The effect of verbal working memory load in speech/gesture integration processing

Kendra Gimhani KANDANA ARACHCHIGE[1], Laurent LEFEBVRE[1], Isabelle SIMOES LOUREIRO[1], Wivine BLEKIC[1], Mandy ROSSIGNOL[1], Henning HOLLE[2]

1 Department of Cognitive Psychology and Neuropsychology, University of Mons, Belgium

2 Department of Psychology, School of Life Sciences, University of Hull, England

Co-speech gestures are characterized by a formal relationship between hand movements and the verbal units accompanying them [1]. Even though they retain a certain meaning despite the lack of context, they rely on the latter to be understood in a conversation [2]. Several studies agree on an impact of co-speech gestures on language comprehension [3, 4, 5, 6, 7]. A study among aphasic patients showed improved comprehension following the presentation of congruent co-speech gestures compared to incongruent ones [8]. Furthermore, they were perceived and processed by brain regions linked to semantic information [7]. The observed gestures modulated the neural activation, suggesting an attempt of comprehension by the listeners. Because of this online processing between co-speech gestures and verbal utterances during speech, the working memory (WM) is likely to be involved in their integration [9, 10]. In 2014, Wu & Coulson [11] have investigated how verbal (VWM) and visuospatial WM capacity influence the processing of speech/gesture integration. They have highlighted better performances on gender classification task (task in which participants should discriminate whether they hear a man or woman's voice while watching gestures enacted by a man or woman; a gender congruent condition being the voice of a man heard simultaneously of a gesture enacted by a man) among participants with a higher ability in processing visuospatial information. However, they failed to show this with its verbal counterpart. Nevertheless, given the nature of iconic gestures (i.e. their associativity with verbal information), its involvement would have been expected. One explanation relied in the potential lack of complexity of the VWM task (i.e. remembering 1 to 4 digits) to cause an interference between the tasks. In order to assess this lack of effect, we suggested to conduct a similar study, in which the VWM task would be of increased difficulty. The aim of our study was to observe a reduced benefit of semantic congruency on gesture/speech integration when increasing the load on VWM. We thus expected: (1) a main effect of semantic congruency, shown by reduced reaction times (RTs) for the semantically congruent

(SC) condition compared to the semantically incongruent (SI), (2) a main effect of gender congruency, shown by faster RTs in the gender congruent (GC) condition, and (3) an interaction between the VWM and SC. In the latter, we expect the difference in RTs between SI and SC to be reduced in the high load condition compared to the low load. For this, 53 participants (27 females, age $M = 23.75$; $SD = 8.76$) took part in an hour long study. They were students from the University of Hull and had no reported sensitive or psychic disorders. All participants were fluent in English and gave written informed consent. They received 8£ for taking part in the experiment. This study was approved by the Ethics Committee of the University of Hull. The study included a reading span test (RST) and a computerized task (main task) composed of a gender classification task (GCT) as the primary task and a word span test (WST) as a secondary task. The RST [12] required participants to read out loud blocks of semantically unrelated sentences and remember the final word in each. The test was of increasing difficulty, with each level containing 3 blocks of either 2, 3, 4, 5 or 6 sentences. Participants were thus asked to retain 2, 3,4, 5 or 6 words. An individual's reading span was the highest level where the final words from 2 consecutive sentences were recalled correctly. Performances on this task were used to create our grouping variable. The main task (fig.1) was composed of a GCT embedded in a WST. For the GCT, stimuli consisted on the one hand of video recording of 16 simple acted actions (e.g. zipping up a coat or breaking a bar) which have been used in a previous study [13]. Each action was either completed by a man or a woman. On the other hand, verbal utterances describing each action were recorded separately. Video and voice recordings were then paired, the audio following the onset of the video from 200ms, to create an audio-visuo stimulus which was either congruent in gender (i.e. the person completing the action and the voice heard belonged to either a man or a woman) and/or gesture (i.e. the seen action matched the verbal utterance) or incongruent in gender and/or gesture. This task was completed by a secondary word span task. Stimuli consisted of 1280 English words, retrieved from subtlex-uk [14]. All the words contained either 1 or 2 syllables and were selected based on ratings for familiarity, concreteness, imageability. The Zipf scale was considered as the measure for word frequency. The words were then randomly ordered and seperated into 4 groups: high load targets and high load distractors (4 words in each group), and low load targets and low load distractors (1 word in each group). The experimental task comprised 256 trials consisting of words and gesture videos. Each trial began with either one (low

load) or four (high load) written words at an ISI (inter stimuli interval) of 750 ms. Participants were asked to remember these words for futher recognition. They were then presented with an audio-visuo stimulus and were asked to indicate whether the voice heard was spoken by a male or female by clicking on the right or left button on the mouse (conditions were conter-balanced). If they responded incorrectly or failed to answer within 2000ms, they received a 500ms feedback. Past the 2000ms, or after a given response, participants were then confronted with two (low load) or eight (high load) written words, displayed around the center of the screen. They were asked to click on the word(s) previously presented, at the beginning of the same trial, in the order of presentation. Trial were separated by an ITI (inter trial interval) of 500, 750 or 1000 ms. All participants were asked to reply as quickly and accurately as possible. Participants were grouped into 2 groups (RST low and RST high) according to their spans on the reading span task (span of 2 = 1 ; > 2 = 2). Following this, 2 participants were excluded from the experiment for outliers. We conducted a 3 way repeated-measures ANOVA (semantic(2)*load(2)*gender(2)) with the RST variable as inter-subject factor. Results showed a main effect of semantic congruency ($F(1,49) = 5.03$; $p = 0.03$), indicating that RTs were lower in the semantic congruent (SC) condition ($M = 619.16$; $SD = 15.78$) than incongruent (SI ($M = 625.73$; $SD = 15.9$)). A main effect of gender congruency (GC) was also found ($F(1,49) = 71.12$; $p < 0.01$), indicating that RTs were lower in the gender congruent (GC) condition ($M = 609.48$; SD = 15.66) than incongruent (GI ($M = 635.42$; $SD = 10.04$)). A triple interaction semantic*gender*RST ($F(1,49) = 4.76$; $p = 0.03$) was also highlighted. For both groups (low RST ; high RST), performances were slower in the GI condition compared to the GC condition in both the SC and SI conditions ($M_{SC-GI-RST-low} = 668.1$; SD = 24.3 ; $M_{SC-GC-RST-low} = 651.08$; SD = 23.61 ; $M_{SC-GI-RST-high} = 595.05$; SD = 21.16 ; $M_{SC-GC-RST-high} = 562.45$; SD = 20.57 ; $M_{SI-GI-RST-low} = 678,43$; SD = 24,37 ; $M_{SI-GC-RST-low} = 650.38$; SD = 23.92 ; $M_{SI-GI-RST-high} = 600.1$; SD = 21.23 ; $M_{SI-GC-RST-high} = 574.01$; SD = 20.84). Furthermore, an interaction load*semantic*gender*RST was also found ($F(1,49) = 4.029$; $p = 0.05$). Finally, we also calculated the difference between SI and SC in each condition, to determine the advantage of congruency. Participants with a high RST performed significantly slower ($F(1,28) = 5.54$; $p = 0.03$) in the SI condition compared to SC ($M = 8.31$; $SD = 3.53$) when in high load and GC condition. This difference was not found among the RST low participants ($F(1,29) = 0.98$; $p = 0.33$). In the lack of significant differences between RTs for SI and SC according to the load on the WST,

these results suggest an absence of VWM loading effect on gesture/speech integration. However, when considering the performances on RST, we do observe significant differences between performances according to the groups. Therefore, although gesture/speech integration doesn't seem affected by the secondary task on its own, participants with a higher RST show significantly slower RTs in the SI condition compared to SC in the high load condition. Participants with a low RST tend to not be disturbed. This could indicate an interference effect for the high RST participants when faced with semantically incongruent stimuli in a high load condition. Hence, the WST could have engaged specific verbal WM resources for high RST participants, slowing their RTs in the main task. In conclusion, it seems that VWM plays a role in gesture/speech integration among participants presenting high RST performances. However, these results need to be further investigated and a deeper analysis is required to better understand its role.