The Missing Power: Language Mediates Sensorimotor-related Beta Oscillations during On-line Comprehension of Different Types of Co-speech Gesture

Yifei He (yifei.he@staff.uni-marburg.de)
Department of Psychiatry and Psychotherapy, Philipps-University Marburg,
Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany
Department of General Linguistics, Johannes Gutenberg-University Mainz,
Jakob-Welder-Weg. 18, 55122 Mainz, Germany

Helge Gebhardt (helge.gebhardt@psychiat.med.uni-giessen.de)
Centre for Psychiatry, Justus-Liebig University Giessen,
Ludwig Str. 23, 35390 Giessen, Germany

Isabelle Rondinone (rondinon@students.uni-marburg.de)
Department of Psychiatry and Psychotherapy, Philipps-University Marburg,
Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany

Benjamin Straube (straubeb@med.uni-marburg.de)
Department of Psychiatry and Psychotherapy, Philipps-University Marburg,
Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany

Abstract

We used Electroencephalography (EEG) to investigate the processing difference between co-speech emblematic gestures (EM) and tool-use gestures (TU). We found that TU shows beta power decrease against EM in a foreign language condition (Russian) but this effect is missing in the native language condition (German). With regard to the beta power effect, we reasoned that beta power decrease is a neural marker for recruitment of the sensorimotor system. However, with regard to the missing beta effect in the German condition, we suggested two proposals: on the one hand, it may suggest that semantic integration process of gesture and speech could also be related to beta power oscillations; on the other, the missing power could be considered as an indication of a shared and interactive neuronal network by both sensorimotor system and higher-level semantic system.

Keywords: emblematic gesture, tool-use gesture, EEG, beta power, semantic integration, sensorimotor system

Introduction

Co-speech gesture forms an important part during inter-personal communication and it has been intensively investigated via various methods (e.g. Andric & Small, 2012; Straube, Green, Jansen, Chatterjee, & Kircher, 2010). In the current study, we focused at two distinct types of co-speech gesture: emblematic gesture (EM) and tool-use gesture (TU) and we investigated how both types of gestures interact with speech.

An emblematic gesture is also called an emblem (Ekman & Friesen, 1969); it conveys meaning which is usually not related to its physical form (e. g. the thumb up sign for “good job”). Emblems are often used together with speech in daily conversations and they usually refer to abstract concepts about social life or interpersonal situations. A tool-use gesture is, by its name, a type of gesture which conveys meaning corresponding to its physical form (Higuchi, Imamizu, & Kawato, 2007; Johnson-Frey, Newman-Norlund, & Grafton, 2005). For example, actions such as “using a dagger to stab” or “using a hammer to hammer” could be easily transferred to related tool-use gestures. In contrast to emblematic gestures, tool-use gestures usually refer to concrete images, especially images that depict motoric movements.

Although bearing the similarity that both emblematic and tool-use gestures convey identifiable meanings (even without speech), the two types of gesture differ in terms of their position on the two ends of an abstract-social/concrete-tool continuum: tool-use gesture bears the closest relationship to hand motions whereas emblematic gesture carries little indication of motoric movements and is more abstract.

A previous imaging study suggests that there are distinct processing mechanisms for co-speech gestures that differ in concreteness (Straube, Green, Bromberger, & Kircher, 2011); however, a caveat of its design paradigm (comparing the same multi-meaning gesture with different speech contexts) makes it impossible to directly compare the processing pattern of different types of gestures *per se*. More importantly, studies with fMRI do not provide us with satisfactory time resolution which is crucially informative concerning how, and more importantly, WHEN language interacts with different types of gestures.

To compensate these remaining questions, we used Electroencephalography (EEG) to investigate co-speech emblematic and tool-use gestures. Specifically, we focused at the time-frequency domain of the EEG. Alpha and beta frequency bands are reported closely related to both action execution as well as observation (Avanzini et al., 2012; Quandt, Marshall, Shipley, Beilock, & Goldin-Meadow, 2012). More importantly, it is reported that both alpha and beta frequency bands are sensitive to iconic vs. deictic gestures (not co-speech) in the sense that iconic gestures showed suppression for both alpha and beta power because they engage more richly descriptive motion than deictic gestures (Quandt et al., 2012). Taken together, alpha and beta oscillations are considered as potential neural markers of the recruitment of human sensorimotor system.

In the current study, we took a step further by testing emblematic and tool-use co-speech gestures accompanied by two different languages (interpretable German and un-interpretable Russian). By comparing co-speech gestures differing in the degree of concreteness and involvement of motoric movements (EM vs. TU), we make the following hypotheses:

1. Given that TU is more closely related to hand motions and that EM is more abstract, we expect suppression (power decrease) of alpha and beta frequency bands for the TU condition.
2. The alpha and beta suppression will remain unaffected by the language of the co-speech because the effect originates from the motoric nature of the gesture.
3. In the German condition, we may expect some form of integration difference between two gesture types, but this effect will be late because it is dependent on the establishment of semantic content of both speech and gesture.

Methods

Participants

Twenty subjects (12 female, Mean age=23.45, SD=2.54) participated in this experiment. All were native German speakers and received 30€ compensation.

Materials

Video clips of combinations of speech (German and Russian) together with emblematic (EM) and tool-use (TU) gestures were recorded, resulting in four experimental conditions: Emblematic German (EMG), Tool-use German (TUG), Emblematic Russian (EMR), and Tool-use Russian (TUR). The co-speech gestures were performed by a German-Russian bilingual actor in a natural and spontaneous way. All video clips were 5 sec long with at least 0.5 sec blank screen before and after the sentence onset and offset, during which the actor neither spoke nor moved (for the same procedure, see Straube et al., 2011; Straube et al., 2010).

Procedure

An experimental session comprised 104 experimental trials (26 for each condition) together with 52 trials with only German speech (no gesture) and additional 26 trials with meaningless gestures with Russian as fillers, which resulted in 182 trials in total divided into two experimental blocks. Crucially, for both German and Russian conditions, the gestures are performed identically. A sample trial for the EM and TU conditions (German) is illustrated in **Figure. 1**.

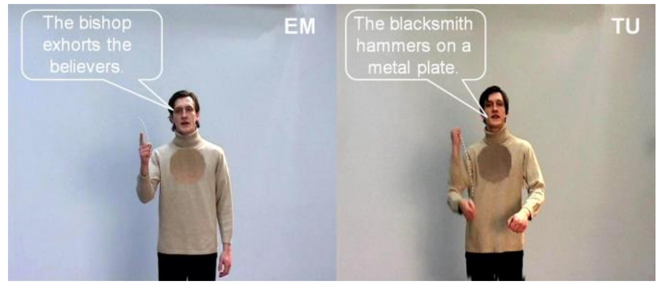

Figure. 1: Stimulus material (German condition)

EEG Recording

EEG data were collected from 29 standard channels (10/20) system EasyCap (EasyCap GMBH, Munich) referenced to the FCz electrode (AFz ground). Signals were magnified using Brain Vision (Brain Vision GMBH, Munich) amplifier and recorded with a sampling rate of 500 Hz. The impedance of all electrodes was reduced to below 5KΩ.

Data Analysis

All analyses were carried out using the Fieldtrip toolbox for EEG/MEG analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011). The raw EEG were segmented into 4 sec segments after the onset of each gesture. The data were high-pass filtered at 0.1 Hz and low-pass filtered at 125 Hz. EOG and muscle Artifacts were automatically detected and rejected based on the amplitude distribution across trials and channels. Cutoffs for the EOG and muscle artifact rejection were set at z = 4 and z = 9 and performed in frequency bands of 1-14 Hz and 110-140 Hz respectively.

In order to reveal event-related power oscillations for the different frequency components of the EEG, time-frequency representations (TFRs) of the single trial data were computed. In order to optimize the trade-off between time and frequency resolution, TFRs were computed in two different but partially overlapping frequency ranges. In the low frequency range (2-30 Hz), a 400 ms Hanning window was used to compute power changes in frequency steps of 1 Hz and time steps of 0.1 sec; in the high frequency range (25-80 Hz), the time-frequency analysis was carried out using Morlet wavelets (seven cycles) with frequency steps of 5 Hz and time steps of 0.1 sec. For statistical analyses, a mass cluster permutation test was carried out on the baseline-corrected (-0.5 to -0.15 ms) absolute power change in the frequency space between 2 Hz and 30 Hz because the method elegantly handles multiple-comparisons problem. The procedure is briefly described here (for an elaborate description of the approach, see Maris & Oostenveld, 2007): firstly, for every data point (time-frequency-channel point), a simple dependent-samples *t* test is performed and resulted in uncorrected *p* values. Secondly, all significant data points ($p < .05$) are grouped as clusters. For each cluster the sum of the *t* statistics is used in the cluster-level statistics. Finally, a Monte-Carlo simulation with 1000 repetitions was used to identify significant clusters in the time-frequency-channel space. In the current experiment, we firstly identify time-frequency bins that showed a significant effect at $p < .025$, we then searched for similar time-frequency-channel clusters based on the identification of a minimum of 4 neighboring electrodes as a cluster.

Results

As we were interested in the processing difference between the two gesture types, we directly compare EM and TU gestures within the German and Russian conditions. For the comparison within the German condition (EMG vs. TUG), the contrasts in both low (2-30 Hz) and high (25-80 Hz) frequencies did not elicit any significant clusters. When comparing EMR vs. TUR, in the high frequency band, no significant clusters were observed; in the low frequencies, the contrast showed one significant positive cluster ($p = .0001$), and this effect is shown in **Figure 2**.

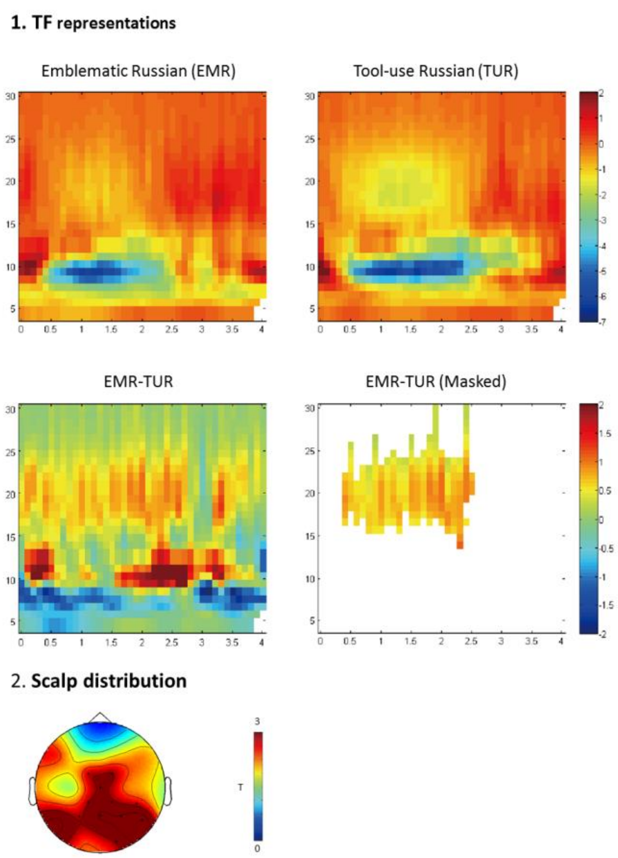

Figure 2: Results of the TF analysis in the EMR vs. TUR conditions. (1) The upper panel shows the TF representations of the EMR (left) and the TUR (right) conditions at electrode Pz for the low frequency range. The lower left panel shows the raw contrast between the EMR and TUR conditions, and the lower left panel shows the statically masked contrast. (2) The scalp distribution in the 20 Hz band in the 0.2-2.5 ms time interval. The electrodes that show significant differences between the two conditions are marked with "*" on the topographic plot.

Discussion

Partly in-line with our hypotheses, our results showed beta power suppression for the TUR condition in comparison to the EMR condition, the effect starts early (around 300 ms after the onset of gesture) and elapses until 2.7 sec. This effect has a central-parietal scalp distribution. However, contrary to our hypothesis, no effects of frequency oscillations were observed in the German condition.

We interpret the beta power suppression for the TUR vs. EMR condition as an indication of higher degree of action observation, echoing what is proposed by Avanzini et al., (2012) and Quandt et al., (2012). Both studies suggest that the observation of hand movement (such as gesture) elicits frequency oscillations which are analogous to that occurring during action execution. Although both emblematic and tool-use gestures have motoric basis, given that tool-use gestures entail richer involvement of motoric movements than emblematic gestures, the beta power change is most naturally interpreted as resulted from higher degree of hand movement observation.

However, what is more interesting is the absence of any effects in the German condition, despite the fact that identical gestures have been presented. If beta (and alpha, as reported in other studies, e. g. Quandt et al., 2012) power suppression reflects the different degree of involvement of sensorimotor system, the effect will be independent from the specific language which accompanies the observation of gesture, be it interpretable or not. To account for our results, we propose two tentative explanations.

A first viable explanation states that during the comprehension of co-speech gestures, the observer automatically integrates all audio-visual semantic information. In the current experiment, this semantic integration process is present only in the German condition because no speech-based semantic information is available in the Russian condition. Given that previous imaging studies suggest distinct semantic integration mechanisms for gestures differing in concreteness (Straube et al., 2011), if we assume that such integration difference, as observed in the fMRI studies, is related to a suppression of beta power for the more abstract condition. As a result, in the current study, we would expect, for the EMG vs. TUG condition, a suppression of beta power either. On the other hand, as TUG vs. EMG may elicit potentially motoric beta suppression, then the two effects could potentially cancel out each other and thus remain unobservable. However, considering the latency of the beta effect in the Russian condition, another crucial assumption has to be made: the multi-modal integration process has to be initiated early (around 300 ms) to fully mediate the motoric beta suppression effect, and this would suggest that the semantic integration of audio-visual information is not only fast and automatic, but also early. In another word, it suggests that observers are able to, or at least trying to identify and incorporate the semantic content from both input modalities even when the meaning of gesture and/or speech is not fully determined. After all, this explanation, being a viable account, assumes gesture observation and semantic integration as separate modules but sharing similar frequency oscillation characteristics.

The second account does not consider observation of gesture and integration of multimodal input as distinct neuronal processes. It has been suggested that the comprehension of gesture recruits similar brain networks as speech (Xu, Gannon, Emmorey, Smith, & Braun, 2009); meanwhile there is evidence showing that the online comprehension of speech also recruits sensorimotor networks (c. f. Pulvermuller, 2005). Similarly, the results of the current experiment may be indication of an interactive brain which is not modality-specific: only when the comprehension of German speech together with the integration of multi-modal input involves the sensorimotor system, the potential motoric beta oscillation difference, as reflected in the Russian condition might be erased. This account, being tempting though, would require further imaging investigations to be validated.

To summarize, the current study found that observation for tool-use gestures against emblematic gestures is related to beta power decrease when gesture is accompanied by an un-interpretable foreign speech; however, this power difference is missing when the co-speech is understandable. We propose that the missing beta power could either stem from 1) the fact that the semantic integration process is also related to beta power, or from 2) an interactive neuronal network shared by sensorimotor system and higher-level semantic system. However, both accounts would require further validations.

Acknowledgments

This research project is supported by a grant from the 'Von Behring-Röntgen-Stiftung' (project no. 59-0002). Y. H. and H. G. are supported by the 'Von Behring-Röntgen-Stiftung' (project no. 59-0002), B.S. is supported by the BMBF (project no. 01GV0615).

References

Andric, M., & Small, S. L. (2012). Gesture's neural language. *Frontiers in psychology, 3*.

Avanzini, P., Fabbri-Destro, M., Dalla Volta, R., Daprati, E., Rizzolatti, G., & Cantalupo, G. (2012). The Dynamics of Sensorimotor Cortical Oscillations during the Observation of Hand Movements: An EEG Study. *PLoS ONE, 7*(5), e37534.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 88-106.

Higuchi, S., Imamizu, H., & Kawato, M. (2007). Cerebellar Activity Evoked By Common Tool-Use Execution And Imagery Tasks: An Fmri Study. *Cortex, 43*(3), 350-358.

Johnson-Frey, S. H., Newman-Norlund, R., & Grafton, S. T. (2005). A Distributed Left Hemisphere Network Active During Planning of Everyday Tool Use Skills. *Cerebral Cortex, 15*(6), 681-695.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Intell. Neuroscience, 2011*, 1-9.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. [10.1038/nrn1706]. *Nat Rev Neurosci, 6*(7), 576-582.

Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., & Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: An EEG study of action production and gesture observation. *Neuropsychologia, 50*(12), 2745-2751.

Straube, B., Green, A., Bromberger, B., & Kircher, T. (2011). The differentiation of iconic and metaphoric gestures: Common and unique integration processes. *Human Brain Mapping, 32*(4), 520-533.

Straube, B., Green, A., Jansen, A., Chatterjee, A., & Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia, 48*(2), 382-393.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences, 106*(49), 20664-20669.