« The pig with the pink hat »: An experimental study on speech/gesture coordination during development

Coriandre Vilain (coriandre.vilain@gipsa-lab.fr), Anne Vilain (anne.vilain@gipsa-lab.fr)
GIPSA-Lab & Grenoble University, BP 25, 38040 Grenoble cedex 9
Jeanne Clarke (jeanne.nora.clarke@gmail.com)
University Medical Center Groningen - Graduate School of Medical Sciences, University of Groningen, The Netherlands.

Abstract

This paper presents two experimental pilot studies on the coordination between speech and pointing gestures in adults vs children, in a "find the odd one" game. Experiment 1 tests the effect of the length of the name of the target, and experiment 2 the place of the informative focus in the noun phrase that is used to designate the target. Both experiments reveal similar patterns of coordination in adults and children: (i) gesture adapts to the length of the spoken utterance; (ii) gesture starts before speech (all the more so for adults); (iii) the apex of the pointing gesture is aligned with the beginning of the name of the target, and not with the crucial informative feature in the utterance; and (iv) the end of the gesture is reached after the end of the spoken utterance.

Introduction

Manual gesture is now widely considered as a major linguistic modality, as evidenced in particular by its close and consistent temporal relationship with speech in language production.

In adult speakers, a number of studies have shown that the production of manual gestures is temporally connected with the production of speech (since Levelt et al. 1985). In particular, these studies reveal that the onset of speech production is coordinated with the onset of the manual gesture. In his 1998 study, De Ruiter explored more specifically the coordination between specific landmarks in speech and gesture, and showed that the onset and the apex of the pointing gesture are coordinated with the focus, i.e. the distinctive element in a sentence (e.g. the green crocodile as opposed to the yellow crocodile, vs the green crocodile as opposed to the green lizard). These findings are consistent with McNeill's (1992) hypothesis that gesture and speech are co-expressive, meaning that "the gesture synchronizes with speech segment(s) that, in the context of speaking, express the same underlying idea unit as the gesture -these speech segments may or may not be 'lexical affiliates'".

As concerns the ontogenesis of language, manual gesture has been shown to play a crucial role in the development of language. Studies on early development reveal that the manual and the vocal motor behaviours evolve with a similar timing and with similar rhythmic structures (see Iverson & Thelen, 1999 for a review), in a form of « co-activation ». At the end of the first year, both modalities are recruited for the first symbolic behaviours of the babies, their first steps into referential communication: discrete pointing gestures first, then discrete word productions, both being most of the time produced conjointly (Lock et al. 1990). During the second year of life, the combined use of manual and vocal symbolic units is a major trigger for the development of language (Özçalışkan and Goldin-Meadow, 2005).

Not much is known, however, about the temporal relationship between speech and manual gesture after the second year of life. In a study on pointing gesture use in 4-5 year-old children vs adults, Pechmann & Deutsch (1982) show that in both groups, manual and vocal productions tend to overlap, manual gesture onset tends to precede speech onset, and "pointing occurred mostly at the instant the distinctive features in the verbal descriptions were uttered (i.e., features that are indispensable for the listener to differentiate the target from context)".

The present paper describes a pilot study aimed at systematically testing the temporal coordination between speech and manual gestures, as a function of the length of the lexical affiliate, and as a function of the position of the distinctive features in the utterance, in adults vs children.

Experiments

Two experiments were run, to test for the effect of varying (i) the length of the lexical item to be designated (experiment 1), and (ii) the structure of the noun phrase to be used (experiment 2). Both experiments are based on a "find the odd one" game.

Subjects

Ten subjects (5 children aged 7-12 years, 5 adults aged 21-37 years) participated in the pilot study. All were volunteers and naive as to the purpose of the study. No financial or other incentive was given to the participants for their participation.

Procedure

The subjects were seated near the experimenter in front of a large screen showing 6 picture stimuli of animals. They had to check whether there was an odd picture (target) among the stimuli and, in that case, were asked to point at the target picture while naming it; in the other case (filler stimuli, all the pictures are the same) no action was expected. To name the target picture, the subjects were asked to use the simple syntactic structure: "c'est le/la (*it is the*) + *target name*".

Material

In experiment 1, we modified the number of syllables in the target name in order to increase the duration of the lexical affiliate. We used names ranging from 1 to 3 syllables (e.g. from "chat" (*cat*) to "papillon" (*butterfly*)).
In experiment 2, we modified the syntactic structure of the target names in order to increase the syntactic complexity and to shift the distinctive element in the sentence. Figure 1 gives an example of stimuli for both experiments.

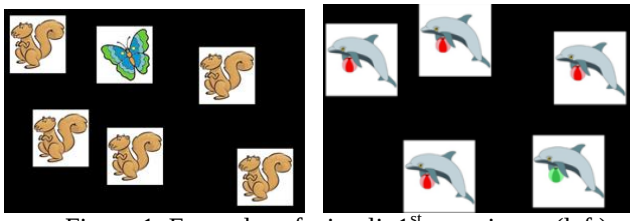

Figure 1: Examples of stimuli. 1st experiment (left), expected answer: "c'est le **papillon** (*it's the butterfly*)"; 2nd experiment (right), expected answer :"c'est le dauphin avec le ballon vert" (*it's the dolphin with the green ball*)"

Five different syntactic structures were used (see Figure 2). In this figure, the uppercase item corresponds to what we define as the informative focus element (IF), i.e. the element that makes the target picture different from the others. In this experiment, we made sure to use only IF elements with the same number of syllables. We used the following disyllabic words : "cochon (*pig)"*," poussin (*chick*)"," mouton (*sheep)"*," panda" and "dauphin (*dolphin*)".

S1 - c'est le COCHON *(it's the pig)*

S2 - c'est le cochon NOIR *(it's the black pig)*

S3 - c'est le GROS cochon *(it's the big pig)*

S4 - c'est le cochon avec le GROS gateau *(it's the pig with the big cake)*

S5 - c'est le cochon avec le chapeau ROSE *(it's the pig with the pink hat)*

Figure 2: The five syntactic structures used in experiment 2.

Apparatus

The stimuli were presented using the Presentation Software[2] running on a PC computer. Each trial was launched by the experimenter once she had checked the correctness of the previous trial.

The pointing gestures were recorded with a video camera recorder, and the oral productions with a digital audio recorder. The two signals were then synchronized.

Data analysis

Audio

Audio data were segmented using the Praat software[5].

In the first experiment we annotated the onset of the sentence (SO), the onset of the target picture name, that we called informative focus onset (IFO) for the sake of homogeneity with experience 2, and the end of the informative focus (IFE) that also corresponds in this specific case to the end of the sentence (SE).

In the second experiment we annotated the onset of the sentence (SO), the onset of the noun phrase (NPO), the onset of the noun (NO), the end of the noun (NE), the onset of the informative focus (IFO), the end of the informative focus (IFE), the end of the sentence (SE) (see figure 3 for an example).

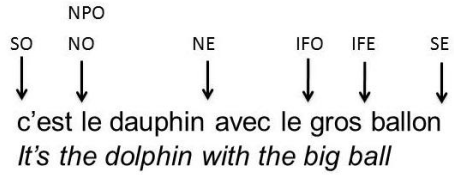

Figure 3: Marks used for audio annotation in a S4 type of syntactic structure.

Video

Video data were analyzed using the ELAN software[6]. In both experiments, we annotated the onset of the pointing gesture (GO), the apex of the gesture (GA), the return of the gesture (GR) and the end of the gesture (GE), once the hand had come back to its initial position.

Results and discussion

A mixed ANOVA model was used to analyze the data. Two variables were taken into account in each experiment: an intra-subject variable: the Number of Syllables (NS, for exp. 1) or the syntactic complexity (SC for exp. 2) and the same inter-subject variable for both experiments: the age group (AG: Children/ Adult). Considering the small number of subjects per group in this pilot study, we did not measure the interactions between the intra- and inter- subject variables.

Experiment 1

The first question in experiment 1 is whether the length of the lexical affiliate affects the duration of the manual gesture. Unsurprisingly, the duration of the target image name is shown to increase with the number of syllables ($p<0.001$). And as shown in Figure 4, the duration of the Gesture Hold (GH = GR-GA) also increases significantly with the number of syllables ($p<0.05$). A difference is also found for the duration of the Informative Focus (IFE-IFO) as a function of the Age Group ($p<0.01$), as well as an important difference in the duration of the Gesture Hold as a function of the Age Group ($p <0.01$). In the latter case, gesture duration is shorter in adults than in children.
So, the number of syllables has an effect on both gesture and speech duration: gesture *is* coordinated with speech.

[2] www.neurobs.com
[5] http://www.fon.hum.uva.nl/praat/
[6] http://tla.mpi.nl/tools/tla-tools/elan/

The second question is then *how* gesture is coordinated with speech.

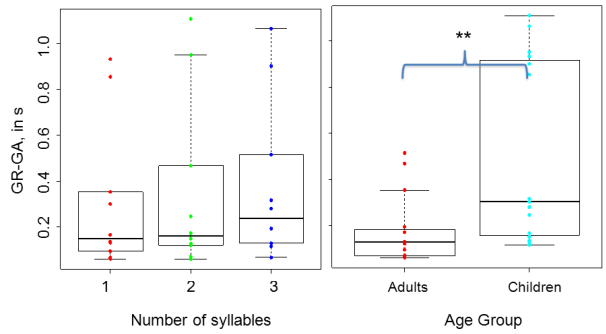

Figure 4: Gesture hold duration (GR-GA) as a function of the number of syllables and of the age group.

Figure 5 illustrates the distance between Speech Onset and Gesture Onset (SO-GO) as a function of the number of syllables and as a function of the Age Group (Figure 6). SO-GO is significantly higher than 0 ($p<0.001$), it does not vary with the Number of Syllables but it does vary with the Age Group. This means two things: (1) the anchoring between speech and gesture seems independent of the syllabic structure of the lexical affiliate and (2) in both adults and children, Gesture Onset anticipates Speech Onset and this anticipation is larger for adults than for children.

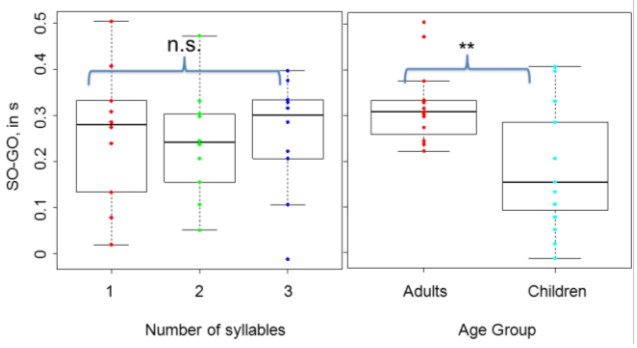

Figure 5: Speech Onset - Gesture Onset (SO-GO) as a function of the Syntax (left) and the Age Group (right).

We then tested the relation between the Gesture Apex and the Informative Focus Onset. IFO-GA does not vary with either the number of syllables or the Age Group. IFO-GA is also not significantly different from 0. IFO and GA thus appear as good candidates for synchronization in both adults and children.

Lastly, we compared the Informative Focus End with the Gesture Return. IFE-GR varies with the Number of Syllables ($p<0.001$), indicating that these two events are not coordinated.

In sum, the outcome of this first experiment is that children and adults have the same pattern of coordination: (i) Gesture Onset anticipates Speech Onset; (ii) Gesture Apex is synchronized with Informative Focus Onset; (iii)No coordination is shown between Gesture Return and Informative Focus End.

The second experiment is going to test whether the syntactic structure of the sentence can influence the coordination between speech and gesture. In other words, what happens for gesture if we vary the position of the informative focus in the noun phrase?

Experiment 2

Speech and gesture durations

Before answering this question, we first confirm the influence of the syntax and the age group on speech duration (SE-SO, see Figure 6) and gesture duration (GE-GO, see Figure 7).

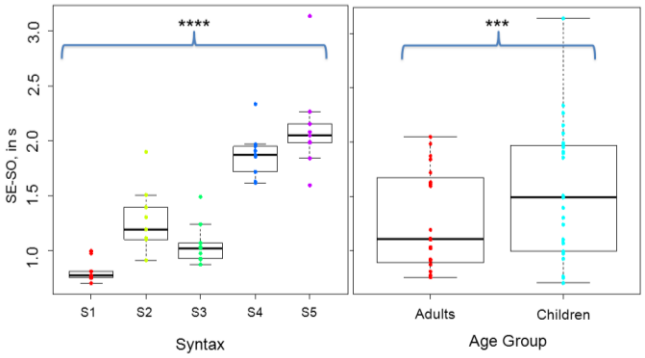

Figure 6: Speech duration (SE-SO) as a function of Syntax (left) and Age Group (right)

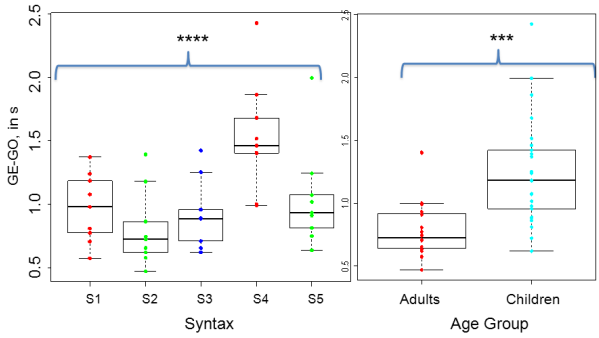

Figure 7: Gesture duration (GE-GO) as a function of Syntax (left) and Age Group (right)

Both gesture and speech durations vary with syntax and age group. This can sound trivial, given the kind of syntactic structure we chose, but it indicates that although the number of syllables is the same for all animal names in the noun phrases, the gesture duration also varies (note that in experiment 1, the gesture duration varied with the number of syllables): so gesture duration is correlated with the duration of the whole noun phrase, and not simply with the duration of the first noun.

Coordination between Speech and gesture

To explore the coordination between speech and gesture we start with the onset events. We compute the differences between the Gesture Onset and four different speech events that could be good candidates to be synchronized with GO: Sentence Onset (SO), Noun Phrase Onset (NPO), Noun

Onset (NO) and Informative Focus Onset (IFO). SO is the event that appears consistently closer to GO (figure 8) than the other ones.

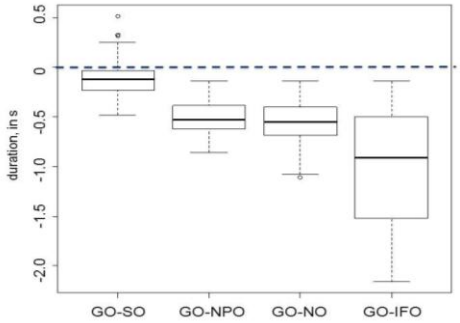

Figure 8: Distance between GO and 4 speech events: SO, NPO, NO and IFO.

So we carried out an ANOVA on SO-GO as a function of Syntax and Age Group (figure 9). The results show that (1) SO-GO does not depend on Syntax and that (2) it does depend on Age Group ($p<0.001$). Result (1) is of first importance because it means that wherever the informative focus is placed in the noun phrase, it does not affect the Gesture Onset timing. Result (2) shows that adults anticipate their Gesture Onset more than children, for whom it seems more synchronized with Speech Onset.

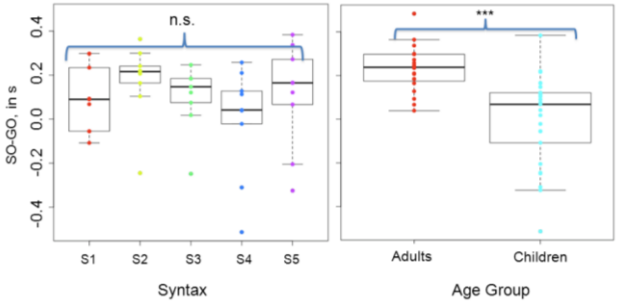

Figure 9: SO-GO as a function of Syntax (left) and Age Group (right)

To test the anchoring between gesture apex and speech events, we compared Gesture Apex (GA) to Informative Focus Onset (IFO), Noun Phrase Onset (NPO) and Noun Onset (NO): 3 speech events that were confounded with each other in experiment 1. IFO-GA is found to vary significantly with syntax ($p<0.001$) whereas both NPO-GA and NO-GA are not.

This confirms the previous result obtained with Gesture Onset: Gesture Apex is consistently aligned with the beginning of the noun phrase that designates the whole target image, and not with the informative focus element that gives the distinctive feature in that image.

As concerns the gesture end, no clear synchronization could be found with any speech event. The only result is that GE is significantly superior to SE ($p<0.005$), indicating that the gesture system waits for the end of the noun phrase before reaching its initial position.

Conclusion

These two experiments explore the coordination between speech and gesture events, as a function of the length and nature of the lexical affiliates, in adults vs children.

The first experiment shows that: in both groups, (i) gesture adapts to the length of the spoken utterance; (ii) gesture starts before speech (all the more so for adults); (iii) the apex of the pointing gesture is aligned with the beginning of the name of the target; (iv) no clear coordination is shown between the end of the gesture and the end of the utterance.

The second experiment confirms these results and gives more precision about the coordination. It reveals that the speech/gesture alignment does not vary as a function of where exactly in the sentence the distinctive feature is: the pointing gesture is not used to reinforce the focus in the utterance, but instead, it appears to designate the whole visual target. It is the object that shapes the pattern of gesture/speech coordination, and not the informational structure (as in Roustan, 2012). Finally, the second experiment shows that the final point of the gesture is protracted until the end of the sentence.

In sum this pilot experimental study provides promising results showing that, although they are slower and they anticipate less than adults, 7-12 year-old children use the same pattern of speech/gesture coordination as adults. This study will be pursued with more subjects and more age group, in order to understand when and how children adopt this adult-like pattern.

References

de Ruiter, J. (1998), *Gesture and speech production*, PhD thesis. Netherlands: Catholic University of Nijmegen, 1998.

Iverson, J. & Thelen, E. (1999), Hand, mouth and brain: The dynamic emergence of speech and gesture, *Journal of Consciousness Studies*, vol. 6(11-12), p. 19–40.

Levelt, W., Richardson, G. La Heij, W. (1985), Pointing and Voicing in Deictic Expressions, *Journal of Memory and Language*, 24.2 (1985), p. 133–164.

Lock, A., Young, A.W., Service, V and Chandler, P (1990), The origins of infant pointing gestures, in V. Volterra and C. Erting (Eds.) *From gesture to language in hearing and deaf children*. New York: Springer-Verlag. Pp. 42-55

McNeill, D. (1992), *Hand and Mind : What Gestures Reveal about Thought*, University Of Chicago Press.

Özçalışkan Ş, Goldin-Meadow S. (2005), Gesture is at the cutting edge of early language development, *Cognition*, 96(3): B101–B113.

Pechmann, T. et W. Deutsch (1982), The development of verbal and nonverbal devices for reference, *Journal of Experimental Child Psychology*, vol. 34, p. 330–341.

Roustan, B. 2012, *Étude de la coordination gestes manuels/parole dans le cadre de la désignation*, PhD thesis, Université de Grenoble.