Gesture

High verbal working memory load impairs gesture-speech integration: Evidence from a dual task paradigm

--Manuscript Draft--

Manuscript Number:	GEST-20028R2
Full Title:	High verbal working memory load impairs gesture-speech integration: Evidence from a dual task paradigm
Short Title:	
Article Type:	Article
First Author:	Kendra Gimhani Kandana Arachchige
Other Authors:	Henning Holle
	Mandy Rossignol
	Isabelle Simoes Loureiro
	Laurent Lefebvre
Corresponding Author:	Kendra Gimhani Kandana Arachchige Université de Mons Mons, Walloon Region BELGIUM
Funding Information:	
Keywords:	Gesture-speech integration; verbal resources hypothesis; Iconic gestures; gesture-speech comprehension; verbal working memory
Manuscript Classifications:	comprehension; memory
Abstract:	While previous studies have shown the importance of visuo-spatial working memory in the processing of co-speech iconic gestures, clear evidence for a potential involvement of the verbal working memory (vWM) is currently lacking. To address this issue, participants in the present study were presented with a dual task paradigm. The main outcome variable was the performance on a Stroop-like gesture task which provides a behavioural index of gesture-speech integration. Participants performed this task under conditions of either high or low concurrent vWM load. Unlike in previous studies, the number of words to remember in the high load condition was determined by their individual verbal span rather than being fixed. Results showed longer reaction times for semantically incongruent gesture-speech combinations as compared to congruent combinations. However, this semantic congruency effect disappeared when the vWM load increased. This result suggests a causal involvement of verbal working memory capacity in gesture-speech integration.
Author Comments:	A "Response to the reviewers" document has been attached as a manuscript document but separately given the presence of images that do enter fit in the allocated space below.
Order of Authors Secondary Information:	

Abstract

While previous studies have shown the importance of visuo-spatial working memory in the processing of co-speech iconic gestures, clear evidence for a potential involvement of the verbal working memory (vWM) is currently lacking. To address this issue, participants in the present study were presented with a dual task paradigm. The main outcome variable was the performance on a Stroop-like gesture task which provides a behavioural index of gesture-speech integration. Participants performed this task under conditions of either high or low concurrent vWM load. Unlike in previous studies, the number of words to remember in the high load condition was determined by their individual verbal span rather than being fixed. Results showed reaction time costs in the form of longer reaction times for semantically incongruent gesture-speech combinations as compared to congruent combinations. However, this semantic congruency effect disappeared when the vWM load increased. This result suggests a causal involvement of verbal working memory capacity in gesture-speech integration.

Keywords: Gesture-speech integration; verbal resources hypothesis; iconic gestures; gesture-speech comprehension; verbal working memory

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Introduction

Daily communication often requires us to integrate between a variety of information received in order to make sense of the speaker's intent. Not only does one need to understand the spoken message, but this also needs to be combined with the visual information conveyed by the speaker's gestures. Iconic gestures, which are one of the most frequent gesture types, can be described as gestures that carry featural resemblance to the objects they refer to, or that illustrate actions or spatial relationship between objects (McNeill, 1992). These gestures are produced exclusively during speech and are temporally aligned to the verbal utterance they refer to (Hadar & Butterworth, 1997; Kita & Özyürek, 2003). Numerous studies using behavioural (e.g., Holler et al., 2009), electrophysiological (e.g., Kelly et al., 2009) and imaging techniques (e.g., Dick et al., 2009) have shown that the iconic gestures produced by a speaker are indeed subject to processing by the listener and that the latter identifies their meaning and utilizes it in their understanding of the intended message.

Addressees are usually unable to recall whether a specific piece of information was conveyed in the gesture channel or the speech channel, suggesting that rather than maintaining separate speech and gesture memory traces, the information is integrated into a single unified semantic representation (Kelly et al., 1999; McNeill et al., 1994; Zhao et al., 2018). The concept of *integration* refers to an implicit cognitive process allowing the combination of audio-visual information in a single representation (Green, et al., 2009). One way of observing this integration is by highlighting a transfer of information from a speaker's gesture to the listener's response. In a study evaluating the effects of iconic gestures on verbal comprehension, Kelly et al. (1999, Exp. 4) observed that participants recalled information that was solely presented through iconic gestures. While participants rarely remembered from where they had gotten that information from (Kelly et al., 1999), two other studies showed that some participants were able to tell that the additional stemmed from the speakers' gestures (Exp. 1 & 2, Alibali et al., 1997; Ovroye & Storm, 2019).

Another way of observing evidence for gesture-speech integration is by showing an interference effect of incongruent information presentation. A decrease in performances (shown by longer reaction times or an increase of incorrect responses) following the presentation of incongruent gestural and verbal information (i.e., when the gestural information does not match its verbal counterpart), is therefore suggestive of a failed integration attempt of the information contained in both modalities (Holle & Gunter, 2007). The nature of this integration is still subject to debate. While some studies have suggested this integration to be entirely automatic (McNeill et al., 1994; Kelly et al., 2010), others have taken a nuanced stance (Holle & Gunter, 2007; Kelly et al., 2007).

Because gesture-speech integration appears to happen on-line (Kelly et al., 2004; Özyürek et al., 2007 ; Wu & Coulson, 2014) working memory (WM) is likely to play a role in the integration of information from both modalities. According to the classical model of WM put forward by Baddeley & Hitch (1974), the WM is essential for online processing, as it allows to temporarily retain information. The model describes a *central executive* dedicated to regulate, retrieve and process the flow of information and two sub-components; a phonological loop, devoted to maintain and refresh phonologically coded information, and a visuospatial sketchpad, involved in the maintenance of material with a visual or spatial component. Following research by Smyth & Pendleton (1990) that suggested the existence of a kinaesthetic or movement-based sub-system used in gesture, an updated version of the classical WM model (Baddeley, 2012) considers kinaesthetics as part of the visuospatial sketchpad. More recently however, Wu and Coulson (2015) suggested the existence of a distinct *kinaesthetic working memory*, to maintain and manipulate bodily representations and allow listener's to buffer vague body movements until they can be combined with related concepts uttered in speech.

With respect to which working memory components underpin gesture-speech integration, two (not mutually exclusive) hypotheses have been put forward.

One the one hand, according to the *visuo-spatial hypothesis*, the main function of co-speech gestures is to express spatially encoded information in a verbal medium, thus drawing heavily on the visuo-spatial sketchpad. There is an emerging body of evidence supporting the visuo-spatial hypothesis (for a review, see Alibali, 2005; Hostetter, 2011). First, in gesture production, gestures are preferentially produced when needing to express spatial (Beattie & Shovelton, 2002) or motor information (Feyereisen & Havard, 1999) and have been suggested to allow coordination of spatio-motoric aspects of a message with its linguistic content (Kita, 2000). Furthermore, using a gesture eliciation task, Chu et al. (2014) showed that visual working memory capacity predicted the frequency of deictic (i.e. pointing) and depictive (i.e. iconic) gestures. Participants presenting lower visual working memory capacities gestured more frequently compared to participants with a higher visual working memory capacity (Chu et al., 2014). With respect to gesture-speech comprehension, Wu and Coulson (2014) explored whether taxing the visuospatial sketchpad in a secondary task would impair participants on a primary speech-gesture comprehension task. The logic behind such a dual task paradigm is that if both the secondary and primary tasks share the same cognitive resources, the presence of the secondary task will interfere and lead to a decrease in performances on the primary task. Doing this, Wu and Coulson (2014, Exp. 2) showed that the presence of a secondary visuospatial task reduced participants ability to perform on a primary gesture/speech integration task (i.e., where they asked participants to explicitely judge the relatedness of a picture probe to a previous utterance). The authors interpreted these results as demonstrating an important role for visuospatial resources in gesture/speech integration. This is in line with previous studies showing similar neuronal activity during the interpretation of iconic gestures and fixed-images (Wu & Coulson, 2011) suggesting common underlying processing for both elements and that participants used the visuospatial cues contained in these gestures to build a more precise representation of the related verbal utterance (Wu & Coulson, 2007).

On the other hand, the *verbal ressources hypothesis* emphasizes the close link of iconic gestures with the speech they accompany, including close temporal and semantic alignment (McNeill, 1992). According to this hypothesis, semantic analysis of gestures depends heavliy on verbal resources. However, evidence for a causal involvement of vWM in gesture-speech integration is currently sparse. This is surprising, given how deeply gesture and speech are intertwined, both in language production (McNeill, 1985) as well as comprehension (Kelly et al., 2010). In language production, Wagner et al. (2004) showed that not only did gesturing have a similar impact on visuospatial WM than it does on verbal but also that the propositional content of the gesture (i.e., the gesture semantics) mattered. Other authors have also suggested that higher gestures rates among individuals with a lower vWM could be interpreted as a facilitating effect of gestures in language production (Gillepsie et al., 2014) but of course such correlational approaches do not allow inferences of causality. In fact, in another study, Chu et al. (2014) found no association between vWM capacity and iconic gesture production. In gesture-speech comprehension, to the best of our knowledge, Wu and Coulson (2014, Exp. 3) are the only authors having attempted to highlight a causal relationship between gesture-speech integration and vWM. Using the same dual task paradigm as described above (Wu & Coulson, 2014, Exp. 2) in combination with a verbal secondary task, the authors failed to demonstrate a link between gesture-speech integration and vWM load. In their study, performances on the primary gesture-speech integration task was not affected by an increase of vWM load.

One possible explanation for the currenty paucity of evidence in support of the verbal resource hypothesis could be that previous dual paradigm studies did not take individual differences in verbal working memory into account when designing their high-load condition. The need of taking individual differences into account is consistent with a recent review highligting the limited research in this field (see Özer & Göksun, 2020a). For example, Wu and Coulson (2014, Exp. 3) used a fixed number of 4 digits in their high load condition. This may not have been a sufficient number to fully tax verbal WM capacity in high-span individuals, reducing the effectiveness of the load condition and therefore increasing the risk of a negative finding. An alternative approach where the number of items in the high load condition is individually adjusted depending on a person's digit span could increase the statistical power in this case. Özer and Göksun (2020b) considered individual differences and investigated how visuospatial and verbal abilities affect gesture and speech processing. The authors observed that the sensitivity to gestures and to speech depends on the indidivual WM capacity. In other words, an individual with a higher visuospatial or verbal WM capacity will be more sensitive to, respectively, the gestural or verbal information (Özer & Göksun, 2020b).

Another important consideration is the type of task used to assess the impact of gesture during comprehension. In Özer and Göksun's (2020b) task, participants were asked to make a conscious decision about gesture-speech relationship. Wu and Coulson (2014) also required participants to explicitely judge the relatedness of a picture probe to a previous utterance. Kelly et al. (2009) developed a more implicit gesture comprehension paradigm. In this Stroop-like task, participants were presented with an audiovisual stimuli (spech and gesture) and asked to judge the gender of the voice heard (i.e. Gender Classification Task ; GCT). Audiovisual stimuli differed with respect to their gender congruency (same vs different) and the semantic congruency (congruent vs incongruent). The authors found that even though the task did not require to attend the gestures, the semantic congruency between gesture and speech affected their performance on the gender classification task, with longer reaction times for semantically incongruent gesture-speech combinations. The advantage of using such an implicit paradigm is that it is less likely to be influenced by demand characteristics as gestural information is irrelevant to the behavioral task (Kelly et al., 2007).

As mentioned, previous gesture studies have already pointed out the importance of considering individual differences (Gillepsie, et al., 2014; Özer & Göksun, 2020a,b; Wu & Coulson, 2014). However, there are concerns that the Reading Span Task (Daneman & Carpenter, 1980), as used in Wu and Coulson's study (2014), is more likely to assess working memory capacity (Engel, 2002), represented by the Baddeley's *central executive* (Delaloye et al., 2008) rather than the integrity of the phonological loop. This component is mostly investigated with a verbal memory-span procedure (Baddeley, 1992). In the present study, we, therefore, used the digit span task, representing a classical measure of an individual's verbal working memory span (Wechsler, 2008; Boehringer et al., 2013)

Because of the current paucity of evidence in support of the verbal resource hypothesis, the present study thus investigated the impact of vWM load (low vs high) on gesture/speech integration using an implicit measure of integration, as proposed by Kelly et al. (2009) during a dual-tak paradigm. As their primary task, participants completed an established reaction time task (i.e. a gender classification task) providing a behavioral index of gesture-speech integration (Kelly, Özyürek, & Maris, 2010 ; Zhao et al. 2018). Participants were asked to perform this task under conditions of either high or low concurrent verbal working memory load (i.e., secondary task). Importantly, the number of words to remember in the high load condition was determined by their individual verbal span. In the primary task, participants were presented with co-speech gestures (e.g., gesturing writing with an imaginery pen while saying "write"), with gender and semantic congruency of audiovisual stimuli being experimentally manipulated. Participants had to identify the gender of the spoken voice. The typical pattern of results in this paradigm is that, though task irrelevant, gestural information strongly influences RTs, with participants taking longer to respond when gestures are semantically incongruent with speech (Kelly et al., 2009 ; Zhao et al., 2018).

We predicted that participants would show: (1) a main effect of semantic congruency, shown by longer RTs for the semantically incongruent condition (SI) compared to the semantically congruent (SC) condition, reflecting the reaction time costs associated with gesture-speech integration, (2) a main effect of gender congruency, shown by faster RTs for the gender congruent condition (GC), and (3) an interaction between the vWM and semantic congruency. In the latter, we expect that the reaction time costs associated with gesture-speech integration are reduced in the high load condition compared to the low load condition.

Methods

Participants

Twenty-eight healthy French-speaking students or alumni from age 18 to 33 were recruited from two universities in Belgium to take part in this study. Exclusion criteria included neurological and current psychological disorders as well as visual and/or auditory impairments. Because previous studies have demonstrated that vWM can be impacted by anxiety levels (Vytal et al., 2013) and that although task accuracy does not seem impacted, reaction times (RTs) could suffer from high levels of anxiety (Hadwin et al., 2005), state-anxiety measures were also taken, through the State Trait Inventory Anxiety (French-Canadian adaptation: "Questionnaire d'auto-évaluation de C.D.-Spielberger et al., 1966 – Inventaire d'Anxiété Etat-Trait", Gauthier, & Bouchard, 1993). Participants completed the questionnaire prior to testing and were not to present high rates of state-anxiety (STAI cut-off rate: 56). A descriptive analysis of the State-STAI score revealed an average score of 34.00 (SD = 1.95). The subjects were therefore not considered as anxious while taking part in the experiment. Boxplot analysis however highlighted one outlier with a score of 57. This subject was removed from further statistical analysis. In addition, two participants were removed for medical reasons (i.e., presence of pre-existing neurological conditions) and a third due to a technical failure. Thus, the final sample consisted of twenty-four right-handed participants (6 men; M_{age} = 22.71; Min

= 18; Max = 33; SD = 0.82). Although our sample is unbalanced regarding gender, no significant differences according to gender were highlighted in the results. Participants received 5€ for taking part in the experiment (funded by the University of Mons). All participants gave written informed consent before taking part in this study.

The Digit Span score determined which version of the vWM task was presented during the computerized Gender classification task. Across the sample, the Digit Span scores were distributed as follows: 5 participants obtained a score of 5, 12 a score of 6, 6 a score of 7 and 3 a score of 8 (M_{span} = 6.25; SD = 0.19).

Materials

Digit Span task

The Digit Span task (DS), from the WAIS-IV (Weschler, 2008), is a classical measure of an individual's verbal working memory span. Participants are asked to repeat increasing sequences of numbers. Each level contains 2 items. The individual's span is determined by the highest level at which both items were successfully repeated.

Dual Task Paradigm

- *Creation of stimuli*

A) Gender classification task. The final stimuli used during the experimental task consisted of video recordings of 32 simple actions or shapes (e.g., closing a book or rowing a boat). A list of the gestures used for this study (and their English translation) can be found in Appendix 1 of this paper. Each action was either completed by a man or a woman while simultaneously uttering the corresponding speech token. The videos were then edited so that (1) all gestures started at the preparation phase (i.e., the actor having their hands on their knees) and (2) the audio file was separated. Video and voice recording were then combined and paired, with the audio following the onset of the video with a delay of 700ms. The combination of video and audio material allowed to realise the experimental manipulations of Gender congruency and Semantic congruency. For the Semantic congruency manipulation, a gesture was paired with a seemingly incongruent speech token (e.g., gesturing *steering* while saying *throw*). The reverse combination was also realized. For the Gender manipulation, a gesture enacted by a man was paired with the speech token of a women.

Because previous studies have been carried out among English-speaking participants, our stimuli had to undergo a 3-step validation process. First, forty-nine healthy French-speaking participants (13 men; M_{age} = 23,7; SD = 2,68) were recruited. They were asked to judge the semantic congruency and incongruency of 102 videos (i.e., 34 videos associated with a sound x 2 gender, and 34 videos for incongruent condition; the incongruent condition was only presented in one way) on a 5-level Likert scale ranging from 1 (totally incongruent) to 5 (totally congruent). They were also asked to judge the voice of the 34 recordings as belonging to either a man or a woman. After analysis, 16 pairs (i.e., 32 gestures) were chosen, the congruent pairs having been considered congruent at an average of 4.5/5 and incongruent pairs at an average of 1.75/5. One incongruent pair (*shake-hammer*) was removed from the stimuli set after being judged too similar (average recognition rate of 2.33/5). Furthermore, participants classified the gender of the voice recordings correctly in 100% of cases. Second, thirteen new healthy French-speaking participants (2 men; M_{age} = 19,7; SD = 4,05) judged the iconicity of each video presented with no sound. They were asked to name the gesture seen (interpretative task). Results showed a 60% recognition rate, replicating results from Zhao and collaborators (2018). Given the specific nature of iconic gestures, that contain meaning per se but also depend on context to be understood (Holle & Gunter, 2007), we can safely assume that these results support the iconicity of our gestures (all video and sound files can be found in the supplementary material accompanying the manuscript submission). Third, eleven new healthy French-speaking participants took part in the last validation step. This task was carried out using Presentation® software (Version 20.1, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) where video and voice recordings were paired, with audio onset following the onset of the video with a delay of 700ms, to create an audio-visual stimulus. Participants were asked to answer as fast and accurately as possible whether they heard the voice of a man or a woman and their reaction times were recorded. A 2 (Semantic congruency) x 2 (Gender congruency) repeated-measures ANOVA yielded a significant main effect of Semantic Congruency ($F_{(1,10)} = 7.468$; $p = 0.02$) and a significant main effect for Gender Congruency ($F_{(1,10)} = 7.208$; $p < 0.02$) with faster RTs for semantically congruent (Sc) and gender congruent (Gc) pairs (respectively, $M_{Sc} = 562.12$; SD = 34.97 and $M_{Gc} = 591.82$; SD = 35.56) compared to semantically incongruent (Si) and gender incongruent (Gi) pairs (respectively, $M_{Si} = 647.29$; SD = 47.50 and $M_{Gi} = 617.59$; SD = 42.13). This suggests that although gestures are not task-relevant, they nonetheless elicit a reaction time cost reflecting a tendency to integrate gesture and speech, replicating earlier findings using English language materials (Kelly et al., 2010; Zhao et al., 2018). The results of these three validation studies allowed us to create a database with 256 different combinations of iconic gestures and sound (32 gestures x 2 pairs x 2 semantic congruency x 2 gender congruency).

B) Verbal working memory (vWM) task. In order to assess individualised verbal working memory capacity on gesture/speech integration, five versions of the vWM task were created (span 4, 5, 6, 7, and 8). For this purpose, French words were retrieved from the Lexique3 database (lexique.org, consulted on 26th March 2018) using the following criteria. The words required for this study could start by any letter (ortho = µ*), must contain 1 or 2 syllables (nbsyll

< 3), should be nouns (cgram = NOM), and have to be familiar (freqlemfilms2 <3). Following further manual filtering, 8960 words were randomly ordered and separated into 5 files (one per assessed span) which were then used to create the 5 versions of the computerized task. The Span 4 version contained 1280 words, Span 5 contained 1536 words, Span 6 contained 1792 words, Span 7 contained 2048 words and the Span 8 version contained 2304 words.

In each file, words were separated into 4 groups: low load targets (1 word to remember), low load distractors (1 word that serves as distractor during the recognition task), high load targets (4 to 8 words to remember) and high load distractors (4 to 8 words that serve as distractors during the recognition task).

- *Dual Paradigm Task*

This task was performed using Presentation® software (Version 20.1, Neurobehavioral Systems). During the computerized task, participants were asked to sit in front of a computer and keyboard. The version of the task depended on the participants individual vWM span, as assessed through the Digit Span task. The experimental task comprised 512 trials (256 videos x 2 load), separated in 8 blocks of 64 trials each. The videos were presented randomly (coded in Presentation) across the blocks. A graphical illustration of the trial structure is provided in Figure 1.

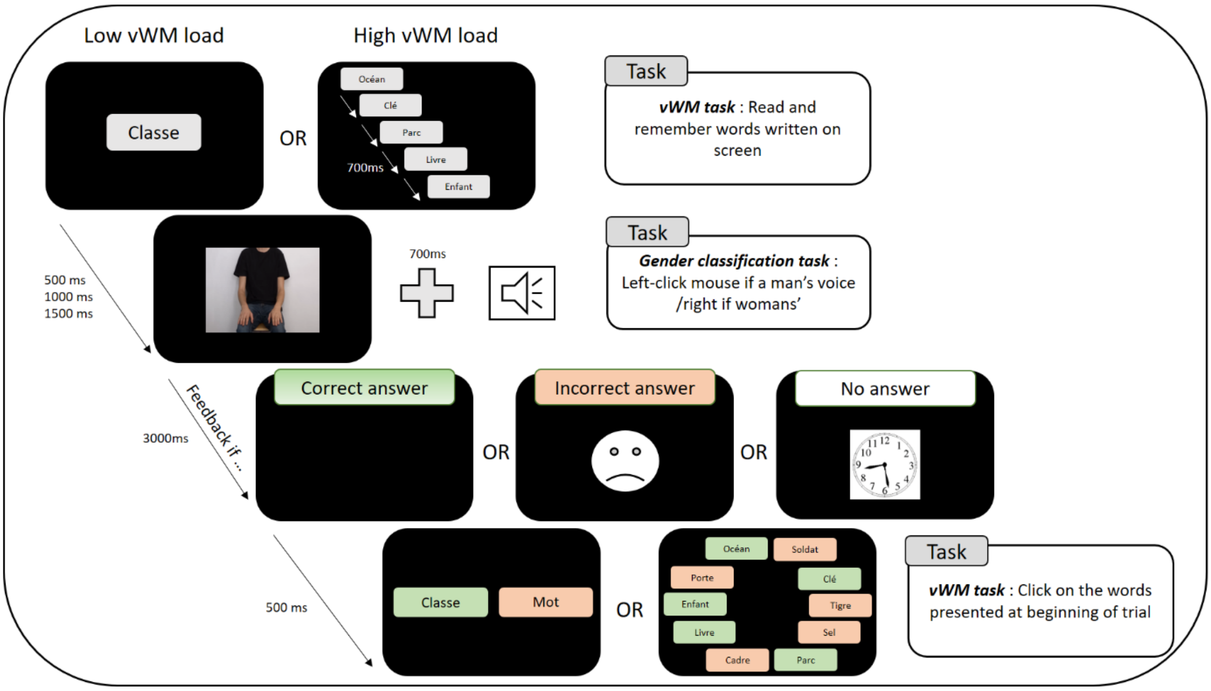

Figure 1 - Construction of Gender Classification Task embedded in verbal Working Memory task

Each trial began with either one (low load) or several (high load) written words at an inter-stimulus interval of 700ms. The number of words presented in the high load condition depended on an individual's digit span (e.g., if a participant's digit span was 5, they were presented with the version of the task where they would need to remember 5 words in the secondary vWM task). This ensured a maximum load on the vWM for each participant individually. Participants were then presented with an audio-visual stimulus and were asked to indicate, as fast as possible whether the voice heard was spoken by a male or female by clicking on the right or left button on the mouse. Key assignment (i.e., whether they would need to click on the right/left side of the mouse for the male/female voice) was counterbalanced across participants. If they responded incorrectly or failed to answer within 3000ms, they were shown, respectively, a frowning face or a clock, for 500ms. Subsequently, participants were presented with two (low load) or several (high load) written words, displayed in a spherical manner around the centre of the screen. Half of all words in each display were targets. They were asked to click on the word(s) previously presented, at the beginning of the same trial, in the order of presentation (see Figure 1). They were not provided with any feedback on their results on the vWM task. Trials were separated by an inter-trial interval of 500, 750 or 1000ms (randomly assigned). For the word recognition part, they were advised to take their time to answer (no RT data was collected for this part) and ensure they clicked inside the allocated answer area. Responses were made using the left click of the mouse.

Experimental design and statistical analysis.

We explored whether increasing the load on the vWM leads to a reduction of the semantic congruency effect. In a within-subject design, participants underwent one session where they completed a digit-span task to assess their individual span and a computerized task consisting of a gender classification task embedded in a verbal working memory task matching their individual span. The full experimental design was a 2 (high load/low load) x 2 (semantic congruency) x 2 (gender congruency) factorial design, and a corresponding 2x2x2 repeated-measure ANOVA was used to analyse the RT data. Following the ANOVA, follow-up t-tests were conducted.

Results

Main analysis

Gender Classification Task

After removing incorrect or missed trials (2.47%) and outliers at 2.5 SD (2.7%), the repeated-measures ANOVA yielded a main effect of Semantic congruency ($F_{(1,23)} = 5.15$; $p = .03$; $d_z = 2.27$), with faster RTs for semantically congruent pairs (M = 631.65; SD = 29.88), compared to semantically incongruent ones (M = 644.79; SD = 33.27). No main effect of Gender congruency was observed ($F_{(1,23)} = 0.73$; $p = .4$). Importantly, the interaction of Load x Semantic congruency was also significant ($F_{(1,23)} = 4.42$; $p = .04$; $d_z = .99$), reflecting a reduction of the semantic congruency effect when verbal working memory load is high (see Figure 2). A summary of the ANOVA results can be found in Table 1.

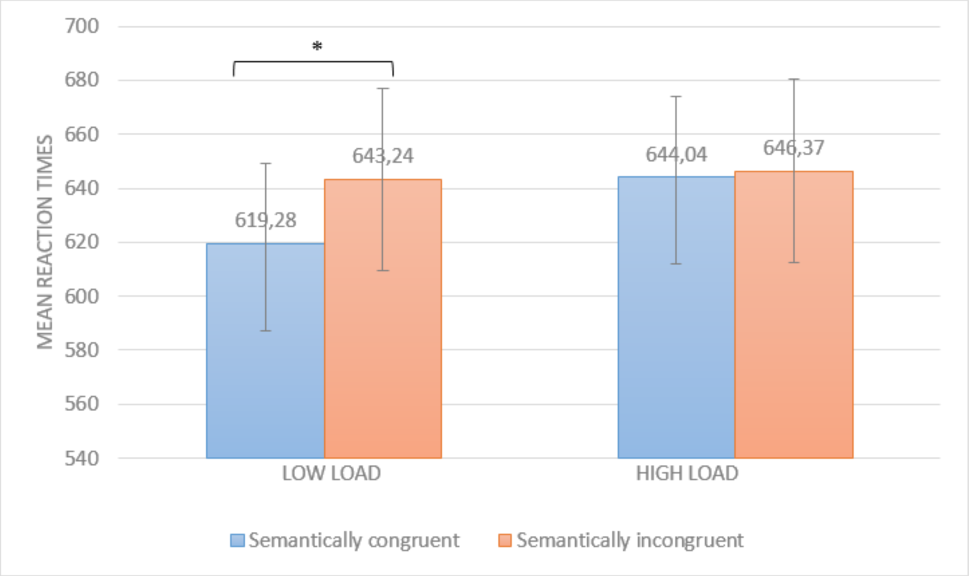

Figure 2 - vWM load effect on Semantic congruency differences. Error bars show 1 standard deviation
** representing a significant difference at 0.001*

Follow-up paired t-tests were conducted to further clarify the nature of the vWM by Semantic Congruency interaction. These follow-up tests indicated a significant difference ($t_{(23)}$ = -3.16 ; *p* < *.01*) between the semantically incongruent and congruent pairs in the low load condition (respectively, M = 643.24 ; SD = 33.90 and M = 619.27 ; SD = 29.96). No difference was found between the semantically incongruent and congruent pairs in the high load condition.

Secondary verbal Working Memory task

The behavioural data from the secondary vWM task was separated according to the Span (Span 5, Span 6, Span 7 and Span 8) and analysed.

In the low load condition, participants had an average accuracy (i.e., they correctly identified the presented word in the beginning of the trial) of 97.6% (Span 5: 96.9%; Span 6: 97.3%; Span 7: 97,2%; Span 8: 99.2%). A one-way ANOVA analysis yielded no significant differences between the span groups ($F_{(3,23)}$ = 0.83; *p* = *.5*).

In the high load condition, the first analysis consisted of determining the accuracy of a completely correct serial recognition (i.e., when the participants recognized all the presented words in the correct order of appearance). The average accuracy was of 2% (Span 5: 2.5%; Span 6: 5%; Span 7: 0.5% and Span 8: 0%). A one-way ANOVA analysis yielded no significant differences between the span groups ($F_{(3,23)} = 1.163$; $p = .35$). The second analysis consisted of determining the accuracy of a completely correct random recognition (i.e., when the participants recognized all the presented words irrespective of their order of appearance). The average accuracy was of 22.32% (Span 5: 31.1%; Span 6: 26.8%; Span 7: 18.5% and Span 8: 13%). No significant differences between the span groups were found ($F_{(3,23)} = 1.9$; $p = .16$). The final analysis consisted of verifying the accuracy of a correct random recognition at N-1 (i.e., when the participants recognized one item less than their span irrespective of their order of appearance). This analysis was conducted to ensure that participants' span was indeed saturated in the high load condition. The average accuracy was of 44.82% (Span 5: 47.3%; Span 6: 46%; Span 7: 40% and Span 8: 45.8%). Again, no significant differences between the span groups were found ($F_{(3,23)} = 0.5$; $p = .66$).

Discussion

The aim of this study was to test for a potential involvement of verbal working memory (vWM) in gesture/speech integration. The main finding was an interaction effect between vWM load and semantic congruency driven by slower reaction times for the semantically congruent pairs in the high load vWM condition compared to the low load vWM condition. This suggests that vWM is causally contributing to the integration of gesture and speech during comprehension.

The semantic congruency effect in the Gender Classification Task replicates earlier results in English-speaking populations (Kelly et al., 2009; Zhao et al., 2018). Indeed, it seems that although participants were not explicitly asked or required to pay attention to the meaning of gestures to complete the task, they did so anyway. This led to faster responses in classifying the gender of the spoken voice when the information conveyed in the speech matched that of the one conveyed in the gesture. While Kelly et al. (2009) showed this for action iconic gestures, this study extended it to both action iconic gestures and iconic gestures representing object attributes (i.e., shapes). This observation supports McNeill's (1992) and Kelly et al.'s (2010) hypothesis of iconic gestures and speech as being part of an integrated system where both gestural and verbal meanings are obligatorily combined in order to create an integrated concept.

However, this semantic congruency effect disappeared after increasing the load in vWM on the secondary task. Indeed, a smaller difference between the semantically incongruent pairs and the semantically congruent pairs (SI-SC) is observed in the high vWM load condition compared to the low load condition, regardless of the gender congruency. It therefore appears that when the demand on vWM is high (as shown by the decrease in performance in the high load condition compared to the low load condition), the advantage of a congruent bimodal presentation (Holler et al., 2009) is lost. This supports the hypothesis that a shared limited capacity resource is needed to complete the secondary vWM task and the primary reaction time task. It has been suggested that when two tasks are conducted simultaneously, performances slow down if the tasks share the same resources for response selection (Navon & Miller, 2002). According to this, it can be assumed that the resources involved in the vWM task are also involved in gesture/speech integration, the latter being hindered by increasing demand on the vWM.

This finding of high vWM impairing gesture-speech integration adds to the growing literature questioning the notion of "automatic processing" (McNeill et al., 1994; Kelly et al., 2010) of iconic gestures during speech comprehension. While Kelly et al. (2004) first suggested an automatic integration of speech and iconic gestures from early stage of processing, later studies provided evidence that the context in which participants were required to affects the degree to which gesture and speech interact (Kelly et al., 2009). The automaticity of gesture/speech integration seems, therefore, subject to variation according to contextual variables (Kelly et al., 2009), but also perceived intentionality of the addresser (Kelly et al., 2007), and the quantity of semantic overlap between the iconic gesture and speech (Holle & Gunter, 2007; Kelly et al., 2004; Özyürek et al., 2007). To this list we add that of cognitive load on vWM. One alternative would therefore be to consider gesture/speech integration as *strongly affected by context* rather than automatic, as suggested by Besner and Stolz (1999). For these authors, automaticity is not synonym of uncontrollable, but rather "affected (*controlled*) by context and distribution of spatial attention" (Besner & Stolz, 1999).

Finally, although our results did not put forward a gender-effect (i.e., faster RTs for gender congruent pairs compared to incongruent pairs, as shown as in Kelly et al.'s study, 2009) this could be consequent to the nature of the task involved. Indeed, while their study involved the completion of a single task, the current study utilises a dual-paradigm task which is more cognitively demanding. A change in the task modalities could alone explain the differences in obtained results (Wolf et al., 2017). Another methodological aspect that could explain the absence of gender effect is our stimuli. Because our actors were presented in the most neutral way possible (i.e., with a black t-shirt, jeans, and no distinguishable feature), it may have made their gender less apparent. In Kelly et al.'s (2004) study, actors were wearing two different elements of clothing, while in Zhao et al.'s (2018) study, the female actor wore a prominent belt. The change in methodology could also explain the discrepancy between these results and those obtained by Wu and Coulson (2014). In their study, the task was to explicitly judge the relatedness between a picture probe and the speaker's utterance. In contrast, the current study used a Stroop-like task, such as suggested by Kelly et al. (2009), involving an implicit judgement of relatedness.

Following this study, one point of discussion concerns the effect of anxiety-levels on performance. Unlike in previous research investigating working memory processes in gesture-speech integration (Momsen et al., 2020; Wu & Coulson, 2014), state-anxiety levels were considered in the present study for participants' selection. As mentioned above, performance on vWM tasks can be impacted by anxiety levels (Vytal et al., 2013) and reaction times (RTs) could suffer from high levels of anxiety (Hadwin et al., 2005). Furthermore, Owens et al. (2014) highlighted that trait-anxiety interacts with working memory capacity to predict cognitive performances. The authors suggest that performance of participants with a low working memory capacity were impaired by the presence of anxiety (Owens et al., 2014). Further research could investigate more directly how participants with different levels of anxiety perform on the task by adding a post-testing measure of anxiety levels, to see how the task affected anxiety levels.

Overall, this study thus supports the *verbal resources hypothesis* put forward by Wu and Coulson (2014) which suggests a relationship between the vWM capacity and the impact of iconic gesture on verbal comprehension. By experimentally manipulating verbal working memory load, we were able to provide causal, not merely correlational, evidence for vWM contributing to gesture-speech integration. Although gestures and their lexical counterparts tend to be temporally synchronized, the degree of synchrony is variable (Morrel-Samuels & Krauss, 1992) and large temporal offsets between gesture and speech tend to reduce the communicative impact of gesture (Habets et al., 2010; Obermeier et al., 2011). One possibility here is that vWM is required as a transient storage for verbal/phonological items to facilitate potential linkage with co-expressive gestures. This is in line with numerous studies showing the reliance of iconic gestures on their co-occurring verbal utterance to be understood (Dick et al., 2009; Hadar & Pinchas-Zamir, 2004; Holle & Gunter, 2007; Holle et al., 2008; Kelly et al., 2010; Krauss et al., 1996; Krauss et al., 1991).

Conclusion

In conclusion, the present study suggests an involvement of verbal resources in gesture-speech integration. Not only did healthy participants classify semantically related gestures and words faster compared to unrelated pairs, but this congruency advantage disappeared when increasing the load on the verbal working memory. This study goes against previous suggestions that verbal working memory is not involved in gesture/speech integration compared to visuo-spatial working memory. Rather, they both interact in the construction of meaning when observing co-speech iconic gestures. These findings are consistent with the literature showing the tight link between iconic gestures and verbal information.

References

Alibali, M. (2005). Gesture in spatial cognition: Expressing, Communicating and Thinking about spatial information. *Spatial Cognition and Computation, 5*(4), p.307-331.

Alibali, M.W., Flevares, L.M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology, 89*(1), p.183-193

Baddeley, A.D. (1992). Working memory. *Science, 225*(5044), p.556-559.

Baddeley, A.D. (2012). Working memory: Theories, Models and Controversies. *Annual Reviews of Psychology, 63*, p.1-29.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G. Bower, *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol.8, p.47-89). London: UK: Academic Press.

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous hand gestures? *British Journal of Social Psychology, 41*, p.403-417.

Besner, D., & Stolz, J. (1999) Unconsciously controlled processing: The Stroop effect reconsidered. *Psychonomic Bulletin & Review, 6*(3), pp. 449-455

Boehringer, A., Macher, K., Dukart, J., Villringer, A., & Pleger, B. (2013). Cerebellar transcranial direct stimulation modulates verbal working memory. *Brain Stimulation, 6*, pp. 649-653

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology: General, 143*(2), p.694-709

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* pp. 450-466

Delaloye, C., Ludwig, C., Borella, E., Chicherio, C., & de Ribaupierre, A. (2008). The reading span as a measure of working memory capacity: Norms based on a French speaking population of 775 younger and older adults. *Revue européenne de Psychologie Appliquée, 58*, pp. 89-103

Dick, A., Goldin-Meadow, S., Hasson, U., Skipper, J., & Small, S. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping, 30*, pp. 3509-3526

Engel, R. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), pp. 19-23

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

Gauthier, J., & Bouchard, S. (1993). A French-Canadian adaptation of the revised version of Spielberger's State-Trait Anxiety Inventory. *Canadian Journal of Behavioural Science, 25*(4), pp. 559-578

Gillepsie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition, 132*, pp. 174-180

Green, A., Straube, B., Weis, S., Jansen, A., Wilmes, K., Konrad, K., & Kircher, T. (2009). Neural integration of iconic and unrelated co-verbal gestures: a functional MRI study. *Human Brain Mapping, 30*(10), pp. 3309-3324

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2010). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*(8), pp. 1845-1854

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery and word retrieval in speech. *Semiotica, 115*(1/2), pp. 147-172

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture. Implications for gesture classification and function. *Journal of Language and Social Psychology, 23*(2), pp. 204-214.

Hadwin, J., Brogan, J., & Stevenson, J. (2005). State anxiety and working memory in children: A test of processing efficiency theory. *Educational Psychology, 25*(4), pp. 379-393.

Holle, H., & Gunter, T. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*(7), pp. 1175-1192.

Holle, H., Gunter, T., Rueschemeyer, S., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage, 39*(4), pp. 2010-2024.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in face-to-face context? *Journal of Non-verbal Behavior, 33*(2), pp. 73-88.

Hostetter, A. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*(2), pp. 297-315

Kelly, S.D., Barr, D.J., Church, R.B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Langage, 40*, p.577-592

Kelly, S. D., Creigh, P., & Bartolotti, J. (2009). Integrating speech and iconic gestures in a stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience, 22*(4), pp. 683-694.

Kelly, S., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, pp. 253-260.

Kelly, S., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance communication. *Psychological Science, 21*(2), pp. 260-267.

Kelly, S., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language, 101*, pp. 222-233.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill, *Language and Gesture* (pp. 162-185). Cambridge, UK: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for and interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*(1), pp. 16-32.

Krauss, R., Chen, Y., & Chawla, P. (1996). Non-verbal behavior and non-verbal communication: What do conversational hand gesture tell us? In M. Zanna, *Advances in experimental social psychology* (pp. 389-450). San Diego: Academic Press.

Krauss, R., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61*(5), pp. 743-754.

McNeill, D. (1985). So you think gestures are non-verbal? *Psychological Review, 92*(3), pp. 350-371.

McNeill, D. (1992). *Hand and mind.* Chicago: Chicago University Press.

McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction, 27*(3), pp. 223-237.

Momsen, J., Gordon, J., Wu, Y.C., & Coulson, S. (2020). Verbal working memory and co-speech gesture processing. *Brain and Cognition, 146,* 105640

Morrel-Samuels, P., & Krauss, R. (1992). Word familiarity predicts asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), pp. 615-622.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology, 44,* pp. 193-251.

Navon, D., & Miller, J. (2002). Queuing or Sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology, 44*, pp. 193-251.

Obermeier, C., Holle, H., & Gunter, T. (2011). What iconic gesture fragments reveal about gesture-speech integration: When synchrony is lost,memory can help. *Journal of Cognitive Neuroscience, 23*(7), pp. 1648-1663.

Ovroye, A.L., & Storm, B.C. (2019). Remembering what was said and done: The activation and facilitation of memory for gesture as a consequence of retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition, 45*(3), 526

Owens, M., Stevenson, J., Hadwin, J.A., & Norgate, R. (2014). When does anxiety help or hinder test performance? The role of working memory capacity. *British Journal of Psychology, 105*(1), p.92-101

Özer, D., & Göksun, T. (2020a). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology, 11,* 573555

Özer, D., & Göksun, T. (2020b). Visual-spatial and verbal abilities differentialy affect processing of gestural vs spoken expressions. *Language, Cognition and Neuroscience, 35*(7), 896-914

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*, pp. 605-616.

Smyth, M., & Pendleton, L. (1990). Space and movement in working memory. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 42*(2), pp. 291-304.

Spielberger, C. (1966). Theory and research on anxiety. In C. Spielberger, *Anxiety and Behavior.* New York: Academic Press.

Vytal, K., Comwell, B., Letkiewicz, A., Arkin, N., & Grillon, C. (2013). The complex interaction between anxiety and cognition: insight from spatial and verbal working memory. *Frontiers in Human Neuroscience, 7*(93), pp. 1-1.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (n.d.). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*, pp. 395-407.

Wechsler, D. (2008). *Wechsler adult intelligence scale- forth edition.*

Wolf, D., Rekittke, L.-M., Mittelberg, I., Klasen, M., & Mathiak, K. (2017). Perceived conventionality in co-speech gestures involves the fronto-temporal language network. *Frontiers in Human Neuroscience, 11*(573), pp. 1-19.

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain & Language, 101*, pp. 234-245.

Wu, Y. C., & Coulson, S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain & Language, 119*, pp. 184-195.

Wu, Y. C., & Coulson, S. (2014). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychologia, 153*, pp. 39-50.

Wu, Y. C., & Coulson, S. (2015). Iconic gestures facilitate discourse comprehension in individuals with superior immediate memory for body configurations. *Psychological Science, 26*(11), pp. 1717-1727

Zhao, W., Riggs, K., Schindler, I., & Holle, H. (2018). Transcranial Magnetic Stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. *The Journal of Neuroscience*, *38*(8), pp. 1891-1900

Appendix

List of the 32 congruent and incongruent Gesture-Speech Pairs

	Congruent		*Incongruent*	
Speech	*English translation*	*Gesture*	*Speech*	*Gesture*
Soulever	*Lift*	Soulever	Soulever	Déchirer
Déchirer	*Tear*	Déchirer	Déchirer	Soulever
Casser	*Break*	Casser	Casser	Touiller
Touiller	*Stir*	Touiller	Touiller	Casser
Composer	*Dial*	Composer	Composer	Tirer
Tirer	*Pull*	Tirer	Tirer	Composer
Fermer	*Close*	Fermer	Fermer	Toquer
Toquer	*Knock*	Toquer	Toquer	Fermer
Ouvrir	*Open*	Ouvrir	Ouvrir	Coudre
Coudre	*Stich*	Coudre	Coudre	Ouvrir
Couper	*Cut*	Couper	Couper	Peser
Peser	*Weigh*	Peser	Peser	Couper
Poignarder	*Stab*	Poignarder	Poignarder	Essuyer

Essuyer	*Wipe*	Essuyer	Essuyer	Poignarder
Conduire	*Steer*	Conduire	Conduire	Lancer
Lancer	*Throw*	Lancer	Lancer	Conduire
Peindre	*Paint*	Peindre	Peindre	Plier
Plier	*Fold*	Plier	Plier	Peindre
Pousser	*Push*	Pousser	Pousser	Ecrire
Ecrire	*Write*	Ecrire	Ecrire	Pousser
Verser	*Pour*	Verser	Verser	Ramer
Ramer	*Raw*	Ramer	Ramer	Verser
Cacheter	*Stamp*	Cacheter	Cacheter	Aspirer
Aspirer	*Vacuum*	Aspirer	Aspirer	Cacheter
Peler	*Peel*	Peler	Peler	Donner
Donner	*Give*	Donner	Donner	Peler
Cercle	*Circle*	Cercle	Cercle	Rectangle
Rectangle	*Rectangle*	Rectangle	Rectangle	Cercle
Triangle	*Triangle*	Triangle	Triangle	Carré
Carré	*Square*	Carré	Carré	Triangle

Nuage	*Cloud*	Nuage	Nuage	Cœur
Cœur	*Heart*	Cœur	Cœur	Nuage