

Speaking while moving: Does the head compensate for the hands not being able to move?

Hélène Serré¹, Marion Dohen¹, Susanne Fuchs², Amélie Rochet-Capellan¹

¹Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

³Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

helene.serre@grenoble-inp.fr, marion.dohen@grenoble-inp.fr, fuchs@leibniz-zas.de,
amelie.rochet-capellan@grenoble-inp.fr

Abstract

Some studies tested the gesture-speech link by investigating whether speech production was affected when speakers are not able to gesture with their hands. Their results are mixed. However, when hand motion is blocked or constrained, speakers may move other body parts as a compensation. Here, we investigate a potential compensation by enhancement of natural head motion. We explored this question by using data from a motion capture experiment in which people had to recall stories in different movement conditions with hands being either free, blocked or constrained by circular motion using a minibike. A within-subject design was used with 25 native speakers of German. Our results are contrary to our predictions and suggest that head motion may actually be inhibited when the hands are prevented from gesturing. They also suggest that motion capture could be a good way to explore the question. These preliminary findings are also interpreted with respect to an overall stiffening of the upper body, and a potential head-hand coordination in motion.

Keywords: head motion, hand gestures, speech production

1. Introduction

Coverbal gestures are ubiquitous in oral communication. They interact with speech in different ways (McNeill, 1992; Kendon, 2004; Alibali, 2005). In particular, there is evidence that coverbal gestures may support language-related cognitive processes such as lexical access (Krauss *et al.* 1999).

In this direction, previous work reported that not being able to move the hands while speaking can reduce vocabulary diversity and precision (Rimé 1984; Hostetter, Kita & Alibali 2007). It could also increase disfluency (Finlayson *et al.* 2003; Hostetter, Kita & Alibali 2007). This latter effect was however not reproduced by Hoetjes *et al.* (2014). Co-verbal gestures are not limited to the hands (e.g. Kendon 2004). For example, people also move their head and eyebrows when speaking. These movements were also shown to possibly synchronize with speech prosodic parameters (Graf *et al.* 2002; Hadar 1983).

Finlayson *et al.* (2003) suggested that participants could compensate for not being able to move their hands by means of moving (or moving more than typically) other parts of their body. This compensation might avoid effects of the hand being blocked on speech production. For example, Rimé *et al.* (1984) observed an increase of eyebrow movements when speakers

cannot gesture with their hands. This could also be the case for other body parts such as the head.

Most of the work that investigated co-verbal gestures or movements of different parts of the body use qualitative descriptions and analyses based on video recordings. These methods enable to distinguish between co-verbal gestures and other non-communicative movements (such as, for example, scratching), as well as to attribute communicative functions to co-verbal gestures (Colletta *et al.* 2008). However, they entail subjectivity and have a substantial annotation cost. We trust that automatic analysis of motion using motion capture could help analyzing larger data sets and provide complementary evidence to qualitative approaches (Zhao *et al.* 2012).

As a supplementary step in this direction, we tested this idea by analyzing the kinematics of head movements in different hand movement conditions during a narrative task. The data are part of a corpus initially recorded to investigate speech-limb-breathing interactions. This corpus allows a quantitative analysis of head movements using motion capture and to explore the hypothesis of head compensation using this method.

2. Method

Twenty-five native speakers of German were involved in the corpus, aged from 20 to 29 years old (mean=23.3, standard deviation=2.5), 19 females. Their task was to watch short cartoons while sitting on a chair. They were then invited to retell the stories in different conditions: hands free (HF), hands blocked (HB), arms biking on a mini-bike (ABi) vs. legs biking (LBi) on the same mini-bike (Figure 1). Biking conditions were performed at a comfortable rhythm. For the hands blocked condition, we asked participants to keep their hands still under the chair. For the hands free condition, as we did not want to trigger any bias, we did not explicitly ask participants to gesture with their hands. Participants were recorded twice in the same conditions, on two different days.

Speech and movement were recorded synchronously, respectively using a microphone and an Optitrack® motion capture system consisting of 12 Prime 13 cameras. Sample frequency was 200 Hz. To track the body movements, we used rigid bodies made up of three markers linked by a rigid plastic material, so their relative positions among each other are fixed. Rigid bodies allowed us to label each part of the body, making detection more reliable than one marker which is not recognized

each time it is hidden. Six rigid bodies were positioned on different parts of the speaker’s body: one on the forehead, one on each shoulder, one at the top of the back and one on each hand. Head motion was characterized by calculating the average displacement over each trial.

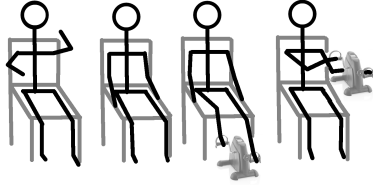


Figure 1: Experimental conditions, from left to right: Hands Free (HF), Hands Blocked (HB), Legs Biking (LBi), Arms Biking (ABi)

Due to its biomechanical nature, the head is mainly involved in rotation movements relatively to the torso (Figure 2). To investigate head motion, we quantified the rotation movement on every axis as well as the 3D motion of the rigid body on the head (Graf *et al.* 2002). For each axis, we took the coordinates of the gravity center of the rigid body and calculated the variation of movement (displacement) from one sample to the next one. To get the average displacement, we divided the sum of these displacements by the duration of the trial.

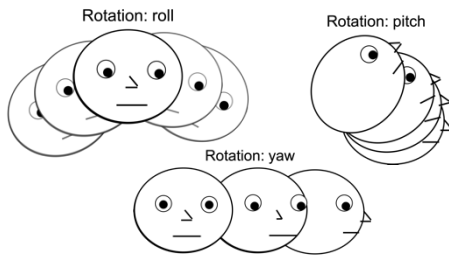


Figure 2: Rotations axes of the head

To avoid biased movements corresponding to an entrainment resulting from motion of the upper part of the body, we switched from the fixed world coordinate system of the motion capture system to the moving coordinate system of the participant, which was encompassed by the back and the shoulders (Livingstone & Palmer 2016). However, even in this new coordinate system, due to anatomical constraints, the head may follow the movement of the back with some phase-lag, introducing dependent head motion. To check for that, we looked at the coherence (Schmidt & O’Brien 1997) between the movement of the back and the movement of the head at the dominant frequency of the back. As seen in the equation (1), the coherence measures the degree of correlation between two signals, in the frequency domain. It is an indicator of the synchronization of two signals at a given frequency.

$$C(f) = \frac{\partial_{back\&head}(f)}{\partial_{back}(f) * \partial_{head}(f)} \quad (1)$$

f is the dominant frequency of the back, $\partial_{back\&head}(f)$ is the cross-correlation spectrum of the back and head signals for the frequency f , and $\partial_{back}(f), \partial_{head}(f)$ are respectively the auto-correlation spectral functions of the back and the head signals

at the frequency f . Like correlation, coherence is a ratio from 0 (no synchronization) to 1 (perfect synchronization). Values from 0.5 upwards reflect good coherence (Schmidt & O’Brien 1997).

For each session, we calculated the coherence on the three axes. Then, we averaged the values of the coherence over the participants and the days. Figure 3 shows that in the hands biking condition, coherence is over 0.5 on the antero-posterior axis and over 0.75 on the mediolateral axis. These results suggest that the head is strongly following the upper body in the hands biking condition with a phase lag, which seems coherent. But in almost all the other conditions, coherence is below 0.5 on every axis. This suggest that most of the entrainment of the head by the upper body has been discarded when the back became the referential coordinate system. Therefore, as we want to compare HF and HB only, the 3D displacement quantity of the head should be a good measure of independent head motion.

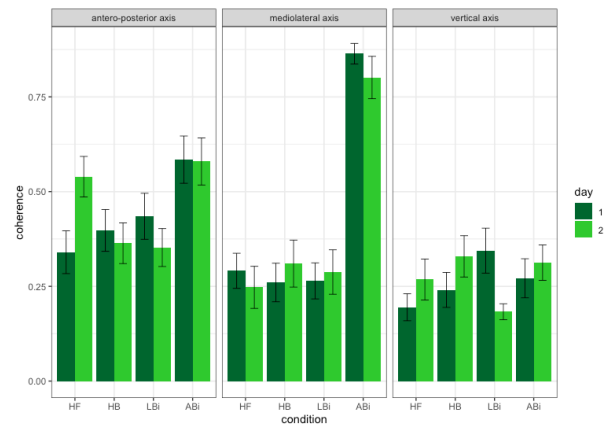


Figure 3: Coherence between head and back movements in the different experimental conditions.

We expected more movements of the head if speakers compensate for having the hands blocked (HB) in comparison to the hands free (HF) condition. The other two conditions (hands or legs biking) will be shown here for completeness and as a way to have more data to evaluate our method.

3. Results

Average displacement of head motion is displayed in Figure 4 showing no clear change in the amount of head motion between HF and HB conditions, while head motion appears to have been affected by cyclical limb movements, as the coherence measure was suggesting.

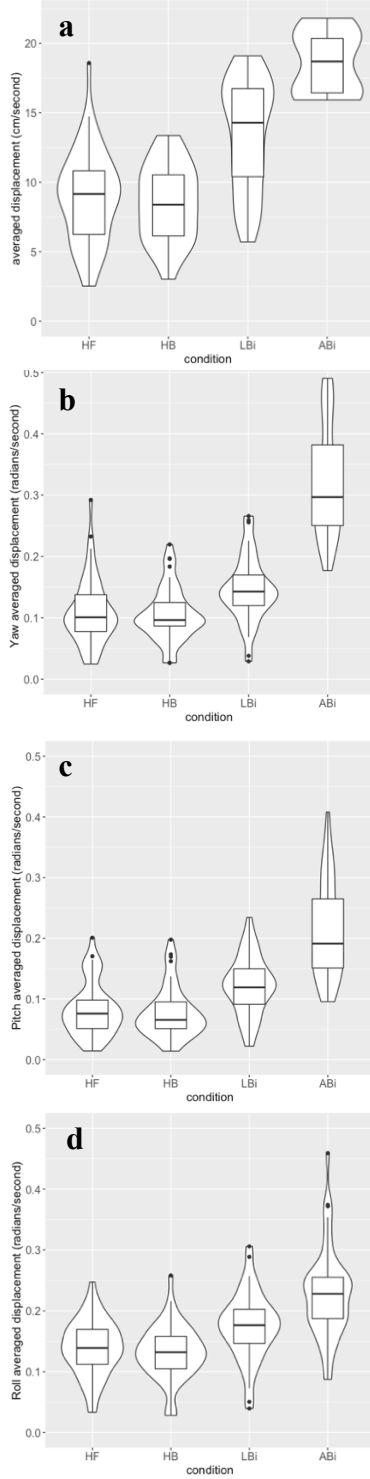


Figure 4: Violin plots with included boxplots of the average displacement of the head (y-axis) as a function of movement conditions (x-axis): a. 3D head displacement; b. Yaw rotation; c. Pitch rotation; d. Roll rotation

We applied a linear mixed model on the logarithm of the 3D head displacement quantity as the dependent variable and

condition as the fixed factor with the lme function from the nlme package (all the analyses were made on R version 1.2.1335). This function allows to consider potential differences in residual variances across effects. In order to select the best model, we used likelihood ratio tests, with the anova function in R. The effect of adding random slope was analyzed first. We then tested the fixed effect.

Once the final model was validated, we made a post-hoc analysis with the glht function of the multcomp package. Table 1 summarizes the output of the models. As visible from descriptive statistics, no significant difference is observed between HF and HB and head movements were increased in biking motion conditions and in particular when biking with the hands.

	Estimate	Std. Error	Z value	Pr(> z)
HB-HF	-0.08	0.08	-1.1	0.7
HB-LBi	-0.5	0.08	-6.1	<1e-05
HB-ABi	-1.3	0.08	-10.6	<1e-05

Table 1: Summary of the linear mixed model (quantity movement-condition+random(participant)) and Tukey post hoc analysis.

4. Discussion

We assessed the hypothesis that head movements may compensate for the impossibility to move the hands in a narrative task using motion capture analysis. A similar averaged displacement of the head was observed in Hands Free and Hands Blocked conditions. Without surprise, the clearest effect on head motion was observed when biking with the hands. In this condition, there was a co-activation of arms and head that support to some extent the validity of our analysis in the other conditions.

Several interpretations could explain the lack of effect between the HF and the HB conditions. First, most of the participants did not move their hands much in the HF condition, with large between-subject variability. If the hands do not move in the HF condition, there is nothing to compensate for in the HB condition. To further assess this idea, we investigated whether the average displacement of the head correlated with the one of the hands in the HF and HB conditions. In other words, do the participants who gesture a lot with their hands in HF compensate more with the head in the HB condition than those who gesture less? If they do, the difference of average displacement of the head between HB and HF should be close to 0 for those who do not gesture a lot and increase as the displacement of the hands in HF increases (the head moves more in HB than in HF). Figure 5 shows the opposite. In particular, the two participants who gestured the most with their hands in the HF condition were the ones who decreased the most their head movements in HB as compared with HF.

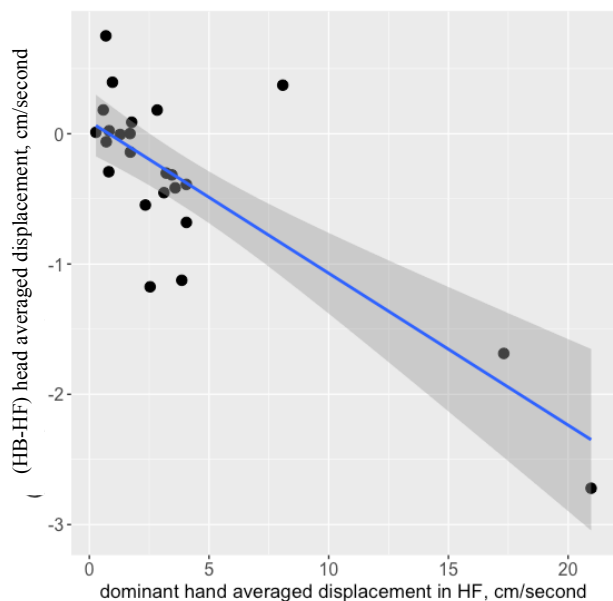


Figure 5: Scatterplot with blue regression line and confidence bands (grey) of the difference of displacement of the head between HB and HF (y-axis) as a function of the averaged displacement in HF (Pearson correlation is -0.8 , $t = -6.21$, $df = 22$, $p\text{-value} = 2.83e-06$).

This suggests that head movements do not compensate for the lack of hand gestures, rather having hands blocked inhibits head motion, at least for the two participants who gestured the most. However, our results should be taken carefully as we analyzed an existing corpus rather than conceived an experiment specifically dedicated to the research question. The first limit is that narrative sessions were really short in our study as compared with previous work. Rimé *et al.* (1984) found that when the head, arms, hands, legs, and feet are blocked, the activity of eyebrows, eyes, mouth, and fingers increases. Participants were engaged in a 50-minute conversation, whereas in our study, their task was a 2-minute story retelling with a much higher cognitive load. Therefore, even if the experimenter was listening to the participants and providing non-verbal backchannels, this short monologue context may have involved less communicative gestures.

Motion capture may also affect speakers' motion: it is less natural to move the hands when Optitrack markers are fixed on them. It is thus possible that in the HF condition, the hands are blocked due to the unnatural settings. Speakers could thus already move their head to compensate in this condition. Future work should investigate the effect of motion capture markers on speaker's motion behavior. Similarly, speakers also wore equipment on their heads. One can expect other types of compensations such as eyebrow movements that were not influenced by any Optitrack marker. One way to reduce this effect would be to train speakers to get used to the situation. Also, it would be good to distinguish between gestures and other non-communicative movements using methods from artificial intelligence such as deep learning even though this may require larger sets of data.

Finally, it is possible that entrainment between body parts rather than compensation supports speakers' behaviors.

Despite its limits, this preliminary study suggests that motion capture solution should be further included and adapted to the study of co-verbal gestures and motions.

5. Acknowledgments

This research is jointly supported by the French National Research Agency (ANR-17-FRAL- 0005-01) and the German Research Foundation (DFG) as part of the SALAMMBO project (Spoken Language in Motions: Learning and Adaptation of Speech Communication in the context of Body Motions). The authors thanks François Bérard for his help with 3D motion analysis.

6. References

- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial cognition and computation*, 5(4), 307-331.
- Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. *Gesture, speech, and sign*, 93-116.
- Rimé, B., Schiaratura, L., Hupet, M., & Ghysseleincx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, 8(4), 311-325.
- Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does sitting on your hands make you bite your tongue? The effects of gesture prohibition on speech during motor descriptions. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29), 1097-1102.
- Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. *Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5-8 September 2003, Göteborg University, Sweden. Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 90, ISSN 0349-1021, p 21-24*.
- Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, 57, 257-267.
- Graf, H. P., Cosatto, E., Strom, V., & Huang, F. J. (2002). Visual prosody: Facial movements accompanying speech. In *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition* (pp. 396-401).
- Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Human Movement Science*, 2(1-2), 35-46.
- Colletta, J. M., Kunene, R. N., Venouil, A., Kaufmann, V., & Simon, J. P. (2008, May). Multi-track annotation of child language and gestures. In *International LREC Workshop on Multimodal Corpora* (pp. 54-72). Springer, Berlin, Heidelberg.
- Zhao, W., Chai, J., & Xu, Y. Q. (2012). Combining marker-based mocap and RGB-D camera for acquiring high-fidelity hand motion data. In *Proceedings of the ACM SIGGRAPH/eurographics symposium on computer animation* (pp. 33-42).
- Livingstone, S. R., & Palmer, C. (2016). Head movements encode emotions during speech and song. *Emotion*, 16(3), 365-380.
- Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, 9(3), 189-206.

