. & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science 12*, 182-7.

8 Multimodal marking of information structure in 4-year-old German children

Sofia Koutalidis[1], Friederike Kern[1], Anne Németh[1], Ulrich Mertens[2], Olga Abramov[3], Stefan Kopp[3], and Katharina J. Rohlfing[2]

[1]Bielefeld University, Germany; [2]Paderborn University, Germany; [3]CITEC, Bielefeld University, Germany

Abstract

Children are known to make use of language differently than adults as their linguistic and communicative competence is still developing. This applies to verbal as well as multimodal means. The present paper focuses on information structuring, which refers to the organisation of information in our utterances in order to deliver the intended message, and how it is signalled through different modalities, with focus on gesture. Usually, we make use of various modalities to highlight what is important in our utterances and help progress the discourse. The present work tries to capture the means that children use in order to mark information structure during an utterance and how those differ from the ones used by the adults, focussing on the use of beat gestures in different pragmatic contexts.

1 Introduction

In order to understand the complexity and variation of meaning-making in communication, both verbal and nonverbal content has to be taken into consideration (Wagner, Malisz, & Kopp, 2014). Modalities like gesture or body movements (henceforth BM) do not only accompany speech but seem to have an impact on the mental process of information packaging and the conceptual planning of utterances as well (Alibali, Kita, & Young, 2000; Wagner et al., 2014). Modalities can enhance or modify the verbal content of an utterance, which attributes them pragmatic functions as well (Freigang & Kopp, 2015; Kendon, 2004; Wagner et al., 2014). Mc Neill's „pragmatic synchrony rule" states that all modalities serve the same pragmatic function (Mc Neill, 1992). So, when verbal mechanisms highlight information, all other modalities would serve the same purpose. Beat gestures and head nods are usually the main way to highlight information in an utterance next to prosodic marking (McNeill, 1992, Wagner et al., 2014).

When delivering pragmatic information on a verbal level, we mark what we consider important and what will progress the communication further. Information structure focuses the way we organize information, and it entails the way we "package" information to deliver the intended message (Chafe, 1976). The term 'focus' is used to describe new or contrastive information (Kiss, 1998), while the terms 'topic' and 'comment' are used to describe what an utterance is about and what is said about (Pittner & Berman, 2013). In the West Germanic family, topic and focus are marked either prosodically or syntactically in discourse (Pittner & Berman, 2013).

Freigang, Klett and Kopp looked at how pragmatic information is multimodally carried. Functions like emphasizing and deemphasizing (foc+ and foc -) as well as uncertainty (epi+) can be conveyed in various modalities (Freigang, Klett, & Kopp, 2017). Although the definition of focus (what is emphasized) is not identical with the one used within information structure, it is clear that information can be highlighted multimodally.

From a developmental perspective, information structuring –as an aspect of pragmatic competence –, takes more time to develop (Höhle et al., 2014; Musolino & Lidz, 2006). Children will rather make use of prosody rather than syntactic structure to mark saliency and therefore apply the simplest way to mark focus (Sauermann et al., 2011). Likewise, gestures of toddlers carry mostly deictic and semantic properties (Rohlfing, Grimminger, & Lüke, 2017; McNeill 1992). However, young children may use protobeats to highlight information. Protobeats are like beats in that they have no referential properties and disappear as soon as beats appear; however, their temporal and rhythmic alignment is not as sophisticated yet. (McNeill, 1992).

2 Data corpus and annotation

The present work will look at the modalities 4-year-old German children use to highlight information in their utterances, with special focus on the use of BM that may serve as protobeats. The study uses data from the Eco-Gest corpus1 where 40 German children were recorded in four different experimental settings (explaining, retelling, correcting and reporting). The data from 2 children in the explanation and report setting are going to be analyzed. The verbal utterances were transcribed according GAT 2 (Selting et al., 2009). Annotation of the information structure (focus, topic and

comment) will follow Fery et al. (2007) and will be conducted in ELAN. The categories of focus+, focus- as described by Freigang, Klett and Kopp are going to be annotated separately since they are defined as the emphasizing function (Freigang et al., 2017). This will provide a more detailed view on the organization of information within children's utterances. Next, the modalities information is visibly highlighted with will be documented. Focus is on children's body movements that may serve as beats or protobeats.

3 Preliminary results and discussion

First analyses suggest that children employ other modalities apart from prosody as well to signal saliency. Even though children use prosody to enhance the "important" parts of the utterances, various body movements (e.g., shoulder shrugs, wrist and arm movements, hand claps) often coincide with those accentuated constituents of speech, as they frequently make use of their whole body to enhance the verbal content. Especially interesting is the fact that there seems to be a temporal and rhythmical alignment with the verbal content uttered and with the focused constituent of the utterance. This supports the hypothesis that they carry pragmatic properties and may be regarded as protobeats since they do not seem to carry referential properties. To record what forms this process takes during acquisition is the aim of the paper.

References

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15(6)*, 593–613.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and the point of view. In Subject and Topic.

Freigang, F., Klett, S., & Kopp, S. (2017). Pragmatic Multimodality: Effects of Nonverbal Cues of Focus and Certainty in a Virtual Human. In *17th International Conference on Intelligent Virtual Agents* (pp. 142–155). Stockholm.

Freigang, F., & Kopp, S. (2015). Analyzing the Modifying Functions of Gesture in Multimodal Utterance. *Gespin 4, (MaMuD 3)*, 107–112.

Höhle, B., Hörnig, R., Weskott, T., Knauf, S., & Krüger, A. (2014). Effects of focus and definiteness on children's word order: Evidence from German five-year-olds' reproductions of double object constructions. *Journal of Child Language, 41(4)*, 780–810.

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

Kiss, K. E. (1998). Identificational Focus versus Information Focus. *Language, 74(2)*, 245.

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

Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics, 44(4),* 817–852.

Pittner, K., & Berman, J. (2013). *Deutsche Syntax.* Tübingen: Narr Francke Attempto Verlag.

Rohlfing, K. J., Grimminger, A., & Lüke, C. (2017). An interactive view on the development of deictic pointing in infancy. *Frontiers in Psychology, 8*: 1319.

Sauermann, A., Höhle, B., Chen, A., & Järvikivi, J. (2011). Intonational Marking of Focus in Different Word Orders in German Children. *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 313–322.

Selting, M., & Auer, P. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion, 10(10)*, 353–402.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57,* 209–232.

9 Constructional potential of flat-hand – palm-lateral – away-body gestures: a crosslinguistic corpus-based study

Jakub Jehlička and Eva Lehečková
Charles University, Prague, Czech Republic

Abstract
Multimodal constructions are conventional pairings of meaning and a combination of speech and gesture. As is the case with unimodal constructions, some multimodal constructions are more constructionalized that others - exhibiting various degrees of entrenchment, formal variety and frequency. In our paper, we analyzed the constructional potential of a certain gesture family defined by flat handshape with lateral palm orientation and away-body movement. We did so by a qualitative and quantitative assessment of the distribution of the target gesture co-occurring with grammatical constructions in two languages - English and Czech. Our data come from the multimodal corpora of language interactions in naturalistic settings. We identified several functional clusters of multimodal constructions with the target gesture. Certain clusters were significantly more frequent in specific languages, suggesting typologically motivated tendencies of the same gesture families to occur in different types of constructions.

10 Investigating the coordination of patients' and therapists' conceptual phases in hand movements that accompany speech during psychotherapy sessions

Katharina Reinecke, Niklas Neumann and Hedda Lausberg
German Sport University Cologne, Germany

Abstract
Aim: The coordination of patients' and therapists' hand movement behaviour in psychotherapy, also referred to as nonverbal synchrony (NVS) is considered an indicator for positive therapeutic rapport and therapeutic success. Across research studies, time-based or matching-based NVS definitions are used. The present study aims to present the measurement of NVS with NEUROGES-ELAN, which considers both definitory aspects – behaviour matching and coordination in time.
Methods: 42 video recordings of first and next-to-last psychotherapysessions from the *Social Phobia Psychotherapy Research Network Project* were analysed. Using the *Liebowitz Social phobia Scale* (LSAS), patients were grouped according to improved symptoms (n = 10) and stagnated symptoms (*n* = 11). Two independent certified raters analysed hand movements with NEUROGES-ELAN. NVS between patients' and therapists' complex movement phases that accompany speech was assessed.
Results: In the stagnated symptoms group, synchronous complex phase movements had a longer duration and proportion of time than in the remitted symptoms group. In comparison to the beginning of psychotherapy, at the end of psychotherapy nonverbal synchrony frequency was higher for the stagnated symptoms group.
Conclusion: Results show, that this fine-grained movement analysis reveals increased synchrony durations and proportions of time in synchronous complex-phase movement for the stagnated symptoms group. Future studies referring to NVS should therefore deliberately choose their NVS operationalization and measurement parameters. Besides timing, they should consider behavioural matching as definitory aspect.

References

Altmann, U. (2011). Investigation of movement synchrony using windowed cross-lagged regression. *In Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues* (pp. 335-345). Springer, Berlin, Heidelberg.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behaviour as predictors of interpersonal consequences: Ameta-analysis. *Psychological bulletin, 111(2)*, 256.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. AmericanPsychiatric Pub.

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behaviour matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), *Studies in emotion & social interaction. Fundamentals of nonverbal behaviour* (pp. 401-432). New York: Cambridge University Press; Paris.

Bernieri, F. J., Davis, J. M., Rosenthal, R., & Knee, C. R. (1994). Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and social psychology bulletin, 20(3)*, 303-311.

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behaviour patterns. *Journal of nervous and mental disease, 143(4),* 338 -347.

Davis, M. (1982). *Interaction rhythms: Periodicity in communicative behaviour*. New York: Human Sciences Press.

Davis, M., & Hadiks, D. (1990). Nonverbal behaviour and client state changes during psychotherapy. *Journal of Clinical Psychology, 46(3)*, 340-351.

Davis, M., & Hadiks, D. (1994). Nonverbal aspects of therapist attunement. *Journal of Clinical Psychology, 50(3)*, 393-405.

Dvoretska, D., Denissen, J., & Lausberg, H. (2013). Kinesic Turn Taking and Mutual Understanding in Interactive Dyads*. Conference Proceedings, TiGeR 2013*, Tilburg.

Holle, H., & Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behaviour Research Methods, 47(3),* 837-847.

Koole, S. L., & Tschacher, W. (2016). Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance. *Frontiers in psychology, 7,* 862.

Kreyenbrink, I., Joraschky, P., Konstantinidis, I., Neumann, N., & 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(3)*, 297-313.

La France, M., & Broadbent, M. (1976). Group rapport: Posture sharing as a nonverbal indicator. *Group & Organization Studies, 1(3)*, 328-333.

Lausberg, H. (2013). *Understanding Body Movement: A Guide to Empirical Research on Non-verbal Behaviour With an Introduction to the NEUROGES Coding System*. Frankfurt/M: Peter Lang GmbH.

Lausberg, H. (2019). *Understanding Body Movement: A Guide to Empirical Research on Non-verbal Behaviour With an Introduction to the NEUROGES Coding System*. Frankfurt/M: Peter Lang GmbH.

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., & Sloetjes, H. (2009). Coding gestural behaviour with the NEUROGES-ELAN system. *Behaviour Research Methods, 41(3),* 841-849.

Lausberg, H., & Sloetjes, H. (2016). The revised NEUROGES–ELAN system: An objective and reliable interdisciplinary analysis tool for non-verbal behaviour and gesture. *Behaviour Research Methods, 48(3),* 973-993.

Leichsenring, F., Salzer, S., Beutel, M. E., von Consbruch, K., Herpertz, S., Hiller, W., ... & Konnopka, A. (2009). SOPHO-NET–Forschungsverbund zur Psychotherapie der sozialen Phobie. *PPmP-Psychotherapie·Psychosomatik. Medizinische Psychologie, 59(03/04*), 117-123.

Leichsenring, F., Salzer, S., Beutel, M. E., Herpertz, S., Hiller, W., Hoyer, J., ... & Ritter, V. (2013). Psychodynamic therapy and cognitive-behavioural therapy in social anxiety disorder: a multicenter randomized controlledtrial. *American Journal of Psychiatry, 170(7),* 759-767.

Luborsky, L., Barber, J. P., Siqueland, L., Johnson, S., Najavits, L. M., Frank, A., & Daley, D. (1996). The revised helping alliance questionnaire (HAq-II): psychometric properties. *The Journal of psychotherapy practice and research, 5(3)*, 260.

Neumann, N., Kreyenbrink, I., Reinecke, K. C. H., & Lausberg, H. Hand Rest Positions of patients with social phobia and therapists in psychodynamic psychotherapy sessions (SOPHO-NET project). *Journal of Multimodal Communication Studies, 4(1-2)*, 50.

Paulick, J., Deisenhofer, A. K., Ramseyer, F., Tschacher, W., Boyle, K., Rubel, J., & Lutz, W. (2018). Nonverbal Synchrony: A New Approach to Better Understand Psychotherapeutic Processes and Drop-Out. *Journal of Psychotherapy Integration, 28(3),* 367.

Paulick, J., Rubel, J. A., Deisenhofer, A. K., Schwartz, B., Thielemann, D., Altmann, U., ... & Lutz, W. (2018). Diagnostic Features of Nonverbal Synchrony in Psychotherapy: Comparing Depression and Anxiety. *Cognitive Therapy and Research, 42(5),* 539-551.

Ramseyer, F., & Tschacher, W. (2010). Nonverbal synchrony or random coincidence? How to tell the difference. *In Development of multimodal interfaces: Active listening and synchrony* (pp. 182-196). Springer Berlin Heidelberg.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of consulting and clinical psychology, 79(3),* 284.

Ramseyer, F., & Tschacher, W. (2014). Nonverbal synchrony of head-and body-movement in psychotherapy: different signals have different associations with outcome. *Frontiers in Psychology, 5,* 979.

Reinecke, K. C. H., Dvoretska, D., Joraschky, P. & Lausberg, H. (under review). Diagnostic potential of motor signs: Fidgeting signalizes co-morbid depression in patients with social phobia.

Reich, C. M., Berman, J. S., Dale, R., & Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *Journal of Social and Clinical Psychology, 33(5)*, 481-494.

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

Stein, M. B., & Stein, D. J. (2008). Social phobia. *The Lancet, 371(9618),* 1115-1125.

Tomicic, A., Martínez, C., Pérez, J. C., Hollenstein, T., Angulo, S., Gerstmann, A., ... & Krause, M. (2015). Discoursevoice regulatory strategies in the psychotherapeutic interaction: a state-space dynamics analysis. *Frontiers in psychology, 6*, 378.

Tschacher, W., Rees, G. M., & Ramseyer, F. (2014). Nonverbal synchrony and affect in dyadic interactions. *Frontiers in psychology, 5*, 1323.

Tschacher, W., & Pfammatter, M. (2016). Embodiment in psychotherapy—A necessary complement to the canon of common factors. *European Psychotherapy, 13,* 9-25.

Tschacher, W., Ramseyer, F., & Koole, S. L. (2018). Sharing the now in the social present: Duration of nonverbal synchrony is linked with personality. *Journal of personality, 86(2)*, 129-138.

Von Glischinski, M., Willutzki, U., Stangier, U., Hiller, W., Hoyer, J., Leibing, E., ... & Hirschfeld, G. (2018). Liebowitz Social phobiaScale (LSAS): Optimal cut points for remission.

11 The relation between individual differences in speech-gesture behaviour of 4-year-olds across three different experimental tasks

Olga Abramov[1], Ulrich Mertens[3], Anne Németh[2], Katharina Rohlfing[3], Friederike Kern[2], and Stefan Kopp[1]
[1]CITEC, Bielefeld University, Germany; [2]Bielefeld University, Germany; [3]Paderborn University, Germany

Abstract

Concerning the role of iconic gestures in language development, it is still not well understood whether they 'just' accompany or supplement language production and to which extent. Work by Alibali et al. (2009) suggests that children in the age of 5-10 produce more complementary gestures than adults. Apparently, child's speech-gesture system is less integrated than those of adults. We studied semantic coordination of speech and accompanying iconic gestures in a corpus of 36 4-year-olds recorded performing different experimental tasks (narrating, explaining and reporting). Since *iconic gestures* (McNeill, 1992) represent meanings that can be also verbalized, relating iconic gestures to speech allows us to have a closer look at the individual variation among children's use of speech and gesture. Related work in this area coded gestures as a whole (i.e., overlapping vs. non-overlapping with speech). However, iconic gestures often add more or less novel information to speech. That is, to be precise, we should be able to identify the degree of overlap between speech and gesture. This is what the present approach aimed at. More specifically, we asked what kinds (and how much) of meaning is represented in speech and gesture separately and redundantly? We coded semantic features (Bergmann and Kopp, 2006) for each utterance or/and gesture related to the event of the task. Comparing semantic features present in gesture and the related utterance, we were able to quantify the amount of information present in *speech-alone*, *gesture-alone* and *speech-gesture* (per age group and task). Three different experimental tasks were analysed. Our results reveal significant positive correlations between cognitive skills of the children and speech-gesture coordination. In addition, high cognitive skills correlate with speech-gesture coordination in some tasks and in others not (regarding the same children). Overall, speech-gesture integration of 4-year-olds is not the same for children with different cognitive skills and also for the same children fulfilling different tasks. The presented approach of assigning semantic features to speech and gesture allows for a more accurate view on speech-gesture integration quantifying the amount of information in each modality separately (and in the overlap).

References

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., and Mainela-Arnold, E. (2009). Gesture–speech integration in narrative: Are children less redundant than adults*? Gesture, 9(3)*, 290–311.

Bergmann, K. and Kopp, S. (2006). Verbal or Visual? How Information is Distributed across Speech and Gesture in Spatial Dialog. In Schlangen, D. and Fernandez, R. (Eds.): *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (SemDial-10)*, 90–97, Potsdam.

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

12 An Open Source Dataset of Human Gestures Through Human-Robot Interaction

Jan de Wit, Mirjam de Haas, Emiel Krahmer and Paul Vogt
Tilburg University, The Netherlands; [j.m.s.dewit,mirjam.dehaas,e.j.krahmer,p.a.vogt]@uvt.nl

Abstract
This work documents a new open-access dataset of gestures recorded and represented using Kinect. The gestures were acquired through a playful human-robot interaction, in the form of a game of charades, through which a humanoid robot is able to learn how to produce and recognize gestures by interacting with human participants.

1 Introduction
The use of gestures is not only evident in inter-human communication, but also occurs widely in, for instance, human-robot interaction (HRI). To facilitate such interactions, robots should be able to recognize gestures produced by humans and – ideally – learn to produce their own gestures from human's demonstration (Argall, Chernova, Veloso, & Browning, 2009). In order to develop such a robot, we collected a large data set of human-produced gestures for 35 different objects, gathered through a game of charades with a robot and recorded using a Kinect sensor. The resulting data set can be used in various ways, for instance to study how machine learning techniques can be developed to automatically recognize gestures, or to investigate how humans from different ages vary in the gestures they produced in this task.1

2 Data collection
Data was collected by inviting visitors of two public events to play a game of charades with a robot as follows: After completing a practice round, the robot started the game by performing a gesture from its set of examples, previously recorded from other participants. The tablet then displayed a picture of the item that the robot tried to enact, along with three distractors (Figure 1, left). If the participant guessed incorrectly, the robot performed a gesture for the same object once more for another guess. Next, the roles were reversed and the participant was shown an object on the screen, which (s)he then described using an upper-body gesture (Figure 1, centre). The robot tried to recognize the object that was portrayed, and if it guessed incorrectly the participant was asked to perform a gesture for the object again for a second attempt. Each game session lasted five rounds of the robot and participant taking turns guessing, covering ten objects – five performed by the robot, five by the participant – out of a total set of 35, which included animals, static objects (furniture, buildings), tools (e.g., cup, book, toothbrush), musical instruments, and vehicles. The gestures were recorded using a Microsoft Kinect sensor, yielding motion capturing videos (Figure 1, right). For the game, gesture recognition was implemented by extracting the *gist* of the gesture, inspired by the work of Cabrera and Wachs, 2017, and using a k-nearest neighbors approach to find the object that was most likely depicted by the current gesture (see Anonymous, 2019 for more details).

Figure 1: Left: The participant correctly guesses a gesture performed by the robot (*glasses*). Centre: The participant is performing a gesture for the robot to guess (*ball*). Right: Four examples of recorded gestures for *guitar* – first and second are by children, third and fourth by adults.

Data was collected at two locations: 1) NEMO Science museum, mostly visited by children and teenagers with their parents from different countries, and 2) Lowlands music festival where most visitors were adults. The system ran for fourteen days at NEMO, and for three days at Lowlands; all recorded data were cleared between the two events, so that the robot would have to start learning from scratch again. Table 1 shows the demographics and number of gestures gathered from each location.

Table 1: Description of datasets

	NEMO dataset	Lowlands dataset
Participants	311	116
Gender	157 Male	49 Male
(6 Unknown)	148 Female	67 Female
Average age (years)	12.9 (*SD* = 10.6)	28.3 (*SD* = 8.7)
Number of gestures	Approx. 2,500	998

The recorded gestures were stored in the form of a CSV file containing the 3D coordinates of the participant's tracked joint positions, sampled at approximately 30 frames per second from the Kinect sensor, as well as a movie file containing a 2D render of the gesture (Figure 1, right). Furthermore, gestures can be linked to participants and their demographic information by a unique identifier. During the games, the robot learned to recognize gestures with an average of around 23%, which was well above chance level (approximately 3%).

3 Conclusion

We collected an extensive and varied data set that we believe is a useful resource for future gesture research, with applications in the field of HRI, automatic gesture recognition and gesture coding. We intend to conduct further analyses on the recorded gestures (e.g., which strategies were used, whether these changed between first and second attempts, differences between participant groups), and aim to further improve the robot's ability to recognize and produce gestures. The data sets of gesture recordings, together with the demographic information, as well as the source code of the system will be made publicly available after our further analyses are complete.

References

Argall, B. D., Chernova, S., Veloso, M., & Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and autonomous systems, 57(5)*, 469-483.

Cabrera, M. E., & Wachs, J. P. (2017). A human-centered approach to one-shot gesture learning. *Frontiers in Robotics and AI, 4*, 8.