1ST INTERNATIONAL MULTIMODAL COMMUNICATION SYMPOSIUM

BOOK OF ABSTRACTS

April, 26-28 2023

Universitat Pompeu Fabra Barcelona

INDEPENDENT RESEARCH FUND DENMARK

mmsym.org

TABLE OF CONTENTS

WELCOME

We are delighted to welcome you to the *1st International Multimodal Communication Symposium*, MMSYM 2023, to be held at the Universitat Pompeu Fabra in Barcelona on April 26-29. The symposium aims to provide a multidisciplinary forum for researchers from different disciplines who study multimodality in human communication as well as in human-computer interaction. It is organised and supported financially by the GrEP Research Group (Prosodic Studies Group), from the Department of Translation and Language Sciences of the University of Pompeu Fabra, Barcelona, Catalonia, in conjunction with the research network on GEstures and Head Movements in Language (GEHM).

The symposium follows up on a tradition established by the Swedish Symposia on Multimodal Communication held from 1997 until 2000, and continued by the Nordic Symposia on Multimodal Communication held from 2003 to 2012. Since 2013 the event has acquired a broader European dimension, with editions held in Malta, Estonia, Ireland, Denmark, Germany and Belgium. This year it will be held in Catalonia for the first time and has a truly international ambition, hence the new name.

This year the call for papers focused on three **research themes** of particular interest to the GEHM network. The first is ***language-specific characteristics of gesture-speech interaction***, which seeks to account for how speakers' ability to process and produce gesture and speech is affected and changed by their language profile. The second is ***multimodal prominence***, which investigates the theoretical question of how linguistic prominence is expressed through combinations of kinematic and prosodic features. The third is ***conceptual and statistical modelling of multimodal contributions***, particularly head movements and the use of gaze. An outcome of the network that cuts across the three research areas, and which will be presented at the symposium, is an annotated corpus of online zoom meetings.

Aspects relating to all three themes will be discussed by the three keynote speakers who have accepted our invitation to share their important research results with us. In the first keynote, Prof. Jelena Krivokapić examines to what extent the temporal alignment between co-speech gesture and prosodic structure is linguistically driven, and presents three kinematic studies based on electromagnetic articulometry, motion capture and video data. One of the issues investigated is whether gesture-prosody coordination changes depending on the structural properties of a language. Therefore, the keynote relates to two of the GEHM network's themes: multimodal prosody and the language specificity of gesture-speech interaction. The talk by Prof. Catherine Pelachaud deals with how multimodal behaviour can be modelled to support the development of

Socially Interactive Agents, in other words software agents that can communicate with their human interlocutors using speech and gesture in complex and socially competent ways. Important aspects addressed in the talk include the role played in multimodal behaviour modelling by different methodologies, such as corpus analysis, user-centred analysis and motion capture, and the use of symbolic as well as deep learning approaches. In the last keynote, Prof. Alan Cienki discusses self-adaptors, a type of hand gesture that is often excluded from gesture studies, but which the use of new technology such as modern travel and web cams allows us to study in more detail. Cienki argues in his talk that self-adaptors may have pragmatic functions and that they can be used in mimicry behaviour, but also that they are relevant to thinking for speaking, and consequently relate to the first of GEHM's themes.

As for the MMSYM 2023 program itself, we are very happy for the enthusiastic response that the symposium has generated, a response that has allowed us to put together a program of very high quality with 30 oral presentations and 51 poster presentations. All the presentations have been carefully selected, being the top-most rated by three abstract reviewers. We would really like to thank the work of the reviewers and the program committee, together with the local organizing committee, in putting the program and the event together. We hope that through this exciting program we will be able to further strengthen the ties within groups in our research community, and that we wil be successful in creating a friendly and scientifically inspiring atmosphere to share our research.

Finally, thanks for attending the conference and coming to our vibrant city of Barcelona.

We hope that you have time to enjoy Catalan culture, food and traditions. You will be able to appreciate Catalan *ball de bastons* on the first day while you enjoy some Catalan wine. In the meantime, we wish you a profitable, collaborative and exciting stay in Barcelona for the MMSYM 2023.

Patrizia Paggio, Coordinator of the GEHM Research Network, Department of Nordic Studies and Linguistics, University of Copenhagen, and Institute of Linguistics and Language Technology, University of Malta.

Pilar Prieto, Coordinator of the Prosodic Studies and Gesture Group, Department of Translation and Language Sciences, ICREA-Universitat Pompeu Fabra.

A GEHM network initiative: the GEHM Zoom corpus collection

Patrizia Paggio, *University of Copenhagen, University of Malta*
Manex Aguirrezabal, *University of Copenhagen*
Bart Jongejan, *University of Copenhagen*
Costanza Navarretta, *University of Copenhagen*
Leo Vitasovic, *University of Copenhagen*

One of the outcomes of the research collaboration in the GEHM network presented at MMSYM 2023 is an annotated corpus of online Zoom meetings. The GEHM network, which is funded by the Danish Research Council, has the goal of fostering new theoretical insights into the way hand gestures and head movements interact with speech in face-to-face multimodal communication. It is a cooperation among leading research groups working in the area of gesture and language at a number of European universities and research bodies, i.e. Kiel University, KTH Royal Institute of Technology, KU Leuven, Linnaeus University, Lund University, Trinity College Dublin, Pompeu Fabra University, the University of Malta and the University of Copenhagen. The corpus of online Zoom meetings, which is a tangible product of this cooperation, will be made available to the research community to support studies of the way multimodal interaction works in video conferencing.

Due to the restrictions on social interaction imposed by the COVID-19 pandemic worldwide, and to the necessity of cutting CO2 emissions deriving from travelling, we have seen an increase in the use of video conferencing for group meetings, teaching, international conference organisation, etc. Empirical evidence of the way gesture and speech are used in online meetings, however, is scarce (Koh et al. 2022). Our multimodal corpus of Zoom meetings will contribute to fill out this gap. The corpus consists of 12 video recordings of meetings held on Zoom by a group of researchers in the context of a collaborative research project. The meetings have an average duration of about 40 minutes each, for a total of 8 hours. The language used is English. Participants are native and non-native speakers (5-8 per meeting), who all gave their informed consent that the annotated corpus would be made available for research purposes. The audio-visual recordings will be distributed together with an orthographic transcription as well as face and hand position coordinates.

To create the orthographic transcription, we performed an initial evaluation of a number of models from the Google speech-to-text API by measuring their error rate on a manually transcribed extract. We considered models for British and American English, given the fact that speakers in the meetings speak different varieties of English, as well as models trained on telephone and video interactions. We opted for the US video model, which produced the lowest word error rate (15.89%). Each speaker's speech output in each video was consequently transcribed using that model. The output was converted into the Praat TextGrid format (Boersma & Weenink 1992–2022), where the spoken contribution of each speaker is transcribed in a separate tier and time aligned with the video through time stamps before and after each word. Speakers' names were replaced by unique identifiers. The automatic transcription will be revised manually to correct errors.

Before extracting visual position coordinates, the video recordings had to be processed manually to create separate video files for each speaker. The space taken up by the individual speakers in the videos varies due to different numbers of participants in each meeting. It was decided to keep a constant size of 1920-1080 pixels in the extracted single videos. OpenPose (Cao et al. 2018) was then run on each video to extract position coordinates of nose, eyes, cheekbones, neck and hands. The positional coordinates (keypoints) were saved in JSON files, one file per video frame per speaker. The visual coordinate extraction process is quite demanding (it was run on high-performance NVIDIA GPUs with 40GB of VRAM). Therefore, we believe that making the results of this processing step available will be of great service to the community and ultimately avoid unnecessary energy consumption.

Acknowledgements: We would like to acknowledge the members of the GEHM network that provided input on the construction of the corpus and accepted to be recorded.

References

Boersma P. and Weenink, D. (1992–2022): Praat: doing phonetics by computer. Version 6.2.06, retrieved 23 January 2022 from https://www.praat.org.

Cao, Z., Hidalgo, G. Simon, T., Wei, S. and Sheikh, Y (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, https://10.48550/arkiv.1812.08008.

Koh, Jung In, et al. (2022) Show of Hands: Leveraging Hand Gestural Cues in Virtual Meetings for Intelligent Impromptu Polling Interactions. *27th International Conference on Intelligent User Interfaces*, pp. 292–309 https://doi.org/10.1145/3490099.3511153

COMMITTEES

Local committee

Pilar Prieto (coordinator, ICREA-Universitat Pompeu Fabra)

Florence Baills (Universitat Pompeu Fabra-Universität zu Köln)

Sara Coego (Universitat Pompeu Fabra)

Joel Espejo Álvarez (Universitat Pompeu Fabra)

Júlia Florit-Pons (Universitat Pompeu Fabra)

Celia Gorba (Universitat Pompeu Fabra)

Mariia Pronina (Universitat Pompeu Fabra)

Patrick L. Rohrer (Universitat Pompeu Fabra-Nantes Université)

Paula G. Sánchez-Ramón (Universitat Pompeu Fabra-Goethe-Universität Frankfurt)

Ïo Valls (Universitat Pompeu Fabra-Southern Denmark University)

Ingrid Vilà-Giménez (Universitat de Girona)

Xiaotong Xi (Universitat Pompeu Fabra)

Ting Yao (Universitat Pompeu Fabra)

Yuan Zhang (Universitat Pompeu Fabra)

Program committee

Patrizia Paggio (coordinator, University of Copenhaguen)

Jens Edlund (KTH Royal Institute of Technology, Stockholm)

Marianne Gullberg (Lund University)

David House (KTH Royal Institute of Technology, Stockholm)

Pilar Prieto (ICREA-Universitat Pompeu Fabra)

Maria Koutsombogera (Trinity College Dublin)

Carl Vogel (Trinity College Dublin)

Margaret Zellers (Kiel University)

Geert Brône (KU Leuven)

Celia Gorba (Universitat Pompeu Fabra)

Florence Baills (Universitat Pompeu Fabra- Universität zu Köln)

Patrick L. Rohrer (Universitat Pompeu Fabra-Nantes Université)

PROGRAM AT A GLANCE

	Wednesday, April 26th	Thursday, April 27th	Friday, April 28th
9h - 9h20		**Keynote Speech II** Catherine Pelachaud	**Oral Session** Multimodal annotation & corpus
9h20 - 9h40			
9h40 - 10h			
10h - 10h20		**Oral Session** Non-typical development	
10h20 - 10h40			
10h40 - 11h			**Coffee Break (40 min)**
11h -11h20		**Coffee Break (40 min)**	
11h20 - 11h40			**Oral Session** Pragmatics
11h40 - 12h		**Oral Session** Development	
12h - 12h20			
12h20 - 12h40			
12h40 - 13h			
13h - 13h20	**REGISTRATION** **(opens at 13h)**	**Lunch Break (1h20)**	**Lunch Break (1h20)**
13h20 - 13h40			
13h40 - 14h			
14h - 14h20			
14h20 - 14h40	**Opening Ceremony** (homage to Kendon)	**Oral Session** Gesture types & representation	**Keynote Speech III** Alan Cienki
14h40 - 15h	**Keynote Speech I** Jelena Krivokapić		
15h - 15h20			
15h20 - 15h40			**CLOSING**
15h40 - 16h	**Coffee Break (40 min)**	**Coffee Break +**	
16h - 16h20		**Poster Session**	
16h20 - 16h40	**Oral Session** Multimodal prominence & synchronization		
16h40 - 17h			
17h - 17h20			
17h20 - 17h40		**Oral Session** L2 & multimodal pedagogy	
17h40 - 18h			
18h - 18h20	**Coffee Break +**		
18h20 - 18h40	**Poster Session**		
18h40 - 19h			
19h - 19h20			
19h20 - 19h40			
19h40 - 20h	**WELCOME DRINKS**		
20h - 20h20		**CONFERENCE DINNER**	
20h20 - 20h40			
20h40 - 21h			

INVITED SPEAKERS

Jelena Krivocapić

Associate Professor in Linguistics at University of Michigan

Caherine Pelachaud

CNRS Research director; Institute of Intelligent Systems and Robotics Sorbonne Université

Alan Cienki

Full professor of Language Use & Cognition and English Linguistics at Vrije Universiteit Amsterdam

Prosodic structure and speech planning in speech and manual gestures

Jelena Krivokapić, *University of Michigan, Haskins Laboratories*

Co-speech gestures and prosodic structure are often temporally aligned, but the precise details of this alignment are not known. The mechanisms by which it arises are also unclear. Specifically, it is not understood if the alignment is the result of general coupling principles or if it is—at least to some extent—linguistically driven.

I address these questions from different angles in three kinematic studies using electromagnetic articulometry, motion capture, and video data. The first two studies examine what the landmarks of coordination are for both speech and co-speech gestures, and if this coordination differs depending on the structural properties of a language. A third study examines the effect of speech planning on speech and co-speech gesture coordination. The results of the studies are discussed from the point of view of how co-speech gestures are recruited in the process of expressing prosodic structure and how the temporal coordination between speech and co-speech gesture arises in the process of language production.

Keywords: prosodic structure; speech planning; manual gestures

Socially Interactive Agents, their communicative behaviours and their adaptation mechanisms

Catherine Pelachaud, *CNRS-ISIR, Sorbonne University*

Our aim is to develop Socially Interactive Agents SIAs able to communicate verbally and nonverbally with their human interlocutors. To this aim, we have conducted research along two main research directions: 1) develop richer models of multimodal behaviours for the agent; 2) make the agent a more socially competent interlocutor.

We have conducted various studies to simulate communicative behaviours, social attitudes and behavioural expressiveness. We have worked to enrich the palette of multimodal behaviours of SIAs by applying different methodologies, based on corpus analysis, user-centred analysis, motion capture and by proposing symbolic and deep learning approaches. One particular focus was on metaphoric gestures that are linked to the expression of abstract concepts. Based on work in embodied cognition, we have proposed an approach using the image schema representation (Ravenet et al., 2018) to capture the underlying physical action carried by the gestures. First, the semantic information is extracted from the analysis of the verbal content. It is then characterized in terms of image schemas which are, in turn, represented by means of gesture primitives. The last step relies on the notion of Ideational Units introduced by Geneviève Calbris (1991) to compute how successive communicative gestures evolve in shape and time. Lately, we turned our attention on capturing the expressive behaviour style of interlocutors (Fares et al., 2023); everyone having their own gesturing style. We have developed a generative model to capture it. It allows us to perform zero-shot multimodal style transfer on new speakers without requiring any further training.

During an interaction, we adapt our behaviours at several levels: we align our ways of speaking (vocabulary, syntax, level of formality), but also our behaviours (we respond to the smile of our interlocutor, we imitate the posture, the gestural expressiveness...), our conversational strategies (to be perceived as warmer or more competent), etc. This multilevel adaptation can have several functions: to reinforce engagement in the interaction, to emphasize our relationship with others, to show empathy, to manage the impression we give to others.... We have proposed several models of adaptation working on different levels such as the level of conversational strategies or of multimodal behaviours (Biancardi et al., 2021). To capture the reciprocal adaptation between partners in an interaction, we have developed a recurrent neural network with attention mechanisms. We validated our model by conducting human-agent perceptual studies.

Keywords: Socially Interactive Agents; communicative gesture; adaptation

References

Biancardi, B., Dermouche, S., & Pelachaud, C. (2021). Adaptation Mechanisms in Human–Agent Interaction: Effects on User's Impressions and Engagement. *Frontiers in Computer Science*, 3, 696682.

Calbris, G. (2011). *Elements of Meaning in Gesture*, 1-398. John Benjamins Publishing Company - Gesture Studies - ISBN: 9789027285171

Fares, M., Pelachaud, C., & Obin, N. (2023, January). Zero-Shot Style Transfer for Multimodal Data-Driven Gesture Synthesis. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition* (FG) (pp. 1-4). IEEE.

Ravenet, B., Pelachaud, C., Clavel, C., & Marsella, S. (2018). Automating the production of communicative gestures in embodied characters. *Frontiers in psychology*, 9, 1144.

Woo, J., Pelachaud, C., & Achard, C. (2023, March). ASAP: Endowing Adaptation Capability to Agent in Human-Agent Interaction, in *2023 ACM 28th Annual Conference on Intelligent User Interfaces* (IUI). ACM.

Self-adapters: Renewed interest in an often ignored category

Alan Cienki, *Vrije Universiteit Amsterdam*

For about the past 30 years of research in gesture studies, self-adapters have largely been excluded as an object of research (with a few notable exceptions, e.g. the NEUROGES system [Lausberg, 2013] and the guide for "Annotating multichannel discourse" [Kibrik & Fedorova, 2020]). The term "self-adapters+" will be used here as a cover label for phenomena that in other research have variously been called self-adapters (British spelling: self-adaptors), postures, poses, positions, fidgeting, or idiosyncratic movements (e.g. Żywiczyński et al., 2017).

Several 'data contexts' in which the technology allows the use of self-adapters+ to be seen more prominently have prompted looking at this topic more closely (in both a literal and figurative sense). One is a project on simultaneous interpreters' gesture involving the use of a GoPro camera on the table in front of the interpreters; the closeness to the hands has afforded unanticipated new forms of data. A second project concerns interaction between tutors and students via video calls on Microsoft Teams (PhD research by Paloma Opazo Reyes, VU Amsterdam & KU Leuven); the close-up yet constrained view afforded by the webcams, focusing on the face and shoulder area, again brought self-adapters+ to the fore. A third one is the increasingly frequent use of video from television (via YouTube, etc.) for gesture analysis, e.g. the UCLA Library Broadcast NewsScape analyzed via tools from the Distributed Little Red Hen Lab.

The contexts of talk in these recording have special characteristics in that they involve some restraint on the part of the speaker. Many of the interpreters in our study had been trained not to move too much while interpreting, so that they would be less conspicuous. The tutors and students are in a 'problem-solving' setting, discussing the writing in the students' essays. In the TV setting, the genres of the news, interviews, or talk shows involve varying degrees of restraint, both physical (if seated at a desk) and cultural (if more formal in nature).

Several phenomena have come to the fore in our research on self-adapters+ that give rise to some new research questions. One concerns the category of self-adapters+ in relation to what are traditionally considered other parts of gesture units or other gesture functions. For example, in our data there are various types of post-stroke holds that become self-adapters+. In addition, what many researchers might consider to be 'rest positions' may involve hand-internal dynamicity, raising questions about what constitutes 'rest'. Another issue concerns the many instances of pragmatic gestures being embedded between self-adapters+ or within a self-adapter+. Small extensions outward of fingers or hands in the data often constitute beats serving pragmatic functions of emphasis or presenting an idea, as if miniature versions of the palm-up open-hand (Cienki, 2021). Another phenomenon concerns the frequency with which mimicry (alignment, mirroring, etc.) occurs with self-adapters+ in the context of the video calls. These are being studied in relation to speaker roles (tutor vs. student) and participants' subsequently evaluation of rapport (or lack of it) in the interaction.

In conclusion, it could be worth paying more attention to self-adapters+ not only because of the specific questions raised above, but also in order to investigate broader topics. One of these is how speakers manage a heavy cognitive load or stressful situations (e.g. Densing et al., 2018). Another is how self-adapters+ relate to thinking for speaking (à la Slobin, 1987). Most gesture research on this topic has concentrated on representational gestures, but the research on the interpreters, for example, shows frequent use of self-adapters during the resolution of disfluencies in speech. Additionally, the study of self-adapters+ can yield new insights in research on alignment (mimicry, etc.) in interaction.

In general, looking more closely at self-adapters+ takes us back to some of the origins of modern gesture studies in the 1960s and 70s, e.g., in the analyses of movement by Freedman, Kendon, and Sheflen. While currently available technologies for motion tracking clearly allow for new kinds of analysis of micro-movements, the present studies illustrate how even technologies as simple as GoPro cameras and webcams for video conferencing allow for renewed investigation of self-adapters+ within observational research.

Keywords: self-adapters; interpreting; video conferencing

References

Cienki, A. (2021). From the finger lift to the palm-up open hand when presenting a point: A methodological exploration of forms and functions. *Languages and Modalities*, *1*, 17–30. https://doi.org/10.3897/lamo.1.68914

Densing, K., Konstantinidis, H., & Seiler, M. (2018). Effect of stress level on different forms of self-touch in pre- and post-adolescent girls. *Journal of Motor Behavior, 50*(5), 475–485.

Kibrik, A. A., & Fedorova, O. V. (2020). *Annotating multichannel discourse: A guide book.* RAS Institute of Linguistics.

Lausberg, H. (2013). *Understanding body movement: A guide to empirical research on nonverbal behavior. With an introduction to the NEUROGES coding system.* Peter Lang.

Slobin, D. I. (1987). Thinking for Speaking. *Annual Meeting of the Berkeley Linguistics Society*, *13*, 435-445. https://doi.org/10.3765/bls.v13i0.1826

Żywiczyński, P., Wacewicz, S., & Orzechowski, S. (2017). Adaptors and the turn-taking mechanism: The distribution of adaptors relative to turn borders in dyadic conversation. *Interaction Studies*, *18*(2), 276-298.

ORAL SESSIONS

(In order of appearence in the program)

	Topic	Chair
Day 1	**Multimodal Prominence & Synchronization**	David House
Day 2	**Non-typical development**	Júlia Florit-Pons
	Development	Maria Graziano
	Gesture Types & Representation	Margaret Zellers
	L2 & Multimodal Pedagogy	Geert Brône
Day 3	**Multimodal Annotation & Corpus**	Ada Ren-Mitchell
	Pragmatics	Maria Koutsombogera

DAY 1. SESSION "MULTIMODAL PROMINENCE & SYNCHRONIZATION"

When the beat drops – beat gestures recalibrate lexical stress perception

Ronny Bujok, David Peeters, Antje Meyer & Hans Rutger Bosker

Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement

Lara Pearson & Wim Pouw

Multimodal signatures of prosodic prominence in habitual and loud speech

Lena Pagel, Simon Roessig, Márton Sóskuthy & Doris Muecke

Contextual influences on multimodal alignment in Zoom interaction

Sho Akamine, Mark Dingemanse, Antje Meyer & Aslı Özyürek

Not only prominence, but also phrasal prosodic structure guide gesture-speech alignment patterns

Patrick L. Rohrer, Elisabeth Delais-Roussarie & Pilar Prieto

When the beat drops – beat gestures recalibrate lexical stress perception

Ronny Bujok, *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
David Peeters, *TiCC Tilburg University, Tilburg, The Netherlands*
Antje Meyer, *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
Hans Rutger Bosker, *Donders Institute, Radboud University, Nijmegen, The Netherlands*

Speech is highly variable and noisy. Listeners may therefore use the visual modality to disambiguate ambiguous speech sounds. For instance, when repeatedly presented with an ambiguous sound /a?a/ midway between /aba/ and /ada/, paired with a video of a talker producing either /aba/ or /ada/, listeners adjust their perception of the ambiguous sounds based on the visual cues (Bertelson et al., 2003). That is, after this audiovisual exposure, listeners were biased to perceive an audio-only /aba/-to-/ada/-continuum as /aba/, if /a?a/ had been paired with a video of a talker producing /aba/. Conversely, when paired with an /ada/ video, they were more likely to perceive /a?a/ as /ada/. This effect is called *recalibration*, with listeners adjusting their perceptual categories based on the visual context with lasting consequences for later audio-only perception.

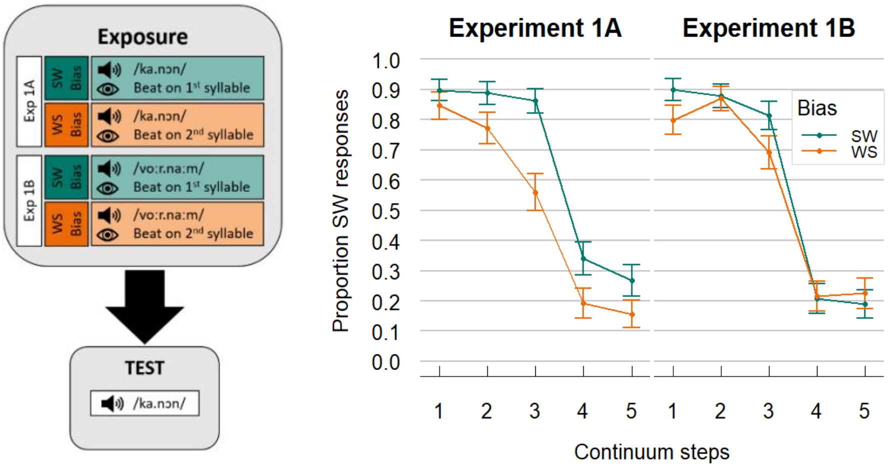
Here we tested whether manual beat gestures can also recalibrate listeners' perceptual categories, specifically focusing on lexical stress. Beat gestures, which usually align to stressed syllables, can influence lexical stress perception immediately (Bujok et al., 2022). That is, participants are more likely to perceive lexical stress on the syllable indicated by the beat gesture. However, this study tested whether beat gestures can have a more long-lasting effect and can recalibrate perception of lexical stress in subsequently presented audio-only stimuli.

In Experiment 1A, we tested 36 participants using a recalibration paradigm, including an (audiovisual) exposure and (audio-only) test phase. In the exposure phase, participants were repeatedly presented with an ambiguous token /ka.nɔn/, midway between Dutch *CAnon* [strong-weak (SW); "canon"] and *kaNON* [weak-strong (WS); "cannon"], disambiguated by a beat gesture either aligned to the first (SW-bias group) or second syllable (WS-bias group). The SW-bias group was thus expected to learn that the ambiguous stress cues on /ka.nɔn/ indicated initial stress, while the WS-bias group learned that the same ambiguous cues indicated final stress. In a subsequent audio-only test phase, participants were asked to categorize five ambiguous tokens from a *CAnon – kaNON* continuum (manipulating F0) as either SW or WS. Results from GLMM models indicated that participants from the SW-bias group gave more SW responses in the test phase, while participants from the WS-bias group gave more WS responses (Figure 1). In Experiment 1B, we tested if another 36 participants could generalize this effect to different words. Participants were exposed to a different item in exposure (*VOORnaam* [SW; “first name"] vs. *voorNAAM* [WS; “respectable”]; /voːr.naːm/), while tested on the same *CAnon-kaNON* continuum. Here, the group effect was not significant.

This study demonstrates a *multimodal recalibration effect* in lexical stress perception, showing an effect of beat gestures beyond immediate perception: they have a lasting impact even on later audio-only perception. Ongoing experiments investigate the generalization of gesture-driven recalibration. To conclude, we propose that listeners use the timing of seemingly meaningless hand gestures to adapt to suprasegmental variability in speech.

Keywords: beat gesture, lexical stress, recalibration, prosody, speech perception

Figure 1. *Design and results. In the exposure phase, the SW-bias group (beat on 1st syllable) learned that the ambiguous /ka.nɔn/ (Experiment 1A) or /voːr.naːm/ (Experiment 1B) indicated initial stress. The WS-bias group (beat on 2nd syllable) learned the same auditory token indicated final stress. In the subsequent audio-only test phase in Experiment 1A, the SW-bias group perceived a CAnon-kaNON continuum as more CAnon-like, while the WS-bias group perceived it as as more kaNON-like. However, this group effect was not significant in Experiment 1B, suggesting limited generalization to new words.*



References

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect. *Psychological Science*, *14*(6), 592–597. https://doi.org/10.1046/j.0956-7976.2003.psci_1470.x

Bujok, R., Meyer, A., & Bosker, H. R. (2022). *Audiovisual Perception of Lexical Stress: Beat Gestures are stronger Visual Cues for Lexical Stress than visible Articulatory Cues on the Face* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/y9jck

Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement

Lara Pearson, *Max Planck Institute for Empirical Aesthetics*, *Frankfurt am Main*
Wim Pouw, *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen*

Across a wide range of musical styles worldwide, vocalists tend to gesture manually while they sing. In existing research, such co-singing gesturing has been analyzed with regard to communication, expressivity, iconicity and effort (e.g., Davidson, 2001; Clayton, 2007). However, fundamental questions remain unanswered regarding coupling between gesture and sound—namely, what features of vocal sound and gesture kinematics are most closely coupled, and in what way. In this study, we address these questions for the insight provided into why performers gesture as they do. This is explored in the context of a South Indian musical practice, Karnatak music, where vocalists tend to gesture spontaneously while singing (Pearson, 2016). The study's theoretical background lies in work on gesture-speech coupling, showing that peaks in kinematic features such as speed and acceleration tend to align with emphasis in speech (Wagner et al., 2014) and that vocal aspects of speech are affected by gesture accelerations producing forces onto the respiratory–vocal system (Pouw et al., 2020). It is an open question however what aspects of gesture and vocalization couple in music making.

Motion capture (Xsens MVN-Link) and audio-visual recordings were made of 4 Karnatak vocalists singing *ālāpana* (extemporisation) across 8 different *rāgas* (melodic types), producing a total of 35 recordings. Acoustic measurements, including pitch (F0) and smoothed amplitude envelope (ENV), were extracted. We located kinematic peaks in motion-tracking measurements using a peak-finding algorithm, and performed generalized additive modelling (GAM), a type of non-linear mixed regression, to assess whether the kinematic measures reliably coupled with the acoustic measures around peaks in kinematics. To determine the optimal kinematic variables that predicted acoustic fluctuations, we assessed through GAM model comparisons whether vertical velocity (Δz), speed, or acceleration peaks were more strongly temporally predictive for changes in vocal acoustics (ΔF0, ΔENV). Using mixed-linear regression, we further assessed whether the magnitude of kinematic peaks predicted the magnitude of change in vocal acoustics.

Acceleration was the most predictive model for ΔF0, showing clear non-linear scaling relations such that higher deceleration (negative acceleration) or acceleration (positive acceleration) predicted higher ΔF0 peaks. We observe that acceleration had the most reliable magnitude coupling with vocal acoustics, showing a 1/3 power relation whereby each doubling of acceleration magnitude co-occurred with about 33% increase in vocal acoustics change.

An interesting implication of acceleration being maximally predictive for F0 in this analysis, particularly in the case of magnitude coupling, is that acceleration and deceleration peaks mark moments in movement where forces are produced onto the body, which is not the case for peaks in speed or vertical velocity. This suggests that such force production might be particularly salient in gesture-vocal coupling in this style. These findings are in line with theories that highlight the physical connection between gesturing and vocal production (e.g., Pouw et al., 2020).

Music is fundamentally multimodal and embodied: to produce sound we must move, and qualities of movement that create a sound can be perceived from the resulting sound (Vanderveer, 1979). If movement is viewed as a dimension of musical meaning, then gesture and sound in this style can be understood as parts of a single aesthetic entanglement, exhibiting an array of cross-domain mappings. The implications of this study are that mappings between force and acoustic change could be particularly relevant for further inquiry into co-singing gesturing.

Keywords: cross-modality; gesture–speech physics; vocal music

References

Clayton, M. (2007). Time, Gesture and Attention in a Khyāl Performance. *Asian Music*, *38*(2), 71-96.

Davidson, J. W. (2001). The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. *Musicae Scientiae*, *5*(2), 235-256.

Pearson, L. (2016). *Gesture in Karnatak Music: Pedagogy and Musical Structure in South India.* PhD. Durham University, Durham.

Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*, *117*(21), 11364-11367.

Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds*. Dissertation Abstracts International. University Microfilms No. 8004002.

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

Multimodal signatures of prosodic prominence in habitual and loud speech

Lena Pagel, *University of Cologne*

Simon Roessig, *Cornell University*

Márton Sóskuthy, *University of British Columbia*

Doris Mücke, *University of Cologne*

Prosodically prominent entities are temporally coordinated with events in the stream of co-speech gestures, such as head motion and manual gestures (Ambrazaitis & House, 2017; Krivokapić et al., 2017). We investigate signatures of prosodic prominence in terms of head movements in habitual and loud speech. Former research has shown that speaking loudly is associated with a decreased differentiation of prominence degrees in terms of F0 (Roessig et al., 2022) but maintained or even enhanced contrasts in lip and tongue kinematics (Pagel et al., 2022). However, it has not yet been studied if prominence-related modulations can be observed in co-speech movements in loud speech.

We recorded the acoustic and kinematic signals of 20 German speakers with 3D Electromagnetic Articulography, capturing movements of the articulators and the head. Participants interacted with a virtual avatar and produced embedded target words, which were either in *broad* or in *corrective* focus. These two focus types elicit two degrees of prominence, since words in corrective focus are more prominent than in broad focus (Breen et al., 2010). Utterances were produced in *habitual* and *loud* speech. In a two-step analysis, 3D head motion was first assessed by fitting GAMMs to the movement trajectories of three sensors on the head (nose, right and left ear), as illustrated in fig. 1. In a second step, the nose sensor, which showed the strongest movement, was analysed three-dimensionally with respect to displacement and velocity in a time window including the target word and surrounding 100 ms. Displacement was examined using LMMs, velocity using GAMMs.

The results show that 3D movement displacements are overall greater in loud than in habitual speech (cf. fig. 2, comparison between speaking styles). However, in both speaking styles (though especially in loud speech), head movements are larger when accompanying more prominent entities (cf. fig. 2, comparison between focus types). In terms of 3D velocity, head movements are overall faster in loud than in habitual speech (cf. fig. 3, comparison between speaking styles). Nevertheless, prominence degrees are differentiated in both speaking styles: in habitual by the shape of the velocity profile, in loud by its height (cf. fig. 3, comparison between focus types).

In summary, the data suggest that prosodic prominence degrees are reflected by co-speech head movements *across speaking styles* – movements are larger and exhibit a different velocity profile in more prominent positions. These spatiotemporal modulations reflect a gradient increase of biomechanical effort (Nelson, 1983) as a concomitant of speaking style and prosodic prominence. This underlines the robustness of the multimodal system in the marking of prominence under varying communicative demands.

Keywords: prosodic prominence; co-speech head movements; focus marking; loud speech

Figures

Figure 1. *Visualisation of 3D head motion as a first analysis step, based on nose and ears.*

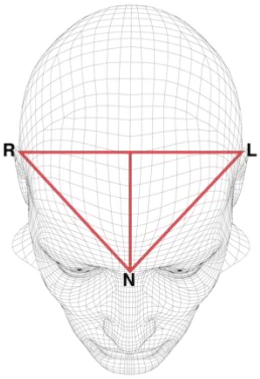


Figure 2. *3D displacement of head movements between speaking styles and focus types.*

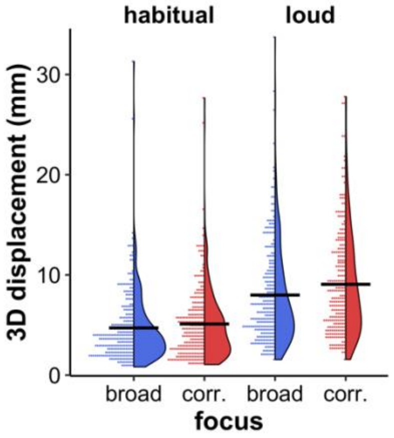
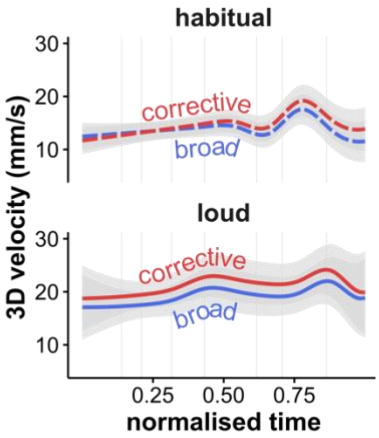


Figure 3. *GAMMs for 3D velocity of head movements between speaking styles and focus types.*



References

Ambrazaitis, G., & House, D. (2017). Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication*, *95*, 100–113. https://doi.org/10.1016/j.specom.2017.08.008

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology*, *8*(1), 1–26. https://doi.org/10.5334/labphon.75

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, *46*, 135–147. https://doi.org/10.1007/BF00339982

Pagel, L., Roessig, S., & Mücke, D. (2022). How to achieve a prominence GOAL! in different speaking styles. In C. Gianollo et al. (Eds.), *Paths through meaning and form. Festschrift offered to Klaus von Heusinger on the occasion of his 60th birthday* (pp. 200–204). Universitäts- und Stadtbibliothek Köln. https://doi.org/10.18716/omp.3.c52

Roessig, S., Pagel, L., & Mücke, D. (2022). Speaking loudly reduces flexibility and variability in the prosodic marking of focus types. *Proceedings of Speech Prosody, 23-26 May, Lisbon, Portugal*. https://doi.org/10.21437/SpeechProsody.2022-102

Contextual influences on multimodal alignment in Zoom interaction

Sho Akamine[1], *Max Planck Institute for Psycholinguistics*

Antje Meyer[1], *Max Planck Institute for Psycholinguistics*

Mark Dingemanse[2], *Center for Language Studies, Radboud University*

Aslı Özyürek, *Max Planck Institute for Psycholinguistics*

In daily conversation, people repeat or mimic each other's communicative behavior, such as words and gestures. This cross-participant repetition of communicative behavior is called *alignment*, which most frequently occurs multimodally (Rasenberg et al., 2022). Investigating the mechanisms of alignment is crucial in understanding interactive language use, as alignment pervades interactive communication (Dideriksen et al., 2020).

Theoretical approaches to studying alignment can be classified into two major perspectives: *priming* and *grounding* (Rasenberg et al., 2020). Priming accounts argue that alignment at each level of linguistic representation can occur between two interlocutors and within the speaker, which is driven by automatic priming mechanisms (e.g., Pickering & Garrod, 2004). In contrast, grounding accounts propose that speakers strategically coordinate their behavior to establish common ground and mutual understanding (e.g., Holler & Wilkin, 2011). Here, we explore how these two processes interact and influence speech and gestural alignment by varying the context of interaction (i.e., visibility of two interlocutors on Zoom).

Experiment

Design: Manipulating visibility on Zoom creates three conditions. In the SymAV condition, speakers can see each other's gestures. In the AsymAV condition, speaker A is visible to speaker B but not vice versa. In the AO condition, speakers cannot see each other (Figure 1).

Procedure: Each Dutch-speaking dyad will perform a referential communication task on Zoom. In each trial, participants will be a director or matcher. The director will describe the target Fribbles indicated by a red square, and the matcher will identify them through free interaction. The entire session will be video-recorded via a video camera on each person. Further, before and after the communicative referential task, they will give names to each Fribble (naming task) and rate them based on conceptual features (e.g., round; features task).

Coding: The speech and co-speech gestures will be annotated in ELAN. We operationalize lexical alignment as content words that refer to the same Fribble subparts. For co-speech gestures, iconic gestures referring to the same subparts will be coded as gestural alignment.

Analysis: A (generalized) linear mixed-effects model (FE: conditions; RE: participants and items) will be performed on alignment rate and naming/feature similarity scores.

Predictions: We predict the gestural alignment rate to be lower in the AsymAV condition than in the SymAV condition. This is because Speaker B in the AsymAV condition should not align in gesture strategically (grounding) as they are aware that their partner cannot see their gesture. Also, Speaker A in the AsymAV condition should not be primed to align in gesture as A will not see B's gestures. Further, speakers in the AO condition should show less gestural alignment than in the other conditions due to no visual cues available (Figure 2). Finally, if visual cues help participants converge conceptually, naming and feature similarity scores should be highest for the SymAV condition, followed by AsymAV and AO conditions.

In summary, we aim to investigate the mechanisms of multimodal alignment. As alignment is an important aspect of communicative interaction, this study would be another step toward revealing the mechanisms underpinning all aspects of language use.

Keywords: alignment, multimodality, Zoom interaction

Figures

Figure 1. *Experiment setup across three conditions*

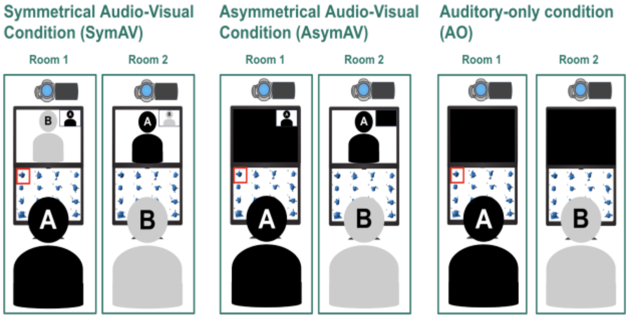
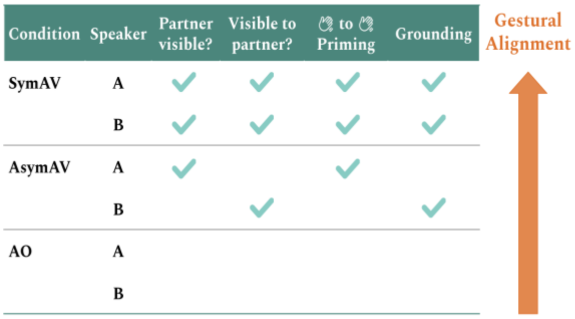

Symmetrical Audio-Visual Condition (SymAV)
Asymmetrical Audio-Visual Condition (AsymAV)
Auditory-only condition (AO)
Room 1
Room 2
Room 1
Room 2
Room 1
Room 2
A
B
A
B
A
B

Figure 2. *Predictions for gestural alignment*



Condition	Speaker	Partner visible?	Visible to partner?	✋ to ✋ Priming	Grounding
SymAV	A	✓	✓	✓	✓
	B	✓	✓	✓	✓
AsymAV	A	✓		✓	
	B		✓		✓
AO	A				
	B				

Gestural Alignment

References

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2020). *Quantifying the interplay of conversational devices in building mutual understanding*. PsyArXiv. https://doi.org/10.31234/osf.io/a5r74

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

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Rasenberg, M., Özyürek, A., Bögels, S., & Dingemanse, M. (2022). The primacy of multimodal alignment in converging on shared symbols for novel referents. *Discourse Processes*, *59*(3), 209–236. https://doi.org/10.1080/0163853X.2021.1992235

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/10.1111/cogs.12911

Not only prominence, but also phrasal prosodic structure guide gesture-speech alignment patterns

Patrick Louis Rohrer, *GrEP, Universitat Pompeu Fabra, Barcelona & LLING - UMR6310, Nantes Université, France*

Elisabeth Delais-Roussarie, *LLING - UMR6310, Nantes Université, France*

Pilar Prieto, *ICREA & GrEP, Universitat Pompeu Fabra, Barcelona*

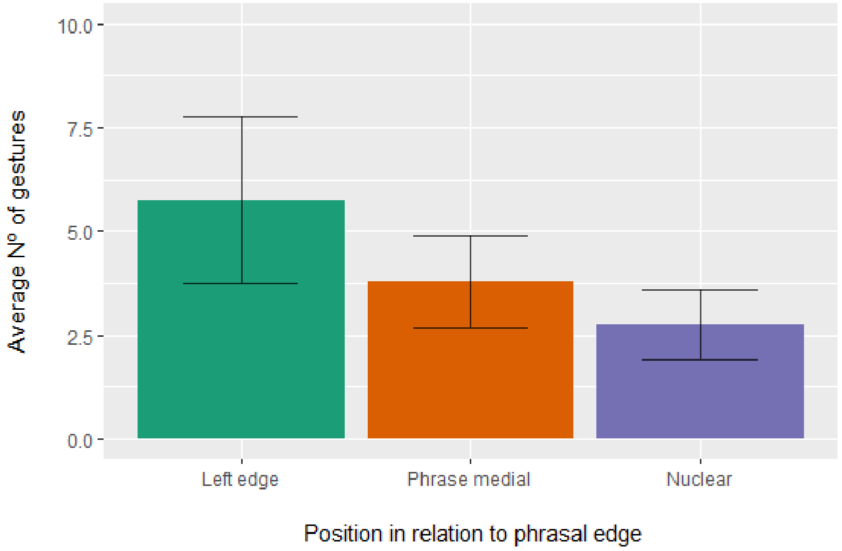
Gesture and prosodic prominence are closely temporally coordinated (see Shattuck-Hufnagel & Ren, 2018 for a recent review). Research has shown a tight temporal relationship between prominence-lending tonal movements (i.e., pitch accentuation) and prominence in gesture, that is, gesture strokes and apexes (i.e., the interval or point in time respectively in which the peak of effort in the gesture occurs, see Kendon, 1980; Loehr, 2012). However, prosodic structure consists of not only prosodic heads (e.g., pitch accentuation) but also of prosodic edges (loosely understood here as initial and final positions within a prosodic phrase). Initial evidence has suggested that prosodic phrasing indeed plays a role in the temporal execution of gesture, namely in that gestures tend to begin in coordination with intermediate phrase onsets (Loehr, 2012), gesture strokes tend to lengthen under prosodic prominence (Krivokapić et al., 2017), and and that both apex and pitch accent peaks are co-produced earlier in the stressed syllable when the syllable is followed by a boundary tone (Esteve-Gibert & Prieto, 2013). However, to our knowledge, no previous studies have assessed the value of prosodic edges (in terms of nuclear vs. early prenuclear pitch accentuation) in the attraction of manual gestures while at the same time controlling for the relative degree of prominence associated with the pitch accents in an independent manner. The current study adds to our knowledge of how gestures temporally associate with speech by assessing the following three questions, namely (a) whether the strokes and apexes of manual gestures associate with pitch-accented syllables; (b) whether gesture strokes align more with nuclear than prenuclear pitch accents at the intermediate phrase level; and (c) whether this relationship is driven by prominence relations or by phrasal position.

A prosodic and gestural analysis of the English M3D-TED corpus was carried out (Rohrer et al., 2021), which contains a total of 5 academic lectures with over 23 minutes of multimodal speech. Results revealed that while the majority of strokes of manual gestures (85.99%) overlapped a pitch-accented syllable, similar to rates that have been reported before, apex alignment was shown to occur at relatively lower rates (50.4%). Crucially, our results showed that at the phrasal level, strokes tend to align with phrase-initial prenuclear pitch accents over phrase-medial or nuclear accents, and this relationship is not driven by prominence relations between the pitch accents. All in all, these findings show that not only prosodic heads, but also prosodic edges (referring to the first prenuclear pitch accent), act as strong attractors of manual gestures, and that future research about gesture-speech temporal association should take this modulating factor into account.

Keywords: Gesture-speech synchrony; Multimodal prominence; Phrasal prosodic structure

Figures

Figure 1. *Gesture association as a function of phrasal position of the pitch accent.*



References

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864. https://doi.org/10.1044/1092-4388(2012/12-0049)

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). Mouton.

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology*, *8*(1), 3. https://doi.org/10.5334/labphon.75

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1), 71–89. https://doi.org/10.1515/lp-2012-0006

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, A., Shattuck-Hufnagel, S., & Prieto, P. (2021, February 24). The MultiModal MultiDimensional (M3D) labeling system. https://doi.org/10.17605/osf.io/ankdx

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01514

DAY 2. SESSION "NON-TYPICAL DEVELOPMENT"

Eyetracking the children's use of gesture and prosody to infer meanings when linguistic abilities are impaired

Albert Giberga, Alfonso Igualada, Nadia Ahufinger, Mari Aguilera, Ernesto Guerra & Núria Esteve-Gibert

Gesture use in late talking and typically developing children during picture naming

Caterina Verganti, Nathalie Frey, Alessandra Sansavini & Carina Lüke

Playing a guessing game to study multimodal syntactic complexity in children with and without Developmental Language Disorder

Corrado Bellifemine

Eyetracking the children's use of gesture and prosody to infer meanings when linguistic abilities are impaired

Albert Giberga[1], Alfonso Igualada[1], Nadia Ahufinger[1], Mari Aguilera[2], Ernesto Guerra[3], Núria Esteve-Gibert[1]

[1]*Universitat Oberta de Catalunya*

[2]*Universitat de Barcelona*

[3]*Universidad de Chile*

Linguistic prosody and body gestures help children accessing pragmatic and discursive meanings (e.g., Armstrong et al., 2018; Morett et al., 2021; Trott et al., 2019; Wiedmann & Winkler, 2015). The mastery of pragmatics has shown to be affected by structural language abilities (Katsos et al., 2011), so the presence of multimodal prosodic and gesture cues might especially support pragmatic comprehension when structural components of language are compromised (such as in children with Developmental Language Disorder or DLD; Norbury et al., 2008). Here we compare children with and without DLD to see whether they benefit from multimodal cues to process pragmatic meaning when structural components of language are impaired.

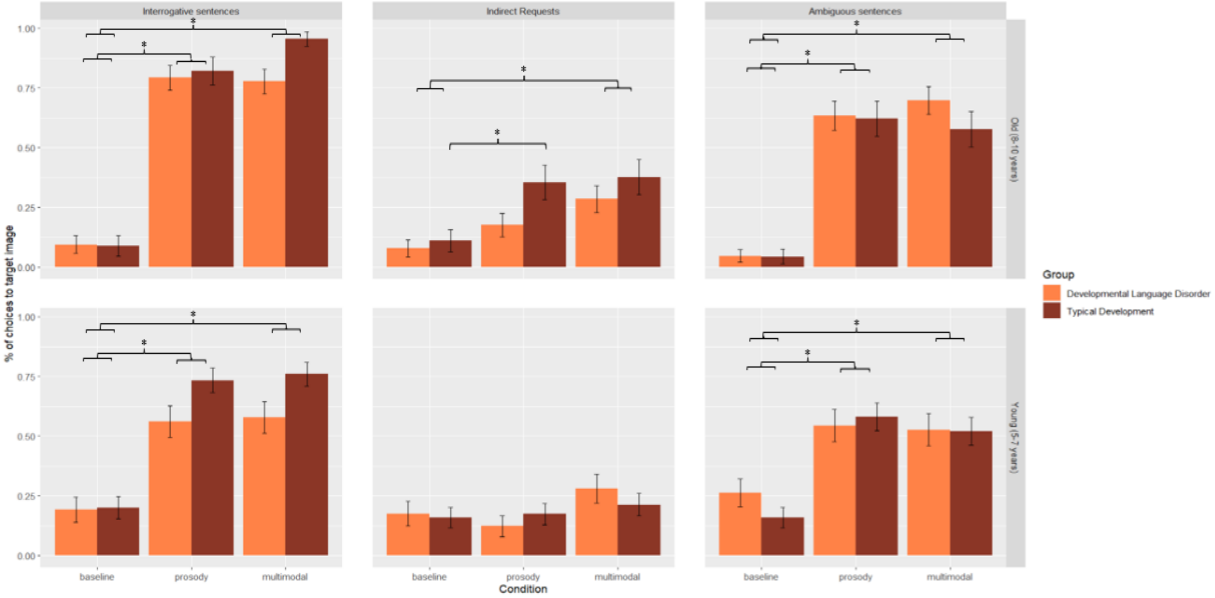
A sample of 39 children with DLD and 39 TD children aged 5 to 10 were first assessed for their cognitive and linguistic abilities. Then, they underwent a visual-world eye-tracking task in which, upon hearing a target sentence, they had to point at the image representing the target pragmatic meaning they just heard. We manipulated the meanings to be processed (within-subjects: interrogativity; indirect requests; discourse structure), and the multimodal cues accompanying the sentence (within-subjects: prosodically-enhanced, multimodally-enhanced, and no-enhancement). Our hypothesis was that multimodal cues boost the accuracy of target image selection among children with DLD, compared to the TD group, and that this pattern would emerge gradually, as children encounter more complex pragmatic meanings.

Results of the offline task indicate a significant effect of condition by which the presence of prosodic and multimodal cues enhanced the processing of interrogativity and discourse structure (χ^2 = 36.96, p > 0.001; χ^2 = 37.66, p > 0.001, respectively), with no interaction with age or participant group (see Figure 1). Notably, in indirect requests we found a 3-way interaction wherein multimodal cues were particularly helpful for older children with DLD (χ^2 = 16.99, p > 0.001). Currently, we are analyzing the proportion and timing of fixations to the target image from the eye-tracking data (results expected by the end of April 2023). This research will help determine whether prosody and gesture aid in the comprehension of pragmatic meanings, and will shed light on the appropriate cues to promote successful interaction in children with DLD.

Keywords: Developmental Language Disorder; Pragmatics; Prosody; Gestures

Figures

Figure 1. *Offline results for the three experimental blocks and two age groups. Y axis shows the proportion of selections to the target image for the three pragmatic meanings: interrogative sentences, indirect requests and syntactically ambiguous sentences. X axis shows the three experimental conditions: no-enhancement (baseline), prosodically-enhanced and multimodally-enhanced. * **represents a significance of p < 0.05.***



References

Armstrong, M., Esteve Gibert, N., Hübscher, I., Igualada, A., & Prieto, P. (2018). Developmental and cognitive aspects of children's disbelief comprehension through intonation and facial gesture. *First Language, 38*(6), 596–616.

Katsos, N., Roqueta, C.A., Estevan, R.A., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, *119*, 43–57.

Morett, L. M., Fraundorf, S. H., & McPartland, J. C. (2021). Eye see what you're saying: Contrastive use of beat gesture and pitch accent affects online interpretation of spoken discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(9), 1494–1526.

Norbury, C., Tomblin J., & Bishop D. V. M. (2008) A note on terminology. In Norbury C. F., Tomblin J. B., Bishop D.V.M. (Eds). *Understanding Developmental Language Disorders: From Theory to Practice* (pp. xii-xv). Taylor and Francis.

Trott, S., Reed, S., Kaliblotzky, D., Ferreira, V., & Bergen, B. (2022). The role of prosody in disambiguating English indirect requests. *Language and Speech*, *66*(1), 118–142. https://doi.org/10.1177/00238309221087715

Wiedmann, N., & Winkler, S. (2015). The influence of prosody on children's processing of ambiguous sentences. In S. Winkler (Ed.), *Ambiguity: Language and Communication* (pp. 185-198). De Gruyter Mouton. https://doi.org/10.1515/9783110403589-009

Gesture use in late talking and typically developing children during picture naming

Caterina Verganti, *Department of Psychology "Renzo Canestrari", University of Bologna*
Nathalie Frey, *Special Education and Therapy in Language and Communication Disorders, Faculty of Human Sciences, University of Würzburg*
Alessandra Sansavini, *Department of Psychology "Renzo Canestrari", University of Bologna*
Carina Lüke, *Special Education and Therapy in Language and Communication Disorders, Faculty of Human Sciences, University of Würzburg*

A considerable amount of research has examined the use of communicative gestures in typically developing (TD) children (e.g., Bates & Dick, 2002). Findings highlighted that language and gesture develop in parallel during the initial stage of life (Iverson et al., 2003), and that a transition between the gestural and the spoken modality of expression occurs during the first two years of life. Due to interindividual differences in gestures' use, other studies investigated this topic in populations of children with atypical language development, such as late talking children (LT). Thal, Tobias, and Morrison (1991) found lower imitation abilities of symbolic behaviours in ten LT, but this finding could not be replicated (Thal & Tobias, 1994). Recent findings showed that LT produce gestures mostly in isolation during a naming task (Rinaldi et al., 2022). Research obtained partial and conflicting findings on gesture use in LT, probably due to differences in the age of children involved, types of gestures investigated and complexity of the task used. Therefore, there is still debate about how frequently, for which purpose, and how accurately LT, compared to TD children, produce gestures.

This study aims to explore the use of spontaneous gestures, either produced in isolation or with the accompanying spoken answers, in LT, compared to TD children, during a structured picture naming task.

Eighty-five mono- and multilingual children (M_{Age} = 24.07 months, *SD* = 0.37; 48% male; 47% multilingual), growing up in Germany, participated in the study. Based on their lexical and morphosyntactic skills at a standardized test and expressive vocabulary size at a parental questionnaire they were divided into two groups: TD (*n* = 67, 46% male, 46% multilingual) and LT (*n* = 18, 56% male, 50% multilingual).

The children's use of spontaneous gestures has been investigated during the administration of a German standardized nouns' picture naming task. Total number of gestures produced, modality of expression (unimodal gestural, bimodal spoken-gestural, unimodal spoken), and the relationship between gestures and accuracy of spoken answers (correct or incorrect), were coded and analysed.

No significant differences emerged between the two groups for the total number of gestures produced. Differences were found in the modality of expression, where LT produced significantly more unimodal gestural answers (M_{LTs} = 7.11, *SD* = 5.03; M_{TD} = 2.09; *SD* = 2.66; *U* = 212.500, *p* < .001), and significantly fewer unimodal spoken (M_{LT} = 3.56, *SD* = 4.19; M_{TD} = 11.54; *SD* = 7.02; *U* = 211.500, *p* < .001) and bimodal spoken-gestural answers (M_{LT} = 2.94, *SD* = 2.80; M_{TD} = 7.70; *SD* = 6.38; *U* = 328.500, *p* = .003) than TD children. Furthermore, LT produced significantly less gestures in combination with correct spoken answers than TD children (M_{LT} = 0.61, *SD* = 0.85; M_{TD} = 3.99; *SD* = 4.14; *U* = 270.500, *p* < .001).

Results are consistent with previous findings (Rinaldi et al., 2022) showing that even if LT use as many gestures as TD children in the task, their gestures are rarely accompanied by speech. These findings suggest the presence of a delay in the transition between the gestural and the spoken modality in LT at the end of their second year of life.

Keywords: Spontaneous gestures; picture naming task; late talking children; typically developing children

References

Bates, E., & Dick, F. (2002). Language, gesture and the developing brain. *Developmental Psychobiology*, *40*, 293–310. https://doi.org/10.1002/dev.10034

Iverson, J.M., Longobardi, E., & Caselli, M.C. (2003), Relationship between gestures and words in children with Down's syndrome and typically developing children in the early stages of communicative development. *International Journal of Language and Communication Disorder*, *38*(2), 179-197. https://doi.org/10.1080/1368282031000062891

Rinaldi, P., Bello, A., Lasorsa, F.R., & Caselli, M.C. (2022). Do spoken vocabulary and gestural production distinguish children with transient language delay from children who will show developmental language disorder? A pilot study. *International Journal of Environmental Research and Public Health*, *19*(7), 3822. https://doi.org/10.3390/ijerph19073822

Thal, D., & Tobias, S. (1994). Relationships between language and gesture in normally developing and late-talking toddlers. *Journal of Speech and Hearing Research*, *37*, 157–170. https://doi.org/10.1044/jshr.3701.157

Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A 1-year follow-up. *Journal of Speech and Hearing Research*, *34*, 604–612. https://doi.org/10.1044/jshr.3403.604

Playing a guessing game to study multimodal syntactic complexity in typically developing children and children with developmental language disorder

Corrado Bellifemine, *Université Sorbonne Nouvelle*

Although children with developmental language disorder (DLD) produce simpler utterances often lacking subordination (De Weck, 1993; Blake et al., 2004) than typically developing children (TD), they strongly rely on gestures, which scaffold speech either by completing utterances or by replacing some of the linguistic segments (Blake et al., 2008). Since speech and gestures are intertwined (McNeill, 1992 ; Kendon, 2004), they are also influenced by the type of activity and its specific genre (François, 2002; Colletta, 2022). For instance, narratives mostly involve the use of iconic gestures and descriptions lead to a higher production of deictic gestures (Colletta & Pellenq, 2005). Fewer studies focused on the use of gestures during playful contexts such as board games with specific settings and rules, showing that DLD children use more gestures than typically developing children (TD) and they rely more on iconic and deictic gestures often replacing speech (De Weck et al., 2010).

The aim of this study is to analyze the multimodal syntactic complexity in 13 French typically developing children and 13 children with developmental language disorder aged 7 to 10. Each child was videorecorded while playing a guessing game with one of their parents and gave clues about 21 items (animals, objects, actions) presented on a computer screen. A pre-recorded voice heard on headphones suggested the word for each item and told to not say the word of the item while giving clues. Children could use whichever modality they preferred but were reminded they could give verbal cues if they only used pantomimes as game strategy. Children's multimodal productions were transcribed and analyzed using ELAN. The modality of each production was observed (verbal, multimodal, gestural), as well as the syntactic nature of utterances (simple clauses, juxtaposition, coordination, subordination, cleft structures, infinitive structures). Gestures were analyzed according to their type (deictic, representational, beat and recurrent gestures) and the clause they accompanied.

Results show that the two groups preferred the verbal modality while giving clues. However, DLD children produced simpler utterances than TD children, who relied on more complex syntactic structures. Moreover, DLD children gave quantitatively more clues, which were less precise than TD children's. At the gestural level, DLD children produced more referential gestures, whereas TD children produced more non referential gestures. Furthermore, DLD children produced more gestures accompanying almost all types of clauses. This means that, overall, almost all the DLD children use gestures mostly to accompany their complex utterances, especially juxtaposed and subordinate clauses. This means that, even though DLD children struggle with subordination, the guessing game – characterized by the coordination of several linguistic and extra-linguistic dimensions that children have to master – created for this study enhanced children's complexification and diversification of syntactic structures, multimodally. during the guessing game.

Thus, TD children are characterized by a linguistic maturity that reflects their developmental trajectory, since their speech is more complex and they use more non-referential gestures that help structure their discourse. On the contrary, DLD children rely more on gestures in order to complexify their speech. This reflects the influence of the language disorder and the fact that gestures support speech complexity thus reducing differences between TD and DLD children at the syntactic level. Moreover, the way gesture and speech articulate also reflects the task difficulties and features of the game set (i.e., game rules, word inhibition, clue-giving, parent-child interaction) that children have to master together with their speech planning. In conclusion, playful sets such as board games could be an interesting way of enhancing language progress and socio-pragmatic development during speech therapy.

Keywords: multimodality; gesture; developmental language disorder; guessing game; syntax

References

Blake, J., Myszczyszyn, D., & Jokel, A. (2004). Spontaneous measures of morphosyntax in children with specific language impairment. *Applied Psycholinguistics*, *25*(1), 29-41.

Blake, J., Myszczyszyn, D., Jokel, A., & Bebiroglu, N. (2008). Gestures accompanying speech in specifically language-impaired children and their timing with speech. *First Language*, *28*(2), 237-253.

Colletta, J. M., & Pellenq, C. (2005). Les coverbaux de l'explication chez l'enfant âgé de 3 à 11 ans. *Actes du 2e Congrès de l'ISGS : Interacting bodies, corps en interaction*, 17.

Colletta, J.-M. (2022). On the codevelopment of gesture and monologic discourse in children. In A. Morgenstern, & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (p. 205-242). American Psychological Association.

De Weck, G. (1993). Langage déviant et orthophonie : L'exemple des dysphasies. *Revue Tranel (Travaux neuchâtelois de linguistique)*, *20*, 69-87.

De Weck, G., Salazar Orvig, A., Corlateanu, C., da Silva, C., Rezzonico, S., & Bignasca, T. (2010). Interactions mère-enfant typique et dysphasique : Comment utiliser les gestes pour formuler une devinette ? *Lidil*, *42*, 159-180.

François, F. (2002). Fonctions, genre et mouvements discursifs. Un essai de clarification. In L. Danon-Boileau, C. Hudelot, & A. Salazar Orvig, *Usages du langage chez l'enfant* (p. 9-27). Editions OPHRYS.

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

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* (p. xi, 416). University of Chicago Press.

DAY 2. SESSION "DEVELOPMENT"

The synchronization of gesture and prosody in French children's multimodal pathway into negation

Pauline Beaupoil-Hourdel, Christelle Dodane, Fanny Catteau & Aliyah Morgenstern

Children use gestures to increase the informativeness of their spatial expressions depending on the complexity of spatial relations

Dilay Zeynep Karadöller, Kevser Kırbaşoğlu, Beyza Sumer & Ercenur Ünal

How politeness shapes speech-accompanying gestures: a developmental perspective

Iris Hübscher & Lucien Brown

Priming effects from sign to spoken word in hearing toddlers exposed to sign-augmented communication

Lena Heine & Nicole Altvater-Mackensen

The synchronization of gesture and prosody in French children's multimodal pathway into negation

Pauline Beaupoil-Hourdel, *Sorbonne Université, INSPE de Paris, CeLiSo, UR 7332*

Christelle Dodane, *Sorbonne Nouvelle University, CLESTHIA EA 7345*

Fanny Catteau, *Université de Poitiers, FoReLLIS UR 15076*

Aliyah Morgenstern, *Sorbonne Nouvelle University, PRISMES EA4398*

The expression of negation is a privileged locus to study multimodal combinations. There is a cross-modal continuity in the expression of negation speech acts, which are first mainly expressed by gestures, then by speech (Bates, Camaioni and Volterra, 1976; Bates et al., 1979). From the end of their first year on, children can express negation with headshakes then palms-ups and index waves. Prosody and gestures are also combined to express refusals, protests, epistemic negations or powerlessness, sometimes before the emergence of the first verbal negation markers. It is therefore crucial to analyze gestures and prosody with an integrative approach before and after the emergence of speech.

The goal of this study was to analyze the synchronization of gesture and prosody in children's expression of negation. We analyzed the longitudinal data of four monolingual children recorded monthly for one hour between the ages of 1;0 and 4;0 in spontaneous interaction with their parents (Paris Corpus, Morgenstern & Parisse, 2012). We studied the children's productions within the MLU range of 1 to 4.

We focused on the multimodal productions containing the word "non" (no) in isolation. Three types of analyses were conducted. First, we coded prosodic properties (direction of the intonation contour, accent range, register, duration, intensity), using PRAAT. Second, we coded nonverbal behavior (hand gestures, joint attention expressed through eye gaze and checking behavior, body movement and facial expressions), using ELAN. Third, we compared the prosodic and gestural analyses to look for directional and temporal synchronization patterns using AlphaPose, and comparing the outputs with our PRAAT extractions.

At the prosodic level, results showed that the first vocal productions of "non" emerged around an MLU of 1.1 and were exaggerated at the prosodic level. Between an MLU of 1.1 and 1.8 (phase 1), "non" was mainly realized with rising intonation contours and increased syllabic duration. Between 1.8 and 2.8, (phase 2) it was mainly produced with rise-fall intonation contours and finally, between 3.3. and 4 (phase 3) with flat or falling intonation contours and reduced syllabic duration. Such an evolution seems to reflect a better control in the expression of negation as of an MLU of 2.8.

At the non-vocal level, body movements were most often produced in coordination with verbal production and their direction was mostly synchronized with the direction of intonation contours (rising contours with rising gestures) during the first phase. The more the children expressed protests against adults, the more they exaggerated both their prosody (higher accent range, register, intensity and duration) and their body movements. During phase 2, they used mostly upper-body gestures and movements (head, chest) with a majority of forward and backward or oscillating movements in close parallel with their prosodic contours. As their mastery of speech developed, they gradually stopped exaggerating their prosody and resorted less to non-verbal behavior.

Gestures, body movements and prosody provide powerful resources that the child integrates to make her multimodal entry into language. If children use each modality (vocal and visual) more and more skillfully thanks to adults' scaffolding in everyday life interactions, both modalities actually develop together. This study therefore gives us insights on how children become experts in face-to-face social interaction, which is multimodal in nature.

Keywords: gesture; prosody, negation, child language acquisition

References

Bates, E., Camaioni, L. and Volterra, V. (1976). Sensorimotor performative. In: Bates, E. (Ed.). *Language and Context: the Acquisition of Pragmatics*. New York: Academic Press.

Bates, E. (1979). Emergence of symbols in language and action: similarities and differences. *Paper and Reports on Child Language Development*, Stanford, n.17, p.106-118.

Morgenstern, A & Parisse, C. (2012). The Paris Corpus. *French Language Studies*, 22(1), 7-12, Cambridge University Press.

Children use gestures to increase the informativeness of their spatial expressions depending on the complexity of spatial relations

Dilay Z. Karadöller, *Max Plank Institute for Psycholinguistics, Nijmegen, The Netherlands*

Kevser Kırbaşoğlu, *Ozyegin University, İstanbul, Turkey*

Beyza Sümer, *University of Amsterdam, Amsterdam, The Netherlands*

Ercenur Ünal, *Ozyegin University, İstanbul, Turkey*

Learning to communicate about viewpoint-dependent spatial relations is challenging for children (Clark, 1973; Johnston & Slobin, 1979; Grigologlou et al., 2019). Within viewpoint dependent relations, Front-Behind are acquired earlier than Left-Right. This has been attributed to differences in the complexity of the relations: unlike Front-Behind, Left-Right are symmetrical and lack features distinguishing one relation from the other (e.g., visibility or occlusion; Clark, 1973). Nevertheless, when children's gestures are taken into account, their expressions about Left-Right relations are informative (Karadöller et al., 2022). Here, we ask whether such expressions that become informative with gesture are sensitive to the complexity of spatial relations (Left-Right vs. Front-Behind) or reflect a general tendency in children.

We elicited descriptions from 24 child (M_{age}=8.6) and 23 adult (M_{age}=35.9) Turkish speakers. Stimuli consisted of 56 displays of 4 pictures depicting the same two objects in Left Right or Front-Behind relations as targets. 28 additional displays with viewpoint-independent relations as targets served as fillers. In each trial, participants described a target picture (indicated by an arrow) to a confederate.

Descriptions with specific spatial nouns (i.e., *Front*, *Behind*, *Left*, *Right*) were coded as **informative in speech** regardless of gesture use as speech already conveyed the spatial relation informatively. Descriptions with general spatial nouns (i.e., *Side*, *Next to*) accompanied by spatial gestures that disambiguated the relative locations of the two objects were coded as **informative with gesture**. The remaining descriptions were **under-informative**.

If children rely on visually-motivated expressions in gestures to convey Left-Right due to the complexity of the relations, then they should use such expressions less frequently for less complex Front-Behind relations. To test this, we focused on informative descriptions only and compared the frequency of descriptions informative in speech vs. with gesture (Fig.1a). A glmer model revealed an interaction between Spatial Relation (Left-Right, Front-Behind) and Age (adults, children) on the binary dependent variable (0=informative in speech; 1=informative with gesture) at the item level (*β*=7.576, *SE*=3.090, *p*=.014). As expected, children were more likely to use descriptions that become informative with gesture for Left-

Right than for Front-Behind (*p*<.001). Adults were already informative in speech and this did not change across Left-Right and Front-Behind (*p*=.627).

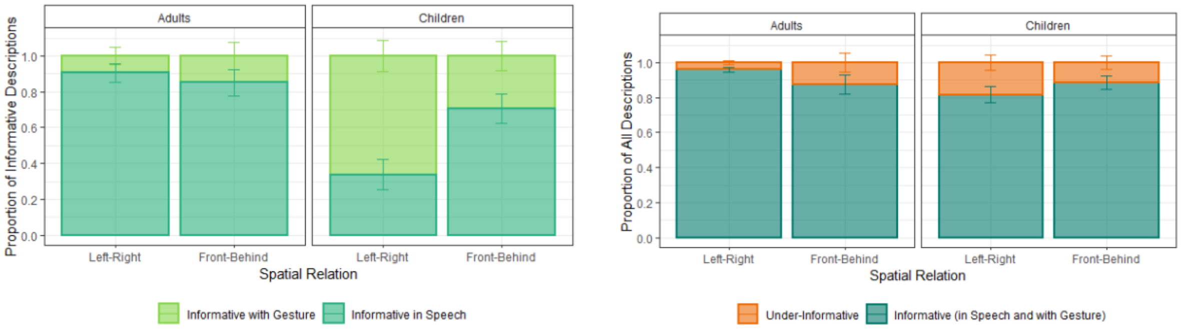
Next, we focused on all descriptions and compared the frequency of informative descriptions (in speech and with gesture) to under-informative descriptions (Fig.1b). A glmer model revealed an interaction between Age and Spatial Relation on the binary dependent variable (0=under-informative; 1=informative) at the item level (β=-1.8524, *SE*=0.369, *p*<.001). For Front-Behind, children produced informative descriptions as frequently as adults (*p*=.698). For Left-Right, children produced informative descriptions less frequently than adults (*p*<.001).

Summarizing, children do not always use gestures to increase the informativeness of their expressions in speech but this is dependent on the complexity of spatial relations. These findings suggest that the challenge in acquiring Left-Right in language, at least partly, comes from having to map symmetrical relations onto arbitrary symbols in speech. Gestures that allow for visually-motivated expressions of Left-Right reduce this challenge to some extent but not completely, pointing to a universal challenge for Left-Right in cognitive development.

Keywords: spatial language; iconicity; multimodal language

Figures

Figure 1. *Distribution of descriptions that are (a) already informative in speech vs. become informative with gesture across (b) informative vs. under-informative*



References

Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. *Cognition*,*2*(2), 161-182. https://doi.org/10.1016/0010-0277(72)90010-8

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language*, *6*(3), 529-545. https://doi.org/10.1017/S030500090000252X

Grigoroglou, M., Johanson, M., & Papafragou, A. (2019). Pragmatics and spatial language: The acquisition of front and back. *Developmental Psychology*, *55*(4), 729-744. https://doi.org/10.1037/dev0000663

Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2022). Sign advantage: Both children and adults' spatial expressions in sign are more informative than those in speech and gestures combined. *Journal of Child Language*. https://doi.org/10.1017/S0305000922000642

How politeness shapes speech-accompanying gestures: a developmental perspective

Iris Hübscher, *Zurich University of Applied Sciences*

Lucien Brown, *Monash University*

Speech and gesture develop in parallel in early childhood and become increasingly more complex over time. Sociopragmatic factors such as common ground, politeness and intersubjectivity have been shown to shape gesture production in adults. For example, research on audience design has demonstrated that gesture is sensitive to different contextual factors, such as whether the interlocutor is visible (Bavelas et al., 2008; Holler & Wilkin, 2009), new to the activity (Galati & Brennan, 2014) or attentive to the speaker (Beattie & Aboudan, 1994; Jacobs & Garnham, 2007). Furthermore, very recently it has been shown that a variety of multimodal cues such as prosody and different facial and body cues are sensitive to the degree of social distance between interlocutors and are adapted systematically in deferential face-to- face interactions (Brown et al., 2023; Winter et al., 2021).

However, little attention has been paid to such pragmatic factors in the development of gesture in language acquisition research. Focusing on one such factor, social distance, this study examines how gestures are performed differently depending on the identity of the interlocutor (friend (same age) or teacher (65 years old, unknown)) during a narrative task in the early school years.

To explore this, data was collected from 14 6-year-old and 15 8-year-old Catalan- speaking children and 14 Catalan adult speakers as a control group. Each speaker produced two retellings of the "Tweety Bird" cartoon: One with a professor/ teacher ("deferential" situation) and one with a close friend ("non-deferential" situation). The video data was annotated in ELAN and analyzed both quantitatively and qualitatively in regard to gestural differences in terms of their frequency, gesture size, hand shape, content (path, manner, ground) and viewpoint (character viewpoint, observer viewpoint).

Results from the Catalan-speaking adults reveal that co-speech gestures show substantial differences between the two situations. When retelling the story to a person with higher status, the speakers gestured less frequently and produced smaller and less animate gestures. Additionally, speakers preferred "observer viewpoint" with the status superior as compared to character-viewpoint with a friend, causing the interaction to become less iconic and playful, resulting into a more serious stance. Preliminary results from children show that along with previous findings, independent of the interlocutor, the use of gesture significantly increased

between the two age groups, however the use at 8 years is still not as high as in adults. Interestingly, in the older age group children not only produced fewer gestures when narrating the story to a teacher than to a friend but also the gesture size was affected, with significantly fewer big gestures and two-handed gestures. These results will be discussed in relation to the more general pattern of polite speech showing decreased animacy across multiple modalities and by comparing adult and child patterns.

Keywords: multimodal politeness; developmental pragmatics; gesture; narrative development

References

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*, 495–520. https://doi.org/10.1016/j.jml.2007.02.004

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, *99*(3–4), 239–272. https://doi.org/10.1515/semi-1994-993-402

Brown, L., Hübscher, I., Kim, H., & Winter, B. (2023). Indexing social distance through bodily visual practices in two languages. In A. H. Jucker, I. Hübscher, & L. Brown (Eds.), *Multimodal Im/politeness. Signed, spoken, written* (pp. 131–161). John Benjamins.

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*, 435–451. https://doi.org/10.1080/01690965.2013.796397

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*, *24*, 267–289. https://doi.org/10.1080/01690960802095545

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*, 291–303. https://doi.org/10.1016/j.jml.2006.07.011

Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., & Grawunder, S. (2021). Rethinking the frequency code: A meta-analytic review of the role of acoustic body size in communicative phenomena. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840), 20200400. https://doi.org/10.1098/rstb.2020.0400

Priming effects from sign to spoken word in hearing toddlers exposed to sign-augmented communication

Prof. Dr. Nicole Altvater-Mackensen, *Department of English, University of Mannheim*

Lena Heine, *Psychological Institute, Johannes-Gutenberg University Mainz*

A growing body of evidence points to the facilitating role of gestures in language learning (Rohlfing, 2019). However, prior research focused primarily on communicative and speech-accompanying gestures. Less is known about how signs as linguistic symbols are integrated in the developing linguistic system of hearing children. To address this question, we asked if signs facilitate processing of spoken language by activating relevant lexical word representations. If so, this would provide evidence that children integrate signs and spoken language in speech processing from early on. We assessed children's use of sign information to facilitate word recognition in a priming task, focusing on hearing children who are exposed to sign-augmented communication (Wilken, 2021) in inclusive daycares. We hypothesized that – if signs activate spoken word representations – children will show facilitated target recognition when being primed by a related compared to an unrelated sign.

40 German-learning children without hearing impairments (mean age = 4.5 years; range = 32-82 months) participated in a modified version of the visual priming task, used with toddlers in Mani & Plunkett (2010). Children were presented with videos of a signer (prime) followed by two side-by-side images of yoked targets and distractors, while their looking time to each picture was measured using eye tracking. The target was named auditorily 50 ms after the appearance of both images (see Figure 1). Each child was presented with a total of 24 trials. In half of the trials (related trials), prime signs and spoken labels coincided in meaning, both referring to the target. In the other half of the trials (unrelated trials), prime and target were unrelated. Word recognition was measured in terms of proportion of target looking (PTL).

A repeated measures ANOVA with prime condition (related vs. unrelated) and block (first vs. second) as within-subjects factors using PTL as dependent measure found a significant main effect of block ($F(1, 37) = 13.764$; $p < .001$), but no effect of prime condition ($F(1, 37) = .105$; $p = .748$) and no interaction between condition and block ($F(1, 37) = 2.296$; $p = .138$). This suggests that the relation between prime and target did not modulate target word recognition. However, post hoc analysis revealed a significant interaction between prime condition and cohort size ($F(1,37) = 46.010$; $p < .001$). Children looked significantly longer at the target in unrelated compared to related trials ($t(37) = -4.223$; $p < .001$) when the item was from a large cohort, indicating an interference effect. In contrast, children looked significantly longer at the target in related compared to unrelated trials when the item was from a small cohort (*t*(37) = 5.639; *p* < .001), indicating a facilitation effect. Cohort size was assessed based on the lexical database *childLex* (Schroeder et al, 2015) and the vocabulary list of the *FRAKIS* (Szagun et al., 2009).

The results demonstrate a similar influence of cohort size on word recognition as has been observed in object-to-word priming in toddlers (Mani & Plunkett, 2011). The different priming effects strongly suggest that signs activate the corresponding spoken word representation in similar ways as implicitly generated spoken labels. In general, the results support the idea that speech-accompanying signs can facilitate speech processing (Yap et al., 2011), and corroborate previous work showing cross-modal co-activation of signed and spoken/written language in bimodal adults (Shook & Marian, 2012).

Keywords: sign augmented communication; priming; multimodal representation

Figures

Figure 1. *Timeline of a trial*

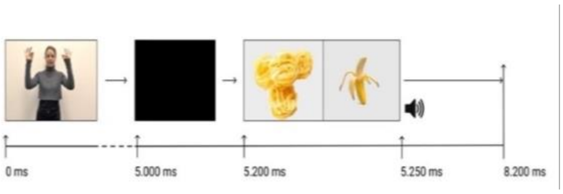
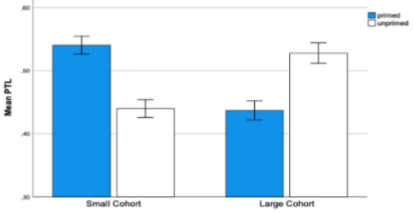


Figure 2. *Cohort effects: Mean proportion of target looking in primed and unprimed trials (+/- 1 SE).*



References

Mani, N., & Plunkett, K. (2010). In the infants minds ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*, 908–913.

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206.

Rohlfing, K. J. (2019). Learning Language from the Use of Gestures. *International Handbook of Language Acquisition*, 213–233.

Schroeder, S., Wörzner, K. M., Heister, J., Geyken, A., & Kliegl, R. (2015). ChildLex - A lexical database for print language for children in German. *Psychologische Rundschau*, *66*(3), 155–165.

Shook, A., & Marian, V. (2012). Bimodal Bilinguals Co-activate Both Languages during Spoken Comprehension. *Cognition*, *124*(3), 314.

Szagun, G., Stumper, B., & Schramm, S. A. (2009). *FRAKIS Fragebogen zur frühkindlichen Sprachentwicklung. FRAKIS (Standardform) und FRAKIS-K (Kurzform)*. Pearson.

Wilken, E. (2021). Unterstützte Kommunikation. Eine Einführung in Theorie und Praxis (6. Auflage). [*Augmentative and Alternative Communication (AAC). An introduction to theory and practice.] Kohlhammer-Verlag, Stuttgart.*

Yap, D. F., So, W. C., Melvin Yap, J. M., Tan, Y. Q., & Teoh, R. L. S. (2011). Iconic Gestures Prime Words. *Cognitive Science*, *35*(1), 171–183.

DAY 2. SESSION "GESTURE TYPES & REPRESENTATION"

PALM-UP: Gesture or Sign—a Corpus Study of Chinese Sign Language

Yuting Zhang & Hao Lin

How Language Modality Influences the Use of Palm-Up Open Hand Gestures: a Comparative Study of German and German Sign Language

Anna Kuder, Sandra Debreslioska & Pamela Perniss

The use of silent gestures to categorise and describe objects and actions in European Spanish

María Morales Pérez, Andrea Ariño-Bizarro & Iraide Ibarretxe-Antuñano

Differences in gestural representations of concepts in blind and sighted individuals

Ezgi Mamus, Laura J, Speed, Gerardo Ortega, Asifa Majid & Aslı Özyürek

PALM-UP: Gesture or Sign—a Corpus Study of Chinese Sign Language

Yuting Zhang, *University of Washington*

Hao Lin, *Shanghai International Studies University*

Palm-up (PU) is prevalent and puzzling across spoken languages and sign languages (Cooporider et al., 2018), and it has been reported ubiquitous in both Chinese hearing communities and deaf communities (Lin, 2019). Its linguistic status in sign language is hotly debated, as whether it is a sign word or a gesture in sign language. In this paper, we explore the developmental journey of PU from a gesture into Chinese Sign Language (CSL) based on the annotated CSL naturalistic data by examining the form-function mappings of each variant of PU. Based on CSL Corpus (Southern China Variant, 2016-), we drew on a subset of conversation data from 52 signer (25 Females), and stratified the sample by age groups according to school policy and sign language change based on Lin (2021). Three age groups were included : '> 38 years old' (N=15), '38-68 years old' (N=18) and '> 68 years old' (N=19).

In our corpus, we identified 793 PUs in total. We coded the context meaning each PU carries and its syntactic role and further annotated each PU's manual features (handshape, number of hands, hand movement) and non-manual features (mouth gesture, mouthing, head movement). We also labelled whether a PU is a gesture or a sign with reference to their annotated features. If a PU form was required by the clause and form a syntactic component in the clause, it was considered as a sign, or otherwise a gesture ; if a PU form was non-truth-conditional, meaning that the removal of the PU does not alter the meaning of the clause it's in, it was considered a gesture (to be more specific, a discourse marker).

We identified four main functions of PU, and according to their different featural clusters, we divided them into: PU-Organizing (lax, one/ two hand, small movement, none mouth action, no head movement/ head tilt; N=76), PU-Modality (expressing helplessness and obviousness) (lax/tight, two hands, moderate/big movement, mouth gesture, no head movement/ head tilt; N=150), PU-Interrogation (tight, two hands, moderate, mouthing, no head movement/ head tilt; N=86), PU-Negation (tight, two hands/one hand, moderate/big, mouthing, headshake/ no head movement; N=481).

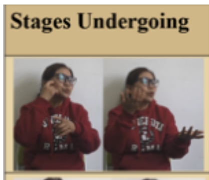
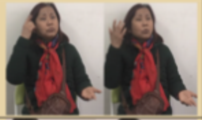
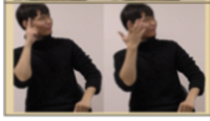
We find that: 1) With its position quite flexible, PU-Organizing is a prosodic cue that may work at the phonological level; 2) PU-Modality appears always at sentence-final, whose form is more stable than PU- Organizing yet still may unpredictably vary across contexts. Working as pragmatic cues, it conveys meaning like 'I cannot help it' 'That's what you see'. We also suspect it such function comparable to gesture with longer duration, emphatic forms; 3) With a stable form, PU-Interrogation is almost always at sentence-final, acting as a one-for-all WH word that can be decoded as ‘what’, ‘who’, ‘where’ etc. 4) PU-Neg when acting as a lexicalized negator, occupies a sentence-final position, with a stable form; when combining with verbs like ‘know’, PU-Neg may work like an affix with mainly two hands in use among the older group, but we observed one-hand expressions in the same cases among the younger group in a more ‘fused’ way (Table 2), which shows an interesting contrast and suggests the undergoing PU grammaticalization. As a result, we believe PU in CSL is undergoing grammaticalization while it is also used as a gesture among the CSL deaf (e.g., PU-Organizing, PU-Modality).

Combing the relevant literature (e.g., Zeshan 2004), we further argue that typologically PU is the preliminary form of WH word across sign languages, the cognitive foundation of such a mechanism may derive from the iconic form of palm-up, suggesting ‘emptiness’ while the essence of WH word is to offer a pure ‘empty’ form for real interactive practice.

Keywords: Palm-up; gesture; sign; grammaticalization

Figures

Figure 1. *The Undergoing Grammaticalization Stages of PALM-UP Negation in CSL*

Stages Undergoing	Number of Hands	Attached Closely or Not
	2	- -
	2	+ -
	1	+ +

References

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. Frontiers in Communication, 0, 23. https://doi.org/10.3389/FCOMM.2018.00023

Lin, H. (2019). The Analysis of Gestural System of the Paralanguage: with the Example of Palm-up. Contemporary Rhetoric, 12(2), 84-95. http://yuxiqbs.cqvip.com/Qikan/Article/Detail?

Lin, H. (2021). Early development of Chinese Sign Language in Shanghai schools for the deaf. Frontiers in Psychology, 12, 702620. https://doi.org/10.3389/fpsyg.2021.702620

Zeshan, U. (2004). Hand, head and face - negative constructions in sign languages. Linguistic Typology, 8, 1-58.

How Language Modality Influences the Use of Palm-Up Open Hand Gestures: a Comparative Study of German and German Sign Language

Sandra Debreslioska, *Lund University, Sweden*
Anna Kuder, *University of Cologne, Germany*
Pamela Perniss, *University of Cologne, Germany*

Spoken language (SpL) users produce gestures together with speech, as an additional visual resource, for utterance construction. Sign language (SL) users gesture too, but there is no modality difference between the grammatical/lexical and gestural elements. Thus, while researchers in the SpL context predominantly analyze speech-associated gestures as spontaneous, non-obligatory movements of the hands/arms, creating meanings on the fly (McNeill, 2002), approaches to defining gestural elements in SL have been more diverse. For instance, some researchers tend to see gestures as being part of a lexicalization or grammaticalization process (van Loon et al., 2014). Others try to draw a parallel to SpL production and propose a sign + co-sign gesture view (Goldin-Meadow & Brentari, 2017). And finally, there are researchers trying to determine to what degree movements are ‘gestural’ versus linguistic depending on categoriality and gradability (Liddell, 2003).

The aim of this study is to shed light on this difference in analysis of SpL and SL by examining the palm-up open hand gesture (PUOH) (Kendon, 2004). The configuration of PUOHs typically includes an open palm, extended fingers (with more or less (ex)tension), and the palm turned upwards, for one or both hands (Müller, 2004). Previous studies have suggested that PUOHs are multifunctional in both SL and SpLs (Kendon, 2004; Cooperrider et al. 2018), being mainly used as discourse structural devices. We focused on the use of PUOH gestures in different discourse contexts in narrative productions in a SL (German SL, DGS) and a SpL (German). We analyzed a corpus of 66 narratives produced by 12 DGS signers and 10 German speakers. All participants watched the same three stimulus videos (excerpts from Charlie Chaplin movies) and retold them to an addressee. We identified 95 PUOHs in German and 54 PUOHs in DGS. Each gesture was assigned to one of five contexts of use: when a) indicating the beginning/ending of a narrative, b) presenting new information, c) searching for a word/sign, d) expressing uncertainty, or e) providing explanations. To determine the context in which a PUOH was produced, we considered the utterance in which a gesture was embedded, the co-occurring non-manuals, mouth actions, and, for German, the speech exactly aligned with the gesture stroke.

Preliminary findings show that DGS signers tend to use PUOHs in contexts of sign search and at narrative endings (35% for both). The third most preferred context of use is when presenting new information (19%). However, signers rarely use PUOHs in contexts of providing explanations or when signaling uncertainty (9% and 4% respectively). Speakers of German, on the other hand, tend to use PUOHs in contexts of presenting new information (45%). The second and third most preferred contexts are when providing explanations and signaling uncertainty (19% for both). Finally, speakers rarely use PUOHs in contexts of word search (10%) or when marking beginning/endings of narratives (7%).

This study is one of the first to compare the use of PUOHs cross-linguistically, cross-modally and based on fully comparable data samples. The findings show that PUOHs play a role in discourse structuring in both SL and SpL. However, the study also suggests that language modality has an important influence on the detailed contexts of use of PUOHs in narrative productions. That is, the patterns are almost reversed for SL versus SpL. While signers predominantly use PUOHs when they have trouble finding the right formulation, as well as when they are signaling the end of narratives, these are the least preferred contexts of use for speakers. Conversely, while speakers typically use PUOHs when introducing new or reactivating important information, this context of use is a lot less frequent for signers. We present examples of the observed phenomena, discuss possible implications for the effect of language modality on the use of PUOH gestures specifically, as well as on the definition of gestures more generally.

Keywords: palm-up open hand gestures, cross-linguistic comparison, German Sign Language, speech-gesture relationship

References

Cooperrider, K., N. Abner & S. Goldin-Meadow. (2018). The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. *Front. Commun*, 3(23).

Goldin-Meadow, S. & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and brain sciences, 40*, E46.

Kendon, A. (2004). *Gesture. Visible action as utterance*. CUP.

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press.

McNeill, D. (2002). Gesture and language dialectic. *Acta Linguistica Hafniensia*, 34: 7-37.

Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family? In C. Müller & R. Posner (Eds.), *The Semantics and Pragmatics of Everyday Gestures* (pp. 233–256). Weidler.

Van Loon, E., R. Pfau, & M. Steinbach. (2014). The grammaticalization of gestures in sign languages. In C. Müller et al. (Eds.), *Body – Language – Communication* (pp. 720–730). De Gruyter Mouton.

The use of silent gestures to categorise and describe objects and actions in European Spanish

María Morales Pérez, *University of Zaragoza*
Andrea Ariño-Bizarro, *University of Zaragoza*
Iraide Ibarretxe-Antuñano, *University of Zaragoza*

Silent gestures are those in which there is an absence of spoken discourse and, although they do not follow any social convention, they show a high degree of systematicity (e.g., to take an imaginary steering wheel to represent the verb to drive). In recent studies, Ortega and Özyürek (2019a, b) have examined which gestural techniques are used by Dutch and Mexican Spanish speakers to represent iconically, through silent gestures, three types of semantic categories: action verbs (writing), manipulable objects (pencil sharpener) and non-manipulable objects (cupboard). Their results show that the gestural technique of acting, i.e., the body reenacts the action, is the one mostly used in both languages for the representation of actions and manipulable objects, while for non-manipulable objects the technique of drawing, i.e., the hands trace the outline of the intended object, is preferred. These authors also point out that, despite being different languages, it does not seem that both speakers behave differently in terms of categorisation and expression of these concepts through silent gesture.

Stemming from these results, the goal of this paper is twofold: (i) to replicate the study of Ortega and Özyürek (2019b) with speakers of European Spanish, in order to establish whether or not there are differences in cases where the language is shared, but not the dialectal and cultural variety, and (ii) to test the functioning of the silent gesture in a new semantic category, abstract actions (e.g. cognition verbs like to think and to feel) and, in this way, to investigate the type of gestural technique used in these hitherto unexplored categories.

Twenty native speakers of European Spanish participated in this study. The task consisted of the silent gestural description of 45 stimuli (3 training + 30 adapted from Ortega and Özyürek (2019b) + 12 own elaboration). In total, the corpus comprised 911 silent gestures, which were been transcribed and classified (McNeill, 1992; Müller, 2016; Hwang et al., 2016).

Results show that acting technique is the most employed gestural strategy in all semantic categories, except for the representation of non-manipulable objects, where drawing is preferred. This reinforces the results of Ortega and Özyürek (2019b) regarding the type of silent gesture and the absence of cross-linguistic differences. However, intergroup differences have been observed with respect to the representation of objects related to new technologies (e.g., telephone) and in relation to the semantic feature used iconically in the representation (e.g.,

drinking). With respect to the second objective, it is shown that the silent gesture is also used in some abstract actions in a systematic way (e.g., thinking), but that, other types of actions (e.g., intuiting) generate difficulty.

Keywords: silent gestures; European Spanish; semantic categories

References

Hwang, S.O., Tomita, N., Morgan, H., Ergin, R., İlkbaşaran, D., Seegers, S., Lepic, R., & Padden, C. (2016). Of the body and the hands: patterned iconicity for semantic categories. *Language and Cognition*, *9*(4), 573-602.

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

Müller, C. (2016). From mimesis to meaning: A systematics of gestural mimesis for concrete and abstract referential gestures. In J. Zlatev, G. Sonesson & P. Konderak (Eds.), *Meaning, mind and communication: Explorations in cognitive semiotics* (pp. 211-226). Peter Lang.

Ortega, G. & Özyürek, A. (2019a). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods*, *52*, 51-67.

Ortega, G. & Özyürek, A. (2019b). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*, 84-113.

Differences in gestural representations of concepts in blind and sighted individuals

Ezgi Mamus[1,2], Laura J. Speed[1], Gerardo Ortega[3], Asifa Majid[4], Aslı Özyürek[2,1],

[1] *Radboud University*

[2] *Max Planck Institute for Psycholinguistics*

[3] *University of Birmingham*

[4] *University of Oxford*

Recent gesture theories have claimed that gestures arise from sensorimotor simulations and have embraced an embodied perspective to explain gesture production (e.g., Hostetter & Alibali, 2008). As gestures are outcomes of our multimodal sensorimotor experience in the world, their forms should reflect gesturers' particular experience with a concept in a representational format suitable for gesture (i.e., visual, spatial, and motoric).

Studies examining gestural representations of semantic concepts (e.g., objects) during speech or when no speech is allowed—i.e., silent gestures (Masson-Carro et al., 2017; Ortega & Özyürek, 2020) have revealed certain regularities in gestures. Concepts that trigger motor imagery—such as manipulable objects like tools—result in the use of an acting strategy (the reenactment of an action with an object). Conversely, when visuospatial information is more pertinent, such as for non-manipulable objects, people tend to use a drawing strategy (tracing the outline of an object). van Nispen et al. (2017) argued that people select salient features of their mental representations for depiction, which are limited by the constraints of the manual-visual modality and the manual affordances of the referent. Mostly motor and visuospatial features of concepts fit these criteria to be depicted in gestures—as other salient features of concepts, such as color, do not lend themselves to gestural forms. If visuospatial and motor cues drive consistent patterns in gesture (Ortega & Özyürek, 2020; van Nispen et al., 2017), a lack of visual experience may result in different gestural forms being selected for depiction of conceptual representations. Also, Fay et al. (2022) showed that communication success in gestures (measured with success at interpreting the meaning of gestures) is greater for sighted than blind gesturers. In the present pre-registered study, we explore whether lack of visual experience alters how specific features of concepts are mapped onto gestural representations of concepts. To address this, we compare silent gestures for simple concepts produced by congenitally blind and sighted individuals.

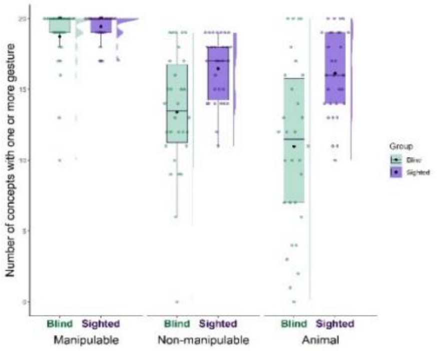
Thirty congenitally blind and 30 sighted Turkish adult speakers were instructed to produce silent gestures for individual concepts from three semantic categories including concepts that predominantly rely on motor information (manipulable objects) versus visuospatial information (non-manipulable objects and animals). We had 60 concepts in total: 20 concepts per semantic category (one concept from non-manipulable object category had to be removed in the final analyses). We coded the strategies (acting, representing, drawing, and personification) for each gesture by following Ortega and Özyürek (2020).

We predicted that both blind and sighted people would produce silent gestures by using an acting strategy for manipulable object concepts as both groups have motor experience with these objects. In contrast, for concepts that rely more strongly on visuospatial information, we predicted a difference in gesture forms between blind and sighted people. Blind people would either not be able to produce any gesture or they would produce fewer drawing and personification gestures for non-manipulable object and animal concepts, respectively. As expected, our preliminary results showed that regardless of the strategy, blind participants produced fewer gestures for non-manipulable object and animal concepts than sighted participants (Figure 1). These results suggest that visual experience plays a role in how certain categories of concepts are mapped onto gesture.

Keywords: silent gesture; blindness; manipulable and non-manipulable objects; animals

Figures

Figure 1. *Frequency of gestures per semantic category*



References

Fay, N., Walker, B., Ellison, T. M., Blundell, Z., De Kleine, N., Garde, M., ... & Goldin-Meadow, S. (2022). Gesture is the primary modality for language creation. *Proceedings of the Royal Society B, 289*(1970), 20220066.

Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495–514.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2017). How what we see and what we know influence iconic gesture production. *Journal of Nonverbal Behavior, 41*(4), 367-394.

Ortega, G., & Özyürek, A. (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods, 52*(1), 51-67.

van Nispen, K., van de Sandt-Koenderman, W. M., & Krahmer, E. (2017). Production and comprehension of pantomimes used to depict objects. *Frontiers in psychology, 8*, 1095.

DAY 2. SESSION "L2 & MULTIMODAL PEDAGOGY"

Linking hand gestures enhances the development of L2 French liaisons

Solene Inceoglu & Ruri Ueda

Stress in motion: Gesture-speech coupling in L2 lexical stress production

Hans Rutger Bosker, Marieke Hoetjes, Wim Pouw & Lieke van Maastricht

Mechanisms underpinning gestural facilitation of L2 word learning

Erin Minton-Branfoot, Richard O'Connor & Henning Holle

Manipulating gestures in motion capture-animated characters for an L2 comprehension study

Valentijn Provè, Bert Oben & Henrik Garde

The effect of temporal alignment of speech and gesture on L2 speakers

Eleni Ioanna Levantinou & Argiro Vatakis

Linking hand gestures enhances the development of L2 French liaisons

Solène Inceoglu, *The Australian National University*
Ruri Ueda, *The Australian National University*

The past decade has seen an increased number of experimental studies examining the effect of co-speech gestures on L2 speech perception and production. The majority of these studies have looked at 'pitch' gestures (e.g., Baills et al., 2019; Morett & Chang, 2015; Yuan et al., 2019), beats (e.g., Gluhareva & Prieto, 2017; Hirata & Kelly, 2010) and handclapping (e.g., Iizuka et al., 2020; Zhang et al., 2020), yet investigations on the effectiveness of gestures on L2 pronunciation learning remain limited to a small set of languages and pronunciation features. Accordingly, one aim of the current study was to expand gesture training research to a phonetic feature that can easily be depicted with hand gestures but that has not yet received attention, French liaisons. Liaison is a process whereby a latent word-final consonant is produced when it precedes a vowel-initial word (e.g., "des [de] amis [ami]" is realized as [de‿zami] with the emerging liaison consonant [z]). Liaisons are a very common feature of the language, however, research shows that they are a difficult aspect of L2 French phonology. In this study, we explored whether pronunciation training with 'linking' hand gesture led to more accurate production of L2 French liaisons than training without gesture.

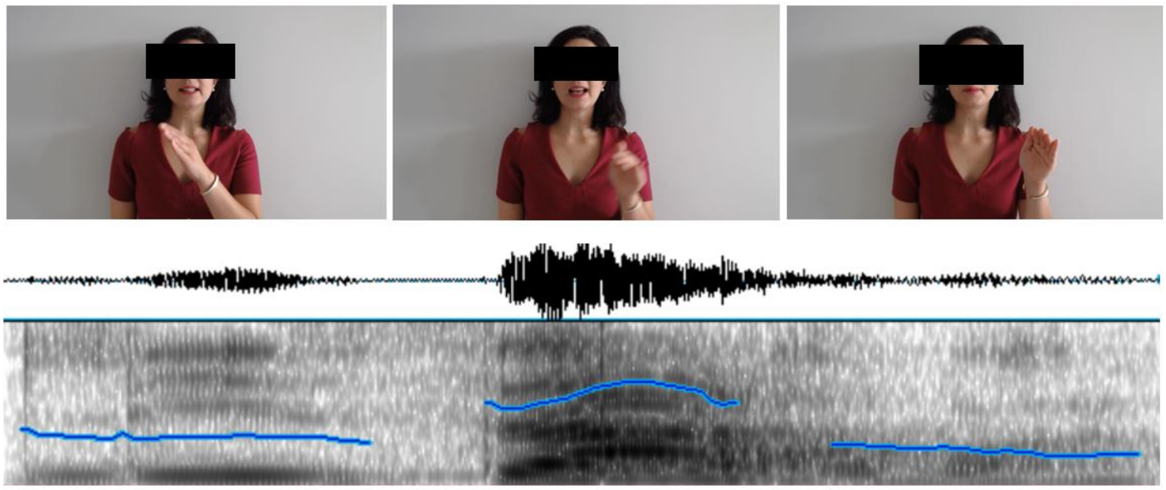
A total of 43 learners enrolled in a first-semester French course recorded a list of 16 words (+ distractors) containing liaisons at pre- and post-test, and 10 additional novel words at a generalization test. The productions were coded for accurate use of liaison consonants and timing. All learners received three weekly 5-min vocabulary/pronunciation video trainings in which they were presented with novel words produced by a female native speaker, followed by a screen with the written word in French and a translation in English. Each training session consisted of 30 words, half containing liaisons. In the Gesture condition (n=23), the speaker moved her hand sideways in a curved fashion (i.e., ‿) to highlight the liaisons (see Figure 1), whereas she remained still in the Non-Gesture condition (n=20).

Results showed production accuracy scores of 38% (Gesture) and 46.9% (NonGesture) at pre-test and 69.8% (Gesture) and 62.5% (NonGesture) at post-test. The interaction Training condition × Time was significant ($p < .001$), indicating that training using gestures enhanced the learning of French liaisons pronunciation. The positive effect of gesture was also observed at the generalization test, with the Gesture group producing liaisons in novel words (77%) significantly better than the Non-Gesture group (66%) ($p = .012$). These findings have strong implications for the pedagogical use of gesture in the teaching of L2 pronunciation.

Keywords: L2 pronunciation learning; co-speech gesture; L2 French

Figure

Figure 1. *Example of training material in the Gesture condition; word: “les arbres” (trees)*



References

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*(1), 33–58.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609–631.

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310.

Iizuka, T., Nakatsukasa, K., & Braver, A. (2020). The efficacy of gesture on second language pronunciation: An exploratory study of hand clapping as a classroom. *Language Learning*, *70*(4), 1054–1090.

Morett, L. M., & Chang, L. Y. (2015). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*(3), 347–353.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, *41*(1), 5–32.

Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, *24*(5), 666–689.

Stress in motion: Gesture-speech coupling in L2 lexical stress production

Hans Rutger Bosker*, *Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen*

Marieke Hoetjes*, *Centre for Language Studies, Radboud University, Nijmegen*

Wim Pouw*, *Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen*

Lieke van Maastricht*, *Centre for Language Studies, Radboud University, Nijmegen*

**shared first authorship among all authors; listed in alphabetical order*

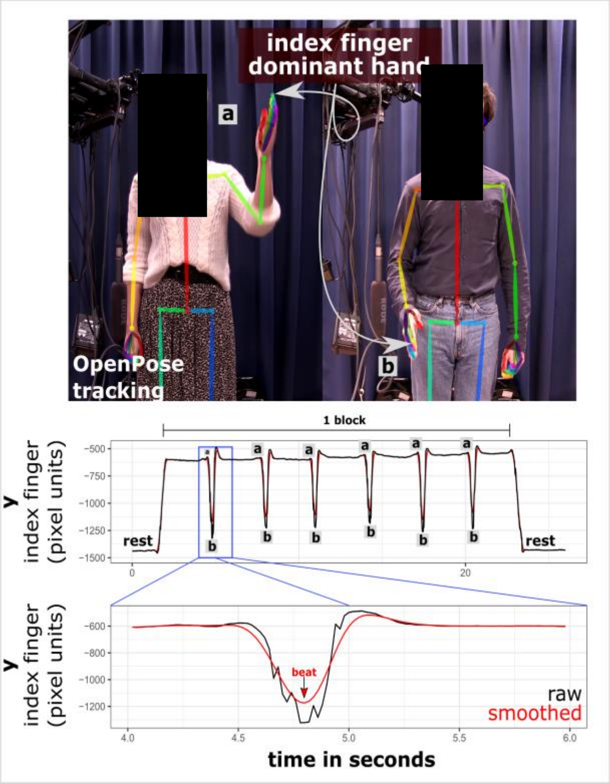
Second language (L2) prosody is difficult to learn, requiring the mastery of various nested multimodal systems, including articulatory and speech-synchronized gestural signals (e.g., Li & Post, 2014). Gesture-speech coupling is so natural that when asked to modulate production in one modality (e.g., by placing acoustic stress on a syllable or increasing hand movement amplitude), one unintentionally increases prominence in the other modality through biomechanical coupling — at least in one's native language (L1) (e.g., Esteve-Gibert & Prieto, 2013; Krahmer & Swerts, 2007). It remains unclear how gesture and speech are coordinated during L2 prosody acquisition, when L1 patterns often interfere. For instance, when a Dutch learner of Spanish wants to produce the cognate 'profeSOR' in Spanish (cf. Dutch 'proFESsor', capitals reflect lexical stress), they need to know that the stress falls on -sor (not -fes), but also temporally coordinate the articulatory movements required to produce stress on the target syllable and align co-speech hand gestures accordingly. In this study, we ask 1) How does gesturing influence the acoustic production of stress by L2 learners? And 2) How is gesture-prosody coupling in the L2 influenced by competition from the speaker's L1?

We conducted a production experiment in which 26 Dutch speakers (sample size defined based on power simulations) of L2 Spanish were video-recorded producing 48 stress-matching (Dutch and Spanish: 'MANgo') and 48 stress-mismatching cognates (Dutch: 'proFESsor', Spanish: 'profeSOR') in Spanish, once with explicit instructions to produce a beat gesture on the word, and once without (Figure 1). Acoustic analyses assessed whether producing a beat gesture helped L2 speakers to stress the target syllable and whether gesturing boosted the acoustic correlates of stress through biomechanic coupling. Motion-tracking and time-series analyses tested whether gesture-prosody synchrony was enhanced for stress-matching vs. stress-mismatching cognate pairs. Statistical analyses are ongoing; results will be ready to be presented at MMSYM (preregistration: osf.io/7dj54/). It is expected that gestural timing is biased either in the L1 or L2 direction (i.e., either boosting or hindering target like L2 prosody production), informing gesture-speech interaction and multimodal L2 acquisition theory.

Keywords: second language prosody; lexical stress; gesture-speech coupling

Figures

Figure 1. *Example OpenPose motion tracking*


[redacted]
[redacted]

Note. For each frame, we computed pose data and constructed a time series collecting information on the vertical position of the dominant index finger (50 Hz sampling rate). Middle panel: Time series for a single block, where the position of the index finger starts from rest, is raised to a start position (a) followed by 6 trials of gesture-speech utterances where a beat (b) is timed with a speech unit. Lower panel: Time series of a single trial showing the original estimate of OpenPose (black line) and the smoothed version of this motion trace (red line).

References

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*, 850-864. https://doi.org/10.1044/1092-4388(2012/12-0049)

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*, 396-414. https://doi.org/10.1016/j.jml.2007.06.005

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, *36*(2), 223-255. https://doi.org/10.1017/S0272263113000752

Mechanisms underpinning gestural facilitation of L2 word learning

Erin Minton-Branfoot, *University of Hull*

Richard O’Connor, *University of Hull*

Henning Holle, *University of Hull*

Vocabulary learning is one very challenging aspect of second language (L2) acquisition, especially for learners with limited access to L2 immersion. Iconic gestures may lend a helping hand here, by improving recall and recognition of L2 words (Macedonia, 2014). Previous research suggests that gestures outperform other cues such as L1 translations (Huang et al., 2019), pictures (Repetto et al., 2017) and meaningless, incongruent gestures (Kelly et al., 2009). However, since most studies presented gestures in combination with an L1 translation (but see Tellier, 2008), it is currently not clear whether this represents a unique gesture advantage.

The present series of four experiments had two goals. First, we wanted to see whether the gesture advantage persists when gestures are the only cue accompanying an L2 word. Second, we wanted to explore through which mechanism gestures boost L2 learning. There are two possible mutually non-exclusive explanations. The first possible explanation is disambiguation. The majority of studies presented isolated action words, which are semantically underspecified (e.g., *push* – this could refer to different contexts, push a button, push someone over.). Here, gestures may boost learning by naturally disambiguating the context of the word. Second, gestures may provide privileged access to action representations and motor traces during learning that are not provided by other learning methods. These representations may deepen the sensory motor image that accompanies the L2 word, capturing the meaning in an embodied way making them more memorable.

In **Study 1**, 45 participants with native or native-like proficiency of English learnt 24 Chinese verbs relating to manual actions (e.g., *dào*, meaning *pour*). Half of the words were taught using L1 translations, the other half were taught using short video clips of iconic gestures – presented *without* any L1 translations (e.g., *left hand in hold position, while right hand makes a pouring movement*). Participants’ learning was assessed by a written L2 to L1 translation task, both one day and seven days after learning. On both assessment days, there was greater learning in the gesture condition than in the translation condition. These results demonstrate that even with only one semantic cue, the gesture advantage is still present. If gestures boost learning through disambiguation, then the gesture advantage should disappear when the control condition also contains disambiguating information. This hypothesis was tested in **Study 2** (n=56) where the English words in the translation learning condition were accompanied with a short example in the infinitive form (e.g., to snap a stick). The gesture learning condition remained the same. We observed no significant differences between these two conditions on the number of correct translations. These findings don't appear to support the privileged access theory as equivalent learning was found despite only the gesture condition having this special capability. But they are in line with the disambiguation hypothesis as the gesture advantage disappeared once equal disambiguation was present across both conditions. In **Study 3** (n=43), we explored whether the effectiveness of gestures can be further enhanced by inclusion of an L1 word. We therefore compared a gesture-only with a gesture+translation condition. Results showed better learning for the gesture+translation condition possibly reflecting that gestures with translations provide an even clearer disambiguating context than gestures alone. Alternatively, it could be the case that multiple learning cues are simply more effective than single cues. **Study 4** (n=41) directly tested whether there is evidence for the privileged access account when experimental conditions are controlled for the amount of disambiguating information and number of cues. Greater learning was evident in the Gesture+translation condition than the translation+example learning condition.

In summary, our results extend existing knowledge by demonstrating that gesture-based learning outperforms simple translations-based learning during early L2 word acquisition. Disambiguation and privileged access seem to be two mechanisms that underpin this gesture advantage. More research is needed to see whether the effects reported here generalise beyond action verbs.

Keywords: Gestures; learning; vocabulary

References

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and Vocabulary Learning in a Second Language. *Language Learning*, *69*(1), 177–197. https://doi.org/10.1111/lang.12326

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334. https:/doi.org/10.1080/01690960802365567

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01467

Repetto, C., Pedroli, E., & Macedonia, M. (2017). Enrichment Effects of Gestures and Pictures on Abstract Words in a Second Language. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.02136

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Manipulating gestures in motion capture-animated characters for an L2 comprehension study

Valentijn Provė, *KU Leuven*

Bert Oben, *KU Leuven*

Henrik Garde, *Lund University, Humanities Lab*

Motion capture-animated characters are well-known for their extensive use in entertainment productions. The main advantage of these animations is that their physical appearance can be fully designed while using realistic body movements of human actors. It is precisely this type of control that has recently attracted researchers in the (digital) humanities (Karuzaki et al., 2021). A feature of particular interest for multimodal studies is that animated characters can be manipulated to create experimental conditions that differ in only one modality. To illustrate, there is a close temporal and semantic coordination involved in the production of speech and gestures, so we might expect that speech-gesture asynchrony is perceived as unnatural. However, it is almost impossible to empirically test this assumption because speech and gesture are difficult to disentangle in the human language production process. Therefore, controlled animations offer a great solution (Nirme et al., 2020). In a similar vein, this paper explores how kinematic properties of hand gestures can be manipulated in animations to create experimental conditions in a speech comprehension study.

Concerning the semantic speech-gesture integration, a hypothesis that requires experimental testing is that (a) gestures with an iconic dimension (McNeill, 1992) enhance semantic transparency and (b) L2 speakers benefit from this effect. Studies supporting this hypothesis have measured a better comprehension of isolated L2 words or sentences (Drijvers et al., 2019). However, in longer stretches of discourse, the effects are equivocal (Kamiya, 2022). In this regard, we identify two issues to be solved by using animated characters. First, in order to keep the variation in the speech modality constant, most studies have used audio or video (face only) versus video (full body) conditions, which is different from conditions with and without gestures being produced. Second, the conditions are always based on a binary distinction between the intensive use of multiple types of gestures and the total absence of gestures. As such, any observed effects cannot be directly attributed to a specific dimension of gesture. Our approach proposes formal manipulations of single gestures so that the measured effect in comprehension can be localized - as in the studies that confirm the hypothesis, but presenting them in a more ecologically valid context.

In our data we had a native speaker of Dutch act out a series of scripts in which a complex picture was described (duration = ca. 60 seconds per description). In each script, all gestures started and ended in the same rest position, contained full preparation, stroke and retraction phases, had a deictic and/or an iconic dimension and semantically corresponded to a lexical item that appeared in the picture. We obtained speech recordings and three-dimensional body/face motion capture data following the procedure outlined by Nirme et al. (2020) to render animated characters. Next, we manipulated the gestures in the animated characters to create more salient (i.e. spatially more articulated) and more reduced (i.e. more beat-like, McNeill, 1992, p. 81) conditions of the same picture description. These manipulations should lead to different degrees of iconicity, while keeping all other factors constant.

In this contribution we first present our method of manipulating the animated characters. This involves a detailed study of the three-dimensional gesture trajectories and the joints that need to be adjusted to render a naturally looking manipulation. Second, we present the results of a manipulation check that will allow us to gauge the naturalness of the characters in the different conditions. Because we want to use the manipulated characters as stimuli in an experiment (which is a follow-up study outside the scope of the present work), we want to make sure we are not confusing saliency with naturalness. In other words, we will test whether the manipulated gestures are perceived as equally natural as the non-manipulated ones. The results of this manipulation check and the implications for our methods of carrying out the manipulations will be discussed.

Keywords: motion capture; gesture; L2 comprehension

References

Drijvers, L., van der Plas, M., Özyürek, A., & Jensen, O. (2019). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *NeuroImage*, *194*, 55–67. https://doi.org/10.1016/j.neuroimage.2019.03.032

Kamiya, N. (2022). The limited effects of visual and audio modalities on second language listening comprehension. *Language Teaching Research*, *0*(0). https://doi.org/10.1177/13621688221096213

Karuzaki, E., Partarakis, N., Patsiouras, N., Zidianakis, E., Katzourakis, A., Pattakos, A., Kaplanidi, D., Baka, E., Cadi, N., Magnenat-Thalmann, N., Ringas, C., Tasiopoulou, E., & Zabulis, X. (2021). Realistic Virtual Humans for Cultural Heritage Applications. *Heritage*, *4*(4), 4148–4171. https://doi.org/10.3390/heritage4040228

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

Nirme, J., Haake, M., Gulz, A., & Gullberg, M. (2020). Motion capture-based animated characters for the study of speech–gesture integration. *Behavior Research Methods*, *52*(3), 1339–1354. https://doi.org/10.3758/s13428-019-01319-w

The effect of temporal alignment of speech and gesture on L2 speakers

Eleni Ioanna Levantinou, *Laboratory of Applied Psychology, Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece*

Argiro Vatakis, *Laboratory of Applied Psychology, Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece*

The integration of speech and gesture is a multisensory act that enables listeners to understand incoming messages using both auditory and visual information. Evidence from studies support that gestures play an important role in the production and comprehension of speech, the transmission of a message (McNeill, 1992), in second language acquisition (Gullberg, 2008), and in memory and recall (Cook et al., 2010). All the above, require the temporal alignment of gesture and speech between the anchor point of speech and the anchor point of gesture for the integration of the two inputs (Habets et al., 2011; Obermeier & Gunter, 2014). To date, however, there are no studies on how temporal alignment affects the integration of information in second language speakers. Thus, there are no evidence on how the temporal misalignment of speech and gesture can affect the process of learning and recall in a language different from the mother tongue. In the present study, therefore, we investigated the limits of temporal alignment of speech and gesture, in a working memory task with multilingual adults. We hypothesized that speech-gesture asynchronies (i.e., ±600, ±200, 0, where – gesture leads speech, while + gesture lags speech) will define integration levels and, thus, affect recall with large speech-gesture asynchronies negatively affecting memory as compared to synchronous conditions. To this purpose, videos were created, where a bilingual actor was performing words accompanied by iconic and beat gestures and words alone. The asynchronies were generated using Adobe Premiere tool by separating gesture from speech. Each word (standardized by frequency, syllables and stress) was edited in every asynchrony. 6 Conditions were designed (Iconic gestures-Beat gestures-No Gestures, in Greek and English) and each condition was comprised of 15 words. Participants were divided in two groups accordingly to their level and everyday usage of the second language in order to investigate also if the level of language acquisition will affect the task. Stimuli were presented to participants at various levels of asynchronies in a 3-back task and the participants had to decide if they have seen the same video 3 rounds before. Prior to the experiment there was a training session of 6 stimuli. Stimuli and conditions were presented counterbalanced. A between and within subject analysis of 21 participants (3 excluded as outliers) showed higher recall of mother tongue as composed to second language stimuli and iconic gestures were found to be more supportive in recall than beat gestures and words alone.

These results are consistent with studies that have investigated memory recall on second language speakers (e.g., Andrä et al., 2020). No significant differences were obtained between the participant's groups, which could be due to the overall familiarization of the participants with the English language. Recall was higher within the integration window which occurs from -200 to +200msc and this comply with relevant studies (e.g. Obermeier & Gunter, 2014). On the borders of this window (-200 and +200msc), recall was most negatively affected compared with all the other conditions. This probably could be due to the confusion caused by the difficulty of the participants to have a clear view whether or not will integrate the two stimuli or not. Outside this window (-600), recall performance was also significantly lower which might be because no integration takes place. Overall, the experiment confirmed the hypothesis that speech-gesture asynchronies compared to synchronous conditions negatively affect memory and recall on second language speakers. More research is needed, however, to clarify more aspects of the temporal relation of speech and gesture on second language.

Keywords: Gestures; speech; temporal coordination; second language

References

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educational Psychology Review*, *32*(3), 815-850.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of memory and language*, *63*(4), 465-475.

Gullberg, M. (2008). Gestures and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition* (pp. 276-305). Routledge.

Habets, B., Kita, S., Shao, Z., Özyurek, A., &Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845-1854.

McNeill, D. (1992). Hand and Mind. *Advances in Visual Semiotics*, 351.

Obermeier, C., & Gunter, T. C. (2014). Multisensory integration: The case of a time window of gesture–speech integration. *Journal of Cognitive Neuroscience*, *27*(2), 292-307.

DAY 3. SESSION "MULTIMODAL ANNOTATION & CORPUS"

MUNDEX: A multimodal corpus for the study of the understanding of explanations

Olcay Türk, Petra Wagner, Hendrik Buschmeier, Angela Grimminger, Yu Wang & Stefan Lazarov

Introducing the 3MT_French Dataset

Beatrice Biancardi, Mathieu Chollet & Chloé Clavel

Building a Chinese Sign Language (Shanghai) Corpus: a progressive report

Huan Sheng, Hao Lin & Yan Gu

Multimodal Annotation: Investigating referentiality and gesture meaning signaling

Ada Ren-Mitchell & Stefanie Shattuck-Hufnagel

MUNDEX: A multimodal corpus for the study of the understanding of explanations

Olcay Türk[1], Petra Wagner[1], Hendrik Buschmeier[1], Angela Grimminger[2], Yu Wang[1], Stefan Lazarov[2]

[1] Bielefeld University

[2] Paderborn University

Interlocutors invariably monitor each other's multimodal behaviour for cues of understanding (multimodal feedback), which are used to adapt the interaction to meet their partner's communicative needs. It should be possible to use these multimodal cues to monitor the level of understanding of a recipient moment-by-moment in co-constructed interactions. The precise interplay of these cues, their functions, their structural distribution and how their use varies between individuals remain largely not well understood.

The *MUltimodal UNderstanding of EXplanations* (MUNDEX) corpus is built to investigate these multimodal interaction dynamics within ongoing explanations. The explanation scenario involves a speaker (the explainer) explaining how to play a board game to a recipient (the explainee). To study individual variation in multimodal feedback use while also avoiding confounding with individual explanation strategies of the explainer, one explainer engages in explanatory dialogues with three different explainees one after another. The interactions are filmed in a professional studio using six camera perspectives (1920x1080, 50fps) and multiple dedicated microphones. This enables the separation of audio sources and the capturing of most bodily movements (facial expressions, head, hands, and torso movements), allowing the collection of a rich multimodal repertoire. Further, to associate these multimodal expressions with different levels of understanding, both interaction partners carry out a retrospective thinking aloud task (i.e., a video recall task) after the interaction session. In this task, the explainees comment on their state of understanding, and the explainers on their belief on the explainees' level of understanding. Time-alignment of these auto-assessments with the multimodal data yields information on regions of (non-)understanding, ultimately informing statistical and computational models.

The corpus will consist of 90 dialogues between German native speaker dyads (30 explainers, 90 explainees, aged 18-40, not controlled for gender) and their corresponding video recall tasks, totalling roughly 150 hours of data. It will contain the semi-automatic transcription of all speech (including utterance, word and syllable level segmentations), and automatic and manual multi-layered annotations of gesture and gaze behaviour, acoustic information, prosody and discourse. Semi-automatic gesture annotation relies on video-based motion-tracking using OpenFace and MediaPipe (Baltrusaitis et al., 2018; Lugaresi et al., 2019). With these, face, hand and upper body landmarks are estimated in order to be used for (1) the calculation of velocity, acceleration and jerk, (2) other signal related measurements (e.g., peak width and prominence), (3) the estimation of head nods and rotation, torso leans, blinking and gaze, hand positions in gestural space, (4) identification and categorisation of facial action units, and hand gesture type. For verbalisations, the MUNDEX corpus will contain acoustic information such as pitch, intensity, durations of segments, and voice quality (cepstral peak prominence) time-aligned to the framerate of the recordings (20ms at 50 fps). Moreover, a subset of data will be manually annotated using DIMA prosodic annotation guidelines (Kügler et al., 2019), which may be interpreted as cues to different levels of understanding.

The corpus will also be annotated for dialogue acts based on well-known manuals such as DIT++ (Bunt 2009). Here, dialogue segments will be tagged with various communicative functions defined based on their potential of affecting the information state of dialogue partners (including (non-)understanding), and with pragmatic relationships between these functions. In addition, these segments will also contain annotations of explanation phases (Roscoe & Chi, 1998) and which game rules they pertain to. Overall, this multidimensional classification provides a foundation for understanding the multi-functionality of dialogue segments, enabling investigations of relationships between fragments of explanations and multimodal behaviour coded in the MUNDEX corpus.

Ultimately, the MUNDEX corpus will provide a very rich resource for modelling the monitoring of (non-)understanding in human-machine interaction, but also for the in-depth understanding of how human interlocutors with their individual strategies are able to monitor and scaffold processes of understanding dynamically in an interactive fashion.

Keywords: Multimodal, explanation, understanding, feedback

References

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition*, 59–66. DOI: 10.1109/FG.2018.00019

Bunt, H. (2009). The DIT++ taxonomy for functional dialogue markup. *Proceedings of the AAMAS 2009 Workshop `Towards a Standard Markup Language for Embodied Dialogue Acts'*, 13–23.

Kügler, F., Baumann, S., Andreeva, B., ... & Wagner, P. (2019). Annotation of German intonation: DIMA compared with other annotation systems. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1297-1301.

Lugaresi, C., Tang, J., Nash, H., ... & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*.

Roscoe, R. & Chi, M.T.H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science*, *36*(4): 321-350. DOI: 10.1007/s11251-007-9034-5

Introducing the 3MT_French Dataset

Beatrice Biancardi, *LINEACT CESI, Nanterre, France*

Mathieu Chollet, *School of Computing Science, University of Glasgow, Glasgow, U.K.*

Chloé Clavel, *LTCI, Télécom Paris, IP Paris, Palaiseau, France*

Public speaking constitutes a real challenge for a large part of the population: estimates indicate that 15 to 30% of the population suffers from public speaking anxiety (Tillfors & Furmark, 2007).

Several existing corpora were previously used to model public speaking behavior. Those created ad-hoc for research purposes (e.g., Wörtwein et al., 2015) often provide a limited number of speakers, and are collected in an experimental setting without a real human audience. In monologues (e.g., Chen et al., 2017), the interaction with the audience is mostly asynchronous, and they are collected in the context of job interviews so the annotations are focused on hireability. TED Talks are a great resource but risk to contain mostly high-quality presentations, making it difficult to investigate the behaviors related to low-quality speeches or to anxious speaking behavior. Moreover, the videos are relatively long and the annotation protocol quite complex.

In most public speaking datasets, judgements are given after watching the entire performance, or on thin slices randomly selected from the presentations (e.g., Chollet & Scherer, 2017), without focusing on the temporal location of these slices. This does not allow to investigate how people's judgments develop over time during presentations, under the perspective of socio-cognitive theories such as Primacy and recency (Ebbinghaus, 2013) or first impressions (Ambady & Skowronski, 2008).

To provide novel insights on this phenomenon, we present the 3MT_French dataset. It contains a set of presentations of PhD students participating in the French edition of 3-minute Thesis competition. The jury and audience prizes have been integrated with a set of ratings collected online through a novel annotation scheme and protocol. Global evaluation, persuasiveness, perceived self-confidence of the speaker and audience engagement were annotated on different time windows (i.e., the beginning, middle or end of the presentation, or the full video).

We aim at providing two types of contributions. First, the 3MT_dataset with its particular properties:

- A relatively large amount (248) of naturalistic presentations;
- The quality of the presentations is highly heterogeneous;

- The presentations all have similar duration (180s) and structure.

On the other hand, we also provide the following methodological contributions:

- A novel annotation scheme, which aims at providing a quick way to rate the quality of a presentation, considering the dimensions in common between other existing schemes;
- The annotations are collected for both the entire video and different time windows.

This new resource would interest several researchers working on public speaking assessment and training, as well as it will allow for perceptive studies, both under a behavioral and linguistic point of view. It will allow for investigating whether a speaker's behaviors have a different impact on the observers' perception of their performance according to when these behaviors are realized during the speech. The automatic assessment of a speaker's performance could benefit from this information by assigning different weights to segments of behavior according to their relative position in the speech. In addition, a training system could be more efficient by focusing on improving the speaker's behavior during the most important moments of their performance. The 3MT_French dataset is available here: https://zenodo.org/record/7603511#.Y_cgMXbMJPY

Keywords: corpus; public speaking; annotation scheme; first impressions; primacy-recency effect

References

Ambady, N., & Skowronski, J. J. (Eds.). (2008). *First impressions*. Guilford Press.

Chen, L., Zhao, R., Leong, C. W., Lehman, B., Feng, G., & Hoque, M. E. (2017, October). Automated video interview judgment on a large-sized corpus collected online. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 504-509). IEEE.

Chollet, M., & Scherer, S. (2017, May). Assessing public speaking ability from thin slices of behavior. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 310-316). IEEE.

Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of neurosciences*, *20*(4), 155.

Tillfors, M., & Furmark, T. (2007). Social phobia in Swedish university students: prevalence, subgroups and avoidant behavior. *Social psychiatry and psychiatric epidemiology*, *42*(1), 79-86.

Wörtwein, T., Chollet, M., Schauerte, B., Morency, L. P., Stiefelhagen, R., & Scherer, S. (2015, November). Multimodal public speaking performance assessment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 43-50).

Building a Chinese Sign Language (Shanghai) Corpus: a progressive report

Huan Sheng, *Shanghai International Studies University*

Hao Lin, *Shanghai International Studies University / Harvard University*

Yan Gu, *University of Essex / University College London*

China has 20 million deaf people while the accurate number of Chinese Sign Language (CSL) users is unclear (Lin, 2021). We are building an open-access, machine-readable, research and socially relevant Chinese Sign Language (Shanghai) corpus (Corpus CSL Shanghai, 2016-). The corpus is the first to document Chinese deaf signers' life stories, and their natural use of the Shanghai CSL. This abstract gives an overview and progress report of the project.

The overarching goals of our project are three-folds: (1) Documenting Shanghai CSL at different times; (2) Developing an online dictionary; (3) Building up a sustainable platform for CSL research, both synchronically and diachronically (e.g., inter-generational variations).

Our team has collected a corpus of narrative and spontaneous conversations, with topics ranging from their personal growth, school education, health, hospital visits, travelling, etc. So far, 110 deaf signers (aged 20–98) have participated in the project, amounting to about 60-hour video archive. All participants are local Shanghai deaf people in the deaf community. The glosses and Mandarin translations of signs are being coded through ELAN by deaf experts.

Despite the shortage of funding and human resources from time to time, we have finished about 30% of the basic annotations for conversations and about 40% for narratives. We have identified 8899 sign types out of 96478 sign tokens, and divided the sign tokens into two main categories: full-fledged lexicalized signs and non-signs. The non-signs include several types that cannot be regarded as sign words: gestures, finger-spelling, classifiers and others. Signs are roughly labelled with Part of Speech (PoS) including noun, verb, adjective, adverb, number, name, pronoun and functional signs. Functional signs refer to the words mainly functioning at the level of grammar, lacking concrete meanings. The followings are included: negators (e.g., NO, NOT, NOT-HAVE), WH question signs, quantifiers (e.g., SOME, ALL), and particles, which often function as discourse markers, like FORGET-IT, or polar question markers, like GOOD-BAD (see a summary in Table 1).

Second, based on part of our corpus data, we have developed the first version of an online Shanghai CSL dictionary Isigner (https://isigner.app/), which offers both IOS and Android apps. It consists of 4318 independent lemmas with some lexical information, including basic phonological information, parts of speech, etc. Examples are even offered for a small portion of the entries.

Third, the corpus provides rich materials for research, especially offering a view from a typologically different sign language than most western sign languages. We conducted corpus-based studies in which a large sample of CSL deaf signers' naturalistic interactions were recorded. Compared to elicitation materials, naturalistic data are more authentic and reliable, which also offer a "visible" context to differentiate form and function. A large corpus offers us additional vital information that cannot be obtained otherwise (e.g. distribution, frequency), as well as some social linguistic information that may affect language production (Johnston, 2012). For example, we have studied temporal expressions, timelines and number representations in CSL, and found that CSL has asymmetric timeline expressions such as the past is signed toward the back (98.63%) whereas the future is signed downward (81.45%). Only old signers sign the past to up, and the past-to-up mappings in CSL are disappearing among young signers (Lin & Gu, in prep), etc. In short, the corpus will contribute to academic research in better understanding the diversity of sign language and gestures, including the culture and cognition.

Keywords: CSL, corpus, frequency

Table 1. *Overview of types and tokens based on the corpus annotation*.

	PoS	TYPES	TOKENS
signs	NOUN	2296	19741
	VERB	2953	30863
	ADJECTIVE	571	8200
	ADVERB	366	8267
	NUMBER	375	3571
	NAME	518	2875
	PRONOUN	32	9110
	FUNCTIONAL	189	11248
non-signs	GESTURE	321	554
	FINGERSPELLING	222	837
	CLASSIFIER	332	354
	OTHERS	671	858
	TOTAL	8846	96478

References

Lin, H. (2021). Early Development of Chinese Sign Language in Shanghai Schools for the Deaf. *Frontiers in Psychology*, *12*, 2409

Lin, H., & Gu, Y. (2022). "Hold infinity in the palm of your hand." A functional description of time expressions through fingers based on Chinese Sign Language naturalistic data. *Language and Cognition*, 1-24.

Johnston, T. (2012). Lexical frequency in sign languages. *Journal of deaf studies and deaf education*, *17*(2), 163- 193.

Multimodal Annotation: Investigating referentiality and gesture meaning signaling

Ada Ren-Mitchell, *MIT Media Lab*

Stefanie Shattuck-Hufnagel, *Speech Communication Group, Research Lab for Electronics, MIT*

The 2022 ISGS9 revealed that a number of gesture researchers have observed traditionally-categorized non-referential beat gestures contributing systematically to the meaning of an utterance, particularly in concert with prosodic and morpho-syntactic cues. These findings highlight the importance of developing multi-modal labeling systems, such as the M3D system and others. To address the challenges of capturing these emerging findings, we invite gesture researchers of all backgrounds and interests to join a discussion of the challenges they have encountered in labeling these phenomena, and their strategies for adopting annotation systems, as a preliminary to possible ongoing discussions in future online sessions and workshops.

Keywords: Multimodal Annotation; Corpus; Semantic and Pragmatic meaning

References:

Kendon, A. (2017). Pragmatic functions of gestures. Gesture, 16, 157–175. http://doi.org/10.1075/gest.16.2.01ken

Lopez-Ozieblo, R. (2020). Proposing a revised functional classification of pragmatic gestures. *Lingua*, *247*, 102870. https://doi.org/10.1016/j.lingua.2020.102870

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, A., ShattuckHufnagel, S., & Prieto, P. (2021, February 24). The MultiModal MultiDimensional (M3D) labeling system. https://doi.org/10.17605/OSF.IO/ANKDX

Rohrer, P. L. (2022). A temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel MultiModal MultiDimensional (M3D) labeling system [Unpublished doctoral dissertation]. Universitat Pompeu Fabra & Nantes Université.

DAY 3. SESSION "PRAGMATICS"

Variation in gestural input related to prosodic phrasing in infant-directed interactions

Victoria Reshetnikova, Roy Hessels & Aoju Chen

How important is it? The role of hand gestures in managing attentional states

Schuyler Laparle

Role of smiles in topic transitions in French conversations

Mary Amoyal, Béatrice Priego-Valverde & Stéphane Rauzy

Laughter Mimicry in Parent-Child and Parent-Adult interaction

Chiara Mazzocconi, Kevin El Haddad, Benjamin O'Brien, Kübra Bodur & Abdellah Fourtassi

Visual cues, affective stance, and irony

Beatrice Giustolisi & Francesca Panzeri

Variation in gestural input related to prosodic phrasing in infant-directed interactions

Victoria Reshetnikova, *Institute for Language Sciences, Utrecht University*
Roy Hessels, *Experimental Psychology, Utrecht University*
Aoju Chen, *Institute for Language Sciences, Utrecht University*

Children learn language through interactions with their caregivers in a multimodal setting. Adults modify their speech and gestures when interacting with infants (for review see Crowe, 2016). Infants, in turn, are sensitive to such adaptations and make use of them (Hollich et al., 2005). Existing research on infant-adult interaction mainly focuses on acoustic properties of infant-directed speech (for review see de Boer, 2011), and their variation across different speakers (Broesch & Bryant, 2015; Ferjan Ramírez, 2022). Little is known about variation in co-speech gestures in infant-adult interaction.

This study therefore aims to answer the question of what the variation is like in infant-directed co-speech gestures related to prosodic phrasing. Live interaction between nine Dutch-speaking mothers and their 5- to 9-month-old infants were elicited in three daily activities: small talk, storytelling, and free play. Variation was operationalised as three types of differences concerning two articulators (i.e. hands, eyebrows) at both between- and within-speaker levels: phonological interconnectedness between intonational phrase (IP) boundaries and gesture types, temporal interconnectedness between IP boundaries and gesture peaks, and variation in gesture intensity peaks. Data annotation consisted of both manual annotation of hand gestures (Rohrer et al., 2021) and automatic annotation of eyebrow gestures using the facial behaviour analysis software OpenFace (Baltrušaitis et al., 2016).

Mixed-effect modelling (logistic regression and linear regression) was conducted on the association of gesture types and IP-final boundaries, distribution of gesture peaks and intensity of gesture peaks using the R package lme4 (Bates et al., 2015). Our analysis yielded evidence for similar phonological interconnectedness in infant-directed interaction and adult-directed interaction. That is, the same types of gestures tended to occur at IP-final boundaries in infant- and adult-directed speech. Moreover, temporal interconnectedness between IP boundaries and gesture peaks was different than previously reported for infant-directed gestures (De la Cruz-Pavía et al., 2020), i.e. gestures peaked closer to the end of an IP end than to its start. Finally, between-speaker variation in infant-directed gestures was observed for all three types of variation, whereas within-speaker variation was observed only for the temporal interconnectedness and variation in intensity peak of eyebrow gestures.

These results raise the question of how variation in infant-directed co-speech gesture influences early prosodic development. Follow-up research concerning the learning of prosodic phrasing in early infancy will be discussed in the presentation.

Keywords: infant-directed interaction; prosodic phrasing; co-speech gesture

References

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

de Boer, B. (2011). Infant-directed speech and language evolution. In K.R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 322-328). Oxford University Press.

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development*, *16*(1), 31-43.

Crowe, T. N. (2016). *Maternal gestural input to young children: The role of age, language ability, context, and pragmatic function.* University of Missouri-Columbia.

de la Cruz-Pavía, I., Gervain, J., Vatikiotis-Bateson, E., & Werker, J. F. (2020). Coverbal speech gestures signal phrase boundaries: A production study of Japanese and English infant-and adult-directed speech. *Language Acquisition*, *27*(2), 160-186.

Ferjan Ramírez, N. (2022). Fathers' infant-directed speech and its effects on child language development. *Language and Linguistics Compass*, *16*(1), e12448.

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child development*, *76*(3), 598-613.

Rohrer, P. L., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Gibert, N. E., Ren, P., Shattuck-Hufnagel, S., Prieto, P. (2021, February 24). *The MultiModal MultiDimensional (M3D) labeling system*.

How important is it? The role of hand gestures in managing attentional states

Schuyler Laparle, *Tilburg University*

With the acceptance of language as a fundamentally multimodal system, there is increasing interest in developing multimodal models of linguistic communication. Thus far, most attention has been paid toward integrated semantic models (Schlenker, 2020). Despite the well-established use of gesture to convey pragmatic and discourse structural meaning (Bavelas et al., 1992), there is surprisingly little work pursuing multimodal models of discourse structure. Working within a goal-oriented, question-based understanding of discourse structure (Roberts 2012), I argue for the legitimacy of such a pursuit, and the potential for such a model to strengthen both our understanding of discourse structure and interactive (i.e. pragmatic) gesture.

In the present work, I explore the multimodal expression of specification in face-to-face interaction, where *specification* simply means segments of discourse that contribute directly to achieving discourse goals. I focus on three interactive gestures, the palm-up open-hand (PUOH) gesture (e.g. Müller, 2004), the precision grip gesture (e.g. Kendon, 2004), and the containment gesture, which I group into a functional class of `presentational' gestures (see Figure 1). I show that these presentational gesture variants can be used in sequence to (i) introduce a topic for discussion, (ii) comment on that topic, and (iii) emphasize the importance of particular information to achieving discourse goals. The data presented is from a study of the co-occurrence of gesture with lexical discourse markers. This larger dataset consists of 350 examples, all from interviews and monologues on the American talk show *The Late Show with Stephen Colbert,* collected through the UCLA's Communication Studies Archive in collaboration with the Red Hen Lab.

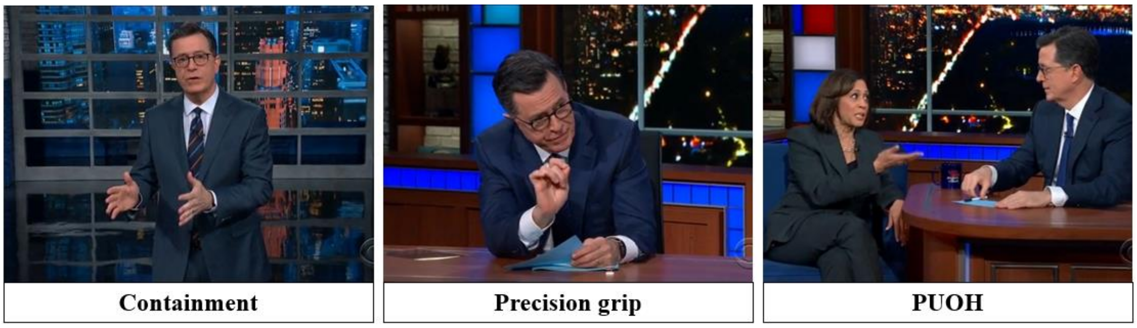
The discursive functions associated with each presentational gesture variant are at least partially motivated by the communicative metaphors they enact (Müller, 2017). All three variants present information as a metaphoric object, and each variant implies distinct metaphoric properties of the object presented. Different formal features are conducive to different use contexts and discursive functions The containment gesture is a two-handed gesture formed by open palms facing inward, as if to hold a rectangular object. The implied metaphoric object is medium sized with clearly bound edges and the potential to contain other metaphoric objects. These affordances are particularly conducive to introducing complex arguments and contrasting those arguments against others. The precision grip gesture is a one-handed gesture with fingers bunched as if to pinch a small object. The implied metaphoric object is small, potentially delicate, and requires close-inspection for identification. These affordances are conducive to emphasizing information as particularly important and worth careful consideration. The PUOH gesture is a one- or two-handed gesture with open palms facing upward as if to hold up a medium-sized object for inspection. These affordances are conducive to the initial introduction of discourse topics.

To demonstrate the discursive capacities of each variant, I look at sequences in which a speaker (i) uses multiple presentational gestures in succession, or (ii) repeats the same presentational gesture non-consecutively throughout a turn. I show that the repetition or succession of these presentational gestures reflects the intended attentional structure of the utterance – each gesture helps us, as interlocutors and discourse analysts, parse incoming information for what is new, what is important, and what to expect next in the discourse.

Keywords: interactive gesture; discourse structure; metaphor

Figures

Figure 1. *Types of presentational gesture*



Containment

Precision grip

PUOH

References

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse processes*, 15(4), 469-489.

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

Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family. *The semantics and pragmatics of everyday gestures*, 9, 233-256.

Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16(2), 277-304.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics*, 5, 1-69.

Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, 38(3), 887-936.

Role of smiles in topic transitions in French conversations

Mary Amoyal, *Aix Marseille Université, CNRS, Laboratoire Parole et Langage UMR 7309*

Béatrice Priego-Valverde, *Aix Marseille Université, CNRS, Laboratoire Parole et Langage UMR 7309*

Stéphane Rauzy, *Aix Marseille Université, CNRS, Laboratoire Parole et Langage UMR 7309*

Through the frameworks of Conversational Analysis (Sacks et al., 1974) and Interactional Linguistics (Selting and Couper-Kuhlen, 2001), this study explores the role of smiles in constructing topic transitions depending on the degree of acquaintance of interactants. During conversations, interactants address various themes and link them with "conversational movements" (Riou, 2015) called thematic transitions. Those moments are described in 3 phases (proposition, acceptation, and ratification) which enable a focus on how the participants use their "multimodal resources" (Birdwhistell, 1968) in the negotiation of the next subject under discussion. As it has been shown that facial signs are particularly observed by the recipient (Barrier, 2006) and that the speaker's smile helps to maintain the recipient's attention (Theonas et al., 2008), this study examines the role of smiles in the proposition and the acceptation phases of transitions. Since smile is described as a "facial gesture" (Bavelas and Gerwing, 2007) that also conveys interactive and pragmatic functions, it is interesting to explore the role of smile in the negotiation of transitions.

The objective of this study is twofold (1) analyze how participants use their smile in thematic transition; (2) explore this smile role depending on the participants' degree of acquaintance. This study has been conducted on two audio-video corpora of face-to-face French conversations. The 20 interactions analyzed brought together 40 participants for 8h: in the corpus CHEESE! (Priego-Valverde et al., 2020) the participants knew each other and in the corpus PACO (Amoyal et al., 2020) the participants met for the first time the day of the recording. The smile annotations were performed with a semi-automatic protocol. The SMAD tool (Rauzy and Amoyal, 2020), developed for the purpose of this study, provides automatic annotations of smiles according to the SIS scale (Gironzetti, Attardo, & Pickering, 2016). We then manually corrected and cross-coded our data.

The results show two major findings. Firstly, during the proposition phase of the transition, the speaker is more likely to drop his/her smile while introducing a new thematic than in any other random moment of the conversation ($\chi2 = 180.44$, DF = 78, $p < 0.01$). In CHEESE! the amplitude of this effect is 4.5 while in PACO it is 5.6. Secondly, when accepting the transition (the second phase), the recipient is more likely to increase his/her smile ($\chi2 = 111.96$, DF = 78, p < 0.01). In CHEESE! the amplitude of this effect is 2.3 while in PACO it is 2.6. These two results on both the speaker and the recipient provides additional evidence that smile shift is closely related to topic change. Moreover, when participants do not know each other, these configurations of smile shift (i.e., reduction of the smile in the proposition phase and increasing of the smile in the acceptation phase) are even more frequent, which suggests an effect of the relationship on their facial gesture. By describing the role of the smile in transitions, this study shows the importance of this facial expression. Smile should then be considered more frequently while analyzing the conversational process.

Keywords: smile; topic transition; conversation

References

Amoyal, M., Priego-Valverde, B. & Rauzy, S. (2020). PACO: A corpus to analyze the impact of common ground in spontaneous face-to-face interaction, In *Language Resources and Evaluation Conference*.

Barrier, G. (2006). *La communication non verbale. Comprendre les gestes: perception et signification*. Issy-les-Moulineaux : ESF Sciences Humaines.

Bavelas, J. B. & Gerwing, J. (2007). Conversational hand gestures and facial displays in face-to-face dialogue. In K. Fiedler (Ed.), *Frontiers of social psychology: Social communication* (p. 283-308). New York: Psychology Press.

Birdwhistell, R. L. (1968). L'analyse kinésique. *Langages*, (10), 101-106.

Gironzetti, E., Attardo, S., & Pickering, L. (2016). Smiling, gaze, and humor in, conversation. Dans L. In Ruiz-Gurillo, *Metapragmatics of Humor: Current research trends* (Vol. 14, p. 235). Amsterdam/ Philadelphia: John Benjamins Publishing Compagny.

Priego-Valverde, B., Bigi, B. & Amoyal, M. (2020). "Cheese!": A corpus of face-to face French interactions. A case study for analyzing smiling and conversational humor, In *Language Resources and Evaluation Conference*, Marseille, France.

Riou, M. (2015). A methodology for the identification of topic transitions in interaction. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics, and computational linguistics*, (16), 3-28.

Sacks, H., Schegloff, E. A. & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, (50), 696-735.

Selting, M. & Couper-Kuhlen, E. (2001). *Studies in interactional linguistics*. John Benjamins.

Theonas, G., Hobbs, D. & Rigas, D. (2008). Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments. *The International Journal of Virtual Reality*, *7*(1), 31-44.

Laughter Mimicry in Parent-Child and Parent-Adult interaction

Chiara Mazzocconi[1], Kevin El Haddad[2], Benjamin O'Brien[3], Kubra Bodur[1], Abdellah Fourtassi[4]

[1] *ILCB - LPL (UMR 7309), CNRS, Aix-Marseille University, France*

[2] *ISIA Lab, University of Mons, Belgium*

[3] *LIA (EA 4128), Avignon Université, France*

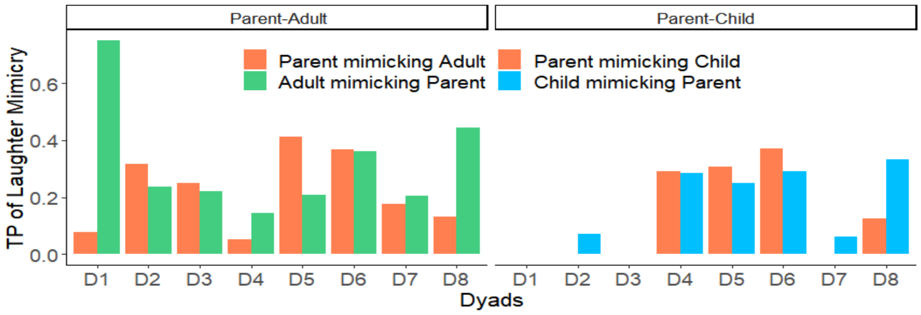
[4] *LIS, CNRS, Aix-Marseille University, France*

Laughter is informative about cognitive and pragmatic appraisals and its use and development begins in the first months of life. Adult studies show that the occurrence of laughter mimicry (i.e., laughter starting after a partner's laugh within 1 second from its offset– El Haddad et al., 2019) is influenced by context and interlocutor (Smoski & Bachorowski, 2003). Babies produce significantly less laughter mimicry in comparison to their caregivers (Nwokah et al., 1994). In comparison to adult-adult interactions, significant differences were also found in caregiver mimicry in response to child laughs over time, where high percentages were reported at initial time points, which subsequently decreased over time (Mazzocconi & Ginzburg, 2022). Less is known about laughter mimicry in middle-childhood. To fill this gap, the current study focuses on the analysis of caregiver-child interactions (6-11y/o) (ChiCo corpus - Bodur et al., 2021). The dataset is composed of video-recorded computer-mediated conversations (mean:17±3min) by 8 Parent-Child (PC) and Parent-Adult (PA - i.e. the parent of each PC dyad interacting with another adult) dyads, all engaged in the same guessing game. Two annotators identified 580 laughs (ELAN 6.4): 337 in PA interactions (per participant: 21±12) and 243 in PC interactions (110 C: 14±14; 133 P: 17±8). Wilcoxon-tests of frequency/minute between PC and PA conversations and between P and C were not significant. Given the variability in laughter production by participants, we measure mimicry in terms of Transitional Probability (TP), i.e. the probability of laughter mimicry given the total laughs produced by the partner. We observe consistently present laughter mimicry in all the PA dyads, however much higher variability in PC interactions (Figure 1). The overall TP means for PA and PC interactions are 0.27±0.17% and 0.14±0.14% (P: 0.13±0.16%; C: 0.16±0.14%) respectively. We observe significantly more laughter mimicry in PA conversations rather than PC ($\chi 2$ 39.82, df=7, p<.001), and significantly higher TP mimicry (W=103, p=0.03). We report no significant differences between P and C and between P laughter mimicry when interacting with their child or another adult. Despite comparable laughter occurrences between children and adults, laughter mimicry is overall significantly less frequent in PC interactions in comparison to PA interactions (the latter being similar to what was observed in adult face-to-face interactions –Mazzocconi et al., 2020). Coupled with the literature on younger babies, these observations suggest that for the caregiver, laughter responsiveness can dramatically change depending on the communicative development of the child and on the nature of the interaction. Children exhibit more laughter mimicry than babies (Nwokah et al., 1994; Mazzocconi & Ginzburg, 2022) and are more balanced in relation to the interlocutors. Our findings support evidence that laughter and its mimicry are not reflexive behaviours and are objects for learning, modulated by the context and the interlocutor. The results suggest that the use of some multimodal elements of communication continue developing through middle-childhood with other pragmatic skills (Cekaite, 2013). Temporal modulation analysis of laughter acoustic features will offer deeper insights on the differences observed in PA and PC interactions.

Keywords: laughter; mimicry; adult-child; multimodal communication development

Figures

Figure 1. *Transitional Probability of laughter mimicry.*



References

Bodur, K., Nikolaus, M., Kassim, F., Prévot, L., & Fourtassi, A. (2021, October). Chico: A multimodal corpus for the study of child conversation. In *Companion publication of the 2021 international conference on multimodal interaction* (pp. 158-163).

Cekaite, A. (2013). Child pragmatics development. *Encyclopaedia of Applied Linguistics*. Blackwell. https://doi.org/10.1002/9781405198431.wbeal0127

El Haddad, K., Chakravarthula, S. N., & Kennedy, J. (2019, October). Smile and laugh dynamics in naturalistic dyadic interactions: Intensity levels, sequences and roles. In *2019 International Conference on Multimodal Interaction* (pp. 259-263).

Mazzocconi, C., & Ginzburg, J. (2022). A longitudinal characterization of typical laughter development in Mother–child interaction from 12 to 36 months: Formal features and reciprocal responsiveness. *Journal of Nonverbal Behavior*, 1-36.

Mazzocconi, C., Tian, Y., & Ginzburg, J. (2020). What's your laughter doing there? a taxonomy of the pragmatic functions of laughter. *IEEE Transactions on Affective Computing*, *13*(3), 1302-1321.

Nwokah, E. E., Hsu, H. C., Dobrowolska, O., & Fogel, A. (1994). The development of laughter in mother-infant communication: Timing parameters and temporal sequences. *Infant Behavior and Development*, *17*(1), 23-35.

Smoski, M., & Bachorowski, J. A. (2003). Antiphonal laughter between friends and strangers. *Cognition and Emotion*, *17*(2), 327-340.

Visual cues, affective stance, and irony

Beatrice Giustolisi, *Department of Psychology, University of Milano-Bicocca*

Francesca Panzeri, *Department of Psychology, University of Milano-Bicocca*

Irony constitutes an apparent contradiction between what is said and the context in which the ironic comments have been uttered, and understanding what the ironist means requires the addressee not to stick to the literal interpretation of a remark. To signal their ironic intent, ironists may make use of several cues. Traditionally, the most studied ironic cues have been acoustic cues; however, highlighting the importance of multimodality for linguistic communication, great attention has been devoted also to the visual cues of irony. Ironic statements have been described as accompanied by a series of visual cues, e.g.: raised or lowered eyebrows, wide-open eyes, or squint (Attardo et al., 2003).

The main goal of the present work is to analyze further the visual cues of irony with an investigation that builds on previous work on the acoustic cues of irony. Having noted that there do not seem to be clear acoustical correlates that univocally characterize ironic criticisms and ironic compliments, Mauchand, Vergis & Pell (2020) investigated whether interlocutors would consider the ‘friendliness’ of speakers as a proxy for the correct detection of their communicative intent. They found general support for the Tinge Hypothesis (Dews et al., 1995) and for the Asymmetry of Affect Hypothesis (Clark & Gerrig, 1984): irony mutes the aggressiveness of criticisms but also attenuates the friendliness of compliments.

Inspired by these studies, our goal is to explore the role of visual cues in transmitting the friendliness of sincere and ironic speakers. To this aim, we presented participants with muted videos of speakers pronouncing literally positive and literally negative remarks sincerely (literal compliments and literal criticisms) and ironically (ironic criticisms and ironic compliments), and we asked them to rate their friendliness on a 5-point Likert scale. Our hypothesis was that in the absence of linguistic cues, raters will base their friendliness ratings on the actors’ facial expressions, which should be influenced both by content and attitude.

We expected friendliness ratings to be higher for comments with positive content than for content with negative content, and to be higher for comments with a positive attitude compared with comments with a negative attitude. An interaction between the two dimensions was also expected. As stimuli, we used 10 pairs of comments produced via a discourse completion task (see Giustolisi & Panzeri, 2021), in which the same remark was produced either sincerely or ironically by four different actors, two females and two males (80 items). We collected data online from 102 raters (76 females, 26 males, mean age=41 yrs, SD=16).

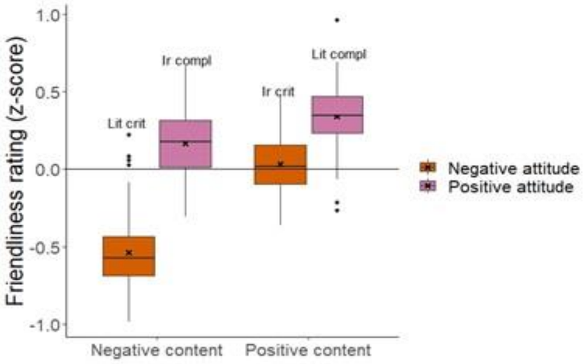
Collected ratings were transformed into z-scores based on each participant's mean rating and standard deviation. The analysis revealed that the effect of content, averaged across the levels of attitude was significant, and specifically comments with positive content received higher friendly ratings than comments with negative content (β = 0.372, SE=0.121, t=3.081, p = .015). The effect of attitude across the levels of content was also significant, with positive attitude yielding higher friendliness ratings than a negative attitude (β=0.505, SE=0.175, t= 2.887, p=.021). The interaction between content and attitude was not significant (p=.13) (Figure 1).

Our results on friendliness ratings based on visual stimuli are in line with that of friendliness ratings based on auditory stimuli, highlighting the multimodal dimension of the Tinge Hypothesis and of the Asymmetry of Affect Hypothesis: irony mutes the aggressiveness of criticisms but also attenuates the friendliness of compliments and this is revealed also by ironists facial expressions.

Keywords: irony; visual cues; friendliness

Figures

Figure 1. *Distribution of participants' mean z-scores across content (x-axis) and attitude conditions (colors). The straight line indicates the median, whereas the cross the mean.*



References

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor, 16(2),* 243–260. https://doi.org/10.1515/humr.2003.012

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General, 113(1),* 121–126. https://doi.org/10.1037/0096-3445.113.1.121

Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Processes, 19(3),* 347–367. https://doi.org/10.1080/01638539509544922

Giustolisi, B., & Panzeri, F. (2021). The role of visual cues in detecting irony. *Proceedings of Sinn Und Bedeutung*, 25, 292–306. https://doi.org/10.18148/sub/2021.v25i0.938

Mauchand, M., Vergis, N., & Pell, M. D. (2020). Irony, prosody, and social impressions of affective stance. *Discourse Processes, 57(2),* 141–157. https://doi.org/10.1080/0163853X.2019.1581588

DAY 1. POSTER SESSION

(In order of appearence in the program)

N	Authors	Title
1	Patrizia Paggio, Manex Aguirrezabal, Bart Jongejan, Costanza Navarretta and Leo Vitasovic	GEHM Network - Creating a Zoom corpus
2	Lara Southern, Tobias Deschner and Simone Pika	The role of multimodality in grooming interactions of chimpanzees in the wild
3	Kayla Kolff and Simone Pika	Communicative repair in nonhuman primates: The role of multimodality
4	Júlia Florit-Pons, Alfonso Igualada and Pilar Prieto	MultiModal Narrative (MMN): An inclusive multimodal narrative-based intervention for boosting preschoolers' narrative and pragmatic abilities
5	Alessandro Panunzi, Luca Lo Re and Valentina Saccone	The CECCO Corpus: a MultiModal Resource of Italian L1 Acquisition, First Surveys on Compositionality Phenomena
6	Ran Gong, Sotaro Kita and Bosen Ma	Gestures in Drawing Instructions by High Functioning Autistic Children
7	Martina Rossi, Maria Graziano and Margaret Zellers	Entrainment through different modalities: a pilot study on spontaneous conversation in Swedish
8	Aliyah Morgenstern, Stéphanie Caët, Claire Danet, Loulou Kosmala and Christophe Parisse	The role of gaze in the choreography of gestures, signs, speech and actions during French family dinners
9	Stefan Lazarov and Angela Grimminger	The relation between multimodal feedback and scaffolding of explanations
10	Malin Spaniol, Alicia Janz, Simon Wehrle, Kai Vogeley and Martine Grice	Investigating the relation between backchannels and gaze in dyadic conversations: A multimodal approach
11	Stamatina Rozou and Marianne Gullberg	Manual gestures of agreement in Greek conversation: The role of gender and familiarity
12	Ellen Fricke, Jana Bressem and Martin Siekfes	Modelling the interplay of speech, gestures and gaze: How empirical gesture studies, eye-tracking, and intensional logic work together in reconstructing joint attention and intention

N	Authors	Title
13	Niklas Krome	Multimodal gesture generation for Social XR
14	Clara Kindler and Jana Junge	Affective Stancetaking in Political Speeches, Some Insights in Processes of Multimodal Meaning-making in Media Specific Contexts
15	Fien Andries, Clarissa de Vries, Katharina Meissl, Kurt Feyaerts, Bert Oben, Paul Sambre, Myriam Vermeerbergen and Geert Brône	Stance-taking in the Visual Modality – a Systematic Literature Review
16	Isa Samira Winter, Kornélia Juhász, Andrea Deme and Reinhold Greisbach	Audio-visual perception of the vocalic rounding opposition in 4 languages
17	David Hernández-Gutiérrez, Romain Pastureau, Anastasia Klimovich-Gray, Mikel Lizarazu and Nicola Molinaro	Syntactic and semantic neural tracking in audiovisual discourse processing
18	Fanny Catteau and Claudia Savina Bianchini	Sub-parametric features of head movements and gaze conveying epistemicity: a study on French Sign Language (LSF) and French co-speech gestures
19	Marisa Cruz and Sónia Frota	Prosodic domains in the head of the signer
20	Sílvia Gabarró-López and Anna Kuder	Comparison across modalities: a case study of the “Away gestures” family in four sign languages
21	Yu Wang and Hendrik Buschmeier	An unsupervised method for the detection of head movements
22	Marion Blondel, Christelle Dodane, Karine Martel and Catteau Fanny	Multimodal focalization processes during French family dinners: a comparison between speaking and signing families
23	Xinyuan Liang	Prelinguistic Deictic Gesture and Co-Speech Sign Acquisition in Deaf Children
24	Joanna Wójcicka and Anna Kuder	Multimodality in linguistic feedback: a study of mirroring in Polish Sign Language
25	Anastasia Bauer, Jana Hoseman, Sonja Gipper and Tobias-Alexander Herrmann	Multimodal feedback signals: comparing response tokens in co-speech gesture and sign languages

The role of multimodality in the grooming interactions of chimpanzees in the wild

Lara Southern, *Osnabrück University*

Tobias Deschner, *Ozouga e.V.*

Simone Pika, *Osnabrück University*

The ability to produce and understand language makes humans unique (Fitch, 2010) yet whether the foundation for language evolution lies in the gestural or the vocal domain remains heavily debated (Corballis, 2002; Hockett & Hockett, 1960). More recently it has become apparent that humans both transmit and receive information through a multimodal lens, and combining the study of visual, audible and tactile sensory channels allows for a more holistic approach when investigating the evolution of this complex communicative system (Fröhlich, Sievers, Townsend, Gruber, & van Schaik, 2019). A growing body of evidence from comparative research demonstrates that non-human primate communication is also inherently multimodal (Slocombe, Waller, & Liebal, 2011). However, traditionally researchers have either studied vocal systems and underlying cognitive complexity or gestural interactions (Liebal, Waller, Burrows, & Slocombe, 2013).

Here, we offer new insights into the role and evolution of multimodality by focusing on communicative interactions of one of our closest living relatives, chimpanzees (*Pan troglodytes troglodytes*). Chimpanzees are known for their cognitive capabilities and their ability to use and combine different signal types both temporally and sequentially in their communicative interactions (e.g., Hobaiter & Byrne, 2014; Liebal, Call, & Tomasello, 2004; Pika, 2014). One such medium to examine these multimodal interactions is social grooming, a context in which chimpanzees mainly rely on short-distance signals, involving gestures and oro-facial sounds (Pika, 2014; Watts, 2016). Additionally, aside from its aesthetic function, grooming is known to maintain and strengthen affiliative relationships in chimpanzees (Mitani, 2009; Watts, 2000) and the role of rank and social relationship factor heavily into interaction choices (Newton-Fisher & Kaburu, 2017).

To investigate this topic, we studied grooming interactions across nine adult males in the Rekambo community in Gabon using recorded video footage. We categorized and coded each interaction part into four distinct categories: actions, oro-facial sounds, visual signals and tactile signals and examined their distribution within the interpersonal context of social grooming. We found that individuals use all sensory channels and combine signals and actions across initiating and response turns. Additionally, we found significant differences across dyads and the degree of variation, in not only the use of certain sensory channels but also the manner in which signals

and actions were combined, was explained by both rank and social bond strength. In conclusion, our findings shed light on the important role of a multimodal approach when investigating the broader topic of language evolution in our primate lineage.

Keywords: chimpanzee; interaction; multimodality; social relationships

References

Corballis, M. C. (2002). *From Hand to Mouth, the Origins of Language*. Princeton, New Jersey: Princeton University Press.

Fitch, W. T. (2010). *The Evolution of Language*: Cambridge University Press.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. *Biol Rev Camb Philos Soc, 94*(5), 1809-1829. https://doi.org/10.1111/brv.12535

Hobaiter, C., & Byrne, Richard W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596-1600. https://doi.org/10.1016/j.cub.2014.05.066

Hockett, C. F., & Hockett, C. D. (1960). The Origin of Speech. *Scientific American, 203*(3), 88-97.

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology, 64*, 377-396.

Liebal, K., Waller, B., Burrows, A., & Slocombe, K. (2013). *Primate Communication: A Multimodal Approach*. Cambridge: Cambridge University Press.

Mitani, J. C. (2009). Male chimpanzees form enduring and equitable social bonds. *Animal behaviour*, *77*, 633-640.

Newton-Fisher, N. E., & Kaburu, S. S. K. (2017). Grooming decisions under structural despotism: the impact of social rank and bystanders among wild male chimpanzees. *Animal behaviour*, *128*, 153-164. https://doi.org/10.1016/j.anbehav.2017.04.012

Pika, S. (2014). Chimpanzee grooming gestures and sounds: What might they tell us about how language evolved?. In *The Social Origins of Language: Early Society, Communication and Polymodality* (pp. 129-140). Oxford: Oxford University Press.

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal behaviour*, *81*(5), 919-924.

Watts, D. P. (2000). Grooming between male chimpanzees at Ngogo, Kibale National Park, Uganda. I. Partner number and diversity and reciprocity. *International Journal of Primatology*, *21*(2), 189-210.

Watts, D. P. (2016). Production of grooming-associated sounds by chimpanzees (*Pan troglodytes*) at Ngogo: Variation, social learning, and possible functions. *Primates*, *57*, 61-72. https://doi.org/10.1007/s10329-015-0497-8

Communicative repair in nonhuman primates: The role of multimodality

Kayla Kolff, *University of Osnabrück*

Simone Pika, *University of Osnabrück*

A prevalent feature of human sociality is engaging in conversations. These conversations are characterized by distinct features with interlocutors taking coordinated turns to communicate and establish common ground (Sacks et al., 1974). However, sometimes communication fails due to problems in hearing, speaking, or understanding the transferred information (Schegloff et al., 1977). Humans can correct these communicative failures through verbal mechanisms, e.g., the "huh? and "what?" expressions (Dingemanse et al., 2013; Enfield et al., 2013) and non-verbal mechanisms, e.g., freeze-look facial expression (Manrique & Enfield, 2015). These mechanisms are known as communicative repair. Based on apparency of the transferred information, unimodal mechanisms may be less successful compared to multimodal mechanisms. Going in line with the Multimodal Advantage Hypothesis, positing that multimodal communication is more effective compared to unimodal communication (Macuch Silva et al., 2020). Albeit, multimodal mechanisms remain understudied despite conversations being inevitably multimodal, especially during face-to-face interactions (Holler & Levinson, 2019; Vigliocco et al., 2014). In non-linguistic species, a limited number of studies have addressed communicative repair (e.g., Cartmill & Byrne, 2007; Genty et al., 2015; Haimoff, 1988; Leavens et al., 2005), yet solely focused on single modalities. Hence, relatively little is known about multimodal mechanisms and the evolutionary trajectory of communicative repair.

In this study, we investigated video footage of grooming interactions in one of our closest living relatives, chimpanzees (*Pan troglodytes schweinfurthii*) living in the Kibale National Park, Uganda. We focused on the failures that occurred during grooming interactions. In chimpanzee societies, grooming interactions are ubiquitous and occur at close distances where a multiplexity of signals can be at play, such as combinations of gestural signals (i.e., visual, auditory, and tactile gestures), or bimodal signals (e.g., combinations of vocalisations or oro-facial sounds with gestures). We present examples where chimpanzees modify or elaborate after their initial signal or use a multiplexity of signals while negotiating, to correct the failures during their grooming interaction, presenting potential repair mechanisms. We then offer an outlook to investigate how individuals correct failures in non-linguistic species to (i) shed light on the evolutionary trajectory of communicative repair, and (ii) to test the Multimodal Advantage Hypothesis (Macuch Silva et al., 2020).

Keywords: communicative repair; multimodality; chimpanzees

References

Cartmill, E. A., & Byrne, R. W. (2007). Orangutans modify their gestural signaling according to their audience's comprehension. *Current Biology*, *17*, 1345-1348. https://doi.org/10.1016/j.cub.2007.06.069

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is huh? a universal word. *PLoS One*, *8*.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., Drew, P., Floyd, S., Gipper, S., Gisladottir, R. S., Hoymann, G., Kendrick, K. H., Levinson, S. C., Magyari, L., Manrique, E., Rossi, G., Roque, L. S., & Torreira, F. (2013). Huh? What?–A first survey in 21 languages. In *Conversational repair and human understanding* (pp. 343-380). Cambridge University Press.

Genty, E., Neumann, C., & Zuberbühler, K. (2015). Bonobos modify communication signals according to recipient familiarity. *Scientific Reports*, *5*(1), 16442. https://doi.org/10.1038/srep16442

Haimoff, E. H. (1988). The organization of repair in the songs of gibbons. *Semiotica*, *68*(1/2), 89-120.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639-652.

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Development*, *76*(1), 291-306.

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science*, *7*(1), 182056.

Manrique, E., & Enfield, N. J. (2015). Suspending the next turn as a form ofr epai rinitiation: Evidence from Argentine sign language. *Frontiers in Psychology*, *6*(1326), 21. https://doi.org/10.3389/fpsyg.2015.01326

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*(4), 696-735.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361-382.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos Trans R Soc Lond B Biol Sci.*, *369*(1651): 20130292.

MultiModal Narrative (MMN): An inclusive multimodal narrative-based intervention for boosting preschoolers' narrative and pragmatic abilities

Júlia Florit-Pons, *Universitat Pompeu Fabra, Barcelona*
Alfonso Igualada, *Universitat Oberta de Catalunya, Barcelona*
Pilar Prieto, *Institució Catalana de Recerca i Estudis Avançats, Barcelona; Universitat Pompeu Fabra, Barcelona*

Narrative and pragmatic abilities are crucial during the preschool years, as they help further develop children's later linguistic, socio-communicative and academic performance (e.g., Demir et al., 2014; Dickinson & McCabe, 2001). Given the key role of these abilities in development, researchers have designed intervention programs for improving narrative and pragmatic skills. Nevertheless, most of them focus on improving either narrative skills (e.g., Spencer et al., 2015) or pragmatic skills (e.g., Kasari et al., 2010) and do not consider them together. Besides, to our knowledge, multimodality (e.g., embodied speech involving hand and bodily gestures, and facial expressions) has not been fully integrated into these existing interventions, despite some evidence of its beneficial role in development (see Hostetter, 2011; Vilà-Giménez & Prieto, 2021, for reviews). In order to assess the role of multimodality, a 3-week multimodal narrative intervention program (MultiModal Narrative, MMN) was designed. MMN trains narrative macrostructure (i.e., structural elements of the narrative) as well as pragmatic and perspective-taking abilities (i.e., understanding characters' perspectives and emotions) with the aid of multimodal strategies. MMN has been tailored to the needs of 93 preschool teachers and speech-language therapists for it to be implemented in the Catalan educational (and also clinical) contexts. Thus, our aim is to assess how the MMN intervention can improve children's narrative and pragmatic abilities in a preschool classroom context.

For this study, we have used a between-subjects pre- and post-intervention design, with 3 groups. Specifically, the experimental groups (i.e., multimodal and non-multimodal) received the MMN intervention, while the control group received treatment as usual. The MMN intervention consists of 9 30-minute sessions where the classroom teacher trains narrative macrostructure and perspective-taking through different strategies such as video cartoons, a storyteller, icons, and question-and-answer sequences. What differentiates the two experimental groups is that while in the non-multimodal group the teacher is asked to act naturally, the teacher in the multimodal group is trained to represent the main actions and emotions of the story and is also asked to encourage the children to embody the stories. Also, the multimodal group watches the video of the storyteller embodying the stories.

Participants were administered different tasks both at pre- and post-test: 2 narrative retelling tasks with different cartoons, and 2 tests for pragmatic skills (receptive pragmatics: PleaseApp; Andrés-Roqueta et al., 2020; expressive pragmatics: APT; Pronina et al., 2019). Part of participants' data has already been already collected (*n* = 47), while another part will be collected in the following 2 months (~*n* = 120-150).

Preliminary results with 47 children have shown a) that participants in the experimental groups significantly improved from pre-test to post-test in terms of receptive pragmatic skills and b) that only the children in the multimodal group were significantly better than those in the control group. However, no significant differences were found for expressive pragmatic skills or narrative macrostructure. Although these findings are still preliminary and will be complemented with data from more than 120 participants (full results will be presented at MMSYM), results seem to indicate that the full multimodal version of the MMN intervention has been more beneficial in boosting preschoolers' receptive pragmatic skills. All in all, initial evidence suggests that multimodality should be systematically introduced in educational interventions. Complementary evidence for this beneficial role will also be provided for the application of MMN in a clinical setting.

Keywords: multimodal intervention; narrative and pragmatic skills; preschool children

References

Andrés-Roqueta, C., Flores, R., & Igualada, A. (2020). *PleaseApp* [App].

Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., & Levine, S. C. (2014). Narrative processing in typically developing children and children with early unilateral brain injury: Seeing gesture matters. *Developmental Psychology*, *50*(3), 815–828. doi: 10.1037/a0034322

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice*, *16*(4), 186-202.

Hostetter, A. B. (2011). When Do Gestures Communicate? A Meta-Analysis. *Psychological Bulletin*, *137*(2), 297–315. doi: 10.1037/a0022128

Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized Controlled Caregiver Mediated Joint Engagement Intervention for Toddlers with Autism. *Journal of Autism and Developmental Disorders*, *40*, 1045–1056. doi: 10.1007/s10803-010-0955-5

Pronina, M., Hübscher, I., Vilà-Giménez, I., & Prieto, P. (2019). A new tool to assess pragmatic prosody in children: evidence from 3- to 4-year-olds. *Proceedings of the International Congress of Phonetic Sciences (ICPhS, 2019)*.

Spencer, T. D., Petersen, D. B., Slocum, T. A., & Allen, M. M. (2015). Large group narrative intervention in Head Start classrooms: Implications for response to intervention. *Journal of Early Childhood Research*, *13*(2), 196-217.

Vilà-Giménez, I., & Prieto, P. (2021). The value of non-referential gestures: A systematic review of their cognitive and linguistic effects in children's language development. *Children*, *8*(2). doi:10.3390/children8020148

The CECCO Corpus: a MultiModal Resource of Italian L1 Acquisition. First Surveys on Compositionality Phenomena

Alessandro Panunzi, *University of Florence*

Luca Lo Re, *University of Florence*

Valentina Saccone, *University of Florence*

In this work we present the CECCO Corpus, a multimodal resource based on audio-video recordings of three infants during their process of acquisition of Italian language as L1. The participants of this collection are three typically developing children (2 first-born boys and 1 first-born girl living in cohousing) unobtrusively videotaped once a month in their home for one year (13 sessions of about 1.5 hours each) in the same situation, i.e. during spontaneous play situations and interactions with a caregiver (the nanny). In the period of the recordings the children were respectively between 14-28, 18-30, and 20-33 months of age.

After the collection, we proceeded with the transcription of the recordings and the annotation of the main phenomena in communication development. The annotation tagset has been based on theoretical studies about language development and then verified on data. The tagset consists of 25 tags, including spoken phenomena (e.g., babbling, proto-words, horizontal/vertical repetition, placeholder, nuclear/extended sentence, etc.) and gestural ones (performative gesture, referential gesture, pointing).

The data are annotated through the software ELAN, with a specific template that includes three tiers for each child: (1) communicative event (performed action or uttered words); (2) phenomenon (a label in the tagset); (3) addressee (caregiver, another child, or per se).

After the corpus compilation, we performed some preliminary analysis of the combinatory forms emerging in the language development, following two main perspectives.

In the first one, we focused on cross-modal compositionality (Morgenstern 2014; Esteve-Gibert & Guellaï 2018). We distinguished oral only (compositional and non-compositional), gesture only (compositional and non-compositional), and cross-modal utterances (combining oral and gesture expressions). First results show that the younger child seems to be influenced by the older ones in developing compositionality at an earlier age (Özçalişkan & Goldin-Meadow 2005). In parallel, we found a decrease of cross-modal phenomena with performative gestures and subsequently the growth of cross-modal phenomena with pointing (Tomasello et al. 2007) between the 20th and the 28th month (even if at different ages for each child). In general, data show us the growing trend of linguistic compositional phenomena along with the growth of cross-modal phenomena with pointing (Capirci et al. 2005).

In the second one, we mostly focused on strictly oral compositionality, and specifically on the prosodic structure of the compositional utterances, in the framework of a perceptive analysis of intonation phenomena (t'Hart et al. 1990; Cresti & Moneglia 2010). We analyzed almost 300 complex dialogic turns with respect to both the number of words and prosodic form. We classified the utterances in 5 structures: (A) simple utterances with internal compositionality, i.e., single prosodic roots containing 2 or more words; (B) sequences of root units, each one containing a single word (separated by a terminal prosodic break, or linked by a non-terminal one); (C) complex utterances with a prefix + root prosodic pattern; (D) complex utterances with a root + suffix prosodic pattern; (E) sequences of 2 independent utterances, following the pattern (root) + (prefix+root).

As an initial outcome, we noticed that the overall tendency shows an incremental complexity within which the lexical and syntactic enrichment are reflected in the development of prosodic structures. These findings suggest that the relevant role played by the prosody is not related only to lexical production (Cavalho et al. 2018; Frota & Butler 2018), but it facilitates the emergence of the first oral compositionality already at 2 years old children.

Keywords: multimodal corpus; language development; compositionality phenomena

References

Capirci, O., Contaldo, A., Caselli, M. C. & Volterra, V. (2005). From action to language through gesture. *Gesture, 5*:1–2 (pp. 155–177). https://doi.org/10.1075/gest.5.1.12cap

Cavalho, A., Dautriche, I., Millotte, S. Christophe, A. (2018). Early perception of phrasal prosody and its role in syntactic and lexical acquisition. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 17–35). Amsterdam, John Benjamins. https://doi.org/10.1075/tilar.23.02car

Cresti, E. & Moneglia, M. (2010). *Informational patterning theory and the corpus-based description of spoken language*. Firenze University Press.

Esteve-Gibert, N. & Guellaï, B. (2018). The links between prosody and gestures: a developmental perspective. *Front. Psychol.* 9:338. doi: 10.1121/1.4986649

Frota, S. & Butler, J. (2018). Early development of intonation: perception and production. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language* (pp. 145–165). John Benjamins.

t' Hart, J., Collier, R. & Cohen, A. (1990). *A perceptual study on intonation. An experimental approach to speech melody*. Cambridge University Press.

Morgenstern, A. (2014). The blossoming of children's multimodal skills from 1 to 4 years old. In C. Müller, A. Cienki, S.H. Ladewig, D. McNeil & J. Bressem (Eds.), *Body - Language - Communication* (pp. 1848-1857). Mouton de Gruyter.

Özçalişkan, Ş. & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*, B101-B113.

Tomasello, M. & Carpenter, M., Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705-722.

Gestures in Drawing Instructions by High Functioning Autistic Children

Ran Gong, *ZheJiang University, University of Warwick*

Sotaro Kita, *University of Warwick*

Bosen Ma, *ZheJiang University*

Autism spectrum disorder is one of pervasive developmental disorders, characterized by persistent challenges in many social communication and social interaction domains, including gestural communication. An area of special importance for autistic individuals is the ability to gesture when knowledge status is asymmetrical between interlocutors. Successful communication in such a case requires autistic individuals to take the interlocutor's perspective and adapt their gestures accordingly. Previous research has well documented autistic children's perspective taking challenges manifested in verbal communication (Baron-Cohen et al., 1986; Tager-Flusberg et al., 2005) while very few studies investigated their gestural communication. So the present study will investigate whether autistic children are comparable to typically developing children in designing their gesture by taking the interlocutor's perspective.

A drawing instruction task was used for gesture elicitation. 18 6-year high functioning autistic children and 18 age-, gender-, full scale IQ- and receptive vocabulary-matched typically developing children described pictures presented on iPad or laptop to the interlocutor. The interlocutor could not see the pictures but was required to replicate the pictures on paper based on children's instruction. Both parties had no constraints on communicative behaviors. It thus provides a natural sample to investigate children's gesture behaviors.

We intend to investigate: (1) Do autistic children produce lower rates of pointing gestures, iconic gestures and tracing gestures than typically developing children? (2) Are autistic children more likely to produce gestures to the screen (vs. to the paper) than typically developing children? (The paper is visible to both the child and the interlocutor, but the screen of the iPad/laptop is visible only to the child.) (3) Are autistic children more likely to produce non-contact (vs. contact) gestures to the paper than typically developing children? (4) Are autistic children less likely to produce index finger (vs. open palm/fist) gestures to the paper than typically developing children? (5) Are autistic children less likely to change their communicative strategy after feedback from the interlocutor than typically developing children?(6) Is autistic children's speech less informative than typically developing children? The reason why we intend to investigate this question is that autistic children probably produce fewer gestures than typically developing children simply because their speech is more informative than typically developing children (i.e., They do not need to gesture to add information). Data results will be presented on the 1st International Multimodal Communication Symposium.

Keywords: gestures; high functioning autism; perspective taking; drawing instructions

References

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioral and international understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, 4: 113-125. https://doi.org/10.1111/j.2044-835X.1986.tb01003.x

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders: Diagnosis, Development, Neurobiology, and Behavior* (pp. 335–364). John Wiley & Sons Inc. https://doi.org/10.1002/9781118911389

Entrainment through different modalities: a pilot study on spontaneous conversation in Swedish

Martina Rossi, *Kiel University*
Maria Graziano, *Lund University*
Margaret Zellers, *Kiel University*

Speakers involved in face-to-face interactions tend to cooperate with each other and coordinate their conversational behaviors in order to achieve common communicative goals. This dynamic can result in conversational entrainment, i.e., in interlocutors adopting synchronized behaviors (Wynn & Borrie, 2022). Entrainment has been observed to occur in different modalities: in speech, though the repetition of lexical items or phrases and the use of similar phonetic patterns of variation between conversational partners, as well as in gesture, through the paralleling or mirroring of a gesture or some aspects of it produced by the previous speaker (Kimbara, 2006; Graziano et al., 2011; Levitan & Hirschberg, 2011; Holler et al., 2011; inter alia). However, studies on entrainment focus mostly on one modality, and investigations including more modalities are only a few, especially those involving lower linguistic levels, such as phonetics (Rasenberg et al., 2020). Therefore, this pilot study aims at observing how different modalities, more specifically parallel gestures and phonetic features, may interact with each other to achieve entrainment between speakers. In particular, we investigate whether, and to what extent, entrainment in one modality is accompanied by entrainment in the other modality.

Using the Spontal multimodal corpus (Edlund et al., 2010), we analyzed three face-to-face spontaneous conversations in Swedish, each involving two adult native speakers, for a total of 15 minutes of conversation. Dialogues were transcribed verbatim and subdivided into turns. We identified all instances of parallel gestures, i.e., those sequences where the speaker repeats, completely or partially a gesture previously made by his/her interlocutor (that is the gesture is very similar in form), independently of its function, and irrespective of the time lag from the original gesture. A total of 17 parallel gestures were found. Gesture annotation was conducted in Elan (Wittenburg et al., 2006). Subsequently, using Praat (Boersma & Weenink, 2022), we subdivided each conversation into equal sections of 1 minute each, in order to observe global entrainment. Then, for each section, we extracted F0 (in semitones) and intensity, both normalized with the speakers' individual means, in order to compare, for each dialogue, the values of the sections where parallel gesturing is present to those where it is not.

Preliminary results point to a cooperation of the two modalities in the achievement of entrainment, as well as a change in entrainment over time. In fact, we observe that, for the mean

F0 in one conversation and the mean intensity in another, speakers' values become more similar to those of the conversational partner in the sections where a parallel gesture is present, especially in sections closer to the end of the dialogue. In the third conversation, speakers entrain in intensity and parallel gesturing, but phonetic entrainment isn't connected to gestural behavior, as intensity values aren't more similar during parallel gesturing. Although it might seem in contrast with what observed in the other dialogues, this could be due to the different way in which gestures are performed by these two interlocutors: their parallel gestures mainly occur during silence (namely at the end of the speaker's utterance), and not together with speech, as in the other two conversations. It could also be the case that other phonetic features (not yet analyzed) might follow the entrainment pattern of the gestural modality.

Although preliminary, these observations suggest that when speakers entrain in conversation, they tend to do so across modalities. Our future investigation will examine more phonetic parameters (e.g., voice quality and speech rate), as well as more data. We will also consider the lexical content of turns and we will try to quantify entrainment using statistical methods.

Keywords: multimodal entrainment; parallel gestures; phonetic variation

References

Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.21, retrieved 1 October 2022 from http://www.praat.org/

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 22(6)*, 1482–1493.

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. *Proceedings of LREC 2010.*

Graziano, M., Kendon, A., & Cristilli, C. (2011). Parallel gesturing in adult-child conversations. In Stam, G. & Ishino, M. (Eds.) *Integrating gestures: The interdisciplinary nature of gesture* (pp. 89-101). John Benjamins.

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

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech 2011.*

Rasenberg, M., Özyürek, A. and Dingemanse, M. (2020), Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*: e12911.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006.*

Wynn, C. J. & Borrie, S. A. (2022). Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics*, *94*, 101-173.

The role of gaze in the choreography of gestures, signs, speech and actions during French family dinners

Aliyah Morgenstern[1], Stéphanie Caët[2], Claire Danet[3], Loulou Kosmala[4], Sophie de Pontonx[5], Lea Chevrefils[1], Christophe Parisse[5]

[1] *Sorbonne Nouvelle University*

[2] *Université de Lille*

[3] *Université Paris 8*

[4] *Université Paris Nanterre*

[5] *CNRS-Paris Ouest Nanterre*

Family dinners grounded in commensality are a collective ritual that plays a key role in family members' cultural heritage (Ochs & Kremer-Sadlik, 2013). They present a perfect opportunity to study the interweaving of language practices and actions in the framework of multiactivity (Haddington et al., 2014) and analyze the coordination of semiotic resources in their natural habitat. Family members collaboratively manage the accomplishments of multiple streams of activity through the embodied performances of dining and interacting (Goodwin, 1984). Because the subtle interweaving of *languaging* (Linell, 2009) and eating fully engages the body, family dinners also offer relevant affordances to study the semiotic differences between participants using a spoken language and a sign language. In this context, gaze plays a crucial role in the organization of these different resources, as it enables all family members to mark their engagement in the unfolding interaction (Goodwin & Goodwin, 1986), manage aspects of turn-taking and draw attention to specific gestures as they are performed in space.

A number of constraints are different for speaking and signing family members: using the mouth to eat and speak is problematic, and it is not easy to cut meat or pour water and be the active addressee of a signer. But there are possible activities one learns to combine - chewing can be synchronous with actively listening and gazing at the speaker or signer.

In order to study language specificities in a multi-activity set-up and more specifically the differences in the management of gaze, we conducted a study on French signing and speaking families. Our aim was to capture the subtle orchestration of the participants' bodies according to their age and to the language they use. We video-recorded dinners in middle-class families speaking French or signing in French sign language (LSF). The families, composed of two parents and two children aged 3 to 10, were filmed twice with two standard and one 360° cameras. The videos were synchronized and coded on ELAN. We annotated all participants' actions, gaze, and *languaging* throughout four dinners in LSF and four dinners in French.

In both signing and speaking families, all participants manage co-activity when needed but the youngest signing children spend less time *languaging* and acting at the same time. The hearing adults master the affordances of the visual and vocal channels to maintain the simultaneity of the two activities and the integration of all participants. The deaf adults skillfully manage to alternate smoothly between dining and interacting in a continuous flow. The deaf and hearing children manifest how they develop their skills to progressively manage multi-activity and multiparty conversations according to their age.

Our quantitative analyses specifically highlight differences in gaze management and uncover a variety of profiles according to participants' language modality and age. Among hearing families, adults might simultaneously monitor their activities without mutual gaze: they can direct their attention towards their current eating activity using their hands while interacting using their voice. Among deaf families, however, mutual gaze is crucial to maintain the simultaneity of the two activities: participants constantly need to secure the gaze of their addressees in order to interact in a continuous flow. Deaf parents can further socialize their children to co-activity thanks to gaze management.

Our study demonstrates how children become expert at coordinating semiotic resources within the framework of everyday experience and how all family members deploy a multitude of skillful multimodal variations, including the affordances of gaze management, in the collective coordination of bodies, activities and artifacts.

Keywords: Gesture; French Sign Language (LSF); Speech; Gaze; Co-activity

References

Goodwin, C. (1984). Notes on story structure and the organization of participation, In *Structures of Social Action: Studies in Conversation Analysis*, J. Maxwell Atkinson, John Heritage (Eds.), London, Cambridge University Press, pp. 225–246.

Goodwin, M., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62 (1–2), 51–76.

Haddington, P., Keisanen, T., Mondada, L., & Nevile, M. (2014). *Multiactivity in Social Interaction: Beyond multitasking*, Amsterdam/Philadelphia, Benjamins.

Ochs, E., & Kremer–Sadlik, T. (Eds.). (2013). *Fast–Forward Family. Home, Work, and Relationships in Middle–Class America*, Los Angeles, University of California Press.

The relation between multimodal feedback and scaffolding of explanations

Stefan Lazarov, *Paderborn University*

Angela Grimminger, *Paderborn University*

An explanation structure might not only be related to the explainers' organisation but also their monitoring of and adaptation to explainees' multimodal feedback (Clark & Krych, 2004). Explainers can provide explainees with additional guidance, also known as scaffolding (Wood et al., 1976), e.g., by making elaborations (Dingemanse et al., 2015). Multimodal signals, like gaze aversions, may indicate resolving difficulties in cognitive processing (Glenberg et al., 1998). Likewise, head nods may signal listeners' continuous engagement with speakers' storytelling (Stivers, 2008). However, little is known about the relationship between explainees' multimodal signals and explainers' structuring of explanations. Therefore, we address this in an explorative data driven analysis.

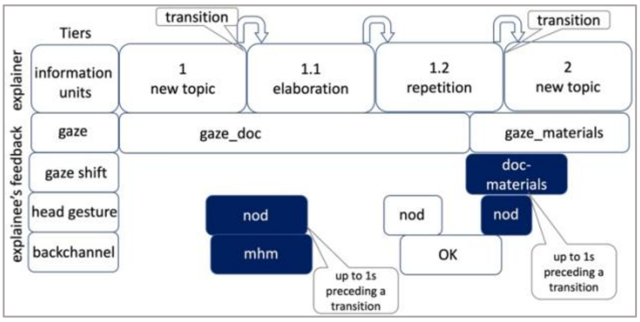
We analysed ten interactions in the health care domain between physicians and parents who were asked about giving an agreement for their children's upcoming surgery after an explanation. We segmented physicians' explanations into episodes of elaborations and topic changes (Roscoe & Chi, 2008) (Fig. 1) and annotated parents' gaze (static, shifting and averting from the interlocutor), head nods and backchannels. Because parents were continuously gazing, but only sporadically nodding and backchannelling, gaze was categorised as a primary signal split into three categories: unimodal, bimodal and multimodal, depending on co-occurrences with head nods and backchannels.

We calculated conditional probabilities of parents' feedback signals related to the physicians' topic structure. The analysis showed: 1) a higher probability of explainee's (multimodal) static gaze being followed by elaborations than topic changes; and 2) similar probabilities of parents' (multimodal) gaze shifts and aversions followed by topic changes and elaborations made by the doctors (Fig. 2). Our finding that gaze behaviour may disambiguate the interpretation of head nods (Stivers, 2008; Gander & Gander, 2020) and backchannels (Buschmeier & Kopp, 2014) contributes to previous studies on gaze aversions as signals of cognitive load (Glenberg et al., 1998; Morency et al., 2006) and turn management signals (Kendon, 1967; Jongerius, 2022). However, this analysis is limited to the explainees' multimodal behaviour, without taking the physicians' multimodal behaviour into account, which may also be related to parents' feedback, e.g. attention guiding via manual gestures.

Keywords: scaffolding; explanations; multimodal feedback

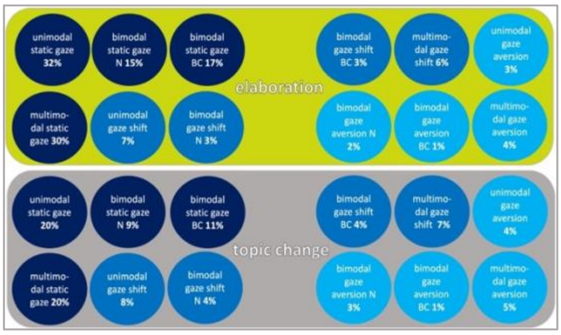
Figures

Figure 1. *Coding model*



Legend: ■ *selected feedback*

Figure 2. *Conditional probabilities*



Legend: N = nod, BC = backchannel
■ *static gaze,* ■ *gaze shift,* ■ *gaze aversion,* ■ *elaboration,*
■ topic change

References

Buschmeier, H., & Kopp, S. (2014). A dynamic minimal model of the listener for feedback based dialogue coordination. *DialWatt - SemDial 2014: Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue.* Edinburgh, UK, 17-25.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81.

Dingemanse M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal Principles in the Repair of Communication Problems. *PLoS ONE* 10:e0136100. https://doi.org/10.1371/journal.pone.0136100

Doherty-Sneddon, G., Phelps, F.G., & Calderwood, L. (2009). Gaze aversion during children's transient knowledge and learning. *Cognition and Instruction*, *27*(3), 225-238.

Gander, A. G., & Gander, P. (2020). Micro-feedback as cues to understanding in communication. *Dialogue and Perception – Extended Papers from DaP2018.* In C. Howes, S. Dobnik, & E. Breitholtz (Eds.) *CLASP Papers in Computational Linguistics* (pp. 1-11). Gothenburg University. http://hdl.handle.net/2077/63998

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*, 651–658.

Jongerius, C., Hillen, M. A., Romijn. J., A., Smets, E. M. A., & Koole, T. (2022). Physician gaze shifts in patient-physician interactions: functions, accounts and responses. *Patient Education and Counceling*, *105*(7), 1-14. https://doi.org/10.1016/j.pec.2022.02.018

Kendon, A. (1967). Some Functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22-63.

Morency, L. P., Christoudias, C. M., & Darell, T. (2006). Recognizing gaze aversion gestures in embodied conversational discourse. *Proceedings of the 8th International Conference on Multimodal Interfaces*. NY, USA, 298-294. https://doi.org/10.1145/1180995.1181051

Roscoe, R. & Chi, M.T.H. (2008). Tutor learning: the role of explaining and responding to questions. *Instructional Science*, *36*(4), 321-350.

Stivers, T. (2008). Stance, Alignment and Affiliation During Storytelling: When Nodding is a Token of Affiliation. *Research on Language and Social Interaction*, *41*(1), 31-57.

Wood, D., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89-100.

Investigating the relation between oral and visual feedback in dyadic conversations: a multimodal approach

Malin Spaniol[1], Alicia Janz[2], Simon Wehrle[2], Kai Vogeley[1], Martine Grice[2]

[1] *Department of Psychiatry, University Hospital Cologne, Germany*

[2] *IfL Phonetik, University of Cologne, Germany*

Very short utterances produced by listeners can function as feedback signals demonstrating understanding and acknowledgement. Such feedback can indicate passive recipiency (PR) or incipient speakership (IS). PR tokens support the ongoing turn of the interlocutor, while IS tokens signal the listener’s intention to take the floor. The properties of oral feedback tokens are complex, but show consistent evidence for the relationship between a token’s lexical form, its intonation contour, and its function (PR vs. IS; Sbranna et al., 2022). In face-to-face conversation, it can be assumed that other communicative channels, such as gaze, help to discriminate between PR and IS tokens. Only a small number of studies have investigated the relation between oral and visual feedback in the context of turn-taking to date.

Direct gaze by the speaker (entailing mutual gaze), creating a so-called “gaze window” (Bavelas et al., 2002), plays an important role in turn-taking (Auer, 2021) and has been proposed to function as backchannel-inviting (Skantze et al., 2014) and turn-yielding signal (Degutyte & Astell, 2021; Kendon, 1967). However, the precise interplay between turn-taking function, oral feedback, and the listener’s gaze is yet to be elucidated.

During both oral and visual feedback, the listener is generally expected to use more directed gaze than the speaker. However, some studies have reported averted gaze at the beginning of turns (Degutyte & Astell, 2021). As IS tokens exclusively occur at the beginning of turns, we can expect that direct gaze during IS tokens will be reduced compared to PR tokens.

We have developed a novel multimodal approach for studying dyadic face-to-face conversation, recording both eye-gaze (using mobile eye-tracking glasses) and speech. We measured oral feedback and gaze, in three different conversational contexts, in dialogues between 8 native speakers of German (four dyads). Speakers first engaged in an introductory conversation, followed by a task-based conversation (Tangram task) and a subsequent discussion thereof. We investigated if and how oral feedback and gaze complement each other during the production of PR and IS tokens. Directed vs. averted gaze was automatically coded using fixation detection and face detection. Speech data were annotated in *Praat*.

Our analysis revealed relatively low amounts of speaker-directed gaze during feedback production, contrary to expectations (IS: 34%, PR: 39%). Still, PR tokens involved slightly more speaker-directed gaze than IS tokens, as predicted. We also observed less speaker-directed gaze in the task-based dialogues (likely due to task demands). Further, we also found clear differences between dyads in the time spent producing oral feedback and the amount of directed gaze, independent of conversational context.

The setup introduced offers opportunities for enriching the study of multimodal communication, and in a second step, can make a contribution to related fields, such as the modelling of human–agent interaction.

Keywords: dyadic multimodal interaction, backchannels, social gaze, feedback, turn-taking

Figures

Figure 1. *Experimental setup.*

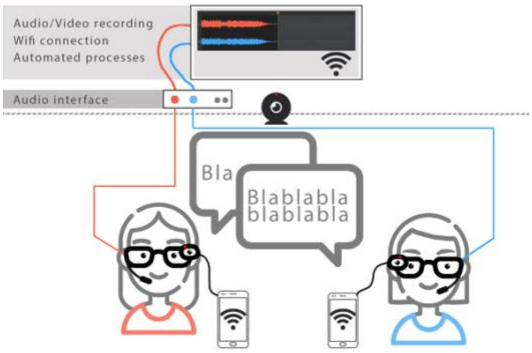
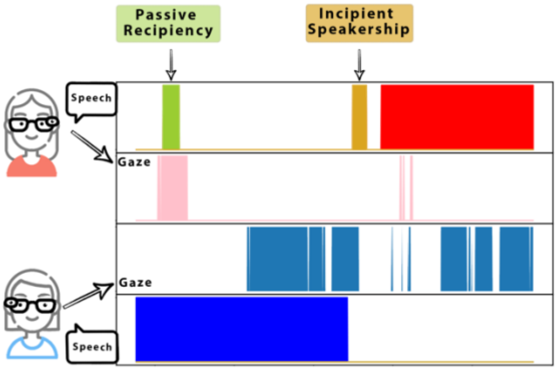


Figure 2. *Example activity plot showing speech and gaze of both interlocutors, featuring an IS token (in yellow) during which gaze is briefly averted, and a PR token (in green) with directed gaze.*



References

Auer, P. (2021). Turn-allocation and gaze: A multimodal revision of the "current-speaker-selects-next" rule of the turn-taking system of conversation analysis. *Discourse Studies*, *23*(2), 117–140.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, *52*(3), 566–580.

Degutyte, Z., & Astell, A. (2021). The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings. *Frontiers in Psychology*, *12*(April), 1–22.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63.

Sbranna, S., Möking, E., Wehrle, S., & Grice, M. (2022). Backchannelling across Languages: Rate, Lexical Choice and Intonation in L1 Italian, L1 German and L2 German. *Proc. Speech Prosody 2022*, 734–738.

Skantze, G., Hjalmarsson, A., & Oertel, C. (2014). Turn-taking, feedback and joint attention in situated human-robot interaction. *Speech Communication*, *65*, 50–66. https://doi.org/10.1016/j.specom.2014.05.005

Manual gestures of agreement in Greek conversation: The role of gender and familiarity

Stamatina Rozou, *Lund University, Lund, Sweden*

Marianne Gullberg, *Lund University, Lund, Sweden*

Agreement between interlocutors is an ordinary aspect of talk-in-interaction. Although previous research has examined agreement in speech (e.g., Makri-Tsilipakou, 1991; Myers, 1998; Pomerantz, 1984), the multimodal expression of agreement has received less attention. Those studies that have investigated multimodal expressions of agreement have predominantly focused on head movements, especially nodding (e.g., Fusaro et al., 2011; Guidetti, 2005; Helweg-Larsen et al., 2004). Less is known about manual gestures of agreement. Regarding gender and familiarity, it has been reported that females tend to gesture more both with familiars and unfamiliars, while males are generally more restrictive (Bente et al., 1998; Friesen et al., 1979; Helweg-Larsen et al., 2004). However, these studies have not examined agreement specifically. This study therefore investigates the expression of agreement asking a) what manual agreement gestures look like, and b) whether interlocutors' gender and familiarity with the speaker affects the frequency and form of such gestures.

We recruited 40 native Greek speakers (20 females) to participate in an elicitation task in which pairs of speakers discussed a set topic. The participants were distributed in eight different groups that each consisted of five speakers. One was selected as the target speaker (8 in total; 4 female) and was paired with two familiar and two unfamiliar interlocutors, one male, one female in each familiarity category (Fig. 1). Speech from 32 conversations was transcribed and agreement utterances were selected. The manual gestures that occurred within these utterances were identified and further coded for articulatory features: number of hands, movement, palm orientation and handshape.

The results show few differences between the genders in agreement speech. In gesture, both genders produce manual gestures in agreement with similar characteristics, mainly the Open Hand Palm Up and Palm Up Oblique gestures (Fig. 2) across all conditions. Turning to gender and familiarity, female speakers gesture more than male speakers in all conditions, as in previous studies. In addition, both female and male speakers produce most gestures with unfamiliar male interlocutors. The results provide new knowledge about multimodal expressions of agreement, and the findings on gender and familiarity suggest a need for further systematic studies to chart the influence of social factors on multimodal pragmatics.

Keywords: agreement; conversation; manual gestures; speech; Greek; gender; familiarity

Figures

Figure 1. *The internal structure of the groups. TS=target speaker, FaF=familiar female, FaM=familiar male, UF=unfamiliar female, UM=unfamiliar male.*

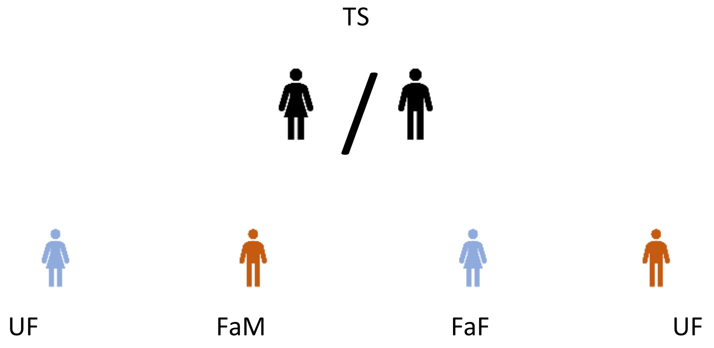
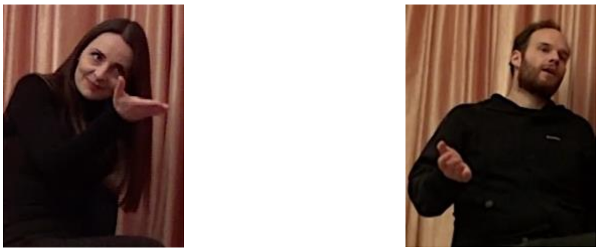


Figure 2. *a) Palm Up gesture as performed by a female target speaker. b) Palm Up Oblique gesture as performed by a male target speaker.*



References

Bente, G., Donaghy, W. C., & Suwelack, D. (1998). Sex differences in body movement and visual attention: An integrated analysis of movement and gaze in mixed-sex dyads. *Journal of Nonverbal Behavior*, *22*(1), 31-58.

Friesen, W. V., Ekman, P., & Wallbott, H. P. (1979). Measuring hand movements. *Journal of Nonverbal Behavior*, *4*, 97-112.

Fusaro, M., Harris, P. L., & Pan, B. A. (2011). Head nodding and head shaking gestures in children's early communication. *First Language*, *32*(4), 438-458.

Guidetti, M. (2005). Yes or no? How young French children combine gestures and speech to agree and refuse. *Journal of Child Language*, *32*(4), 911-924. https://doi.org/10.1017/S0305000905007038

Helweg-Larsen, M., Cunningham, S. J., Carrico, A., & Pergram, A. M. (2004). To Nod or Not to Nod: An Observational Study of Nonverbal Communication and Status in Female and Male College Students. *Psychology of Women Quarterly*, *28*(4), 358-361. https://doi.org/10.1111/j.1471-6402.2004.00152.x

Makri-Tsilipakou, M. (1991). *Agreement/disagreement: Affiliative vs. disaffiliative display in cross-sex conversation* [Doctoral dissertation]. Aristotle University of Thessaloniki.

Myers, G. (1998). Displaying opinions: Topics and disagreement in focus groups. *Language in Society*, *27*(1), 85-111.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson and J. Heritage (Eds.), *Structures of social action* (pp. 57-101). Cambridge: Maison des Sciences de 1'Homme and Cambridge University Press.

Modelling the interplay of speech, gestures and gaze: How empirical gesture studies, eye-tracking, and intensional logic work together in reconstructing joint attention and intention*

Ellen Fricke, *Technical University Chemnitz*

Jana Bressem, *Technical University Chemnitz*

Martin Siefkes, *Technical University Chemnitz*

* This work was funded by the German Research Foundation – Project-ID 416228727 – SFB 1410

Embodied digital technologies are becoming more and more actors in public space. In such hybrid settings, the mutual understanding of intentions and establishing joint attention are crucial. A range of studies highlights that in inter-human communication and human-machine encounters this is shaped and brought about by an interplay of verbal and bodily signs, particularly by verbal deixis, gestural pointing, and gaze (e.g., Brône, & Oben, 2018; Fricke, 2007; Rennert, Pfeiffer, & Wachsmuth, 2014; Staudte et al., 2011; Stukenbrock, 2020; Tomasello, 2008). However, existing studies on human-machine interaction usually base their implementation on experimental setups focusing on tasks with a clear goal (e.g., sandwich making) with mostly one participant.

The present contribution aims to address this gap. By combining empirical gesture studies from a linguistic point of view (Bressem et al. 2013; Fricke 2012) with experimental studies using eye-tracking data, and the formalization of complex sign processes based on intensional logic from a semiotic point of view (Posner, 1993; Siefkes, Fricke, Bressem, & Charoensit 2023), the paper presents a formal approach to support analyses and design processes of complex structures of intending in hybrid interaction scenarios. Using video and eye-tracking data, in which 15 dyads of participants interacted with different digital exhibits in a museum, we show that speech, gestures, gaze, and other bodily behavior establish joint attention and indicate intentions, and that each modality carries a specific relevance in this process. Our data indicate that gaze achieves a particularly relevant function in this process: Gaze alone can be responsible for indicating the communicative intention of a speaker. Moreover, the communicative situation and task influences joint attention and gaze patterns: Even the use of verbal deictics along with a pointing gesture might be disregarded because of the more pressing task of object manipulation. This incongruence of gaze and head direction might be an indicator for multiple attentional targets that result in particular patterns. Moreover, we will show that a formal description, grounded in Posner's (1993) concept of believing, causing, intending provides a unified means for reconstructing joint attention and intention and allows for processes of mutually ascribing

appropriate belief-and-intention configurations on different levels of complexity that may lay the grounds for its later implementation in human-machine interactions.

Keywords: multimodality; joint attention; human-machine interaction

References

Bressem, J., Ladewig, S. H., & Müller, C. (2013). Linguistic Annotation System for Gestures (LASG). In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. (Handbooks of Linguistics and Communication Science 38.1.)* (pp. 1098-1125). De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1098

Brône, G., & Oben, B. (Eds.). (2018). *Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue* (Vol. 10). John Benjamins Publishing Company. https://doi.org/10.1075/ais.10

Fricke, E. (2007). *Origo, Geste und Raum: Lokaldeixis im Deutschen*. Walter de Gruyter. https://doi.org/10.1515/9783110897746

Fricke, E. (2012). Grammatik multimodal: wie Wörter und Gesten zusammenwirken (Vol. 40). Walter de Gruyter. https://doi.org/10.1515/9783110218893

Posner, R. (1993). Believing, causing, intending: The basis for a hierarchy of sign concepts in the reconstruction of communication. In *Signs, Search and Communication* (pp. 215-270). Walter de Gruyter.

Renner, P., Pfeiffer, T., & Wachsmuth, I. (2014). Spatial references with gaze and pointing in shared space of humans and robots. In *Spatial Cognition IX: International Conference, Spatial Cognition 2014, Bremen, Germany, September 15-19, 2014. Proceedings 9* (pp. 121-136). Springer International Publishing. https://doi.org/10.1007/978-3-319-11215-2_8

Siefkes, M., Fricke, E., Bressem, J., & Charoensit, A. (2023). Modelling intentional complexity in hybrid interaction scenarios beyond explicit and implicit communication. *CRC Conference "Hybrid Societies"*.

Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, 120(2), 268-291. https://doi.org/10.1016/j.cognition.2011.05.005

Stukenbrock, A. (2020). Deixis, meta-perceptive gaze practices, and the interactional achievement of joint attention. *Frontiers in Psychology*, 11, 1779. https://doi.org/10.3389/fpsyg.2020.01779

Tomasello, M. (2008). *Origins of human communication*. MIT press.

Multimodal Gesture Generation for Social XR

Niklas Krome, *Social Cognitive Systems Group, Bielefeld University*

Stefan Kopp, *Social Cognitive Systems Group, Bielefeld University*

Human communication is a complex and nuanced process. Creating virtual avatars for interaction in Social XR requires a faithful recreation of the corresponding behaviour. While verbal behaviour can easily be recorded via microphone, capturing non-verbal behaviour requires complex and expensive motion-capture setups. This led to the invention of generative systems, that infer non-verbal behaviour, more specifically co-speech gestures, from the accompanying verbal behaviour. These gesture generation systems take a speech segment, as well as a representation of gesturing style to generate a sequence of joint angles for a predefined humanoid skeleton, resulting in upper or full body motion, fitting the speech segment it is based on.

The human-likeness, being one of the common quality measures of generated gestures, has recently surpassed natural motion. A system submitted to the GENEA Challenge 2022 (Yoon et al., 2022) was ranked first on their human-likeness evaluation scale, even above a motion-capture sequence. It must be said, that both sequences were not classified as “completely humanlike”, as they were both limited by the ability to accurately visualize the recorded or generated motion on a virtual avatar. Still, these results are very promising and may shift future research away from chasing motion quality to other problems, still far from being solved.

Another common measurement that in contrast still calls for improvements, is the appropriateness of generated gestures. Earlier evaluations were optimistic when it came to the state of the art in gesture appropriateness, but recent findings paint a dire picture. The GENEA Challenge 2020 (Kucherenko et al., o. J.) recognized substantial improvements in terms of both human-likeness, and appropriateness of generated gestures. The appropriateness scores evaluated as part of the GENEA Challenge 2022 (Yoon et al., 2022) however, only marginally improved, with no submissions coming close to ground truth motion, even though motion quality was much better than before.

While it is hard for humans to pinpoint what exact gesture would be “correct” for a given context, we still seem to have a good intuition for what looks right and what seems off. This begs the question, why current solutions fare so poorly against real human motion and one possible factor, that is negatively impacting generative systems, may be the underlying input modality. Current gesture generation models gather information from audio, as well as textual information, trying to extract the necessary semantic information to predict a gesture that is perceived as "correct" for a given utterance. Communication, however, is a multimodal process, extending far beyond verbal utterances. Also, non-verbal information isn't always redundant, so speech alone may not even contain the information necessary, to infer the "right" gestures.

To alleviate this problem, we propose a system that incorporates facial expressions into gesture generation to leverage the potential of style encodings. These could easily be recorded with a webcam or the front camera of a smartphone in the scenario we envision. While some gesture generation systems use style encodings to mimic the gesturing style of specific speakers, to personalize the generated gestures, other systems use them to adapt the gestures to emotional states or types of conversations. Gestures during an oration for example differ greatly from gestures expressed during dyadic interactions. Other style differences that can be modelled, are between happy and angry gestures, or during an agreement or disagreement (Ghorbani et al, 2022). In a Social XR scenario, analysing facial expressions to detect emotions and automatically adapting the generated gestures accordingly may improve the perceived appropriateness and lead to a sense of agency over how the avatar gesticulates.

As a basis, we have already produced our own data set, capturing dyadic everyday conversations with three different full body motion capture systems (OptiTrack Motive 3, The Captury, HTC Vive), along with videos of the participants' faces, via an iPhone Camera, running Apple's ARKit, to extract facial Blendshapes. We also already implemented a Unity application that enables multi-party communication of virtual avatars via voice, generating co-speech gestures alongside. We are currently planning a user study to investigate the effect these systems have on immersion and embodiment during dyadic interactions, before adding adaptive style changes through facial feature analysis. Future ideas include also considering the interlocutor's facial expressions, as well as using gaze information to gain directional input for the gesture generation.

Keywords: virtual agents; gesture generation; facial features

References

Yoon, Y., Wolfert, P., Kucherenko, T., Viegas, C., Nikolov, T., Tsakov, M., & Henter, G. E. (2022). The GENEA Challenge 2022: A large evaluation of data-driven co-speech gesture generation. https://doi.org/10.1145/3536221.3558058

Kucherenko, T., Jonell, P., Yoon, Y., Wolfert, P., & Henter, G. E. (o. J.). The GENEA Challenge 2020: Benchmarking gesture-generation systems on common data. 9.

Ghorbani, S., Ferstl, Y., Holden, D., Troje, N. F., & Carbonneau, M.-A. (2022). ZeroEGGS: Zero-shot Example-based Gesture Generation from Speech (arXiv:2209.07556). arXiv. http://arxiv.org/abs/2209.07556

Affective Stancetaking in Political Speeches. Some Insights in Processes of Multimodal Meaning-making in Media Specific Contexts

Clara Kindler, *European-University Viadrina (Frankfurt/Oder)*
Jana Junge, *European-University Viadrina (Frankfurt/Oder)*

Current research on (political) stancetaking tends to focus on semantic aspects (Biber & Finegan, 1989) or on conversation- or interaction analysis based perspectives (Du Bois, 2007). Although research on multimodal stancetaking remains scarce (Goodwin et al., 2012; Horst et al., 2014), it points to stancetaking as a vibrant multimodal activity where people are highly engaged affectively. Based on the ongoing DFG/NCN research project "Multimodal Stancetaking: Expressive Movement and Affective Stance" (http://mmstance.home.amu.edu.pl/), this poster takes a closer look at affective stancetaking in political speeches as processes of multimodal meaning-making.

Multimodality is understood in a double sense: Firstly, it addresses the dynamic interplay of hand and body gestures with the spoken utterance, including prosodic features. Secondly, it includes the media-specific contexts in which the political speeches are embedded, concerning the audiovisual orchestration of camerawork, shots, montage and sound. Both levels of multimodality form an inseparable unit that unfolds temporally as *expressive movement* in the moment of perception (Kappelhoff & Müller, 2011). It is the specific perceivable movement quality and rhythm of these *expressive movements* that mobilizes affective stance. Thus, the methodological framework puts the viewer's perception at center-stage.

The research presents a qualitative approach with analysis of two speeches given by members of the German party "Bündnis 90 / Die Grünen": a parliamentary speech delivered in person in the German Bundestag in 2019 and a speech at the national congress in 2020, held digitally. The speeches under scrutiny are 16:13 Min (parliament speech 2019) and 4:20 Min (party congress 2020) long. They are official video recordings from the German Bundestag and the Green Party and were free for download on the official websites. The analysis focuses on the areas of the speeches that show high affective engagement, in total around 3-3:30 Min unfolding within four expressive movement units (EMU) per speech. These EMUs were analyzed with the expressive movement analysis developed by Kappelhoff and Müller (2011; Müller, 2019; Müller & Kappelhoff, 2018) and as far as multimodal utterances are concerned the analysis draws on Müller's Methods of Gesture Analysis (Müller, 2010, Kappelhoff & Müller 2018, Müller, in press). The analysis of stance-taking on the semantic level is based on the approach by Du Bois (2007).

Taking the concept of *expressive movement* as a starting point, we will illustrate that affectivity is a crucial part of stancetaking. The temporally unfolding of affectivity is perceived as *foregrounding* (Müller & Tag, 2010) of certain aspects of a speaker's stance and therefore sets them relevant in the process of multimodal meaning-making. It is concluded that by considering also the specific media-aesthetic dimension of the speeches different forms of *affect mobilization* (*Affektmobilisierung*; Kappelhoff, 2016) can be retraced.

Keywords: affective stancetaking; gesture analysis; multimodal meaning-making; media-aesthetics; political speeches

References

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. Text - *Interdisciplinary Journal for the Study of Discourse, 9*(1).

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.*), Stancetaking in Discourse: Subjectivity, Evaluation, Interaction* (pp. 139–182). Benjamins.

Goodwin, M., Cekaite, A., & Goodwin, C. (2012). Emotion as Stance. In M.-L. Sorjonen & A. Perakyla (Eds.), *Emotion in Interaction* (pp. 16–41). Oxford University Press.

Horst, D., Boll, F., Schmitt, C., & Müller, C. (2014). Gesture as interactive expressive movement: Inter-Affectivity in face-to-face-communication. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body—Language—Communication. Handbücher zur Sprach- und Kommunikationswissenschaft: Vol. 38.2* (pp. 2112–2125). De Gruyter Mouton.

Kappelhoff, H. (2016). *Genre und Gemeinsinn. Hollywood zwischen Krieg und Demokratie*. De Gruyter.

Kappelhoff, H., & Müller, C. (2011). Embodied meaning construction. Multimodal metaphor and expressive movement in speech, gesture, and feature film. *Metaphor and the Social World*, *1*(2), 121–153.

Müller, C. (2010). Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. In: I. Mittelberg (Ed.), *Sprache und Gestik* (Vol. 1, pp. 37–68).

Müller, C. (2019). Metaphorizing as Embodied Interactivity: What Gesturing and Film Viewing Can Tell Us About an Ecological View on Metaphor. *Metaphor and Symbol, 34*(1), 61–79.

Müller, C. (in press). A toolbox of methods for gesture analysis. In A. Cienki (Ed.), *Handbook of Gesture Studies* (pp. 0–33). Cambridge University Press.

Müller, C., & Kappelhoff, H. (2018). *Cinematic Metaphor. Experience – Affectivity – Temporality*. De Gruyter Mouton.

Müller, C., & Tag, S. (2010). The Dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. *Cognitive Semiotics, Spring 2010*(6), 85–120.

Stance-taking in the Visual Modality – a Systematic Literature Review

Fien Andries[1], Katharina Meissl[1], Clarissa de Vries[1], Kurt Feyaerts[1], Bert Oben[1], Paul Sambre[1], Myriam Vermbeerbergen[1,2]

Geert Brône[1],

[1] *KU Leuven (Belgium)*

[2] *Stellenbosch University (South Africa)*

In recent years, the concept of stance, which can be defined as the communication of mental states in interaction, has received increasing attention in different fields of research (Bohmann & Ahlers, 2021; Debras, 2015; Englebretson, 2007, for a review see Takanashi, 2018). Specifically, there has been ample attention for the multimodal nature of stance (Feyaerts et al., 2022) resulting in a large variety of paradigms, stance phenomena, analyses of semiotic resources involved. To date, a structured overview of a multimodal account of stance-taking in interaction is lacking. In the current contribution, we present the results of a systematic literature review with which we aim to offer a synthesis of the state of the art in multimodal stance research. Given the relevant properties of visual semiotic resources for the expression of stance, such as simultaneity with spoken utterances (Debras & Cienki, 2012; Ford et al., 2012), combined with other observations such as that "affect is lodged within embodied sequences of action" (Goodwin & Goodwin, 2000, p. 37), we focus on the visual expression of stance in signed and spoken language. We were guided by the following question: How is stance expressed in signed and spoken interaction? In our review, we take a semasiological approach to the notion of *stance*, using the lexical term stance as the starting point for our search.

Using systematic search protocols (Macaro, 2019), we gathered primary research that investigates the involvement of the visual modality in the expression of stance in spoken or signed interaction. From the final selection of papers, amounting to 104 articles spanning the last 20 years, we synthesized information about analytical frameworks, languages of interaction as well as phenomena and semiotic modes in relation to which stance is studied.

In the critical appraisal, four highly salient strands of inquiry surfaced in the articles: The first strand concerns form-function coupling, tying together (combinations of) specific semiotic resources (such as facial expressions, gaze, gesture and verbal means) with specific expressions of stance. The second strand pertains to the notion of sequentiality, e.g. the emergence of stance over time. The third strand highlights the relation between multimodality and the intensity or foregrounding of stance. The fourth strand relates to the co-construction of stance by the participants involved in the interaction.

By exploring these four major topics of interest in the literature, we aim to offer a broad overview of the last 20 years in multimodal stance research. Moreover, this review will aid the identification of relevant gaps in the research about stance, and possible leads for further research.

Keywords: multimodality; stance-taking; systematic literature review; face-to-face interaction

References

Bohmann, A., & Ahlers, W. (2021). Stance in narration: Finding structure in complex sociolinguistic variation. *Journal of Sociolinguistics*, *26*(1), 65–83. https://doi.org/10.1111/josl.12533

Debras, C. (2015). Stance-Taking Functions of Multimodal Constructed Dialogue during Spoken Interaction. In G. Ferré & M. Tutton (Eds.), *Gesture and Speech in Interaction—4th edition (GESPIN 4)* (pp. 95–100). HAL. https://hal.parisnanterre.fr//hal-01640486

Debras, C., & Cienki, A. (2012). Some Uses of Head Tilts and Shoulder Shrugs during Human Interaction, and Their Relation to Stancetaking. *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, 932–937. https://doi.org/10.1109/SocialCom-PASSAT.2012.136

Englebretson, R. (Ed.). (2007). *Stancetaking in discourse: Subjectivity, evaluation, interaction*. John Benjamins Publishing Company.

Feyaerts, K., Rominger, C., Lackner, H. K., Brône, G., Jehoul, A., Oben, B., & Papousek, I. (2022). In your face? Exploring multimodal response patterns involving facial responses to verbal and gestural stance-taking expressions. *Journal of Pragmatics*, *190*, 6–17. https://doi.org/10.1016/j.pragma.2022.01.002

Ford, C. E., Thompson, S. A., & Drake, V. (2012). Bodily-Visual Practices and Turn Continuation. *Discourse Processes*, *49*(3–4), 192–212. https://doi.org/10.1080/0163853X.2012.654761

Goodwin, M. H., & Goodwin, C. (2000). Emotion within situated activity. In A. Duranti (Ed.), *Linguistic anthropology* (pp. 239–257). Wiley Blackwell.

Macaro, E. (2019). Systematic reviews in applied linguistics. In *The Routledge Handbook of Research Methods in Applied Linguistics*. Routledge.

Takanashi, H. (2018). Stance. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of Pragmatics* (pp. 173–200). John Benjamins Publishing Company. https://www.jbe-platform.com/content/books/9789027263070#chapters

Audio-visual perception of the vocalic rounding opposition in 4 languages

Isa Samira Winter, *IfLPhonetics, University of Cologne*

Kornélia Juhász, *Dept. Phonetics, Eötvös Loránd University Budapest*

Andrea Deme, *Dept. Phonetics, Eötvös Loránd University Budapest*

Reinhold Greisbach, *IfLPhonetics, University of Cologne*

This study investigates the effect of audio-visual information on speech perception. Specifically, we examine how the visual cue of lip shape affects the perception of lip rounding in vowels. We compare two languages having a phonetic system where front vowels exhibit lip shape contrast (German and Hungarian: /e/ vs. /ø/) and two languages where the contrast exists only in front and back vowels (Georgian and Egyptian-Arabic: /e/ vs. /o/). Our question is how visual information on lip shapes influences perception, and whether this influence is universal.

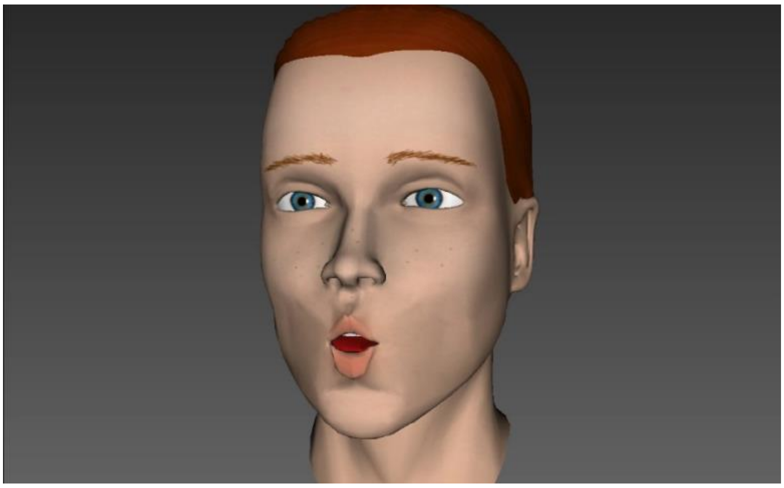
We conducted a perception test including 20 native speakers for each language. Listeners were presented with synthesized vowels. To generate these, we started with an [e:] produced by a German male speaker and lowered its F2 (from 2100 Hz) in several intermediate steps until the timbre of [ø:] (F2 = 1450 Hz), and then the timbre of [o:] (F2 = 600 Hz) was reached using the formant synthesis function of Praat (Boersma & Weenink, 2006). For German and Hungarian, only the signals with F2 between 1450-2100 Hz were used. For Georgian and Egyptian-Arabic, we used all the tokens. To generate visual stimuli (dynamic lip movements), the Talking Head System MASSY (Fagel & Sendlmeier, 2003) was used. Using these videos, we created audiovisual stimuli where rounded or unrounded lips accompanied the audio signal. The experiment was run in 2 blocks repeated twice: i) audio stimuli, ii) audiovisual stimuli. The participants' task was to classify the stimuli as [e:] vs. [ø:] or [e:] vs. [o:] in a forced choice test.

Results for the audio stimuli showed that when the F2 is lowered, there is a continuous transition in perception from the unrounded [e:] to the rounded [ø:] or [o:] in all languages. Visual information shifted the 50% transition point of perception between unrounded and rounded vowels in the expected direction: rounded lips pushed the transition point upwards on the F2 scale, while unrounded lips pushed it downwards. However, we found differences in the location of the transition points and the abruptness of the transition in languages. Results are discussed in the light of the possible origins of language-specific differences.

Keywords: audio-visual perception, perception experiment, vowel perception, rounding opposition, language specific perception

Figures

Figure 1. *Talking Head producing a rounded vowel*



References

Boersma, P., & Weenink, D. (2006): Praat: doing phonetics by computer [Computer program].

Fagel, S., & Sendlmeier, W.F. (2003). Das audiovisuelle Sprachsynthesesystem MASSY – Implementierung und Optimierung. In *Beiträge zur 14. Konferenz Elektronische Sprachsignalverarbeitung*, ESSV, Karlsruhe, 234-241.

Semantic and syntactic neural tracking in audiovisual discourse processing

David Hernández-Gutiérrez[1], Romain Pastureau[1], Anastasia Klimovich-Gray[2],
Mikel Lizarazu[1], Nicola Molinaro[1]
[1] *Basque Center on Cognition, Brain and Language*
[2] *University of Aberdeen*

Ecological language processing typically involves an audiovisual context in which auditory speech is combined with the visual input from the speaker. The visual stream can provide with rich information directly related to the linguistic message, like lip movements or gestures. For communicative purposes, the listener needs to integrate inputs from different modalities to process a single coherent message. In the past years there has been increasing interest on the neural basis of face-to-face language comprehension. Indeed, evidence shows that the brain can synthesize language from visual speech (Bourguignon et al., 2020), and audiovisual information like gestures and verbs can be integrated in a semantic context (Drijvers et al., 2021). However, little is known yet regarding audiovisual processing in more natural linguistic contexts. Thus, it remains unknown how the neural-based semantic and syntactic predictions in a continuous form rely on the speaker's visual cues. Indeed, most multimodal studies in these linguistic domains have used either isolated words or sentence contexts, or have employed time-locked techniques (e.g. Hernández-Gutiérrez et al., 2018). Although the use of larger naturalistic stimuli like audiobooks has become increasingly popular in neurolinguistics (Brennan, 2016; Heilbron et al., 2022), the use of more ecological language like spoken discourse is still scarce.

We use magnetoencephalography (MEG) to study the relationship between the visual input from the speaker and the semantic and syntactic neural processing of speech, in a continuous form. To this aim, we use GPT-2, a state-of-the-art deep neural network, and spaCY, a Natural Language Processing library, to compute the lexical-semantic and part-of-speech (PoS) surprisal of each word (Heilbron et al., 2022). 30 neurotypical adult Spanish native speakers will participate in the study. They will be presented videos (60'' each) of 6 different speakers retelling short animation cartoons. This type of stimuli has been previously used to elicit co-speech gestures (e.g. Graziano & Gullberg, 2018). Both the visual stream (video, no video, video with mask, video without mask) and auditory stream (sound, no sound) are manipulated.

How do semantic and syntactic surprisal entrain with the neural processing of spoken discourse? Is this entrainment sensitive to the audiovisual modality? If so, would visual speech and gestures be differently employed during language prediction? The present study shed light on these questions, adding valuable information to the understanding of language comprehension in face-to-face contexts.

Keywords: neural entrainment; MEG; syntax; semantics, naturalistic comprehension

References

Bourguignon, M., Baart, M., Kapnoula, E. C., & Molinaro, N. (2020). Lip-reading enables the brain to synthesize auditory features of unknown silent speech. *Journal of Neuroscience*, *40*(5), 1053-1065.

Brennan, J. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass*, *10*(7), 299-313.

Drijvers, L., Jensen, O., & Spaak, E. (2021). Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Human Brain Mapping*, 42(4), 1138-1152.

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology*, *9*, 879.

Gwilliams, L., Flick, G., Marantz, A., Pylkkanen, L., Poeppel, D., & King, J. R. (2022). MEG-MASC: a high-quality magneto-encephalography dataset for evaluating natural speech processing. *arXiv preprint arXiv:2208.11488*.

Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119.

Hernández-Gutiérrez, D., Rahman, R. A., Martín-Loeches, M., Muñoz, F., Schacht, A., & Sommer, W. (2018). Does dynamic information about the speaker's face contribute to semantic speech processing? ERP evidence. *Cortex*, *104*, 12-25.

Sub-parametric features of head movements and gaze conveying epistemicity: a study on French Sign Language (LSF) and French co-speech gestures

Fanny Catteau, *Université de Poitiers - Laboratoire UR15076 – FoReLLIS, France*
Claudia S. Bianchini, *Université de Poitiers - Laboratoire UR15076 – FoReLLIS, France*

The aim of the LexiKHuM project is to develop a human-machine interaction system based on a kinesthetic lexicon inspired by natural human gestural communication: co-speech gestures (CSG) and sign languages (SL). As we would like to enable the AI machine to communicate its degree of certainty with regard to the message delivered, we have studied the kinesthetic properties of epistemic gestures.

Previous studies have identified some markers of the epistemic gesture, such as rapid head nods to express certainty (in German and Turkish SL; in French, English and Catalan CSG (Debras, 2017; Herrmann, 2013; Karabüklü & al., 2018; Roseano et al., 2016) or slow head tilts to express uncertainty. However, these studies have only looked at the epistemic gesture as a whole and are focused on SL or CSG. In the LexiKHuM project, we analyze the sub-parametric properties of these gestures, searching for common features in SL and CSG. We assume that (i) some properties of the movement carried by the 3 degrees of freedom (DoF) of the head are involved in the construction of epistemic gestures in SL and CSG (such as the amplitude of the flexion/extension movement or the duration of the abduction/adduction movement of the neck); and (ii) that the direction of gaze associated with the head movements is also relevant to express epistemicity. To investigate these hypotheses, we examined head movements and gaze positions in epistemic gestures in a French CSG and LSF parallel corpus, the DEGELS corpus.

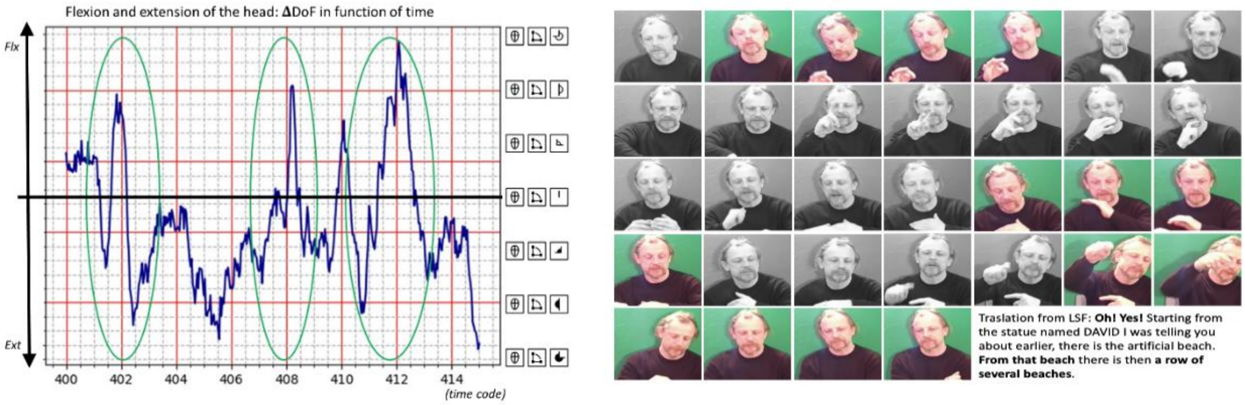
We followed a protocol inspired by the prosodic analysis of SL and kinesiological studies (Boutet, 2018; Puupponen et al., 2015). We (i) manually identified the epistemic sequences of that corpus (with inter-annotator agreement) using *ELAN*; (ii) manually transcribed head and gaze movements using the *Typannot* transcription system (Bianchini et al., 2018) on a sample of 40 epistemic sequences; and (iii) semi-automatically generated the head-movement measurements according to the different DoFs with *AlphaPose* software (see an example in Figure 1).

Our preliminary studies reveal that (i) gestures formed in a context of certainty show a greater amplitude of movement of the flexion/extension of the head in SL and CSG; (ii) in a context of uncertainty, we observe more holding of the flexion/extension movement and greater amplitude of neck rotation movements in LSF. We then combined our observations of neck movements with gaze direction in epistemic contexts. Our first observations show that speakers of French and LSF tend to displace their gaze from their head position when expressing uncertainty: we plan to confirm these initial findings using automatic gaze detection software, such as *OpenFace*. In the near future, LexiKHuM will apply the same protocol presented here to identify more sub-parametric features conveying epistemicity on other articulators (such as shoulders and torso). We will also explore other meanings that will be implemented in our kinesthetic lexicon, such as the expression of urgency or danger.

Keywords: gesture; epistemicity; human-machine interaction; sign language, head movement; gaze

Figures

Figure 1. *Visualization of neck flexion and extension measurements - identification of certainty markers*



Traslation from LSF: **Oh! Yes!** Starting from the statue named DAVID I was telling you about earlier, there is the artificial beach. **From that beach** there is then **a row of several beaches**.

References

Bianchini, C.S., Chevrefils, L., Danet, C., Doan, P., Rébulard, M., Contesse, A., & Boutet, D. (2018). Coding movement in Sign Languages: the Typannot approach. *Proc. "5th Intl Conference on Movement and Computing (MoCo'18)"*, sect. 1 (9), ACM: 1-8.

Boutet, D. (2018). *Pour une approche kinésiologique de la gestualité*. HDR, Université de Rouen.

Debras, C. (2017). The shrug: forms and meanings of a compound enactment. *Gesture, 16*(1), 1-34.

Herrmann, A. (2013). *Modal and focus particles in Sign Languages: a cross-linguistics study*. De Gruyter Mouton (Berlin/Boston).

Karabüklü, S., Bross, F., Wilbur, R., & Hole, D. (2018). Modal signs and scope relations in TİD. *Proc. "Formal and Experimental Advances in Sign language Theory (FEAST)"*, 2: 82-92.

Puupponen, A., Wainio, T., Burger, B., & Jantunen, T. (2015). Head movements in Finnish Sign Language on the basis of motion capture data: a study of the form and function of nods, nodding, head thrusts, and head pulls. *Sign Language & Linguistics*, 18, 41-89.

Roseano, P., González, M., Borràs-Comes, J., & Prieto, P. (2016). Communicating epistemic stance: how speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Processes*, 53, 135-174.

Prosodic domains in the head of the signer

Marisa Cruz, *University of Lisbon*

Sónia Frota, *University of Lisbon*

It is already well-known that sign languages have prosodic constituents and that intonational phrases (IP) are marked by changes in head and/or body position and optional eyeblink (Nespor & Sandler, 1999; Sandler, 1999; Dachkovsky & Sandler, 2009). However, those changes in head position have not been described in detail. Phonological phrases (PhP) are presumably marked by eyes and mouth position (Pfau & Quer, 2010). The main goal of this paper is to explore the role of the head in the prosodic phrasing of Portuguese Sign Language (LGP).

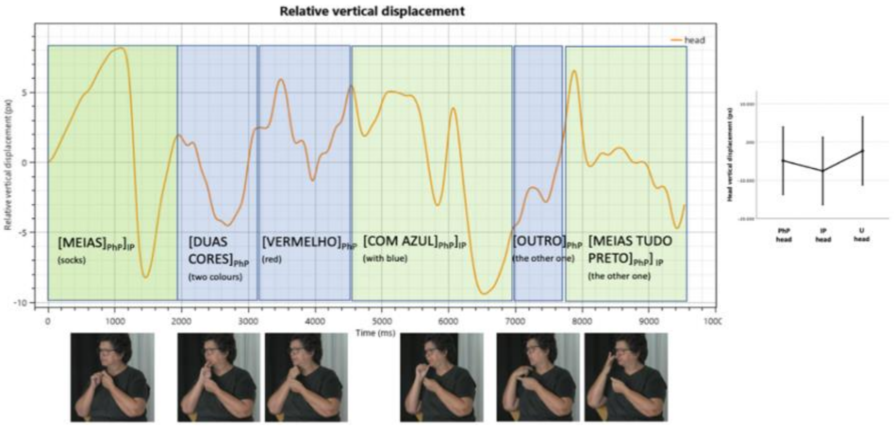
Using a LGP corpus of role-play interviews obtained with an adapted version of the Discourse Completion Task (Félix-Brasdefer, 2009; Billmyer & Varghese, 2000), we selected two ambiguous utterances that can be disambiguated by prosodic phrasing. Following the Prosodic Phonology framework (e.g., Nespor & Vogel, 1986/2007), we analyzed the phrasing of the utterances produced by 4 native signers of LGP. As a general overview, signers used manuals to disambiguate. As for nonmanuals, we observed the occurrence of frequent falling head movements (same movement type), occasionally accompanied by a forward movement of the torso and eyebrow raising. A kinematic analysis of vertical head displacement (pixels) along the time series (ms) was conducted using *Kinovea,* to see whether the amplitude of this nonmanual mattered for prosodic phrasing.

A Generalized Linear Mixed Model (GLMM) was run with participant and utterance as random factors, and prosodic domain - PhP head (head of a Phonological Phrase, internal to an Intonational Phrase), IP head (head of an Intonational Phrase, internal to an Utterance), and U head (final IP head in the Utterance), order in the utterance (1st phrase, 2nd phrase, 3rd phrase), and order per prosodic domain as fixed factors. There was no significant effect of participant or utterance on head displacement (*p*>.05). By contrast, all fixed factors had a significant effect on head displacement (*p*<.001). With respect to the prosodic domain, head amplitude significantly differed between PhP and IP head (*ß*=2.69, *SE*=.62, *t*=4.37, *p*<.001), between IP and U head (*ß*=-5.25, *SE*=.54, *t*=-9.65, *p*<.001), and between PhP and U head (*ß*=-2.56, *SE*=.58, *t*=-4.45, *p*<.001). As shown in Figure 1, falling head movements were more pronounced in IP heads (*M*=-7.57px) than in PhP heads (*M*=-4.89), which resembles the gradience in pre-boundary lengthening usually found in the spoken modality, signaling different levels of the prosodic hierarchy (Frota, 2012). Also interesting was the fact that the IP head that is the head of U is marked by the lowest head amplitude (*M*=-2.33). Since it is the U head, one might expect a larger amplitude of head movement than in other (utterance-internal) IPs. However, such large amplitude could be interpreted as an interrogative, as the falling head movement also plays a role in conveying sentence type meaning (cf. Cruz & Frota, 2022). The two prosodic roles of the head thus seem to be preserved in the prosodic grammar of LGP. Considering phrase order alone, head amplitude significantly increases from 1st to 2nd phrases (*ß*=2.92, *SE*=.59, *t*=4.98, *p*<.001), and from 1st to 3rd phrases (*ß*=1.68, *SE*=.72, *t*=2.32, *p*<.001), but it does not differ between 2nd and 3rd phrases (*ß*=-1.24, *SE*=.77, *t*=-1.61, *p*>.05). Order per prosodic domain showed no differences across IP heads (*p*>.05). However, the amplitude of the head movement significantly differed across PhPs within an IP, which might be explained by the degree of prominence of the PhPs (to be explored in future work). In sum, our results show that the amplitude of the falling head movement is relevant to prosodic phrasing in LGP, distinguishing between phonological and intonational phrases. This finding strengthens the primary role of the head in the prosodic grammar of LGP.

Keywords: sign language; prosodic phrasing; kinematics

Figures

Figure 1. *Head vertical displacement (pixels) extracted with Kinovea for the target utterance [[MEIAS]PhP [AZUIS E ROSA]]IP [[E PRETAS]PhP]IP, produced by participant P9. PhPs are delimited by blue rectangles and IPs by green rectangles, all illustrated by the respective frames.*



Selected references

Cruz, M., & Frota, S. (2022, March 24-26). *"Talking heads" in Portuguese Sign Language* [Conference Presentation]. 35th Annual Conference on Human Sentence Processing, UC Santa Cruz, hybrid format.

Dachkovsky, S., & Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech*, *52*(2/3), 287-314. https://doi.org/10.1177/0023830909103175

Pfau, R., & Quer, J. (2010). Nonmanuals: their grammatical and prosodic roles. In D. Brentari (Ed.), *Sign Languages* (Cambridge Language Surveys, pp. 381-402). Cambridge University Press.

Comparison across modalities:
a case study of the "Away gestures" family in four sign languages

Sílvia Gabarró-López, *Pompeu Fabra University & University of Namur*

Anna Kuder, *University of Cologne*

In this paper, we aim to study four different recurrent gestures which are said to form the "family of Away gestures" in spoken languages (SpLs) (Bressem and Müller, 2014). These forms are "semantically motivated by the effect of actions of removing or keeping away of things" (Bressem and Müller, 2014, p. 1596). To the best of our knowledge, all four forms taken together have not been explored in any signed language (SL) so far. The four manual forms are described as:

1. sweeping away - a (lax) flat hand(s) with the palm(s) facing downwards laterally and being horizontally moved outwards, mostly with a decisive movement (Fig. 1);
2. holding away - a flat hand(s) with the palm(s) vertically facing away in front of the speaker's body (Fig. 2);
3. brushing away - a lax flat hand, palm oriented towards the speaker's body and a movement outwards in a rapid twist of the wrist (Fig. 3); and
4. throwing away - a lax flat hand with the palm facing away from the speaker's body and a movement downwards by bending the wrist (Fig. 4) (Bressem and Müller, 2014).

Our corpus-based study aims to investigate these forms in four SLs: Catalan (LSC), French Belgian (LSFB), German (DGS) and Polish (PJM). We select and analyse a data sample that lasts approximately 3 hours (45 minutes produced by 6 pairs of signers from each of the corpora) in order to address the following research questions:

1. How often are the forms used across the four SLs?
2. Which functions do they express in SL discourse? Are they similar to or different from what has been reported for SpLs?
3. Is the nature of Away forms gestural or conventionalized across the four SLs?

All four Away forms are present in the four SLs with varying frequencies, showing both similarities and differences across the signed and the spoken modality. Although all four forms are used differently across the four SLs, their semantics are related to the expression of negative functions, but to a lesser extent than in SpLs. The most frequent function across all SLs is completion, but most cases (66 out of 89) are found in LSC and are associated with sweeping away. The second most frequent function of this form across all four languages is negation. As for the brushing, holding and throwing away forms, in each SL they most frequently express functions that are neither shared with the other SLs under study nor with the other SpLs for which the forms have been described so far. The second most frequent functions of these three forms are negation, contrast and negative assessment respectively.

The fact that there is a recurring pairing of the three Away forms with some lexical ID-glosses raises the question about the lexicalisation status of these manual forms. We find significant differences between this status across languages. In LSFB and LSC, the degree of lexicalization of manual elements belonging to the Away family exceeds 90%. In DGS and PJM, the Away elements seem to have retained their gestural nature to a greater extent. In DGS approximately 65% of all tokens under inspection were marked as lexicalised, and this number stays at 46% in PJM.

The fact that some functions in the four SLs under study were not only found in the environmental SpLs but also in indigenous SpLs such as Savosavo supports claims that “[t]he documented forms, meanings and functions of sweeping and holding away thus seem not to be restricted to their use in Indo-European languages but might have a much wider cross-linguistic and cross-cultural distribution” (Bressem et al., 2017, p. 200-201).

Keywords: gestures, Away family, sign language, corpus-based study

Figures

Figure 1. *Articulation of the sweeping away gesture*

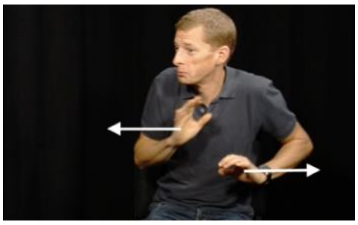


Figure 2. *Articulation of the holding away gesture*

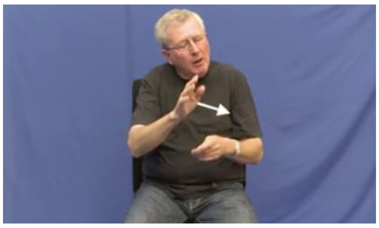


Figure 3. *Articulation of the brushing away gesture*

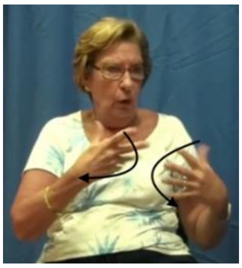
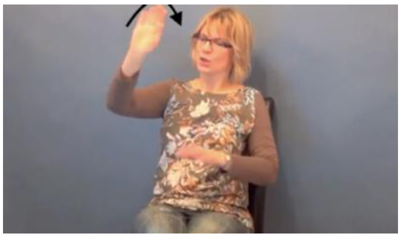


Figure 4. *Articulation of the throwing away gesture*



References

Bressem, J. and Müller, C. (2014). The family of Away gestures: Negation, refusal, and negative assessment. In C. Müller et al. (Eds.), *Body – Language – Communication* (pp. 1592–1604). Berlin: De Gruyter Mouton.

Bressem, J., Stein, N., and Wegener, C. (2017). Multimodal language use in Savosavo. Refusing, excluding and negating with speech and gesture. *Pragmatics* 27(2): 173–206.

An Unsupervised Method for Head Movement Detection

Yu Wang, *Digital Linguistics Lab, Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany*

Hendrik Buschmeier, *Digital Linguistics Lab, Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany*

Head movements are important signals in human-human interaction and head gesture detection has been done quite successfully (Morency et al., 2007; Paggio et al., 2020; Jongejan et al., 2016). Most of the proposed models, however, use supervised learning methods (that requiring a lot of human labeling) and the models are better at predicting non-movement than predicting movement of the head – as evidenced by comparatively higher accuracy and lower F1 scores (Paggio et al., 2020; Jongejan et al., 2016).

As human labeling and segmentation is expensive and models built on supervised learning methods may suffer from unbalanced data, our poster addresses the question whether it is possible to detect head movements in an unsupervised way, based only on movement features. We use OpenFace (Baltrusaitis et al., 2017) to create a spatial model of the head movement and compute the features velocity, acceleration and jerk (Jongejan et al., 2016), resulting in an 18-dimensional feature vector at each point of time. Based on this, we treat the head movement detection problem as a sequence extraction task on a multivariate time series, with sequences being intervals where the head is either in movement or not in movement. Within the given feature space, features are not independent and feature values are dependent on their past.

The approach that we take is built on previous work (Sadri et al., 2017; Deldari et al., 2020) which uses information gain, specifically Shannon entropy, to segment time series. Shannon entropy reflects the variance of a probability distribution function and low entropy values mean a large variance. Given that the information gain of the whole time series is a constant, every subsequence with a relatively large variance value (i.e., a low information gain value) is, potentially, head movement. Adjusting the approach (Sadri et al., 2017; Deldari et al., 2020) to be suitable for head movement detection, we calculate the entropy values for all of the sub-sequence extracted from the time series. For candidates head movements, we then dynamically search a combination of the sub-sequences whose sum of the entropy can be minimized and consider those head movements.

As a first evaluation of the method, we used a three-minute video recording of a person actively listening to an explanation by an interaction partner. Head movements sequences can vary in length, but would normally be regarded as equally long during evaluation, which is not suitable for calculating F1 scores. We instead estimate the similarity between a manual segmentation of the recording by a human annotator and the automatic segmentations resulting from the approach. To do this, we compute the overlapping ratio between the different segmentations as well as the degree of organization (Lücking et al., 2012). The head movement annotations resulting from our algorithm have a total length of 2:38 minutes and 56 annotations are found. This is 51 seconds longer than the annotations resulting from the manual segmentation of the human annotator (1:47 minutes with 46 annotations found), yielding an overlapping ratio of 48.35%. The degree of organization (computed with 20,000 Monte Carlo iterations given that the data for evaluation is quite large, a granularity for annotation length of 10, and significance threshold of $\alpha = 0.05$) yields an agreement score of 0.5591. Agreement between the human annotator and the method is higher than chance, but also shows that there is still a large difference between the manual annotation and the automatic annotation. There are two explanations for this: First, there are differences in the position of the boundaries of some annotations. Second, the automatic method is more sensitive to small movements, as it only looks at the change of information gain value. Breathing and similarly minimal physical movement that affect head position could thus be captured. Still, we consider the proposed method to be a good starting point for unsupervised automatic head movement detection that can be improved upon these points.

Keywords: Head movement detection; Multivariate time series; Unsupervised learning

References

Baltrusaitis, T., Zadeh, A., Lim, Y.C., & Morency, L-P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. doi:10.1109/FG.2018.00019

Deldari, S., Smith, D.V., Sadri, A., & Salim, F. (2020). *ESPRESSO: Entropy and ShaPe awaRe timE-Series SegmentatiOn for Processing Heterogeneous Sensor Data*. doi:10.1145/3411832

Jongejan, B. (2016). Classifying head movements in video-recorded conversations based on movement velocity, acceleration and jerk. *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016)*, 10–17.

Lücking, A., Ptock, S., & Bergmann, K. (2012). Assessing agreement on segmentations by means of Staccato, the segmentation agreement calculator according to Thomann. doi:10.1007/978-3-642-34182-3_12

Morency, L-P., Quattoni, A., & Darrell, T. (2007). Latent-dynamic discriminative models for continuous gesture recognition. doi:10.1109/CVPR.2007.383299

Paggio, P., et al. (2020). Automatic detection and classification of head movements in face-to-face conversations. *Proceedings of ONION 2020: Workshop on PeOple in LaNguage, VIsiOn and the MiNd*, 15–21.

Sadri, A., Ren, Y., & Salim, F. (2017). Information gain-based metric for recognizing transitions in human activities. doi:10.1016/j.pmcj.2017.01.003

Multimodal focalization processes during French family dinners: a comparison between speaking and signing families

Marion Blondel, *CNRS-Paris 8*
Christelle Dodane, *Sorbonne Nouvelle, CLESTHIA EA7345/PRAXILING UMR5267*
Karine Martel, *GRHAPES (UR7287), INSHEA, Suresnes, UPL Université Paris Lumières*
Fanny Catteau, *Université de Poitiers, Forellis*

When interacting, speakers –or signers– structure their discourse according to their communicative intentions. Their discourse combines shared information and new information that is either part of the background content or that is foregrounded (Lambrecht, 1994). When new information is foregrounded, i.e. focused, a variety of processes including speech /sign prosody and (hearing-deaf) shared gestures are used for emphasis. In interaction, focus is implemented via multimodal phenomena: subtle elements and often interwoven/complementary between speech and gesture parameters. Although focalization is relatively well described in spoken dialogue, it remains understudied in signed interactions, and in both signed and spoken multiparty conversations.

Family dinners are a privileged interactive situation as participants need to coordinate two activities: eating – using one's mouth and hands – and successfully transmitting information to the other family members. Our aim is to understand how both adult and child dinner participants, mark information salience in multiparty conversations, in which discourse topics are intermingled. In doing so, we wish to study how focus is shaped across modalities (audio-vocal and visual-gestural), and which prosodic markers are used by speakers and signers. We hypothesize that these markers are articulated in contrastive multimodal patterns and that a parallel can be drawn between spoken French and LSF, in terms of the types of markers mobilized, as well as the prosodic patterns used.

We collected multiparty and multimodal interactions during family dinners in two languages that mostly use contrasting modalities: in spoken French (including its corporeality) and in French sign language (LSF, including mouthing). We then listed the prosodic markers associated with focal occurrences, in order to identify possible patterns, similarities and discrepancies across languages and modalities. The starting point of this inventory is the perceptual identification of focal elements marked by a vocal or gestural accentual prominence in both the dinners in multimodal spoken French and in LSF, completed by an acoustic analysis for spoken French.

In our spoken French dinners, pitch and intensity are two fundamental parameters almost always involved in the realization of prominences. A rapid rise towards high frequency ranges and an amplification of the sonicity are almost systematically observed, in agreement with Rossi (1999) in particular, but, other variables can also contribute to the focus of a speech unit, such as scansion, acceleration after the stressed syllable, glottal stop before and pause after the stressed syllable, micro-pause before the stressed word/syllable, lengthening, etc.

In LSF, and in accordance with what is described in Wilbur (1999) for ASL, van der Kooij et al. (2006) for NGT, Lombart (2021) for LSFB, we observed the use of prosodic markers equivalent to vocal prosodic markers in terms of prominence effect and relative contrast (the distribution of the focused element contributes to their prominence). This inventory for LSF includes manual and nonmanual cues including holds, which cause an elongation of the sign, contrasts in the amplitude of the (manual, facial, body) movement, acceleration, scansion patterns, repetition.

A number of the gestural markers found in the LSF data (including the whole body articulators) are mirrored in the gestural productions of hearing-speakers in multimodal French. In particular, we observe the use of gestures shared by both hearing and deaf participants which contribute to information focusing, such as pointing or presentational gestures, as well as, more generally, a voice-hand-bust coarticulation (Ferré, 2003; a. o.). This reveals that taking LS into account offers a different perspective on the "visual prosody" of spoken languages (Esteve-Gibert & Guellaï, 2018) and allows us to consider spoken languages as embodied languages, by including the relations between vocal prosody and gesture in our description of the linguistic system.

Keywords: Focalization; multimodal; family dinners; speaking; signing

References

Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology*, 9:338.

Ferré, G. (2003). Discursive, prosodic and gestural analysis of focalisation pauses in british english. *Interfaces prosodiques Proceedings* (pp. 265-270). Université de Nantes.

Kooij, E. van der, Crasborn, O., & Emmerik, W. 2006. Explaining prosodic body leans in Sign Language of the Netherlands: pragmatics required. Journal of Pragmatics 38, 1598-1614.

Lambrecht, K. (1994). *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge University Press.

Lombart, C. (2021). Au croisement des ressources orales, gestuelles et signées: Comparaison de la prosodie du français et de la LSFB. In *Papers of the LSFB* (Vol. 15).

Rossi, M. (1999). *L'intonation, le système du français: description et modélisation*. Ophrys.

Wilbur, R. B. (1999). Stress in ASL: Empirical evidence and linguistic issues. *Language and Speech, 42*(2-3), 229-250.

Prelinguistic Deictic Gesture and Co-Speech Sign Acquisition in Deaf Children

Liang Xinyuan, *Department of Linguisitcs, University of Chinese Academy of Social Science*

Gestures serve a facilitating function of language learning and bridge the gap between actions and words though language and motor ability develop in different systems (Iverson & Goldin-Meadow, 2005; Iverson, 2010; Volterra et. al., 2017). The properties of the parameters (handshape, movement, location, etc.) in early gestures retain the later sign language pairing with the correlation of babbling and spoken language (Cheek, 2001).

This research explored the use of gestures in the different language development stages of children and the situation when it is in combination with speech. Different types of Deictic Gesture (DG), Representational Gesture (RG) and sign were distinguished according to the referent or meaning of gesture and speech. The data is from the Child HKSL-Cantonese Bilingual Corpus (http://www.cslds.org/acquisition/en-us/Corpora) on a longitudinal study of a child from 10 to 25 months WT whose mother is deaf and father hearing.

The motoric features were analyzed. The data shows that the child was aware of the morphemes of place of articulation (POA), handshape and orientation. And the phenomena of assimilation and proximation was also found in the data which were compatible with the anatomic characters of children.

RG slightly increses in both speech and sign sessions, wheares DG decreses with clear tendency. The trends were compatible in both speech and sign sessions indicating the synchronization of cognitive development of children language development across models. Furtherly, one of the typical DG, pointing (typical handshape with extended index finger and closed fist), helps with the development of pronouns, catching adults' attention and serves as a scaffold towards the two-word stage cooperating with words or other gestures to express more complicated meanings. Co-speech gesture shows a clear increasing tendency and DG, RG and nods and shakes of the head were combined with the speech of the same meaning, unrelated meaning, and supplementary meaning.

In the interaction between the child and the adults, attention attraction by physical contact was important, especially for the deaf child. And it was observed that the child has already been aware of the morphemes of signs or compositions of the gestures in the prelinguistic stage.

Pointing facilitating the acquisition of language serves a scaffold before the two-word utterance emerges. The development of DG is potential to be an index to the early language stages. The development of gesturing ability paves the way to the sign. Both the motoric practice and cognitive progress are prepared before the use of real language. The errors produced by the child and the compromising process of adults reveal the developing stage and features of the child as well as the adults' strategy when communicating with the child.

Keywords: deictic gesture; co-speech gesture; sign; deaf children acquisition

Figures

Figure 1. *Proportion of Deictic and Representational Gestures (Sign session)*

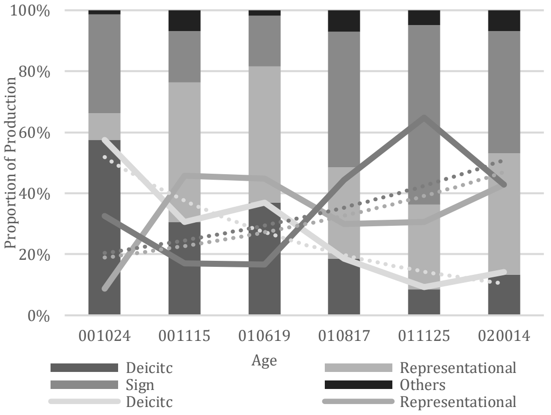


Figure 2. *Proportion of Deictic and Representational Gestures (Speech session)*

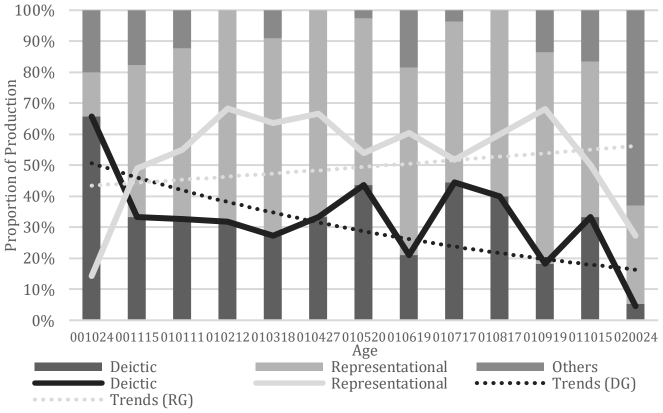
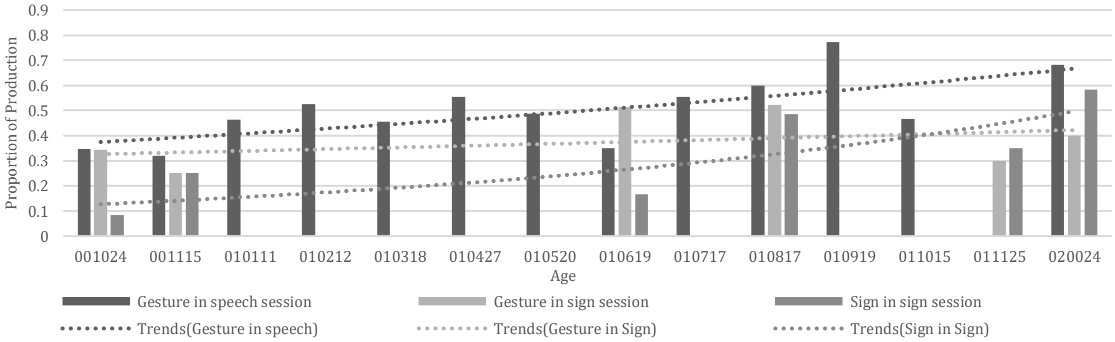


Figure 3. *Proportion of Co-speech Gesture and Sign in Speech and Sign Session*



References

Cheek, A., Cormier, K., Repp, A., & Meier, R.P. (2001). Prelinguistic Gesture Predicts Mastery and Error in the Production of Early Signs. *Language*, 292-323.

Iverson, J.M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science 16*(5), 367-71.

Iverson, J.M. (2010). Developing Language in a Developing Body: the Relationship Between motor Development and Language Development. *Journal of Child Language*, 37(2), 229-261.

Volterra, V., Capirci, O., Caselli, M. C., Rinaldi, P., & Sparaci, L. (2017). Developmental evidence for continuity from action to gesture to sign/word. Language, Interaction and Acquisition, 8(1), 13-41.

Multimodal dimension of linguistic feedback: a study of mirroring in Polish Sign Language

Joanna Wójcicka, *University of Warsaw*

Anna Kuder, *University of Cologne*

This research is a small-scale study of the feedback (precisely: backchanneling) phenomenon called mirroring as utilized by the users of Polish Sign Language (PJM), a natural sign language used by the Deaf community in Poland. Mirroring in social psychology is defined as the behaviour in which one member of the face-to-face interaction unconsciously imitates (matches) gestures, speech patterns, or attitude of another. Research on the topic has so far only been undertaken for nonverbal mirroring used by the speaking population. It is claimed that the primary function of nonverbal mirroring is showing similarity and togetherness (Bavelas et al., 1986) and that there is a connection between nonverbal mirroring and rapport in social interactions (Kendon, 1990).

Our current research stems from these observations and extrapolates them onto the field of sign language. Based on the material coming from the Polish Sign Language Corpus we want to answer the following research questions: (1) How is mirroring realized in PJM?, (2) What is the role of mirroring in natural signed communication?

For the purposes of the study, we choose a sample form the PJM Corpus (Kuder et al., 2022) that contains 12 texts (6 retellings and 6 dialogues) coming from 7 dyads. All analysed texts are interactional, however, the interactions between participants are unbalanced. As the texts are elicited with the use of elicitation materials there are clearly defined turns that the participants are taking while fulfilling the presented task and the turn changes do not happen as often as in a free conversation.

In this dataset we distinguish all cases of repetitive backchannels, which we divide into two types: lexical (manual signs, gestures, mouthings and mouth gestures) and non-lexical (shown either intentionally or unintentionally by the: head movements, facial expressions, mouthing, mouth gestures or manual gestures). Only the unintentional and nonmanual cases are interpreted as cases of mirroring. In the next round of annotation, the functions of all identified cases are interpreted. Those functions include: sentiment matching (understood as reacting to the positive or negative sentiment of the utterance); matching comments (understood as reacting the content of the utterance without repeating the exact signs that were articulated); content matching (understood as exact copying of the text contents) and prosody matching (understood as non-manual following of the utterance rhythm).

The obtained results show that mirroring in PJM is used most often for sentiment matching and most often facial expressions serve as mirroring markers. What is interesting is that what is being matched is usually the emotional load of the utterance rather than the topic of the conversation (e.g. dark jokes about e.g. catastrophes or accidents are usually met with smiling or laughter rather than sadness or upset). Prosody (tempo and rhythm) of signing can be mirrored by rhythmically nodding one’s head. The content of the utterance can be mirrored non manually (e.g. puffing one’s cheeks can be used to match an utterance about a large and round object). Repetitive backchanneling used in sign language discourse has one unique feature that stems from its modality: mouthing can be used to match the manual sign used by the other signer, but whether this is a case of mirroring remains an open question.

Just like in SpLs, mirroring does more to the discourse than just copying what the interlocutor has signed or shown. Our research serves as an argument for the claim that behavioural mirroring is a modality-independent phenomenon allowing the conversation participants to build rapport and togetherness.

Keywords: feedback; mirroring; Polish Sign Language; corpus study

References

Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). “I show how you feel”: Motor mimicry as a communicative act. *Journal of Personality and Social Psychology, 50*, 322–329.

Kendon, A. (1990). *Conducting interaction: Pattern of behavior in focused encounters*. New York: Cambridge University Press.

Kuder, A., Wójcicka, J., Mostowski, P., & Rutkowski, P. (2022). Open Repository of the Polish Sign Language Corpus: Publication Project of the Polish Sign Language Corpus. In E. Efthimiou et al. (Eds). *Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*. (*LREC 2022*), (pp. 118–123). Marseille, France: ELRA.

Multimodal feedback signals: comparing response tokens in co-speech gesture and sign languages

Anastasia Bauer, *University of Cologne*
Jana Hosemann, University of Cologne
Sonja Gipper, University of Cologne
Tobias-Alexander Herrmann, University of Cologne

Feedback signals serve to coordinate interaction, direct the advancement of narrative, manage attention and establish common ground (Sacks et al., 1974.). We define feedback signals very broadly as interactional moves that display some kind of uptake of the information represented by another person's utterance. Feedback signals can be marked by various multimodal cues (vocal (e.g., 'mmh'), manual (e.g., gestures) and/or non-manual (e.g., nods, eye gaze, facial expression) and may indicate active involvement, comprehension or trouble.

The present study focuses on the most frequent type of feedback – *response tokens*, also *backchannels* (Liesefeld & Dingemanse, 2022). Response tokens play an important role in constructing and maintaining shared knowledge in conversation and have been the subject of much work in spoken languages (Gardner, 2001). Research on response tokens in sign languages is extremely sparse to date (Mesch, 2016) and no cross-linguistic studies of response tokens in sign languages or across modalities have been undertaken yet.

We account for response tokens (continuers, acknowledgment tokens, newsmarkers, change-of-state tokens, change-of-activity tokens (Gardner 2001:2)) from a multimodal and cross-linguistic perspective by comparing the use and semiotic composition of response tokens in the discourse of speakers and signers in corpora representing four different languages: spoken German, spoken Russian, German Sign Language (DGS) and Russian Sign Language (RSL). Given the number and the age of speakers/signers (6 people in each corpus) and the contexts, all four corpora can be considered comparable.

The goal of this study is to systematize and to compare the relevance of vocal, manual and non-manual response tokens across languages and modalities. We analyze 30 min of dyadic conversation in each corpus: DGS Corpus (Hanke et al. 2020), RSL Corpus (Burkova 2015), multimodal corpus of spoken Russian (unpublished) and spoken German (unpublished).

Our preliminary results show that non-manuals (mostly head nods, mouth gesture and torso movements) predominate in spoken languages over the vocal (lexical and non-lexical) forms. Head nods mostly co-occur with vocal continuers or acknowledgement tokens but they can also appear on their own. In sign language dyadic discourse non-manual response tokens clearly predominate over manual feedback responses. We account for the following non-manual signals in our data: head nods, head-shakes, smile, eyebrow raise, head turns, change of body posture, nose wrinkles, widened eyes and mouthing. Moreover, we show how the languages compare in the usage of response tokens and which categories are most likely to include which (non)-manual feedback signals.

The current study provides a first cross-modal and cross-linguistic look at feedback mechanisms in four languages. This cross-linguistic and cross-modal element is a critical factor to get a better understanding of the 'human interaction machine'.

Keywords: discourse; signed language interaction; corpus

References

Burkova, S. (2015). Russian Sign Language Corpus. http://rsl.nstu.ru/ (accessed 27 May 2022).

Gardner, Rod. (2001). *When listeners talk: response tokens and listener stance*. John Benjamins.

Hanke, T., Schulder, M., Konrad, R. & E. Jahn (2020). Extending the Public DGS Corpus in Size and Depth. In Efthimiou, Eleni et al. (Eds.), *Proceedings of the LREC2020*, Paris, France: European Language Resources Association, 75-82.

Liesenfeld, A., & Dingemanse, M. (2022). Bottom-up discovery of structure and variation in response tokens ('backchannels') across diverse languages. *Interspeech 2022*, 1126–1130. https://doi.org/10.21437/Interspeech.2022-11288

Mesch, J. (2016). Manual backchannel responses in signers' conversations in Swedish Sign Language. *Language & Communication*, *50*, 22–41. https://doi.org/10.1016/j.langcom.2016.08.011

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696-735.

DAY 2. POSTER SESSION

(In order of appearence in the program)

N	Authors	Title
1	Patrizia Paggio, Manex Aguirrezabal, Bart Jongejan, Costanza Navarretta and Leo Vitasovic	GEHM Network - Creating a Zoom corpus
2	Vivien Lohmer, Lutz Terfloth and Friederike Kern	Explaining the Technical Artifact Quarto!: How Gestures are used in Everyday Explanations
3	Madeleine Long, Aslı Özyürek and Paula Rubio-Fernandez	The role of pointing and joint attention on demonstrative use in Turkish
4	Iraide Ibarretxe-Antuñano, Andrea Ariño-Bizarro, David Moret-Oliver, María Teresa Moret-Oliver and Guillermo Tomás-Faci	Multimodal metaphors in medieval manuscripts: the case of CONTROL IS UP
5	Laura Peiró-Márquez and Iraide Ibarretxe-Antuñano	Multimodal insights into European Spanish: A fine-grained analysis of motion event descriptions
6	Omid Khatin-Zadeh, Hassan Banaruee and Danyal Farsani	A Study of Using Iconic and Metaphoric Gestures with Motion-Based, Static Space-Based, Static Object-Based, and Static Event-Based Statements
7	Hao Lin, Yuting Zhang, Qi Cheng and Yan Gu	Brow and palm reveal the origin of interrogative markers: Evidence from home sign, sign language, and spoken language
8	Marina Zhukova	Drawing parallels between emblems and emoji: a “thumbs-up” case study
9	Fabio Catania, Micol Spitale, Silvia Silleresi and Francesca Panzeri	Joker Face and Voice
10	Setareh Nasihati Gilani and David Traum	Analyzing User’s Mental State and Facial Expressions in Interaction with Different Personalities in a Critical Situation
11	Christina Piot, Julien Perrez and Maarten Lemmens	Talking and gesturing about motion at different L2 proficiency levels
12	Asuman Şimşek Tontuş and Safiye İpek Kuru Gönen	A Micro Analysis of EFL Teachers’ Gesture Use as a Pedagogical Tool in Video-Mediated Interaction

N	Authors	Title
13	Sara Feijoo, Mariona Anglada and Núria Esteve-Gibert	The benefits of multimodal communication in the foreign language classroom: The use of hand gesture to teach morphology and word structure
14	Paula Sánchez Ramón, Alina Gregori, Pilar Prieto Vives and Frank Kügler	The impact of focus types on the prosody-gesture link in two Romance and Germanic languages: a focus elicitation production study
15	Alina Gregori and Frank Kügler	The Distribution of Non-referential Gestures, Information Structure and Prosody: A Corpus Study on Prominence Peak Alignment
16	Clara Lombart	Audiovisual prosody and information structure in French: An investigation of the marking of contrast
17	Gaëlle Ferré	Prosodic features of speech in synchrony with four pragmatic gestures
18	Josua Dahmen	Pointed pronouns: The systematic co-occurrence of pointing gestures and emphatic pronouns in Jaru
19	Elena Nicoladis, Hui Yin and Paula Marentette	Cross-cultural differences in gesture frequency
20	Gilbert Ambrazaitis, Johan Frid and David House	Web-based, audio-visual prominence ratings of Swedish news reading materials: Effects of head movements, rating condition, and hardware
21	Margaret Zellers	A first investigation of the timing of simple and complex co-speech manual gestures in Luganda and their relation to prominence
22	Giorgina Cantalini and Massimo Moneglia	Prosodic cues for Gesture / Speech synchronization and multimodal prominence
23	Ed Donnellan, Yumeng Wang, Levent Emir Özder, Hillarie Man, Kellie Fraser, Amara Jiménez Cañizares, Beata Gryzb, Yan Gu and Gabriella Vigliocco	Audience effects and production demands on timing relationships between representational gestures and speech
24	Anaïs Claire Murat, Maria Koutsombogera and Carl Vogel	Event Chronography in Multimodal Data: a Method for Quantitative Analyses

Explaining the Technical Artifact *Quarto*!:
How Gestures are used in Everyday Explanations

Vivien Lohmer, *Faculty of Linguistics and Literature, Bielefeld University, Germany*

Lutz Terfloth, *Computer Science Education, Paderborn University, Germany*

Friederike Kern, *Faculty of Linguistics and Literature, Bielefeld University, Germany*

In our day-to-day lives, we regularly need to explain something to someone or are the ones to whom something is explained. Prerequisite of our study of everyday explanations is the philosophical theory that technical artifacts can be described by (1) focusing on the architecture and/or (2) focusing on the function or relevance (Vermaas, 2006; Kroes, 2009; Schulte and Budde, 2018). Taking a navigation system as an example, one could (1) explain how it works on the level of data and algorithms or (2) how its features are helpful for relaxed journeys across countries.

The aim of our paper is to explore if and how participants use different gesture types (iconic, deictic, pragmatic) when talking about architectural and/or functional aspects of technical artifacts in order to systematically enrich their verbal descriptions with multiple multimodal resources to generate the most comprehensive explanation (McNeill 1992, 2005). In particular, we assume that (1) participants employ more and predominantly iconic gestures when describing architectural aspects (e.g., the size and shape of game figures, shape of a game board) and (2) while using more pragmatic gestures when talking about relevance.

To answer our hypothesis, we recorded explanations of the two-persons board game *Quarto!* creating a corpus of 25 explanations. The setting is as follows: first the Explainer (EX) explains the game to the Explainee (EE) without the game present. Later, the game is handed over and the participants are instructed to play two rounds while the designated explainer is asked to continue explaining. The interaction was video-recorded from three different camera angles, one perspective on EX and EE respectively, and the third on the table with the game and the gesture space (Kendon 2004) of both participants.

For methods and analytical procedure, two approaches are employed. For a qualitative content analysis (Kuckartz 2018), the explanation talk is coded using the two features of the dual nature, architecture, and relevance, as deductive top-level codes. The resulting corpus of coded segments serves as a foundation for further analyses. Simultaneously, gestures are annotated and then analysed in their sequential context, i.e. in which segments they occur (architectural or functional aspects). The analysis follows the principles of multimodal conversational analysis (short CA) (Mondada 2014; Goodwin, 2017).

So far, we have focused on architecture descriptions. Preliminary results show that (1) the EX predominantly performs iconic and deictic gestures in temporal synchrony with verbal descriptions of architectural aspects of the game (e.g. forms the size and shape of game figures and the game board; locates the imaginary game board on the table); (2) during these descriptions, EXs gaze alternates systematically between their own hands and the EE; (3) on the contrary, in descriptions of relevance the Exs performs predominantly pragmatic gestures. These preliminary results support our hypothesis that the use of gesture is related to the description of architectural or functional aspects of a technical artifact.

Keywords: gesture; dual nature; explanations; architecture; relevance

References

Goodwin, C. (2017). *Co-Operative Action*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139016735

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

Kroes, P. (2009). Engineering and the dual nature of technical artefacts. *Cambridge Journal of Economics*, *34*(1), 51–62.

Kuckartz, U. (2016). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (3., überarbeitete Auflage). Beltz Juventa.

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

McNeill, D. (2008). *Gesture and Thought*. The University of Chicago Press.

Mondada, L (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, *65*, 137–156.

Schmitt, R. (n.d.). Positionspapier: Multimodale Interaktionsanalyse. 10.

Schulte, C., & Budde, L. (2018). A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education. *18th Koli Calling International Conference on Computing Education Research* (Koli Calling '18), 18, 10.

Vermaas, P. E., & Houkes, W. (2006). Technical functions: A drawbridge between the intentional and structural natures of technical artefacts. *Studies in History and Philosophy of Science Part A*, *37*(1), 5–18.

The role of pointing and attention correction on demonstrative use in Turkish

Madeleine Long, *Max Planck Institute for Psycholinguistics*

Asli Özyürek, *Max Planck Institute for Psycholinguistics*

Paula Rubio-Fernández, *University of Oslo*

Pointing gestures and demonstratives (e.g., *this* and *that*) are two of the most early acquired means of referring to objects. However, the exact nature of their relationship has yet to be fully determined. Traditional accounts of demonstrative use have focused on the role of spatial factors (i.e., **distance**) in guiding referential choice (Diessel & Coventry, 2020), whereas more recent accounts argue that psychological factors (e.g., **directing the listener's focus of attention to the correct object)** also play a role (Peeters & Özyürek, 2016). This raises an important question regarding multimodal language use: how are pointing and demonstratives co-ordinated in languages that encode *both* **distance** and **attention correction** such as Turkish?

Prior observations from naturalistic data in Turkish suggest that pointing frequently accompanies "şu" (the middle demonstrative form), directing the listener's attention to the correct object (Özyürek, 1998). However, this has yet to be tested empirically. Here we assessed the role of pointing, distance, and perspective alignment on demonstrative use through carefully controlled video stimuli. Since recent work has shown that "şu" is used more frequently when speaker and listener's perspectives are misaligned (i.e., the listener is looking at an object other than the target) (Rubio-Fernández, 2023) we predicted that a pointing gesture (vs no gesture) would more frequently accompany "şu" to indicate the correct object.

A total of 58 native Turkish speakers were recruited from two sources: university classrooms and Prolific (a crowdsourcing platform). No statistical differences between groups were found thus the data was analyzed together. Participants were shown 48 videos in which a "speaker" and "listener" appeared on opposite sides of a table (for a sample display with link to video see Fig. 1). Participants envisioned that they were the speaker in each video and the listener was their friend. They were told that for each trial there would be four identical fruits on the table, but unbeknownst to their friend only one had been washed. They were instructed to ask their friend to pass them the clean piece of the fruit (which had a red circle around it, indicating where the participant was looking). To do so, they had to complete the phrase "Now I need…" with either "bu" (used for close objects), "şu" (used for objects at mid-distance and to establish joint attention), or "o" (used for distant objects). Here we manipulated the position of the listener and target (Positions 1-4), the perspective of the listener (aligned or misaligned) and the presence or absence of pointing for a fully crossed design.

Using logistic mixed effects regression we modelled Şu Use (Şu=1, Bu/O=0) with Pointing (Pointing vs No Pointing), Position (1-4) and Perspective Alignment (Aligned vs Misaligned) as fixed effects with maximal random effect structure. As predicted, there was a main effect of Pointing (β=.911, SE=.177, p<.001) with şu used more frequently with pointing gestures than without, Perspective Alignment (β=2.100, SE=.272, p<.001) with şu used more often for misaligned than aligned perspectives, and Position (β=.408, SE=.069, p<.001) with şu used more in Positions 2 and 3. In addition, there was a Pointing x Position interaction (β=.255, SE=.115, p=.027) whereby the presence or absence of pointing had a greater influence on şu use in more distant positions (Fig. 2). One possibility is that when the referent is closer to the interlocutors speakers rely less on pointing for attention correction and more on subtle gestures (e.g., eye or head movements) as they are more visible in close proximity. Future work should test this using an eye-tracking paradigm where dyads communicate in a similar set-up.

Overall, our results provide empirical support for prior naturalistic observations, demonstrating that Turkish speakers dynamically integrate demonstrative use and pointing gestures as a function of both **distance** and **attention correction**. These findings add to a growing body of work which shows that pointing and demonstratives are tactically combined to aid referential communication between interlocutors.

Keywords: pointing; demonstratives; joint attention

Figures

Figure 1. *Image of trial with misaligned perspectives and pointing. Video: osf.io/s7dq3*

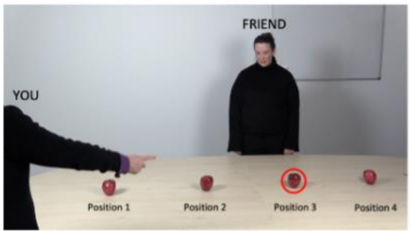
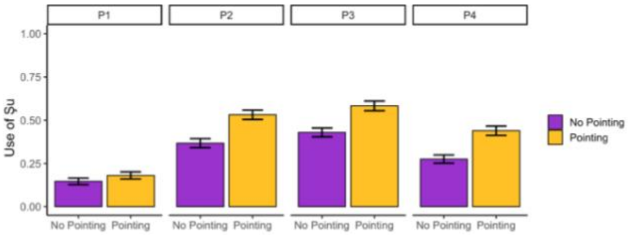


Figure 2. *Şu use across conditions: Pointing and Position (P1-4)*



References

Diessel, H., & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology*, *11*, 555265.

Peeters, D., & Özyürek, A. (2016). This and that revisited: A social and multimodal approach to spatial demonstratives. *Frontiers in Psychology*, *7*, 222.

Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralite et gestualite: Communication multimodale, interaction: actes du colloque ORAGE 98* (pp. 609-614). L'Harmattan.

Rubio-Fernández, P. (2023). Demonstrative systems: From linguistic typology to social cognition. *Cognitive Psychology, 139,* 101519.

Multimodal metaphors in medieval manuscripts: the case of CONTROL IS UP

Iraide Ibarretxe-Antuñano[1], Andrea Ariño-Bizarro[1], David Moret Oliver[2], M.ª Teresa Moret Oliver[1], Guillermo Tomás Faci[3],

[1] *University of Zaragoza*

[2] *Independent researcher*

[3] *Archive of the Crown of Aragon*

Metaphors are powerful cognitive mechanisms that help speakers conceptualising abstract concepts based on more physical and concrete ones. These metaphorical mappings across domains are grounded in our sensory-motor as well as our sociocultural background and are pervasively found in our daily lives (Lakoff & Johnson, 1980; Cienki & Müller, 2008; Valenzuela & Ibarretxe-Antuñano, in press). One case of such as primary metaphor is the so-called CONTROL IS UP. This metaphor captures the idea that anything that is located above exerts more power and control over anything situated below (Schubert, 2005; Valenzuela & Soriano, 2009). This metaphor can be encoded by means of oral metaphorical expressions such as *upper/middle/lower class*, *under their influence*, or *your highness*, but it can also be represented thanks to visual manifestations such as pyramidal structured arrangements and kinetic practices such as genuflexion. Multimodal metaphorical expressions that can be maintained or developed across time.

This paper reports results from an on-going study on the gestural representation of the CONTROL IS UP metaphor in medieval miniatures. The main goal of this study is twofold: (i) to explore how the representation of genuflexions, a gestural vassalage practice started in the Middle Ages, helps us understanding the power relationships between different societal strata at that time, and (ii) to highlight the need to include multimodal expressions in the study of metaphor in order to unveil underlying mappings between conceptual domains.

A set of twenty-six "multimodal genuflexion parameters" was developed to describe both (i) the historical and linguistic background context and (ii) the whole range of kinetic elements involved in the genuflexion such as gestural expressions (face, hands, etc.) as well as distance between characters and body posture measurements (calculated with the CAD software). This set was then applied to a corpus of 34 miniatures selected from the Liber feudorum maior, an Aragonese 12th century manuscript written by Ramón de Caldes and commissioned by King Alfonso II of Aragón in 1192 (Archive of the Crown of Aragon, Chancellery, reg. 1).

Preliminary results seem to indicate that (i) the genuflexion practice includes other kinetic elements beyond the kneeling and bowing such as the physical distance between characters and

the position of their hands, and (ii) the genuflexions represented in these miniatures are different depending on the characters and their vassalage relationships: the closer the relationship with the king, the lower degree of "body bending" (bowing, kneeling) and the shorter distance between their hands.

Keywords: gesture; metaphor; medieval manuscripts

References

Cienki, A., & C. Müller (Eds.). (2008). *Metaphor and Gesture*. John Benjamins.

Lakoff, G., & M. Johnson. (1980). *Metaphors We Live By*. University of Chicago Press.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology*, *89*(1), 1-21.

Valenzuela, J. & I. Ibarretxe-Antuñano. (In press). Chapter 35. Conceptual metaphor in cognitive semantics. In T. F. Li (Ed.), *Handbook of Cognitive Semantics*. Brill.

Valenzuela, J. & C. Soriano. (2009). Is control really UP? A psycholinguistic exploration of a primary metaphor. In J. Valenzuela, A. Rojo & C. Soriano (Eds.), *Trends in Cognitive Linguistics: Theoretical and Applied Models* (pp. 31-50). Peter Lang.

Multimodal insights into European Spanish: A fine-grained analysis of motion event descriptions

Laura Peiró-Márquez, *Universidad de Zaragoza*

Iraide Ibarretxe-Antuñano*, Universidad de Zaragoza*

Iconic co-speech gestures are unconsciously made hand movements which bear a close formal relationship with the semantic content of speech (McNeill, 1922) and which reflect language-specific properties (Kita & Özyürek, 2003; Özçalışkan et al., 2016). One demonstration of this phenomenon is how people speak and gesture about motion events across languages. Multimodal research drawing on Talmy’s (1991) framework of lexicalization patterns and Slobin’s (1996a) thinking-for-speaking hypothesis has revealed that gesture production follows similar typological patterns as identified in speech: speakers consistently gesture about Path and Manner of motion, but there is crosslinguistic variation in how semantic information is distributed across modalities (McNeill, 2000; Özyürek et al., 2008). Nevertheless, the interplay of speech and gesture has mostly been described from a general angle in studies conducted so far, where fine-grained multimodal descriptions are virtually non existent.

This study focuses on language-specific multimodal patterns of motion event representation in European Spanish. The main aim of this research is to provide a granular description of three specific issues that still remain underexplored in the literature: (i) iconicity of non-easily encodable events; (ii) speech-gesture synchronization; (iii) distribution and amount of semantic components across modalities.

Qualitative and quantitative differences across modalities have been analyzed in a corpus consisting of 178 video-taped oral narrations, produced by 12 native speakers of Spanish. Data were elicited using the Tomato Man stimuli (Özyürek et al., 2001) and following Özyürek et al.’s (2008) procedure. Multimodal productions were annotated and transcribed using ELAN 5.9 (Lausberg & Sloetjes, 2009).

Results suggest that: (i) gestures tend to preserve iconicity which is minimized in words (i.e. differences that seem not to play a role in speech are however reproduced in gesture, e.g. distinguishing roll and spin); (ii) the lack of readily accessible linguistic resources to encode an event entails a greater discursive and cognitive effort across modalities (i.e. speakers tend to produce more words and more gestures, and to provide a greater amount of Manner information); (iii) degree of speech-gesture semantic congruency depends on the component conveyed in the verb (i.e. gesture tends to reproduce the Manner information provided in speech, but Path tends to be modified by adding extra details); (iv) the amount of Path information is similar across

modalities, but gesture tends to extend Manner information (i.e. the amount of Manner is greater in gesture than in speech); (v) packaging strategies in gesture might depend on the one used in speech (i.e. speakers are likely to combine a Path-only or a Manner-only gesture with a conflated gesture, but mostly in those cases where they use multiple clauses in speech). These findings support the relevance of fine-grained analyses of iconic gestures to provide a clearer image of the speaker's mental representation of motion events.

Keywords: iconic gestures; motion events; granular analysis

References

ELAN (Version 5.9) [Computer software] (2020), Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. http://archive.mpi.nl/tla/elan

Kita, S. & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16-32. https://doi.org/10.1016/S0749-596X(02)00505-3

Lausberg, H. & Sloetjes, H. (2009), Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, *41*(3), 841-849. https://doi.org/10.3758/BRM.41.3.841

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

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science*, *27*(5), 737-747. https://doi.org/10.1177/0956797616629931

Özyürek, A., Kita, A. & Allen, S. (2001). *Tomato Man movies: Stimulus kit designed to elicit Manner, Path and causal constructions in motion events with regard to speech and gestures [Videotapes]*. Max Planck Institute of Psycholinguistics, Language and Cognition Group.

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R. & Ishizuka, T. (2008). Development of Cross-Linguistic Variation in Speech and Gesture: Motion Events in English and Turkish. *Developmental psychology*, *44*(4), 1040-1054. https://doi.org/10.1037/0012-1649.44.4.1040

Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. Gumperz & S. Levinson (Ed.), *Rethinking Linguistic Relativity. Studies in the Social and Cultural Foundations of Language* (pp. 70-96). Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 17 (pp. 480-519).

A Study of Using Iconic and Metaphoric Gestures with Motion-Based, Static Space-Based, Static Object-Based, and Static Event-Based Statements

Omid Khatin-Zadeh, *School of Foreign Languages, University of Electronic Science and Technology of China*

Hassan Banaruee, *University of Bonn, Germany*

Danyal Farsani, *Norwegian University of Science and Technology*

The differences between the mechanisms of understanding literal and metaphorical statements have been the subject of a large body of works in cognitive linguistics, cognitive psychology, and related fields. Metaphorical statements differ from literal ones because they do not directly refer to their intended meanings. To understand a metaphorical statement, the individual should go beyond the surface or literal meanings of words and figuratively interpret the statement (Banaruee et al., 2017, Khatin-Zadeh & Khoshsima, 2021). In this study, we aimed to examine one specific possible similarity or difference between literal and metaphorical statements. We wanted to know if there was any difference or similarity between literal and metaphorical statements in the ways that gestures are used with these statements. To achieve this objective, we employed the classification of metaphors introduced by Khatin-Zadeh, Farsani, and Reali (2022). We extended this classification to literal statements and made a comparison between each category of metaphors and its corresponding literal category: 1) motion-based metaphorical/literal statements e.g., *He passed through a gate/difficult time*; 2) static space-based metaphorical/literal statements e.g., *She was at the top of the mountain/among her classmates*; 3) static object-based metaphorical/literal statements e.g., *The college had two doors/the college was the door to a new world*; 4) static event-based metaphorical/literal statements e.g., *They cut the paper/their relationship*. In addition, we specifically intended to examine the possible similarity or difference between each category of metaphors and its corresponding category in literal statements in the ways that gestures are used with these statements. The participants in this research listened to five audio short stories in Persian. Each story contained one statement of each metaphoric category and one statement of each literal category. After listening to each story, they retold it in their own language in front of a camera. The data provided from the video recordings were transcribed, classified, and quantified to determine the number of metaphorical and literal statements used during storytelling. Subsequently, the number of metaphoric gestures used with each category of metaphors and the number of iconic gestures used with each category of literal statements were obtained. The gestures produced by participants while retelling the stories were closely examined to ensure that only metaphoric and iconic gestures were selected for further analysis, whereas pointing gestures and beat gestures were removed from the analysis. For metaphorical statements used during the retelling of the stories, chi-square data produced through a contingency table analysis were used to compare the number of metaphoric gestures used with the four categories of metaphorical statements. The results showed that event-based metaphors and event-based literal statements were accompanied by the smallest number of metaphoric and iconic gestures. Furthermore, there was a significant similarity between each metaphorical category and its corresponding literal category in the number of gestures that were used with these categories. This similarity supports the idea that the mechanisms underlying the embodiment of metaphorical and literal statements are essentially similar.

Keywords: iconic gesture; metaphoric gesture; motion-based; static space-based; static object-based; static event-based

References

Banaruee, H., Khoshsima, H., Khatin-Zadeh, O., & Askari, A. (2017). Suppression of semantic features in metaphor comprehension. *Cogent Psychology*, *4*(1), 1–6. https://doi.org/10.1080/23311908.2017.1409323

Khatin-Zadeh, O., & Khoshsima, H. (2021). Homo-schematic metaphors: A study of metaphor comprehension in three different priming conditions. *Journal of Psycholinguistic Research*, *50*, 923-948. https://doi.org/10.1007/s10936-020-09754-z

Khatin-Zadeh, O., Farsani, D., & Reali, F. (2022). A study of using metaphoric and beat gestures with motion-based and non-motion-based metaphors during retelling stories. *Behavioral Sciences*, *12*. https://doi.org/10.3390/bs12050129

Brow and palm reveal the origin of interrogative markers: Evidence from home sign, sign language, and spoken language

Hao Lin, *Shanghai International Studies University*
Yuting Zhang, *University of Washington*
Qi Cheng, *University of Washington*
Yan Gu, *University of Essex/University College London*

Language needs a mechanism to mark questions, which can be realised by morphological forms, syntactic movement, or by prosodic features such as intonation. For example, nearly all modern spoken languages and sign languages have at least one WH word (Dryer & Haspelmath 2013). Despite some variations, brow movement is universally observed in both WH questions and polar questions, as shown in a typological studies on question marker of 37 sign languages (Zeshan, 2006). Many spoken languages and a few sign languages have question particles (ibid.) to mark polar questions. Where are our interrogative markings from? What is the prototype of interrogative marking? This paper argues that brow movements and palm-up are the origin of interrogative markers based on evidence from sign language (including home sign communication system), gesture and speech.

Firstly, we examined the use of brow movements and palm-up (PU) in Chinese Sign Language (CSL) and Chinese home signers. Lin 2019 found that brow movement is the dominant interrogative markers in the interrogative question. We analyzed the data from home signers (N = 9, Mean age =48, 5 females) in a Chinese village, who do not have any WH signs or question particles but BB is the sole interrogative marker. The results showed that BB is the sole marker once there is no other interrogative makers in the signers. PU might be the manual prototype of interrogative marking in users of CSL and Chinese home signers. Here brow-movement is representative of co-articulated facial expression, for example, a brow-raising with eye widening, held-tilting, all of which constitute ‘beseeching brow’ (BB) etc.

Additionally, we studied the behaviour of (1) Deaf CSL teacher who taught CSL to the hearing learners of CSL at beginners’ classes and (2) CSL signers communicate with stranger home signers. We found that 1) PU often co-occurs with BB in interrogatives in the communication among the CSL signers; 2) PU+BB is the main marker for Chinese deaf gestures, which typically appear in the scenario where the CSL signers have to resort to more gestures to communicate with hearing non-signers; and (3) Home signers only rely on BB for interrogative markings.

Finally, we examined the functions of BB and PU in Mandarin spoken language. We looked at their prevalence and frequency in the elliptical questions in speech from publicly available data. For example, there are content questions without WH words, (e.g., "You are going to——" "time——"), whose interrogative marking is realized by a rising intonation, a pause and in particular, also a brow- movement, or occasionally a PU. Here in print, we use a slash to represent all three. The results showed that out of all 193 occurrences, BB appeared in all elliptical questions and co-occurred 24 times with PU (BB+PU).

In summary, supported by the evidence of gestures from the hearing and the deaf, we argue that a brow movement is the prototypical interrogative marker in our language, either for spoken or sign language. The essence of the question marking mechanism is that it needs a placeholder and an indicating cue to trigger the reaction of an answer for the audience. The pause or prolonging of the last word in spoken language or sign language is showing 'there is an information hole'. A PU gesture is semantically 'empty', whose function is similar to a 'pause'. It can be easily produced to enhance the visual prominence of BB as an interrogative marker. When these two co-occur and can hold, PU is originally an emphatic gesture functioning like a slash '——', symbiosis of BB (host) and PU (parasite) continues to work for a long time. Later, PU gets the function of marking interrogatives from its host. Thus, it becomes the manual prototype of interrogative marker in the sign language.

Keywords: brow movement, interrogative marker, palm-up, gesture

References

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info

Lin, H. (2019). Interrogative marking in Chinese Sign Language: A preliminary corpus-based investigation. *Sign Language & Linguistics*, *22*(2), 241-266.

Zeshan, U. (2006). *Interrogative and negative constructions in sign language*. Ishara Press.

Drawing parallels between emblems and emojis: a "thumbs-up" case study

Marina Zhukova, *University of California, Santa Barbara*

Emojis are small digital images or icons used to express emotions or ideas in electronic communication. They have become an increasingly popular way to add personality and emotion to written text, especially in social media and messaging apps (Bai et al., 2019). The first emojis were created in Japan in the late 1990s, and they quickly gained popularity in other parts of the world as well. Today, there are hundreds of different emojis available, including facial expressions, animals, objects, and symbols (Emojipedia, 2022). One way to look at emoji whose images are identical to parts of the human body (e.g., 🖐🤟🏾👏🏿👌🏼) is the opportunity to visualize body movements in digital communication, a form of representation of co-speech gestures and emblems in written speech. In the article, "Emoji as Digital Gestures" (Gawne&McCulloch, 2019), an analogy between emoji and gestures in digital communication was brought up for the first time.

As part of the text message or posts on social media platforms, emojis become frequently included in the court materials. In 2021, there were more than 100 US court opinions that mention emojis (Goldman, 2022). There are no guidelines for the court on identification of the meaning of emoji in a certain context (Murphy Kelly, 2019). To date, the question of the determination of the meaning of emoji with regard to legal cases remains understudied. To explore the question of emoji interpretation with regard to emojis that represent hand gestures, I conducted an online experiment on the platform MTurk where participants were presented with several case descriptions, and in the form of an open-ended question were asked to interpret the message that contained emojis. The survey included descriptions of five court cases whose material was publicly available online and publicly discussed in several news media outlets. For each case report, the use of emojis in text messages or in a social media post was crucial to the interpretation of the case. To be eligible to take the survey, MTurk users had to meet the following criteria: located in the U.S., 18+ years old, native speakers of English. The final sample includes responses from 121 participants in the experiment, 52% males and 48% females. The majority of participants (77.7%) were 25-44 years old.

One of the cases was related to the interpretation of the "👍 Thumbs Up" emoji (Goldman, 2019). The mother moved from Honduras to the United States together with her child, and upon arrival, she sent the father a text message, stating that she and the child arrived safely. The father responded with a "👍 Thumbs Up" emoji. The father claimed that the child's mother took the child to the U.S. without his consent, while the mother argued that the use of the emoji showed the agreement with the child’s relocation to the US. The court ruled in favor of the father, saying that the presence of the thumbs-up was not sufficient to determine the consent.

While interpreting the message, all respondents indicated that the parent did not agree to the child relocation. The “👍 Thumbs Up” emoji was interpreted as a sign of affirmation (e.g., “*responding affirmatively*”), acknowledgment (e.g., “*he is acknowledging receiving her message*”), or indicator of happy feelings (e.g., “*happy they’re saf*e”). One example of the justification is the respondent’s understanding of emoji: “*because I know what the thumbs up means, it’s a confirmation about both parties being on the same page*.” Among the reasons for these interpretations, respondents noted that it was “*based on common sense*”. Most respondents believed that the others would agree with their interpretation; one of the justifications was that “*a majority of people know what thumbs up means in real life*.” Several respondents reflected on their personal use of the “👍 Thumbs Up” emoji: “*I would consider the thumbs up to mean that he understood the message. This is how I personally use the thumbs up emoji*” and the use of the thumbs up in the US: “*that is sort of how we use the thumbs up in this country, an acknowledgment of an accomplishment, or that they agree with something.*”

The findings show that emojis have complex meanings. Emoji can be interpreted in a number of ways, depending on how people use these emoji themselves in texting. At the same time, the parallels can be drawn between the interpretation of a hand emoji and the thumbs up emblem gesture. The study contributes to the research on emoji and multimodality by providing insights about the understanding of the interpretation of the thumbs-up emoji.

Keywords: emblem gestures; emojis; ordinary meaning; experimental methods

References

Aldunate, N., & González-Ibáñez, R. (2017). An integrated review of emoticons in computer-mediated communication. *Frontiers in psychology*, *7*, 2061.

Bai, Q., Dan, Q., Mu, Z., & Yang, M. (2019). A systematic review of emoji: Current research and future perspectives. *Frontiers in psychology*, *10*, 2221.

Emojipedia (2022). In *Emojipedia*. https://emojipedia.org

Gawne, L., & McCulloch, G. (2019). Emoji as digital gestures. *language@ internet, 17*(2).

Goldman, E. (2022). Emoji Case Law Citations.

Murphy Kelly S. (2019, July). Emojis are increasingly coming up in court cases. Judges are struggling with how to interpret them. *CNN Business*.

Joker Face and Voice

Fabio Catania, *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

Micol Spitale, *Department of Computer Science & Technology, University of Cambridge*

Silvia Silleresi, *Department of Psychology, University of Milan - Bicocca*

Francesca Panzeri, *Department of Psychology, University of Milan - Bicocca*

Ironic speakers communicate their contemptuous attitude toward the thought echoed by the ironic remark (Wilson & Sperber, 2012) or towards the hypothetical person who would be so foolish to have this idea (Clark & Gerrig, 1984). To avoid misunderstandings, ironic speakers may display irony markers, i.e., metacommunicative cues that help interlocutors recognize their communicative intent: acoustically, using a characteristic intonational contour (the ironic tone of voice), and visually, with specific facial expressions and bodily movements.

Our study aims to investigate those cues. We used the material prepared by Giustolisi and Panzeri (2021), who videotaped 4 Italian students while they were pronouncing the very same sentences (5 literally positive ones, such as “The party was great fun”, and 5 literally negative ones, such as “Your hands are really dirty”) once ironically and once sincerely, and thus obtained a total of 80 videos. With this material, Giustolisi & Panzeri (2021, Study 1) found that participants could correctly discriminate between ironic and sincere comments, relying on purely acoustic cues (79% accuracy) and on purely visual cues (84% accuracy).

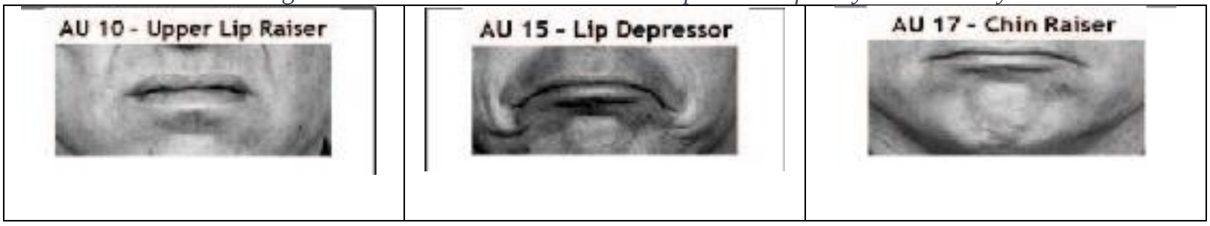
For the visual cues, we used OpenFace, an open-source facial behavior analysis toolkit (Baltrušaitis, Robinson & Morency, 2016), that, thanks to a training with dataset labels by humans, permits to automatically identify the presence and intensity of facial Action Units (AUs, Ekman & Friesen, 1978), which can be connected to facial expressions (De La Torre & Cohn, 2011). We then performed t-tests and found that ironic remarks involved more activation of the AU-10 Upper Lip Raiser (for median: t = 2.046; p = 0.0441), and AU-15 Lip Corner Depressor (standard deviation: t = 2.019, p = 0.047), and less activation of AU-17 Chin Raiser (median t = -3.094, p = 0.003). See Figure 1. For the acoustic cues, we used an artificial intelligence, whose performance is in line with that of humans, to extract the probability that the sincere and ironic audio tracks transmitted joy, sadness, anger, fear, surprise, disgust, and neutrality (Catania, 2022). Then we performed t-tests to verify whether the presence of specific emotions could function as a marker for irony detection. Comparing ironic to sincere remarks, we found that ironic speakers express more disgust (t = 2.083; p = 0.041), less anger (t = -2.869, p = 0.005), and less neutrality (t = -2.875; p = 0.005).

Overall, the analysis of the AUs characterizing ironic speakers thus revealed a major activation of two AUs, AU-10 Upper Lip Raiser and AU-15 Lip Corner Depressor, which had been related to the facial expression of disgust (Darwin, 1872/1965; Izard, 1971; Rozin, Lowery & Ebert, 1994). Disgust also emerged as the emotion primarily associated with the prosodic realization of ironic utterances. Interestingly, this *prima facie* surprising result is in line with the hypothesis proposed by Rockwell (2001) and Haiman (1998) that the visual display of sarcasm evolved from that of disgust and contempt.

Keywords: Irony recognition; Ironic tone of voice; Facial expressions

Figures

Figure 1. *Action units involved in the expression of irony and sincerity*



References

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016). Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision* (WACV).

Catania, F. (2022). *Designing and engineering emotion-aware conversational agents to support persons with neuro-developmental disorders*. (Doctoral dissertation). Politecnico di Milano, Italy.

Clark, H., & Gerrig, R. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*.

Darwin, C. (1872/1965). *The expression of the emotions in man and animals*. New York: Philosophical Library.

De La Torre, F., & Cohn, J.F. (2011). Facial expression analysis. In T.B. Moeslund, A. Hilton, V. Krüger & L. Sigal (Eds.) *Visual analysis of humans. Looking at people*, Springer.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Giustolisi, B., & Panzeri, F. (2021). The role of visual cues in detecting irony. In *Proceedings of Sinn und Bedeutung Vol. 25*.

Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford University Press on Demand.

Izard, C. E. (1971). *The face of emotion*. Appleton-Century-Crofts.

Rockwell, P. (2001). Facial expression and sarcasm. *Perceptual and Motor Skills*, 93(1).

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of personality and social psychology*, 66(5).

Wilson, D., & Sperber, D. (2012). Explaining irony, in *Meaning and relevance*, Cambridge University Press.

Analyzing User’s Mental State and Facial Expressions in Interaction with Different Personalities in a Critical Situation

Setareh Nasihati Gilani, *University of Southern California*

David Traum, *University of Southern California*

The personality of interlocutors plays a crucial role in shaping the structure and the flow of the conversation. There exists a significant body of research on personality and character attributes in dialogue systems, and on how modifying the behaviors of one interlocutor based on the conversant’s personality profile can lead to better outcomes (Yang et al., 2021). But the interaction between the personality profiles of the interlocutors and how they affect one other in the flow of the interaction has not yet been studied widely. In this work, we aim to use realtime user’s facial expressions as well as offline data collected through surveys to explore the mental state of the user and its relation to task performance upon confronting different synthetic personalities in a fast-paced simulation environment.

A wildfire simulation environment was introduced by Chaffey et al., 2019. In this simulation, the operator (human user) acts as a leader in a search and rescue operation to evacuate a town threatened by an approaching wildfire, with the object of saving as many residents as possible. The simulation consists of a series of residents with different personality profiles (e.g. stubborn person or co-operative couple) positioned randomly across the simulation map (illustrated in Figure 1), 10 aerial unmanned drone robots that perform the search & rescue tasks, and 1 transport vehicle that can evacuate residents who cannot evacuate themselves safely. The rescue task involves (1) locating residents with the help of the available swarm, (2) convincing the residents to evacuate (sometimes this requires having a direct conversation with them), and (3) (in some cases) helping them to reach safety using the transport vehicle.

In this work, we first define our method of modeling user’s mental state (using the method introduced in Shao et al., 2019) and task performance (measured as the number of rescued residents). We explore three research questions about the interaction of users with different personality profiles of virtual residents.

1- Whether there is a corelation between user’s mental state and their performance

2- Exploring the relations between different mental states in confrontation with different personality profiles of the residents

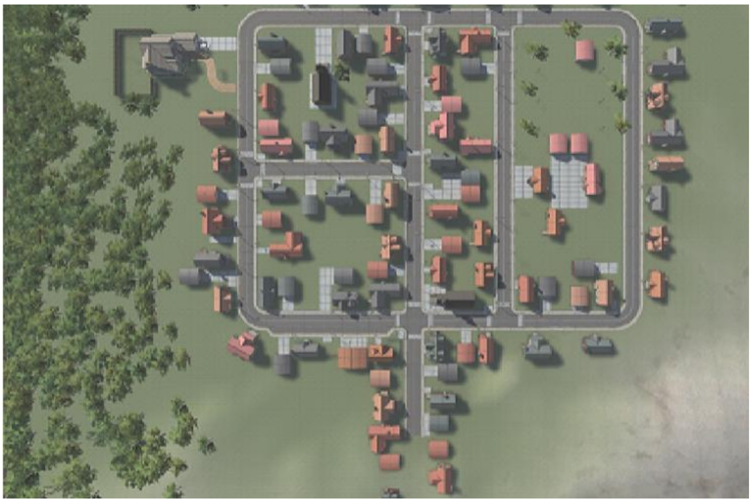
3- Whether the user’s performance is affected by the order of personality profiles that they are confronted with.

To answer our questions, we utilize the facial expressions of the operator as well as the personality models obtained in pre and post study surveys to model user's mental state. We use OpenFace software (Baltrusaitis et al., 2018) to extract the facial expressions of the operator during the course of the interaction with the simulation environment. The recorded information contains a complete log of simulation events, including the operator's timestamped actions and instructions to the spokesperson, their performance, and their frontal video recording while interacting with the system. Facial emotion extraction was done during the episodes of confrontation, which we define as the periods in which the operator has a direct and open line of communication with a resident and is conversing with them.

Keywords: Search & Rescue; Personality Traits; Facial Emotion Detection

Figures

Figure 1. *A bird's eye view of the town from Chaffey et al. (2019)*



References

Chaffey, P., Artstein, R., Georgila, K., Pollard, K. A., Gilani, S. N., Krum, D. M., ... & Traum, D. (2019). Developing a virtual reality wildfire simulation to analyze human communication and interaction with a robotic swarm during emergencies. In *Proceedings of the 9th Language and Technology Conference*.

Yang, R., Chen, J., & Narasimhan, K. (2021, Αύγουστος). Improving Dialog Systems for Negotiation with Personality Modeling. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 681–693. https://doi.org/10.18653/v1/2021.acl-long.56

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, May). Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 59-66). IEEE.

Shao, Z., Chandramouli, R., Subbalakshmi, K. P., & Boyadjiev, C. T. (2019). An analytical system for user emotion extraction, mental state modeling, and rating. *Expert Systems with Applications*, *124*, 82-96.

Talking and gesturing about motion at different L2 proficiency levels

Christina Piot, *University of Liège & University of Lille*
Julien Perrez, *University of Liège*
Maarten Lemmens, *University of Lille*

The typological differences between verb-framed and satellite-framed languages observed by Talmy (2000) have been shown to be reflected in co-speech gestures as well (Gullberg, 2009; Kita & Özyürek, 2003; McNeill & Duncan, 2000). More specifically, differences between the types of language have been observed in terms of the semantic components encoded in gestures and the synchronization between gestures and speech. Such gestures should therefore be taken into account when studying L2 learners' thinking for speaking patterns (Urbanski & Stam, 2022). Against this background, our study aims at determining how motion events are expressed in speech and co-speech gestures by native French speakers, Dutch native speakers, and CLIL French-speaking learners of Dutch.

We conducted an elicitation experiment in which participants recounted the cartoon *Tweety and Sylvester: Tweet Zoo* (Freleng, 1957). Fifteen French speakers, fifteen Dutch speakers, and fifteen CLIL French-speaking learners of Dutch with a proficiency level ranging from A1 to B2 completed the task. We identified the semantic components (manner and path) encoded in the verbs and satellites. Gestures were classified as iconic, beat, deictic, or pragmatic (Kendon, 2004; McNeill, 1992). Iconic and deictic gestures were further analyzed regarding the semantic components of motion they convey (e.g., manner, path, ground, manner + path). Finally, we looked at the synchronization between speech and gestures (Urbanski & Stam, 2022).

So far, 592 utterances and 741 gestures have been analyzed and our results show that French speakers tend to use $\text{PATH}_{\text{VERBS}}$+$\text{PATH}_{\text{SATELLITES}}$+$\text{PATH}_{\text{GESTURES}}$ (see Figure 1) in both their L1 and L2 to describe self-propelled motion events, whereas Dutch speakers prefer using $\text{MANNER}_{\text{VERBS}}$+$\text{PATH}_{\text{SATELLITES}}$+$\text{PATH}_{\text{GESTURES}}$ (see Figure 2). As their proficiency level increases, learners use less often constructions only consisting in $\text{MANNER}_{\text{VERBS}}$ and use more often $\text{MANNER}_{\text{VERBS}}$+$\text{PATH}_{\text{SATELLITES}}$ constructions. Learners with a pre-intermediate level align path gestures with verbs less often and more often with linguistic units that are not core elements of motion events in comparison with French speakers, Dutch speakers, and learners with an intermediate level. Finally, learners with a pre-intermediate level produce more manner fog gestures (which are often compensation gestures, see Figure 3) and location gestures than learners with an intermediate level and native speakers. The number of pragmatic gestures also tends to decrease slightly as the L2 proficiency level increases. These tendencies suggest that

CLIL French-speaking learners of Dutch rely more on gesture than L1 speakers and especially when they have a lower proficiency level.

Keywords: motion events; thinking for speaking; gesturing in L2

Figure 1. *PATH$_{GESTURE}$ co-occurring with the PATH$_{VERB}$ passer and PATH$_{SATELLITE}$ à côté in "Il **passe à côté**" (FR5, ME31) [He passes by]*

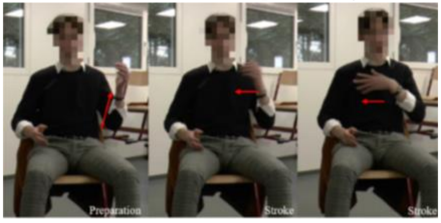


Figure 2. *PATH$_{GESTURE}$ co-occurring with the MANNER$_{VERB}$ lopen and PATH$_{SATELLITE}$ op en neer in "En hij is **op en neer aan het lo**pen" (DU1, ME26) [And he is walking up and down]*

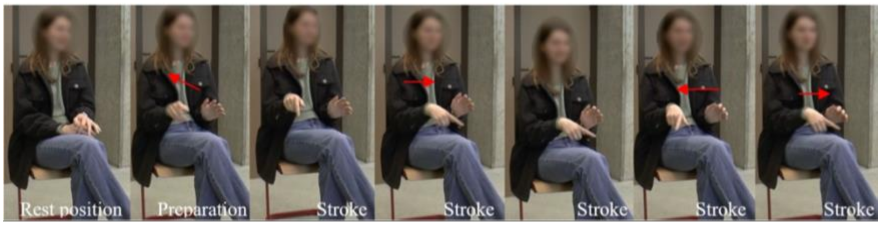
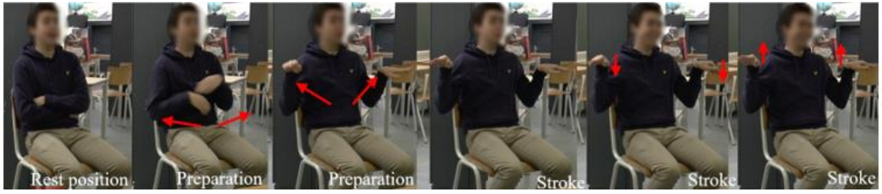


Figure 3. *MANNER$_{GESTURE}$ co-occurring with the kwikwi in "Titi is een vogel dus hij **kwikwi**" (CLIL9, ME66) [Tweety is a bird so he kwikwi]*



References

Freleng, F. (Director). (1957). Tweet Zoo. In *Merrie Melodies*. https://www.youtube.com/watch?v=SqhwmHrdf74

Gullberg, M. (2009). Gestures and the development of semantic representations in first and second language acquisition. *Acquisition et Interaction En Langue Étrangère, Aile... Lia 1*, 117–139. https://doi.org/10.4000/aile.4514

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

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

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

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture* (1st ed., pp. 141–161). Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.010

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 2). MIT Press.

Urbanski, K. (Buescher), & Stam, G. (2022). Overview of Multimodality and Gesture in Second Language Acquisition. In G. Stam & K. (Buescher) Urbanski, *Gesture and Multimodality in Second Language Acquisition* (1st ed., pp. 1–25). Routledge. https://doi.org/10.4324/9781003100683-1

Micro Analysis of EFL Teachers' Gesture Use as a Pedagogical Tool in Video-Mediated Interaction

Asuman Şimşek Tontuş, *Middle East Technical University*
Safiye İpek Kuru Gönen, *Anadolu University*

Providing imagistic thoughts of the messages, gestures are indispensable components of communication. They establish a high degree of intersubjectivity among interlocutors by developing a sense of the shared social, physical, symbolic, and mental space (McCafferty, 2002). By providing 'two simultaneous views of the same process' (McNeill, 1985, p. 350), gesturally enhanced input engenders a greater comprehension and even acquisition in language learning (Gullberg, 2008). Unlike gestures used for everyday communicative purposes, teachers' gestures are pedagogically informed and used for a particular purpose in language classrooms (Stam & Tellier, 2021). As stated by Tellier (2006), teachers' gestures can be classified as information gestures, classroom management gestures, and assessment gestures, indicating that teachers' gestures are multifunctional and pedagogically informed.

Language teaching in video-mediated environments has become a common channel for language teaching all over the world, especially after the sudden outbreak of the pandemic in 2020. As a result, new video-mediated tools such as Zoom and Webex have been predominantly integrated into language teaching, which in turn has paved the way for synchronous video-mediated interaction (henceforth VMI). In addition to verbal cues, teachers' facial expressions, gestures, and body stances are considered to be the key factors affecting students' social presence (Wei et al., 2012) and help students establish a relationship with teachers (Witt & Wheeless, 2001). Studies focusing on VMI revealed that teachers deployed gestures for classroom interaction management (Holt et al., 2015; Malabarba et al., 2022) and explicate lexical items (Codreanu & Celik, 2013). The language classrooms have broadened their physical borders to online environments not just because of the pandemic but also the opportunities which these environments provide. Therefore, it is crucial to understand online interaction to develop a better pedagogical design and exploit the opportunities of these environments (Jakonen et al., 2022). Against this background, this study focused on two research-led questions on (1) the functions and (2) sequential organizations of EFL teachers' gesture use during synchronous VMI English in L2 classrooms. Following the Jeffersonian transcription convention (2004) and Mondada's multimodal transcription convention (2018), the data was analyzed via Multimodal Conversation Analysis. Accordingly, the teachers utilized a variety of gestures based on their pedagogical purposes in video-mediated L2 classroom interaction for language explanations (vocabulary and

grammar) and classroom interaction management (turn-management and giving instruction). This study unveiled that the prominent functions of teachers' gestures were to create a mutual physical and mental space for learners by bringing the physical description of abstract concepts into a shared virtual world across screens. Moreover, it is quite apparent that teachers' gestures share similar and different properties in synchronous VMI and face-to-face classrooms.

Keywords: Language teaching; Teachers' gestures; Video-mediated interaction

References

Codreanu, T., & Celik, C. (2013). Effects of webcams on multimodal interactive learning. *ReCALL*, 25(1), 30-47.

Gullberg, M. (2008). Gestures and second language acquisition. In Nick C. Ellis & Peter Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 276- 305). Routledge.

Holt, B., Tellier, M., & Guichon, N. (2015). The use of teaching gestures in an online multimodal environment: the case of incomprehension sequences. In *Gesture and Speech in Interaction 4th Edition*.

Jakonen, T., Dooly, M., & Balaman, U. (2022). Interactional practices in technology-rich L2 environments in and beyond the physical borders of the classroom. *Classroom Discourse*, *13*(2), 111-118.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In. G. Lerner (Ed.) *Conversation analysis, studies form first generation* (pp. 13-34). John Benjamins.

Malabarba, T., Mendes, A. C. O., & de Souza, J. (2022). Multimodal Resolution of Overlapping Talk in Video-Mediated L2 Instruction. *Languages*, *7*(2), 154. MDPI AG.

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *Modern Language Journal*, *86*(2), 192–203.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*(3), 350–371.

Mondada, L. (2018). Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Research on Language and Social Interaction, 51*(1), 85-106.

Wei, C. W., & Chen, N. S. (2012). A model for social presence in online classrooms. *Educational Technology Research and Development*, *60,* 529-545.

The benefits of multimodal communication in the foreign language classroom: The use of hand gesture to teach morphology and word structure

Sara Feijoo, *Universitat de Barcelona*
Mariona Anglada, *Universitat Oberta de Catalunya*
Núria Esteve-Gibert, *Universitat Oberta de Catalunya*

Previous studies show that children's observation of hand gestures impacts their processing of narrative discourse structure (Vilà-Giménez et al., 2019) and mathematical relations (Goldin-Meadow, 1999). The present study examines whether hand gestures can also improve morphological awareness among foreign language learners. Morphological awareness is the conscious ability to perceive, analyze, and manipulate the morphemic structure of words, and it is positively correlated with reading skills (Carlisle, 2000). The main objective of the present study is to explore whether the use of hand gestures iconically signalling the morphemic structure of words increases learners' morphological awareness when learning new words in a foreign language.

38 British learners of Spanish as a foreign language (grades 10 to 13, age range 14 to 18) were tested in a short intervention study for a total of 3 training sessions. During the training phase, participants were presented with morphologically complex words in one of these four conditions: an audio-highlighting condition (the experimenter marks each morpheme with a pitch accent), an audio-visual-highlighting condition (each morpheme is marked with text in a different colour and with a pitch accent), an audio-gesture-highlighting condition (each morpheme receives a pitch accent and a hand gesture pointing at the boundary between the stem and the morpheme), and a control condition (no specific highlighting). Learners' morphological awareness with derivational morphology and compounding were assessed before and after the intervention by means of a morphological awareness test adapted from Carlisle (2000).

A repeated-measures ANOVA found no significant differences across the four different groups in terms of total learning gains in morphological awareness. However, a repeated measures *t-test* revealed significant differences from pre- to post-test in the gesture group only (*t*=-2.639; *p*=.027), while no significant learning from pre- to post-test was found in any of the other conditions. Furthermore, when compounding and derivation skills were analysed separately, significant differences were found between the gesture group and the other groups in terms of the gains in the compounding subtest (F(3,34)=2.958, *p*=.046), but not the derivational morphology subtest.

Our preliminary results show that the highlighting of the morphemic structure of words with gesture can be an efficient strategy to promote learners' development of morphological awareness, especially in terms of compound words, after only a few training sessions. Second language learners seem to benefit more from multimodal input exposure than from acoustically-enhanced input or audiovisually-enhanced input alone. The evidence of these findings provides new teaching techniques for the foreign language classroom that can help to boost an important skill which is correlated with reading performance.

Keywords: hand gesture; morphological awareness; foreign language learning

References

Carlisle, J. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal*, (12). 169–190. https://doi.org/10.1080/02702710390227369

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, *91*(4), 720–730. https://doi.org/10.1037/0022-0663.91.4.720

Vilà-Giménez, I., Igualada, A., & Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Developmental Psychology*, *55*(2), 250-262. https://doi.org/10.1037/dev0000604

The impact of focus types on the prosody-gesture link in Catalan and German: a focus elicitation production study

Paula Sánchez-Ramón, *Universitat Pompeu Fabra, Goethe University Frankfurt*
Alina Gregori, *Goethe University Frankfurt*
Pilar Prieto, *Universitat Pompeu Fabra, Institució Catalana de Recerca i Estudis Avançats*
Frank Kügler, *Goethe University Frankfurt*

In the last decades, research has shown that gesture and speech are highly interconnected (e.g. McNeill, 1992; a.o.), and that information structure and prosody correlate in terms of prominence (Féry & Kügler, 2008; Dufter & Gabriel, 2016), but the role of focus types in the prosody-gesture link has not been considered previously, nor the role of focus types in the attraction of gesture use in adult speech. For French-speaking children, Esteve-Gibert et al. (2021) found that head gestures rather than prosodic features were used to indicate the informational status of discourse referents, suggesting that those gestures with no referential connection to speech may play a linguistic structural role in communication.

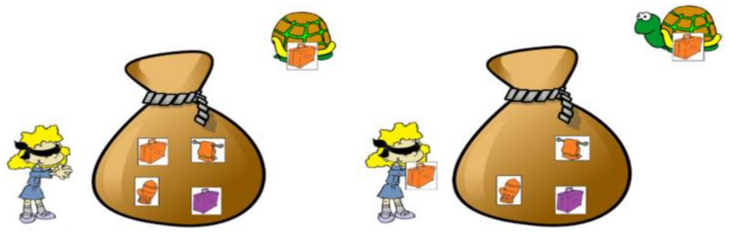
We investigate the impact of prominence degrees (in the form of focus types) on pitch accentual prominence, on gestures, and on their synchronization. Following Krifka (2008), focus conditions are classified as: information focus (most important information), contrastive focus (overt presence of alternatives), corrective focus (disagreement to a previous statement). Contrastive and corrective conditions have a stronger prosodic prominence than broad information focus conditions across languages (Zimmermann, 2008). Thus, we hypothesize that pitch accentuation and gesture will be mostly associated with corrective and contrastive focus constituents, rather than with information focus. A hypothesis for a second research question contemplates that focus will be more extensively marked by gestures in Catalan than in German, as Catalan is less systematic in the prosodic marking through deaccentuation of non-focus and given referents.

A production study is currently being conducted relying on an adaptation of the elicitation method by Esteve-Gibert et al. (2021), which is suitable to investigate the synchrony between prosody and gestures in different focus types. The method consists of pictures prompted in a digital board game (Figure 1). Participants, video-recorded, communicate with an animated conversation partner who is blindfolded and their task is to request certain objects from the digital "speaker". The focus types can be prompted and controlled by the responses of the animated "speaker". For data coding, prosody will be analyzed using ToBI adaptations for each language (Grice et al., 2005; Prieto et al., 2015) by assuming that pitch accents are associated with different levels of prominence (Baumann & Röhr, 2015; Kügler & Calhoun, 2020). Regarding gesture labeling, head nods and hand gestures are expected to be collected. Apexes will be annotated using the M3D labeling system (Rohrer et al., 2020). We expect to analyze the data in the upcoming months for both German and Catalan.

Keywords: non-referential gestures; focus; methodology; prosody-gesture link

Figures

Figure 1. *Example of the elicitation method by Esteve-Gibert et al. (2021). Contrastive focus: Bag contains more items, and two suitcases differ in color.*



1a. "Agafa la maleta TARONJA"
1b. "Nimm den ORANGENEN Koffer"
Take the ORANGE suitcase

References

Baumann, S. & Röhr, C. (2015). The perceptual prominence of pitch accent types in German. 384.

Dufter, A. & Gabriel, C. (2016). Information structure, prosody, and word order. In Fischer, S. & Gabriel, C. (Eds.), *Manual of grammatical interfaces in Romance* (pp. 419–456). De Gruyter.

Esteve-Gibert, N., Lœvenbruck, H., Dohen M. & D'Imperio, M. (2021). Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech. *Developmental Science*. e13154.

Féry, C. & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics 36(4)*. 680–703.

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

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3), 243–276.

Kügler, F. & Calhoun, S. (2020). Prosodic Encoding of Information Structure: A typological perspective. In Gussenhoven, C. & Chen, A. (eds.): *The Oxford Handbook of Language Prosody* (pp. 454-467). Oxford University Press.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. UCP.

Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., Sichel-Bazin, R., & Vanrell, M.M. (2015). Intonational phonology of Catalan and its dialectal varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 9-62). Oxford University Press.

Rohrer, P., Vilà-Giménez, I., Florit-Pons, J., Gurrado, G., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S. & Prieto, P. (2020). *The MultiModal MultiDimensional (M3D) labeling system for the annotation of audiovisual corpora: Gesture Labeling Manual*. UPF Barcelona.

Zimmermann, M. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica, 55*, 347–360.

The Distribution of Non-referential Gestures, Information Structure and Prosody: A Corpus Study on Prominence Peak Alignment

Alina Gregori, *Goethe University Frankfurt*

Frank Kügler, *Goethe University Frankfurt*

This study investigates the impact of pragmatic prominence on prosody-gesture alignment (McNeill, 1992; Loehr, 2012; Im & Baumann, 2020) in spontaneous German speech. The synchronization of non-referential gesture apexes, which provide structural information on the discourse (McNeill, 1992) and pitch accents of different degrees of prosodic prominence (unaccented < L* < !H* < H*+L < H* < L+H*; in accordance with Baumann & Röhr, 2015; Kügler & Calhoun, 2020) is investigated. Given that prosodic prominence varies as a function of information structure (IS, cf. Baumann & Röhr, 2015; Kügler & Calhoun, 2020) we address the research question whether the alignment frequency and (temporal) accuracy of prosody and gestures is mediated by IS in German spontaneous speech.

The data were taken from the SaGA corpus (Lücking et al., 2010). 18 dialogues (204 min.) of task-oriented spontaneous speech were analyzed. The corpus provides word and gesture type annotation. Stroke and apex annotation was done following Rohrer et al. (2020), pitch accent annotation with GToBI (Grice et al., 2005). Following Götze et al. (2007), annotation of IS was done considering information status (Given, Accessible, New) and focus (new-information focus NF, contrastive focus CF, non-focus). The occurrence of gestures in relation to pitch accents and IS categories was extracted.

775 non-referential gesture apexes were found in the corpus, from which 39,5% aligned with IS referents. The remaining 60,5% of apexes occurred on words not coded for IS (Fig. 1). The apexes on IS referents occurred more often with the prominent categories *new* (Fig. 1a) and *focused* (Fig. 1b) than with less prominent categories. Adding pitch accents to this comparison, it stands out that the alignment accuracy correlates with prominence: apexes appear most accurately with L* accents on *given* referents and with L+H* accents on *new* referents. While in general, no correlation of pitch accents and IS was found with regard to prominence (Fig. 2a), considering only referents accompanied by an apex, the alignment of pitch accents L+H*, H* and L* with their IS categories improved (Fig. 2b).

Strikingly, most referents were accented, even though e.g., *given* and *NF* referents are usually prosodically less prominent than *new* and *CF* referents (Kügler & Calhoun, 2020). Presumably, this result might be task-specific behavior of the interlocutors to signal that for memorizing a route (the task of the SaGA corpus), emphasizing every detail, even already active referents, is relevant. Although many of the non-referential gesture apexes were not likely to align with IS referents, these results suggest that prosody-gesture alignment is mediated by information structure: more pragmatic prominence leads to more prosodic and gestural prominence while increasing the accuracy in alignment between these modalities.

Keywords: prosody-gesture link; information structure; corpus study; spontaneous speech

Figures

Figure 1. *Occurrences of gestures on a) levels of information status and b) focus categories in percent.*

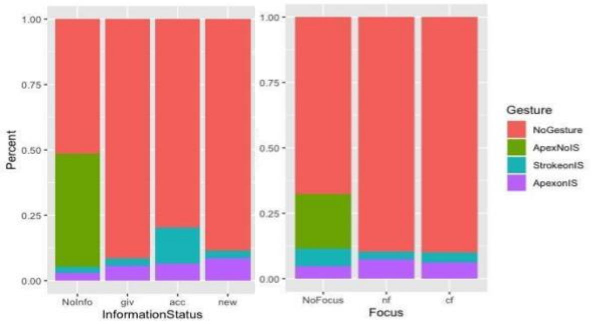
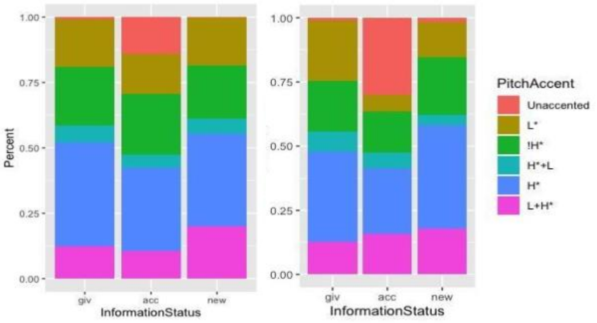


Figure 2. *Occurrences of pitch accents on levels of information status a) in general and b) when an apex is produced.*



References

Baumann, S. & Röhr, C. (2015). The perceptual prominence of pitch accent types in German. *Proc. 18th ICPhS*, Glasgow, paper 384.

Götze, M., Weskott, T., Endriss, C., Fiedler, I., Hinterwimmer, S., Petrova, S., Schwarz, A., Skopeteas, S. & Stoel, R. (2007). Information structure. In Dipper, S., Götze, M. & Skopeteas, S. (eds.), *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics, and Information Structure* ISIS 7, 147–187. Potsdam: Universitätsverlag Potsdam.

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

Im, S. & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proc. of the Linguistic Society of America* 5(1). 685–697.

Kügler, F. & Calhoun S. (2020). Prosodic Encoding of Information Structure: A typological perspective. In Gussenhoven, C. & Chen, A. (Eds.), *The Oxford Handbook of Language Prosody* (pp. 454-467). Oxford University Press.

Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *3*(1). 71–89.

Lücking, A., Bergmann, K., Hahn, F., Kopp, S. & Rieser, H. (2010). The Bielefeld Speech and Gesture Alignment Corpus (SaGA). *ELRA*. 92–98.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. UCP.

Rohrer, P., Vilà-Giménez, I.; Florit-Pons, J., Gurrado, G.., Esteve-Gibert, N., Ren, A., Shattuck-Hufnagel, S. & Prieto, P. (2020). *The MultiModal MultiDimensional (M3D) labeling system for the annotation of audiovisual corpora: Gesture Labeling Manual*. MS, UPF Barcelona.

Audiovisual prosody and information structure in French: An investigation of the marking of contrastive focus and its subtypes

Clara Lombart, *University of Namur and University of Mons*

This presentation explores the interplay between information structure (IS), prosody, and gesture, with a focus on how contrastive focus is marked. Contrastive focus refers to the opposition between several explicit alternatives that form a limited set of possibilities (e.g. Repp, 2016). Previous research has shown that contrast directly influences how utterances are encoded through morphology, lexicon, syntax, or prosody (e.g. Zimmermann & Onea, 2011). However, these effects can only be distinguished by considering different types of contrast (e.g. Umbach, 2004), which are defined differently depending on the researcher. Recent studies have also demonstrated that gestures, either alone or in conjunction with prosodic cues, are used to encode IS (e.g. Debreslioska & Gullberg, 2020; Im & Baumann, 2020). More specifically, given the tight relationships and temporal synchronisation between gestures and prosodic cues (such as pitch accents or phrasing modifications), some scholars have redefined prosody as an audiovisual component (see Shattuck-Hufnagel & Ren, 2018 for a review).

Studies on the relationships between audiovisual prosody and contrast remain limited for French, as they often fail to include different contrast types or mainly concentrate on hand gestures, leaving aside the non-manual cues and combinations of gestural markers (see Ferré 2014 for an exception). This presentation aims to fill these gaps by analysing the encoding of three contrast types in French: discourse opposition (1), selection (2), and correction (3).

(1) Some have [**a square shape**] (…). Some are [**rather triangular**] (…).

(2) A: Dark or light blue? / B: [**Light**].

(3) You told me a triangle-shaped eyebrow. It's more like [**a circumflex**].

From this perspective, we examined the productions of six Belgian French speakers during description or categorisation tasks extracted from the FRAPé Corpus (i.e. *Corpus de Français Parlé*). Approximately 80 contrasts per participant were delimited from the informational perspective. The Inter-Pausal units that contained, preceded, and followed them were annotated for syllabic duration, pitch mean and range, tone (Hirst & Di Cristo 1998), articulation rate, and degree of prominence (based on ANALOR). Hand, head and eyebrow gestures as well as body leans were also taken into account.

Descriptive results indicate that the three contrast types are distinctly marked by prosodic and gestural cues. Contrast is characterised by longer syllabic duration, F0 rise on the first syllable, lower pitch range, and frequent pauses. The degree of prominence is also higher for contrastive foci and corrections than for non-contrastive items and discourse oppositions. Moreover, contrast fosters the production of gestures (except for body leans), and the different types of contrast are related to several gesture forms. For example, non-referential gestures and deictics occur more on discourse oppositions while representational gestures appear more on corrections. Gestures can also combine with one another, and combinations of three or four gestures mostly take place on corrections, while discourse oppositions are accompanied by only one gesture. Finally, some interactions arise between the prosodic and gestural cues. These depend on the contrast type, but a common pattern in the data sample is that the more prominent is a contrastive focus at the prosodic level, the more speakers tend to combine different types of gestures.

Ultimately, considering the multimodality of expression, this research opens new avenues for a more thorough definition of contrast and the functioning of audiovisual prosody in French.

Keywords: French; Information Structure; Gesture; Prosody; Contrastive focus

References

Debreslioska, S., & Gullberg, M. (2020). What's New? Gestures Accompany Inferable Rather Than Brand-New Referents in Discourse. *Frontiers in Psychology*, *11*(1935). https://doi.org/10.3389/fpsyg.2020.01935

Ferré, G. (2014). A multimodal approach to markedness in spoken French. *Speech Communication*, *57*, 268–282. https://doi.org/10.1016/j.specom.2013.06.002

Hirst, D., & Di Cristo, A. (1998). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

Im, S., & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America*, *5*(1), 685. https://doi.org/10.3765/plsa.v5i1.4755

Meurant, L., Lepeut, A., Gabarró-López, S., Tavier, A., Vandenitte, S., Lombart, C., & Sinte, A. (Under Construction). *Corpus de français parlé: Vers la construction d'un corpus comparable LSFB - Français de Belgique.*

Repp, S. (2016). Contrast: Dissecting an Elusive Information-structural Notion and its Role in Grammar. In C. Féry & I. Shinichiro (Eds.), *The Oxford Handbook of Information Structure* (pp. 270–289). https://doi.org/10.1093/oxfordhb/9780199642670.013.006

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, *9*(1514). https://doi.org/10.3389/fpsyg.2018.01514

Umbach, C. (2004). On the Notion of Contrast in Information Structure and Discourse Structure. *Journal of Semantics*, *21*(2), 155–175. https://doi.org/10.1093/jos/21.2.155

Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, *121*(11), 1651–1670. https://doi.org/10.1016/j.lingua.2011.06.002

Prosodic features of speech in synchrony with four pragmatic gestures

Gaëlle Ferré, *University of Poitiers & FOReLLIS lab, France*

This presentation focuses on the prosodic features of speech that accompanies four pragmatic gestures. **Beats** are considered as ‘highlighters’ (Biau & Soto-Faraco, 2013), that help “the listener direct the focus of attention on important information and modulate *how* information is treated”. **Pointing gestures** are also considered as joint-attentional behavior (Enfield et al., 2007), which ensures referential understanding between interactants and regulate interpersonal relationships. Édeline & Klinkenberg (2021) add that pointing allows attention to be focused on a specific portion of space. In this respect, their function is close to that of beats which draw one’s attention to a specific portion of speech. In the literature, Cienki (2021) describes **Palm-Up-Open-Hand gestures** (henceforth PUOH) and the way they emphasize a new point in discourse. These three gestures are illustrated below in Figure 1(a-c). The last two images in the figure illustrate what could be considered as the exact opposite of the PUOH gesture, namely **hands closing**, a movement or gesture which has not been studied so far in the literature to the best of our knowledge, but which can be described as emphasizing the end of a discourse unit.

Although these gestures have therefore been described as emphasizing some part of speech, they do not act at the same discourse level and clearly have **different distributions in the flow of** speech. The hypothesis developed in the present paper is that their focus type is reflected in the **prosody** of accompanying speech, especially **pitch key** (Top, Bottom, Mid), **tone** (Flat, falling tone, rising tone), and **relative pitch range** (Upstepped, Downstepped or Same). The prediction is that PUOH gestures and closing hands are expected to be located at the periphery of Intonation Phrases or Inter-Pausal Units (IPUs), and are more likely to be co-occurrent with **pauses** in speech than beats and points. Because of their focusing functions of smaller speech units, the latter two gestures are expected to co-occur more frequently with **emphatic stress** in speech.

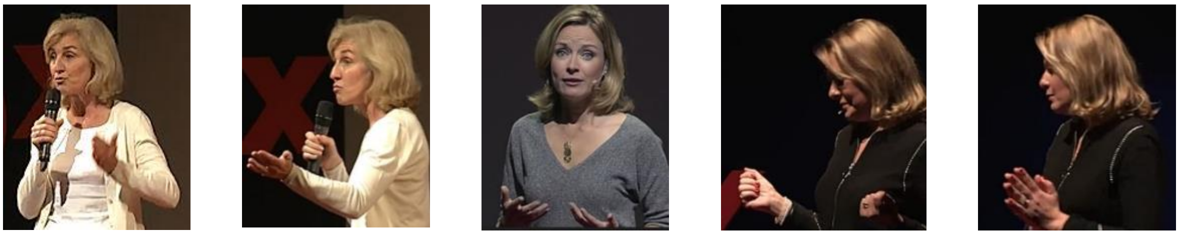
In order to test these hypotheses, 1162 gestures strokes have been coded in **Elan** (Sloetjes & Wittenburg, 2008) in a corpus of **3 TED-Talk videos in French**. The prosodic parameters of the speech that accompanies each stroke, which are listed in the previous paragraph, have been coded using **Praat** (Boersma & Weenink, 2009) and the **Intsint automatic algorithm** (Hirst, 2007) as help to calculate relative pitch range and key.

Chi2 tests were conducted using R v. 4.6.2 (R Core Team, 2012) and confirmed most of the predictions made. They revealed that the four gestures did not occur in the same place in IPUs and IPs of the accompanying speech which is therefore pronounced with varying pitch key, pitch range and tone contours. The tests also revealed that beats show a different distribution regarding pauses and prosodic emphasis from other gesture types.

Keywords: prosody; beats; points; PUOH; closing gestures

Figures

Figure 1. *Gesture types coded in 3 TED Talk videos (total number of occurrences: 1162)*



(a) Beats

(b) Points

(c) PUOH

(d) Closing hands

(e) Closing hands

References

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language, 124(2)*, 143-152.

Boersma, P., Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05).

Cienki, A. (2021). From the finger lift to the palm-up open hand when presenting a point: A methodological exploration of forms and functions. *Languages and Modalities, 1*, 1-14.

Édeline, F., & Klinkenberg, J.-M. (2021). L'index. Un dispositif sémiotique puissant et méconnu. In D. Bertrand & I. Darrault-Harris (Eds.), *À même le sens. Hommage à Jacques Fontanille* (pp. 253-263). Limoges: Lambert Lucas.

Enfield, N. J., *et al.* (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics, 39*, 1722-1741.

Hirst, D. (2007). *A Praat plugin for momel and intsint with improved algorithms for modelling and coding intonation*. In: Proceedings of ICPhS XVI, Saarbrücken, Germany, 1233-1236.

R Core Team (2012). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Computer program: [http://www.r-project.org].

Sloetjes, H., Wittenburg, P., 2008. Annotation by category – ELAN and ISO DCR, in: *Proceedings of 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, 816-820.

Pointed pronouns: The systematic co-occurrence of manual pointing gestures and emphatic pronouns in Jaru

Josua Dahmen, *Australian National University*

Cross-linguistically, pronouns (see Bhat, 2004) are the most common forms that conversationalists use to accomplish reference to co-participants. Because exophoric pronouns are inherently deictic expressions (Levinson, 2006), an accompanying demonstration act is often needed to identify the referent (see, e.g., Kaplan, 1989). Whereas pronominal signs in sign languages are fully grammaticalized pointing signs in themselves (Cormier et al., 2013), exophoric pronouns in spoken languages are often combined with visual-corporal indexical practices (i.e., gaze and pointing gestures) to disambiguate referents and addressees and to calibrate the participation frameworks during a multi-party conversation (see also Dahmen & Blythe, in press). While spoken languages with a single set of pronouns often rely on prosodic features to convey prominence, languages with dual pronoun systems usually exhibit a functional split between the two types of pronouns where free pronouns convey discourse prominence (Schwartz, 1986; Choi, 1999) within an "accessibility hierarchy" (Ariel, 1988). This is also the case for Jaru, a highly endangered Pama-Nyungan language spoken in northern Western Australia, whose personal pronoun system is composed of both free pronouns and pronominal clitics. While the obligatory pronominal clitics in Jaru (e.g., *=n* 'you$_{\text{SG}}$', *=liyarra* 'we$_{\text{DU.EXC}}$') are pragmatically unmarked, optional free pronouns (e.g., nyundu 'you$_{\text{SG}}$', *ngali* 'we$_{\text{DU.EXC}}$') convey discourse prominence. Even though personal pronouns in Jaru encode grammatical distinctions such as clusivity and duality, any pronouns other than first-person singular are still potentially ambiguous when used to refer to co-participants of a multi-party conversation, so speakers commonly draw on visual-corporal resources in conjunction with the referential expressions.

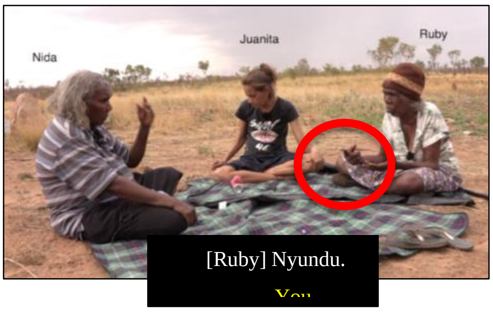
This paper examines how the Jaru pronominal system intersects with participants' visual-corporal conduct to accomplish co-participant reference. The study is based on a recently compiled corpus of video-recorded multi-party conversation between family and friends in Jaru. The recordings were made using lapel microphones and high-definition cameras. In order to obtain an accurate representation of the linguistic and nonverbal practices in the community, the participants did not receive any instructions regarding language choice or conversation topic. Thus far, three hours and thirty minutes of recordings have been transcribed in ELAN. A total of 498 instances of pronouns (both bound and free) indexing co-present participants were collected from eight ten-minute samples with distinct participant constellations.

The study finds a significant correlation between the occurrence of free (emphatic) pronouns and manual pointing gestures in Jaru: Manual pointing gestures that co-occur with pronominal reference forms almost always involve free pronouns, while gaze direction as sole visual-corporal indicator is more common in conjunction with bound pronouns. This highlights an aspect of multimodal prominence in Jaru that is expressed through a combination of linguistic and nonverbal means. However, not all free pronouns co-occur with pointing gestures – only those that may otherwise result in misidentification of the referent. By uncovering a language-specific correlation of emphatic personal pronouns and pointing gestures, this investigation contributes to a more complete understanding of multimodal person reference in interaction and highlights the close link between linguistic and visual prominence markers.

Keywords: pointing; emphatic pronouns; multi-party conversation; multimodal prominence; language-body interface; Australian Aboriginal languages

Figure

Figure 1. *Pointing gesture to a co-participant in conjunction with a free second-person singular pronoun (Note: The name tags and faces are shown in accordance with participants' preferences.)*



References

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24(1), 65–87.

Bhat, D. N. S. (2004). *Pronouns*. Oxford University Press.

Choi, H.-W. (1999). *Optimizing structure in context: Scrambling and information structure*. CSLI Publications.

Cormier, K., Schembri, A., & Woll, B. (2013). Pronouns and pointing in sign languages. *Lingua, 137*, 230–247.

Dahmen, J., & Blythe, J. (in press). Calibrating recipiency through pronominal reference. *Interactional Linguistics*.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford University Press.

Levinson, S. C. (2006). Deixis. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 97–121). Blackwell.

Mushin, I., & Simpson, J. (2008). Free to bound to free?: Interactions between pragmatics and syntax in the development of Australian pronominal systems. *Language, 84*(3), 566–596.

Schwartz, L. (1986). The function of free pronouns. In U. Wiesemann (Ed.), *Pronominal systems* (pp. 405–436). Narr.

Cross-cultural differences in gesture frequency

Elena Nicoladis, *University of British Columbia*

Hui Yin, *Xi'an Jiaotong*-*Liverpool University*

Paula Marentette, *University of Alberta, Augustana Campus*

Some studies have shown cross-cultural differences in how frequently people gesture. For example, So (2010) found that Mandarin speakers gestured less frequently than English speakers. However, not all studies have shown these cultural differences, or at least not in the same direction. Goldin-Meadow and Saltzman (2000) found that Mandarin-speaking mothers gestured more when speaking to their children than did American English-speaking mothers.

In order to account for differing patterns of results across studies, Nicoladis et al. (2018) argued that the cultural differences in gesture frequency might be mediated by storytelling style: a chronicle style (storytellers tell what happened and how it happened) and an evaluative style (storytellers tell what happened and why it happened). They reasoned that a chronicle style would be strongly associated with the production of representational gestures (i.e., gestures that represent a referent through hand movement and/or hand shape) since these gestures can convey what and how events happened. Indeed, they found that Hindi- and Mandarin-English bilinguals gestured less in both of their languages than French- and Spanish-English bilinguals. Also, the Hindi and Mandarin first language speakers were more likely to use an evaluative style than the other participants. The purpose of the present study was to test whether these results generalize to monolinguals. Some previous studies have found differences in gesture frequency between bilinguals and monolinguals (So, 2010).

A total of 41 participants were included in this study: 15 Mandarin monolinguals, 15 English monolinguals, and 11 French monolinguals. Cross-cultural differences have been reported in sample sizes as small as 10 (So, 2010). The participants watched a four-minute segment of a cartoon and recounted the story of what happened. The participants' stories were videotaped for later transcription and coding. Following Nicoladis et al. (2018), we focused on the participants' production of representational gestures since they argued that representational gestures should be particularly strongly associated with a chronicle style of storytelling. The dependent variable was the gesture rate, or the number of gestures used per 100 words. Figure 1 summarizes the gesture frequency results; note that the English monolinguals were more highly variable in gesture frequency than the Mandarin and the French monolinguals. Because of unequal variances between groups, we compared the groups using the Kruskal-Wallis test. This test revealed a statistic of 5.99, *p* = .05. Post-hoc LSD tests showed a tendency for the Mandarin monolinguals to gesture less than the French monolinguals (*p* = .056) but no other differences.

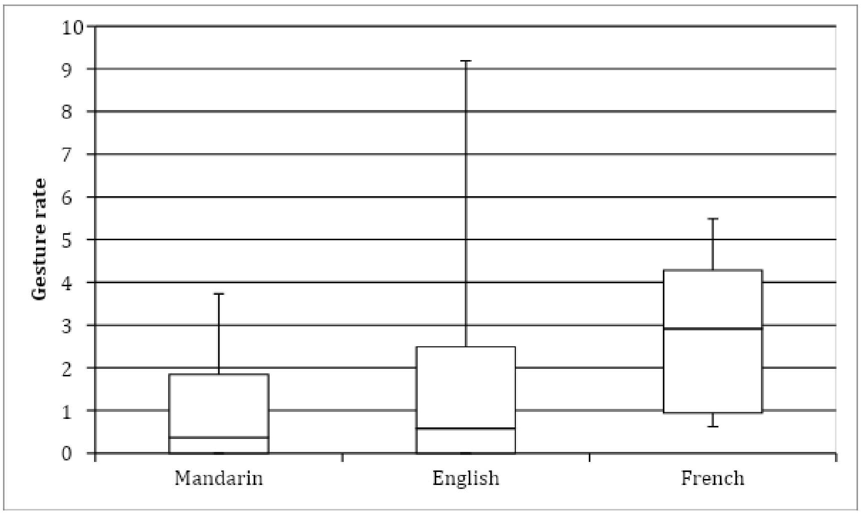
Following Nicoladis et al. (2018), we used a number of different variables to identify participants' storytelling style. On all of these variables except one, there were no differences between language groups. Consistent with the prediction that there are cross-cultural differences in storytelling style, there was a significant difference between groups on the modifier rate. Post-hoc tests revealed that the Mandarin speakers used more modifiers than both the English speakers and the French speakers. Nicoladis et al. (2018) reasoned that greater use of modifiers would indicate an evaluative style of storytelling, since modifiers (i.e., adjectives and adverbs) are optional and convey something about the speaker's perspective on events.

In conclusion, we found, at best, weak evidence to support the argument by Nicoladis et al. (2018) that cross-cultural differences in gesture frequency are mediated by storytelling style. In contrast, we did find evidence for cross-cultural differences in gesture frequency.

Keywords: gesture frequency; cross-cultural differences; discourse style

Figures

Figure 1. *Gesture rate across language groups*



References

Goldin-Meadow, S. & Saltzman, J. (2000). The cultural bounds of maternal accommodation: How Chinese and American mothers communicate with deaf and hearing children. *Psychological Science*, *11*, 307–314.

Nicoladis, E., Nagpal, J., Marentette, P., & Hauer, B. (2018). Gesture frequency is linked to story-telling style: Evidence from bilinguals. *Language and Cognition*, *10*(4), 641-664.

So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Language and Cognitive Processes*, *25*, 1335–1353.

Web-based, audio-visual prominence ratings of Swedish news reading materials: Effects of head movements, rating condition, and hardware

Gilbert Ambrazaitis, *Linnaeus University, Växjö, Sweden*

Johan Frid, *Lund University Humanities Lab, Sweden*

David House, *KTH (Royal Institute of Technology), Stockholm, Sweden*

Although prominence is increasingly recognized as an essentially multimodal phenomenon, relatively little is known about how spontaneous gestures in naturally occurring speech relate to perceived prominence (Jiménez-Bravo & Marrero-Aguiar, 2020). To study these relations, we need to collect prominence ratings for large amounts of ecologically valid audio-visual speech data efficiently, and to this end, we are currently developing a web-based rating set-up (Ambrazaitis et al., 2019; 2020; 2022). In our prototype rating task, 16 clips from Swedish television news (218 words in total) are to be rated by volunteers using a web interface. In our GUI, the text is displayed below the video player. Each word is to be rated as either non-prominent, moderately prominent (yellow), or strongly prominent (red), by clicking on the word button until the desired prominence level is encoded through a specific color. Participants have so far been free to use a mobile phone or a computer, and headphones or loudspeakers.

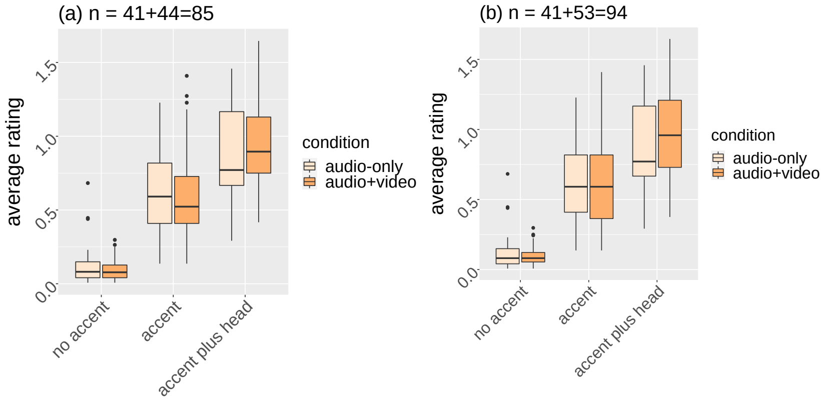
In an initial analysis, we showed that overall rating behavior is affected by hardware choices and screen size (Ambrazaitis et al., 2019). We later also collected data using the same materials in an audio-only condition (Ambrazaitis et al., 2022), and, instead of only measuring overall rating behavior, we compared ratings for words uttered with or without sentence-level pitch accents and head movements of any type (Ambrazaitis et al., 2020; 2022). Figure 1a displays results based on 85 raters, showing, first, a clear effect of the occurrence of accents and head movements, which is highly significant (model comparison: χ^2=322.37, *df*=2, *p*<.001). The higher ratings for 'accent plus head' compared to 'accent' (in both conditions) are explainable, as accents with head movement are often realized stronger acoustically than accents without (Ambrazaitis & House, 2022). The plot also suggests a slight (but not significant) effect of the rating condition (and an interaction with the occurrence of accents and head movements).

In a next step, which will be presented in detail at the symposium, we tested a revised set-up where only a section of the video was shown, displaying only the speaker (i.e., the newsreader), but not the studio background (usually presenting various illustrations), which, we hypothesize, might attract some attention and account for the relatively weak effects of the presence of head movements seen in Figure 1a. So far, we have collected ratings from 29 subjects using this new rating condition. Our preliminary analysis does not reveal any significant difference between the two video-rating conditions. However, it once again reveals an effect of the rating device used (mobile phone vs. computer screen). In Figure 1b, we have subsumed the two video conditions (with or without displaying the studio background) but excluded all raters that used a mobile phone (*n*=20). This plot, again, suggests an interaction between rating condition and the occurrence of accents and head movements, which this time turns out significant (χ^2=4.92, *df*=1, *p*=.027). This preliminary result suggests that words with head movements tend to be rated more prominent if the movement is seen by the rater, as would be expected, but that a considerable number of raters are required to generate this effect, and that a similar rating set-up, using a computer screen instead of a mobile phone, should be used.

Keywords: audio-visual perception; multimodal perception; prominence; pitch accent; head movement; beat gesture; crowdsourcing; web-based

Figures

Figure 1. *Boxplots of average prominence ratings collected in an audio-only (n_{audio}=41) and in an audio-visual condition with **(a) n_{video} = 44** including raters using either computers or mobile phones and **(b) n_{video} = 53** including only raters using a computer screen, but two different video display conditions subsumed (44+29-20, see text).*



References

Ambrazaitis, G., Frid, J., & House, D. (2019, September 9-10). *Multimodal prominence ratings: Effects of screen size and audio device* [Oral presentation/ conference abstract]. The 6th European and 9th Nordic Symposium on Multimodal Communication (MMSYM 2019), Leuven, Belgium. http://mmsym.org/wp-content/uploads/2016/09/MMSYM2019-book-of-abstracts-0905.pdf

Ambrazaitis, G., Frid, J., & House, D. (2020). Word prominence ratings in Swedish television news readings – Effects of pitch accents and head movements. *Proceedings of Speech Prosody 2020*, 314-318. https://doi.org/10.21437/SpeechProsody.2020-64

Ambrazaitis, G., Frid, J., & House, D. (2022). Auditory vs. audiovisual prominence ratings of speech involving spontaneously produced head movements. *Proceedings of Speech Prosody 2022*, 352-356. https://doi.org/10.21437/SpeechProsody.2022-72

Ambrazaitis, G., & House, D. (2022). Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters. *Laboratory Phonology, 13*(1). https://doi.org/10.16995/labphon.6430

Jiménez-Bravo, M., & Marrero-Aguiar, V. (2020). Multimodal perception of prominence in spontaneous speech: A methodological proposal using mixed models and AIC. *Speech Communication,124*, 28-45. https://doi.org/10.1016/j.specom.2020.07.006

A first investigation of the timing of simple and complex co-speech manual gestures in Luganda and their relation to prominence

Margaret Zellers, *Kiel University*

This study investigates the temporal relationship between co-speech gestures and the spoken text in conversation in Luganda, with the aim of gaining more insight into the structure of prominence in Luganda via the temporal alignment patterns that arise. Luganda (ISO 639-3) is a Great Lakes Bantu language spoken in Uganda. It is a tone language, widely reported to have three contrastive tones, High, Low, and Falling, with High and Falling showing substantial phonetic overlap (Myers et al., 2019). The definition of prominence in Bantu languages is somewhat unclear. Focus, a target for phonetic/prosodic prominence in many languages, does not appear to be prosodically marked in many Bantu languages (cf. Hyman, 1999). Penultimate lengthening, that is, a phonological lengthening of the penultimate syllable of the word, has been proposed as a prominence location in Eastern and Southern Bantu languages (Odden, 1999; Hyman, 2013). Luganda does not have penultimate lengthening on a phonological level, although syllables in penultimate position tend to be phonetically longer than their counterparts in other positions (Hyman, 2013). Luganda's tonal structure offers another alternative location for prominence: tonal structure in words is constrained such that a maximum of one transition from High to Low may arise within a word (McCawley, 1970). Such patterns have been argued to underlie reanalysis of tone to stress in some Bantu languages (Ratliff, 2015). Thus, the location of the final high tone or the start of the falling contour is also a possible candidate for a perceived prominence.

It has been shown in many languages that gestures, particularly beat gestures, are closely aligned in time with prominent syllables (e.g., Krahmer & Swerts, 2007; Loehr, 2007). However, prominent syllables and large pitch movements are often confounded in these languages. It is possible that investigating the alignment of beat gestures in Luganda with phenomena such as the penultimate syllable and the word-internal pitch fall can provide evidence for how prominence is organized in Luganda.

The current study uses conversational data from two Luganda conversations. The conversations were recorded in June 2019 at Makerere University in Kampala, Uganda. The speakers were recorded using either head-mounted microphones or a single directional microphone, allowing for channel separation; video recordings were also made. The data were orthographically transcribed and translated to English by native speakers of Luganda. This study uses an initial subset of two conversations, taking a randomly-selected eight-minute chunk of each conversation. All hand gestures from both speakers were annotated using the video without access to the audio or the transcript; the gestures were separated into phases following McNeill (1992). Since gesture strokes often involved complex repetitive movements, these were also further divided into the location of their apices on a separate labelling tier (cf. Prieto et al., 2018).

On the basis of the orthographic transcriptions and the audio data, penultimate syllables and pitch falls in regions where there is ongoing hand gesture will be annotated, and the alignment between these points of interest and the gesture apices will be investigated. It is hypothesized that a clear preference for alignment of gesture strokes with either the penultimate syllable or with the pitch fall will be found. Initial observations suggest that the penultimate syllable is a more promising alignment point for the onset of the gesture stroke. However, many beat gestures involve complex repetitive movements which appear to be produced too quickly to show a simple alignment with penultimate syllables or pitch falls; thus the temporal alignment of complex gestures may pattern differently than simple beats.

Keywords: Luganda; beat gesture; hand gesture; temporal alignment

References

Hyman, L.M. (1999). The interaction between focus and tone in Bantu. In G. Rebuschi, & L. Tuller (Eds.), *The Grammar of Focus* (pp. 151-177). John Benjamins.

Hyman, L.M. (2013). Penultimate lengthening in Bantu. In B. Bickel, L.A. Grenoble, D.A. Peterson, & A. Timberlake (Eds.), *Language typology and historical contingency: In honor of Johanna Nichols* (pp. 309-330). John Benjamins.

Krahmer, E. & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57* (3), 396–414.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7, 179–214.

McCawley, J. (1970). Some tonal systems that come close to being pitch accent, but don't quite make it. *Chicago Linguistic Society*, *6*, 526–531.

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

Myers, S., Namyalo, S., & Kiriggwajjo, A. (2019). F0 timing and tone contrasts in Luganda. *Phonetica*, *76*(1), 55–81.

Odden, D. (1999). Typological issues in tone and stress in Bantu. In S. Kaji (Ed.), *Cross-linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics* (pp. 187-215). ILCAA.

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: a labelling proposal. *Proceedings of the 9th International Conference on Speech Prosody*: 201-205.

Ratliff, M. (2015). Tonoexodus, tonogenesis, and tone change. In P. Honeybone & J. Salmons (Eds.) *The Oxford Handbook of Historical Phonology* (pp. 245-261). Oxford University Press.

Prosodic cues for Gesture / Speech synchronization and multimodal prominence

Giorgina Cantalini, *Civica Scuola di Teatro Paolo Grassi Milano*

Massimo Moneglia, *Università di Firenze*

Gestures are structured in a configurational model. The minimal gestural linear pattern (Gesture Phrase) foresees a compulsory root (the Expressive Phase made by at least one Stroke), constituting the main gestural prominence. Gestural Phrases are packaged within larger Gesture Units (Kendon, 2004; Kita et al., 1998; McNeal, 1992). In speech, syntactic constituents are patterned into Information units, following the flow of thought, each corresponding to Prosodic units (Chafe, 1994). Information units are marked by prosodic prominence and are structured into higher-level Reference units, which are terminated from a prosodic point of view and correlate with Speech acts (Izre'el et al, 2020). The paper presents the results of a corpus-based study on the relations between gestural and prosodic units in spontaneous spoken Italian and focuses on the relation between prosodic and gestural prominences. It is argued that gesture prosody synchronization allows a better definition of gesture scope and meaning.

The dataset comprises three heavily annotated samples of video-recorded interviews with three actors about their profession. Samples (around three minutes each) have been extracted, forming a corpus of 220 Utterances, 732 Information units, and 513 Gesture Phrases.

Gesture and prosody have been annotated independently one from the other and then reconciled in ELAN and PRAAT files. Gesture annotation is based on LASG (Bressem et al., 2013). Co-speech gestures have been segmented at three hierarchic levels, each one aligned to the acoustic source, available to annotators without prosodic annotation: (a) Gesture Units: sequences of gestures between two rest positions; (b) Gesture Phrases: phases of a gesture around a prominence (Stroke); (c) Gesture Phases. A second annotator has replicated the annotation. The rate of overlapped units has been calculated and shows an Average overlap/extent ratio of 0.83 for Units and 0.70 for Phrases. Kappa Cohen's for the categorization of Strokes is over 0.85.

Experts have annotated the Prosodic cues following the L-AcT methodology (Cresti, 2000; Izre'el et al., 2020). Speech acts are identified through correspondence with sequences of prosodic units ending with a Terminal prosodic break. Prosodic units are characterized by Perceptively Relevant Prosodic movements and end with a prosodic boundary ('t Hart et al., 1990; Loehr, 2014). Prosodic units have been annotated with the following Informational values: Comment, Topic, Parenthesis, Appendix, Discourse Connector, and Dialogical.

Results show that co-speech gestures accompany 90% of the speech, and Gesture units are synchronous with prosodic boundaries. Gesture Phrases never cross terminal prosodic boundaries, finding the *utterance* the maximum unit for gesture/speech synchronization (Cantalini & Moneglia, 2020).

The marking of prosodic units and their perceptively relevant movements allows us to understand the linguistic scope of the gesture. Strokes may correlate with all information unit types, by preference with Topic, Comment, and Parenthetical, but not with Dialogic Units (which lack lexical content and work for communication management) and Appendixes (which do not bear a prosodic focus). In these units, Strokes co-occur with the prosodic movement characterizing each type, as expected, but one prosodic unit can guest more than one Stroke. Strokes find their scope at different linguistic levels: a) the word level; b) the information unit phrase; c) the information unit function; d) the illocutionary value of the utterance.

Keywords: Co-speech Gestures; Prosody; Synchronization; Multimodal Prominence

References

Bressem, J., Ladewig, S.H., & Müller, C. (2013). Linguistic Annotation System for Gestures (LASG). In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, (38.1, pp. 1098-1125). De Gruyter-Mouton.

Cantalini, G., & Moneglia, M. (2020). The annotation of Gesture and Gesture / Prosody synchronization in multimodal speech corpora. *Journal of Speech Science*, *1*, 1-24.

Chafe, W. (1994). *Discourse, consciousness, and time*: *The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Cresti, E. (2000). *Corpus di italiano parlato.* Accademia della Crusca.

't Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody.* Cambridge University Press.

Izre'el, S., H. Mello, A. Panunzi, & T. Raso (Eds). (2020). *In Search of Basic Units of Spoken Language*. Benjamins.

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

Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer Interaction* (pp. 23-35). Springer.

Loehr, D. (2014) Gesture and prosody. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (38.2, pp. 1381-1391). De Gruyter Mouton.

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

Audience effects and production demands on timing relationships between representational gestures and speech

Ed Donnellan*[1], Yumeng Wang*[1], Levent Emir Özder[2], Hillarie Man[1], Kellie Fraser[3], Amara Jiménez Cañizares[1], Beata Grzyb[1], Yan Gu[4], Gabriella Vigliocco[1]

[1]University College London, [2]Koç University, [3]University of Aberdeen, [4]University of Essex

**joint first authors*

In naturalistic conversation, speakers often produce co-speech representational gestures, visually depicting properties of referents that they are talking about. These gestures commonly depict similar information as the accompanying speech, e.g., when talking about chopping a carrot, a speaker brings their hand down in a chopping action. In these instances, gestures may aid a speaker package conceptual information to facilitate language production: representing spatio-motoric features of a concept with the hands makes speech about that concept more accessible (Kita et al., 2017). On the other hand, interlocutors may predict upcoming speech from speakers' representational gestures, consistent with theories of multimodal integration of auditory and visual cues by interlocutors (e.g., Zhang et al., 2021). As such, co-speech representational gestures may facilitate language processing for both speaker and interlocutor.

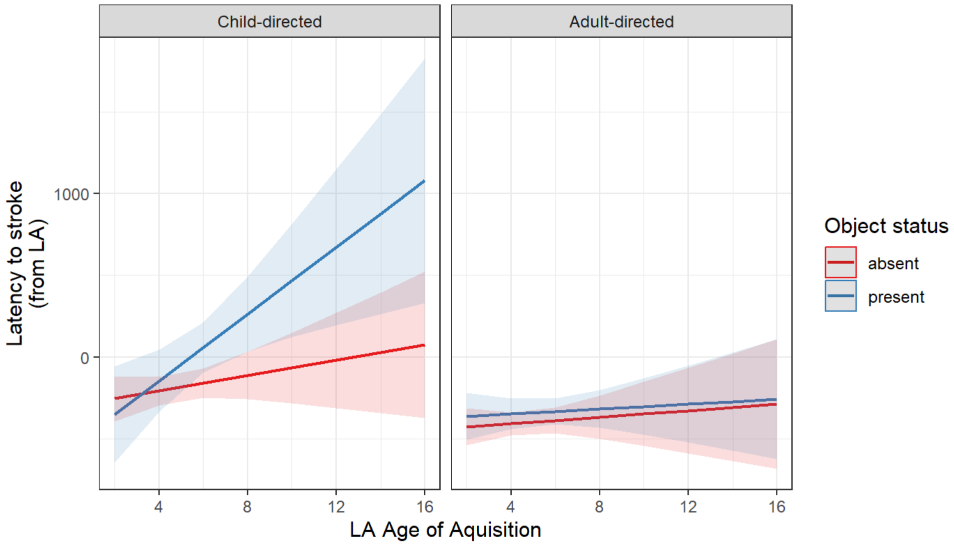
Previous studies investigating the timing relationship demonstrate that, while relatively stable, there is still variability in the latencies between representational gestures and LAs (e.g., ter Bekke et al., 2020). This could be driven by two complementary sources: (i) differing demands on speaker production, e.g., word retrieval or (ii) audience effects, whereby a speaker modifies their communication sensitive to the requirements of their audience. In the current study we determine to what extent gesture-speech timing relationships vary depending on production demands or audience effects. To do so, we focused on co-speech representational gestures from the ECOLANG corpus in which adult-adult (n=33) and adult-child dyads (n=36) engage in semi-naturalistic conversation about objects (Vigliocco et al., unpublished). We compared latencies between gesture stroke onsets (where the meaning of the gesture becomes clear) and LA onsets between gestures produced by adults to other adults (n=1928) and adults to their children aged 3-4 years (n=899). Crucially, to determine audience effects we compared between when the interlocutor was an adult or child. To investigate potential production demands, we considered the age of acquisition of the LA (AoA: indexing word retrieval difficulty) and whether the object was present or not (word retrieval may be easier when objects were physically present).

Our results suggest that audience effects play a large role in the timing of speaker production of representational gestures and their LAs. When interacting with children, gestures were produced later relative to the LA for later-acquired words, with later-acquired words produced concurrently or even before the gesture (Figure 1). This effect was more pronounced when objects being talked about were present (though the three-way interaction was non-significant). In contrast to previous work (e.g., Morrel-Samuels & Krauss, 1992), variability in timing does not seem to be as much due to demands on speaker production (e.g., the inherent difficulty of retrieving unfamiliar LAs). When interacting with other adults, object presence or AoA of the LA only had minor effects on the timing. To our knowledge, our work is the first demonstration that speakers flexibly alter representational gesture-speech timing relationships contingent on their interactional context.

Keywords: audience effects; representational gestures; iconicity

Figures

Figure 1. *Model predictions for Latency between LA and stroke*



References

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. doi: 10.1037/rev0000059

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615–622. doi: 10.1037/0278-7393.18.3.615

ter Bekke, M., Drijvers, L., & Holler, J. (2020). *The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech*. doi: 10.31234/osf.io/b5zq7

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proc. of the Royal Society B: Biological Sciences*, *288*(1955), 20210500. doi: 10.1098/rspb.2021.0500

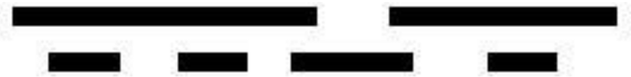
Event Chronography in Multimodal Data: a Method for Quantitative Analyses

Anaïs Claire Murat, *Trinity College Dublin*

Maria Koutsombogera, *Trinity College Dublin*

Carl Vogel, *Trinity College Dublin*

Studying interactions in their complexity requires the exploration of different modalities and their temporal arrangement. Yet, distinct channels (e.g., vocal linguistic content, laughter, gesture, gaze, etc.) do not perfectly align in start or end times, which makes it difficult to count events in cross-modal comparisons. Consider the depiction of sequences of events in two modalities over time:



Answering the simple question “how many events of the lower sort accompany each event of the upper sort?” is not straightforward because the latter spans boundaries of the former. To overcome this difficulty, the literature tends to focus on longer events (such as dialogue episodes), or discontinuous events (e.g., turns with intervening intervals). But such an arrangement is not always relevant, and comparing shorter units of different types can be challenging. For instance, in Murat et al. (2022), overlaps between turns and mutual gazes (MG) (two relatively short annotation types) led longer MGs that spanned through several turns to be counted several times. This honest but necessarily skewed representation of the data limited the account of temporal relationships between the two modalities, notably in the exploration of the durational aspect of MG patterns.

Our ambition here is to contribute a method for studying temporality in multimodal corpora. We detail a new way of individuating instances according to chronologically arranged *events* – rather than time intervals – for quantitative purposes, and show a validating example.

The solution we present highlights the onset and offset of each annotation, and treats them as singular *events*. It then creates a table which chronologically orders these events. Typically, one column accounts for one annotation type, and each row accounts for one *event*: either the *onset* (identified by the letter “B”, for “beginning”) or the *offset* (identified by “E” for “end”) of an annotation. When an *event* precipitates a new line, it also creates a “M” cell (for “middle”) in any already started annotation of another type. “M”s can then be numbered to keep track of events occurring within an annotation. Therefore, a “BMMME” sequence reads “during this annotation, three *events* of another type occurred”. Ultimately, the finalised table comprises a minimum of two columns (minimum two annotation types), and as many lines as events (onsets and offsets) present in the corpus.

With this so-called BME method, we investigate the Multisimo Corpus (Koutsombogera & Vogel, 2018) to illuminate two things: (1) the various possible relations between turns and MG (such as whether MGs tend to be included inside, between, or across turns), and (2) the relevance of event-based analyses over time-based analyses of duration.

Contingency table analysis of the resulting annotations enabled quantification of MG beginnings and endings to occur during middles of turns, suggesting a potential priority of the turn channel. Our second analysis of the MG's Bs, Ms, and Es compared to the amount of repetition by turn illustrates the relevance of event-based duration in addition to time-based duration and suggests that the length of a MG as counted in events is positively correlated to the amount of repetition occurring in between its onset and offset.

Finally, we argue the generality of our method. The strength of the BME method lies in its ability to handle in a similar manner isolated and overlapping annotations, as well as longer and shorter annotations. It pushes time-related matters (such as gaze fixation definition, or cross-annotator agreement time differences) to the background by shifting the focus from a -more or less- arbitrary time duration to a meaningful arrangement of data where the length of an annotation translates the number of other *events* occurring through it. In the future, we aim to use this method to study relationships among words, parts-of-speech and non-verbal aspects of interactions, such as gestures, gaze patterns, and laughter.

Keywords: multimodality; temporal analysis; interaction

References

Koutsombogera, M., & Vogel, C. (2018). Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 2945–2951.

Murat, A. C., Koutsombogera, M., & Vogel, C. (2022). Mutual Gaze and Linguistic Repetition in a Multimodal Corpus. *Proceedings of the Language Resources and Evaluation Conference*, 2771–2780.

Hosted by

Universitat Pompeu Fabra Barcelona

Grup d'Estudis de Prosòdia

Sponsors

Research network on GEstures and Head Movements in language (GEHM) funded by Independent Research Fund Denmark

INDEPENDENT RESEARCH FUND DENMARK

Department of Translation and Language Sciences from Universitat Pompeu Fabra

Universitat Pompeu Fabra Barcelona

Facultat de Traducció i Ciències del Llenguatge 30 aniversari

Contact

Email address: mmsym2023@gmail.com
Postal address: C/Roc Boronat 138, 08018, Barcelona
Twitter: @mmsym2023