



## The effects of pitch accentuation and beat gestures on information recall in contrastive discourse

Olga Kushch<sup>1</sup>, Pilar Prieto<sup>2, 1</sup>

<sup>1</sup> Department of Translation and Language Sciences, Universitat Pompeu Fabra,  
Spain

<sup>2</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA)

olga.kushch@upf.edu, pilar.prieto@upf.edu

### Abstract

Research in audiovisual prosody has shown that typically beat gestures are temporally integrated with prominent positions in speech (e.g., [1, 2]). There is independent evidence that both prosodic prominence (e.g., pitch accents) and gestural prominence associated with words (e.g., beat gestures) facilitate the recall of information (e.g., [3, 4, 5]). However, previous studies did not directly compare the beneficial effects of pitch accentuation without beats with pitch accentuation with beats. This study investigates the role of prosodic prominence (pitch accents) and gesture prominence (beat gestures) on the recall of contrastive information in natural discourse. Twenty Catalan-dominant native speakers were asked to watch 48 short videotaped discourses each containing two contrast sets with two items (e.g., *The fish shop and the grocery shop*). The critical word in the sequence was presented under two experimental conditions: 1) accompanied by prosodic prominence (L+H\* pitch accent); and 2) accompanied by prosodic prominence and gestural prominence (L+H\* pitch accent + beat). The results of the recall task revealed that the presence of beat gestures associated with prosodic prominence favored word recall of contrastive information in discourse in comparison with the condition without beat gestures.

**Index Terms:** prosodic prominence, gestural prominence, pitch accents, beat gestures, contrastive discourse.

### 1. Introduction

Human speech is seldom monotonous. When we speak, we naturally attempt to focus the listener's attention on the most important parts of our discourse, and we can achieve this goal by instantiating prominence patterns in two complementary domains, speech and gesture. Prosodic prominence is one of the strategies languages use to focus information in a discourse. Recent research in audiovisual prosody has shown that prosodic prominence is typically associated with more prominent gestural and articulatory features (e.g., beat gestures, head nods, eyebrow movements, exaggerated articulation, etc.; see, for example, [6, 7]). Beat gestures are rhythmic hand and arm movements that are typically associated with prominent prosodic positions in speech [1]. Not surprisingly, McNeill [1] noted that beat gestures were the most frequent types of gesture found in political discourse and in narratives.

Yasinnik, Renwick, and Shattuck-Hufnagel's [8] study showed that during a narration in more than 90% of instances the gesture apex occurred together with a pitch-accented

syllable. Moreover, a variety of studies have shown that there is a strong temporal connection between pitch accentuation and beat gestures (see [2] for a review). For example, Kraemer and Swerts [9] found that prosodic prominence and visual information work in a complementary fashion and that adding visual beats to prosodic prominence led to stronger production and perception of prosodic prominence.

In spite of the tight integration between prosodic prominence and beat gestures, to our knowledge the impact of these two domains on information recall is not well known. Independent evidence has shown that there is a positive effect of prosodic prominence on both the ability to recall novel information and on the comprehension of information (e.g., [3, 10]). During the investigation of the effects of pitch accent on memory, Fraundorf et al. [3] asked participants to listen to short recorded stories that included two contrasted items, specified in the continuation of the story. In the continuation, the pitch accent of the target item was manipulated so that it ranged from non-contrastive (H\*) to contrastive pitch accentuation (L+H\*). The results of a recognition memory test conducted after the training session showed that the contrastive L+H\* pitch accentuation triggered a facilitation effect on the recall of contrastive focus.

With respect to the mnemonic effect of beat gestures, we find some contradictory results in the literature. Feyereisen [11] argued that beat gestures might not enhance memory recall. In his experiment, he examined the mnemonic effect of meaningful gestures (i.e., representational or iconic gestures<sup>1</sup>) vs. nonmeaningful gestures (i.e., a group of gestures comprising both beat gestures and indefinite gestures which were hard to identify), and detected no effect of nonmeaningful gestures on memory. However, since indefinite gestures and beat gestures were both grouped together as nonrepresentational gestures, it is not possible to analyze their effects separately. A more recent study by So et al. [5] aimed to find out by means of two experiments whether beat gestures would improve word recall in adults and children. In the first experiment, 30 adults were exposed to three different lists of verbs in three conditions (iconic gestures, beat gestures, and no gesture). They were exposed to 10 words in total. In each condition a participant watched a video on a computer screen and then had to recall as many words as possible. The results

<sup>1</sup> According to McNeill [1], gestures can be classified into two large groups, representational gestures (e.g., gestures which represent speech content) and beats (e.g., hand movements produced along with the rhythmical pulsation of speech). Iconic gestures (a form of gesture which bears a close relation to the semantic content of speech) constitute a subset of representational gestures.

of the experiment showed that both iconic and beat gestures enhanced word recall in adults. In the second experiment, the same procedure was applied to children aged 4-5 years. The number of words was reduced from 10 to 5 to accommodate the shorter mnemonic span of the children. The results showed that iconic gestures enhanced memory recall in the children, but no effect of beat gestures was detected. In short, the study proved that beat gestures enhance memory recall in adults but not in children. However, we must note that So et al.'s [5] words accompanied by beat gestures were presented as a sequence of isolated words and not within a discourse context, whereas the crucial property of beat gestures is precisely to highlight the most prominent part of a discourse. Importantly, the two studies mentioned above did not control for the potential effects of prosodic prominence. The specific influence that gesture prominence encoded in beat gestures (as differentiated from prosodic prominence) may have on the encoding of information by listeners has therefore remained largely unknown.

The aim of the current study is to investigate whether naturally produced beat gestures (accompanied by focal pitch accent in speech) have a positive influence on information memorization. In the current study we will present the target items within a contrastive discourse under two experimental conditions: 1) prominence in both speech (pitch accents) and gesture (beat gestures); and 2) prominence in speech alone. Following the study by Krahmer and Swerts [9], who detected that adding the gestural prominence to prosodic prominence enhances prominence perception, we hypothesize that naturally produced beat gestures would be more beneficial for later information retrieval than just prosodic prominence.

## 2. Method

### 2.1. Participants

Twenty Catalan-dominant native speaking students (mean age = 20.5 years, SD = 2.327) from the Universitat Pompeu Fabra participated in the study. Participants were asked to complete a language questionnaire and provided written informed consent to process their data. Participants were each financially remunerated with 5 euros.

### 2.2. Materials

The materials used in the experiment were an adaptation of the materials used by Fraundorf et al [3]. They consisted of 48 video recorded discourses in Catalan containing two parts. The first part was a context passage (1), which establishes two contrast sets containing two items. The context passage was followed by the second part of the discourse, that is a continuation pattern (2), which mentions only one member of each contrast set. The contrast sets of items differed in their grammatical and thematic roles across stories.

(1) *Context passage*: Ahir l'Esmeralda havia planejat anar a la peixateria i a la fruiteria. No obstant això, quan va sortir de casa va veure que la tempesta havia deixat molts carrers coberts de neu i gel.

Translation: Esmeralda was planning to go to the fish and grocery shops yesterday. However, when she left her house, she saw a storm had left many of the streets covered with snow and ice.

(2) *Continuation passage*: L'Esmeralda va aconseguir anar a (la peixateria / la fruiteria), però com que l'altre lloc estava cobert (de neu / de gel), va decidir tornar-hi un altre dia.

Translation: Esmeralda managed to get to the (fish / grocery) shop, but since the other shop was covered with (snow / ice) she decided to go back another day.

The two paired items (e.g., *peixateria / fruiteria* and *neu / gel*) in the continuation passage were presented under two experimental conditions, prominence in both speech and gesture (Condition 1), and prominence in speech alone (Condition 2), such that either the item from the first contrastive pair was assigned as target and presented under Condition 1 or Condition 2 (i.e., accompanied by L+H\* pitch accent and beat gesture or L+H\* pitch accent only) and the item from the second contrastive pair was unaccented, or vice versa, yielding a total of four possible configurations, with target items counterbalanced according to whether they were the first or second item within each pair and whether they were the first or second pair to appear in the sentence.

The beat gestures used in the present study were bimanual palm-up hand movements. The apex of each beat gesture coincided with a pitch-accented syllable of the target item. Figures 1 and 2 represent sample stimuli of the target word *neu* 'snow' inside the utterance "però com que l'altre lloc estava cobert de *neu*, va decidir tornar-hi un altre dia" (see (1)) under the two experimental conditions.

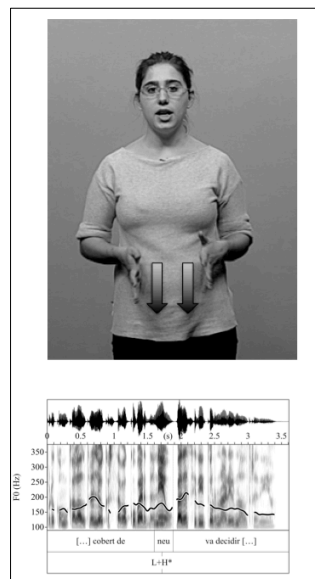


Figure 1: Video still showing speaker making beat gesture while emphasizing the word "neu", with wave form and F0 contours, i.e., Condition 1 (prominence in both speech and gesture).



Figure 2: Video still showing speaker emphasizing the word "neu" without accompanying beat gesture, with wave form and F0 contours, i.e., Condition 2 (prominence in speech alone).

During the preparation of the stimuli special attention was paid to the speaker's control over prosodic prominence. A Catalan native speaker was trained to produce the discourse and target items with the same prosody in both experimental conditions. Table 1 shows that the acoustic measures of

duration and pitch range over the critical word and critical syllables were comparable across conditions.

Table 1: Mean acoustic measures by accent type for experiment materials.

Measure	L+H* (no gesture)		L+ H* (Beat gesture)		F	Sig.
	Mean	SD	Mean	SD		
Critical word (ms)	0.545	0.134	0.614	0.170	1.721	0.195
Stressed syllable duration (ms)	0.236	0.124	0.227	0.990	1.185	0.281
Stressed syllable pitch range (Hz)	36.721	15.511	39.857	20.699	0.669	0.417

### 2.3 Procedure

Each participant was tested individually. They were instructed to listen to 48 video recorded stories and told that they would afterwards be tested on their recall of the stories. The format of the memory test was not specified prior to the listening session. During the listening phase, participants listened to all 48 stories, each story containing one target item. Target items were presented in two within-subject experimental conditions, 24 stories for each condition, half of them chosen from the first contrastive pair and half of them chosen from the second. The order of the discourses was randomized within subjects.

In the test phase, participants were given written transcripts of the 48 stories they had previously heard, in the same order that they had heard them. However, in the written version, the target and unaccented words in the continuation passage were replaced by blanks, as illustrated in (3).

(3) Ahir l'Esmeralda havia planejat anar a la peixateria i a la fruiteria. No obstant això, quan va sortir de casa va veure que la tempesta havia deixat molts carrers coberts de neu i gel. L'Esmeralda va aconseguir anar a \_\_\_\_\_, però com que l'altre lloc estava cobert de \_\_\_\_\_, va decidir tornar-hi un altre dia.

Translation: Esmeralda was planning to go to the fish and grocery shops yesterday. However, when she left her house, she saw a storm had left many of the streets covered with snow and ice. Esmeralda managed to get to the \_\_\_\_\_, but since the other shop was covered with \_\_\_\_\_ she decided to go back another day.

Participants were asked to fill in the blanks according to what they remembered hearing. The total time of the listening and testing sessions was approximately 30 minutes for each participant.

## 3. Results

Responses by subjects for the recall tests were first coded as '0' or '1', with 1 indicating that they had recalled the correct word from the contrastive discourse and '0' that they had not. All such responses obtained (960 trials in total) were submitted to a Generalized Linear Mixed Model (GLMM), using IBM SPSS Statistics 23. The dependent variable was Response (1-right reply; 0-wrong reply). The fixed factors were CONDITION (two levels: prominence in both speech and gesture, and prominence in speech alone), ITEMPOSITION (two levels: first position and second position), as well as their

interactions. PARTICIPANT, ITEM, and DISCOURSE were set as random factors.

The GLMM results revealed a significant main effect of CONDITION ( $F(1,956) = 10.308$ ,  $p < .05$ ), confirming that participants performed better in the recall task when the critical word was accompanied by prominence in both speech and gesture. Interestingly, a main effect of ITEMPOSITION was found ( $F(1,956) = 8.731$ ,  $p < .05$ ), showing that the first position induced a different pattern of recall than the second position. Pairwise comparisons revealed that on the one hand, there was a significant effect of ITEMPOSITION (second position > first position) inside prosody ( $F(1,956) = 10.434$ ,  $p < .05$ ), showing that target items accompanied by only prosodic prominence were remembered better when they were placed towards the end of a discourse. On the other hand, in the first item position a significant difference of CONDITION was found (prominence in both speech and gesture > prominence in speech) ( $F(1,956)=11.423$ ,  $p < .05$ ), showing that beat gestures significantly improved word recall in the first item position.

Figure 3 shows the mean proportion of recalled words across the two item positions (first and second) separated by the two prominence conditions: 1) prominence in both speech and gesture and 2) prominence in speech alone.

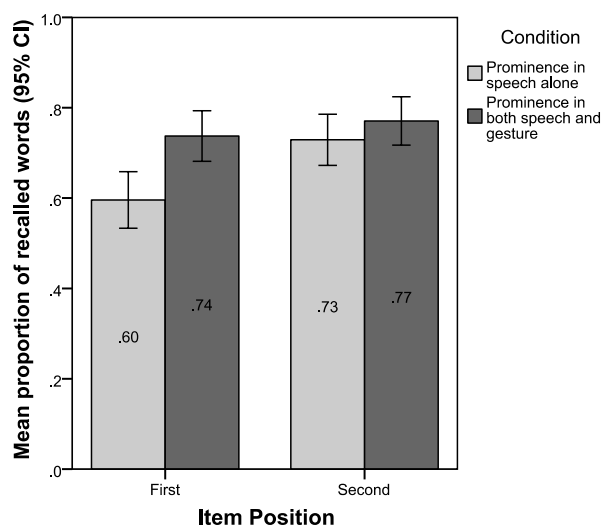


Figure 3: Mean proportion of recalled words across the two experimental conditions in the first and second target item positions.

## 4. Discussion and conclusions

The aim of our study was to investigate the potential positive effects of gestural prominence encoded in beat gestures on information memorization during a contrastive discourse, controlling for the effects of pitch accentuation. We wanted to test whether adding a beat gesture to a prosodically accented item would enhance memorization of the target word. While previous research in the field of gesture has centered on representational gestures and their role in information memorization in a first or second language (e.g., [12, 13]), little is known about the potential effects of naturally produced beat gestures. Previous studies investigating the effects of beat gestures on information recall (e.g., [5], [11]) did not control for the potential effects of prosodic prominence. Our study is the first one to test whether naturally produced beat gestures

accompanying prosodic prominence are beneficial in the context of information recall in one's native language.

The results of our study reveal that beat gestures produced together with prosodic prominence significantly enhance information recall in comparison to prosodic prominence. These results are in line with Krahmer and Swerts's study [9], which found that adding visual beats to pitch accentuation leads to a stronger perception of prominence. We thus interpret our findings to mean that the beat gesture condition leads to a stronger perception of prominence, which in turn induces higher recall rates. The study conducted by Hubbard et al. [14] seems to back up this interpretation. This study investigated whether the presence of beat gestures impacted speech perception at the neural level. Thirteen participants underwent an fMRI while being exposed to videos with spontaneously-produced speech accompanied by either beat gestures, nonsense hand movements, or a still body. The bilateral non-primary auditory cortex showed greater activity when speech was accompanied by beat gestures than when speech was presented alone. Additionally, the left superior temporal gyrus/sulcus revealed stronger activity when speech was presented together with beat gestures in comparison to when the speech was presented with nonsense hand movements. Thus, Hubbard et al.'s [14] results show that beat gestures may help focus the viewer's attention on speech prosody, and in this way increase intelligibility and comprehension. The same idea is confirmed by Biau and Soto-Faraco's [15] study, which aimed to investigate the time course of beat-speech integration in speech perception. In this study, participants were asked to watch a discourse by a Spanish politician producing a wide range of beat gestures, in two conditions, audio-visual and audio only. By measuring participants' ERPs, the researchers found a positive shift in ERPs at an early sensory stage in the audiovisual condition. There was no ERP difference when the same words were heard without viewing the video. The results of this study thus suggest that beat gestures are integrated with speech early on in time and that they modulate word-evoked potentials in early stages of speech processing. The results also support the idea that beat gestures can serve as a highlighter, as they help the listener to direct his or her focus of attention.

Moreover, in our study a positive effect of beat gestures was stronger in the first item position, where the general word recall was lower. Interestingly, in Fraundorf et al. [3] no significant difference was detected in item position. We suggest that a possible explanation of our findings could be related to the fact that contrastive focus in Catalan is typically expressed by focus-fronting, that is, by moving the contrastive focus material to the beginning of the sentence, which is a privileged syntactic position for expressing prominence (e.g., [17]). By contrast, in English contrastive focus is typically expressed in situ.

All in all, our results show that naturally produced beat gestures have a beneficial effect on information recall within a discourse. This confirms previous research in the field that shows that beat gestures aid word recall in both adults [5] and children [4] and that their effects are added to the ones induced by prosodic prominence. Nonetheless, more work is needed on the effects of gestural prominence independent from speech prominence and the question of whether they form part of an integrated system.

## 5. Acknowledgments

We would like to thank the students from the Universitat Pompeu Fabra who participated in the experiment. We are grateful to Judith Llanes for her assistance with the recording of the stimuli. We are also thankful to the technical support team at the Universitat Pompeu Fabra, especially to Ignasi Jacobo, who provided the technical assistance during the stimuli recording. We would also like to thank Santiago González-Fuente, Alfonso Igualada, and Joan Borrás for their support and assessment during the preparation of the study. Finally, this research would not have been possible without funding from the Spanish Ministry of Science and Innovation grant FFI2012-31995 ("Gestures, prosody and linguistic structure"), and a grant awarded by the Generalitat de Catalunya (2014SGR-925) to the *Prosodic Studies Group*. The first author also acknowledges a grant awarded by the Generalitat de Catalunya AGAUR (2015 FI\_B 00094).

## 6. References

- [1] D. McNeill, "*Hand and mind: What gestures reveal about thought*," Chicago: University of Chicago Press.
- [2] N. Esteve-Gibert and P. Prieto, "Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 850, pp. 850–865, 2013.
- [3] S.H. Fraundorf, D. G. Watson, and A. S. Benjamin, "Recognition memory reveals just how contrastive accenting really is," *Journal of Memory and Language*, vol. 63, no. 3, pp. 367–386, 2010.
- [4] A. Igualada, N. Esteve-Gibert, and P. Prieto, "Does the presence of beat gestures help children recall information?" Oral presentation at *Laboratory Approaches to Romance Phonology VII*, Aix-en-Provence, September 3-5, 2014.
- [5] W. C. So, C. Sim Chen-Hui, and J. Low Wei-Shan, "Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?" *Language and Cognitive Processes*, vol. 5, pp. 665-681, 2012.
- [6] M. Swerts and E. Krahmer, "Facial expression and prosodic prominence: Effects of modality and facial area," *Journal of Phonetics*, vol. 36 no. 2, pp. 219–238, 2008.
- [7] P. Prieto, C. Puglesi, J. Borrás-Comes, E. Arroyo, and J. Blat, J, "Exploring the contribution of prosody and gesture to the perception of focus using an animated agent," *Journal of Phonetics*, vol. 49, pp. 41-54, 2015.
- [8] Y. Yasinnik, M. Renwick, and S. Shattuck-Hufnagel, "The timing of speech-accompanying gestures with respect to prosody," *Proceedings of From Sound to Sense*, MIT, MIT, 2004.
- [9] E. Krahmer and M. Swerts, M, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *Journal of Memory and Language*, vol. 57, no. 3, pp. 396–414, 2007.
- [10] J. K. Bock and J. R. Mazzella, "Intonational marking of given and new information: some consequences for comprehension," *Memory and Cognition*, vol. 11, no 1, pp. 64–76, 1983.
- [11] P. Feyereisen, "How could gesture facilitate lexical access?" *Advances in Speech-Language Pathology*, vol. 8, pp. 128-133, 2006.
- [12] S. Goldin-Meadow, *Hearing gesture: How our hands help us think*, Cambridge, MA: Harvard University Press.
- [13] L. Q. Quinn-Allen, "The effects of emblematic gestures on the development and access of mental representations of French expressions," *The Modern Language Journal*, vol. 79 pp. 521–529, 1995.
- [14] A. L. Hubbard, S. M. Wilson, D. E. Callan, and M. Dapretto, "Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception," *Human Brain Mapping*, vol. 30, no. 3, pp. 1028–1037, 2009.
- [15] E. Biau and S. Soto-Faraco, "Beat gestures modulate auditory integration in speech perception," *Brain and Language*, vol. 124, no. 2, pp.143–152, 2013.
- [16] E. Vallduvi, "The role of plasticity in the association of focus and prominence," *ESCOL 7* pp. 295-306, 1991.