Temporal alignment of gesture and speech in Catalan-babbling children

Núria Esteve-Gibert [1], Pilar Prieto [1,2]

[1] Group of Prosodic Studies, University Pompeu Fabra, Barcelona, Spain
[2] ICREA, Barcelona, Spain
nuria.esteve@upf.edu, pilar.prieto@upf.edu

Abstract

This study aims at investigating the nature of intentional communicative acts where gesture and speech occur together at the babbling stage in language development. Results suggest that as early as the babbling stage, communicative acts combining gesture and speech are produced more often than gesture-only acts, and that adequate alignment is shown to emerge in the late babbling and early one-word period.
Index Terms: gesture, pointing, alignment, prosody, language acquisition

1. Introduction

In recent years, several studies have worked on the relationship between gesture and speech in communicative acts, showing that both are tightly related [14,6,7,11,12] and that they might form a single system. Five main reasons are given to justify this: that (1) gestures occur with speech in 90% of cases, (2) gesture and speech develop together in children, (3) gesture and speech are phonologically synchronous, (4) gesture and speech are semantically and pragmatically co-expressive, and that (5) gesture and speech break together in aphasia [14]. Research in this field has focused on the third issue, i.e. the temporal or phonological synchronization between gesture and intonation in adult communication. It is suggested, for instance, that body and pitch movements act in parallel in a 'metaphor of up and down' [3], that there is an alignment of gesture hierarchy with prosodic hierarchy [9], and that the stroke of the gesture precedes or ends at, but does not follow, the stressed syllable of speech [9]. Other studies observed an alignment between gestural phases and boundary tones, and between beats and nuclear tones [13], and that apexes of gestures were aligned with pitch accents [12].

Yet, whereas in adult communication most gestures were found to occur during speech, it is not until the end of the one-word period that children primarily use gestures in combination with speech for communicative purposes [4]. In [4], the authors found that it is not until the end of the one-word period that gesture and speech were temporally synchronized, i.e. that gestural strokes (or peaks of effort in gestures) occurred during or after the stressed syllable of the word they produced.

Before their first words, children are able to communicate by means of gestures and vocalizations which are sometimes combined, and they are also able to demonstrate intentionality [1,8,17]. Yet, as far as we know, there are no studies focusing on the temporal relationship between prosodic and gestural patterns in the babbling period, nor on the precise [temporal alignment between intonation peaks and gestural strokes.

The aim of this study is three-fold: first, to investigate how co-speech gestures develop during the babbling stage; second, to study how children at the babbling stage temporally align deictic gestures and speech; and third, to compare children's results on temporal alignment with adult Catalan-speakers' results. Two studies were carried out to this aim: first, a pilot study with adults to compare the results obtained with data from children, and second, an analysis of a longitudinal corpus of spontaneous speech of four Catalan-babbling children.

2. Study 1 (pilot)

2.1. Method

Participants. Five adult Catalan-speakers were audio-visually recorded using a Panasonic HD AVCCAM. In the final study, 20 Central Catalan-speakers are going to be recorded. All speakers are right-handed.

Procedure. Following [15], each participant was seated in a chair that was approximately 50 cm far from the screen. A target (smiley symbol) and a word were projected on the screen at the same time (*see* figure 1). Participants did a pointing-naming task in which they had to point at the smiley face while reading the word at the precise moment when the smiley face turned from red into green. The target words had a different number of syllables and stress position: /pá/, /papá/, /pápa/, /tá/, /tatá/, /táta/, /má/, /mamá/, /máma/, /ná/, /naná/, and /nána/.

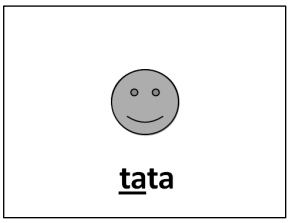

Figure 1: Example of an image projected on the screen during the pointing-naming task.

Coding. 240 instances of deictic gesture-speech combinations were acoustically and gesturally analyzed. As for prosodic analysis, the pitch accent and the pitch peak point in the fundamental frequency line were annotated using Praat software package [2], and their position was then imported to ELAN [10]; as for gestural analysis, phases (preparation, pre-stroke hold, stroke, post-stroke hold, and retraction) were coded using ELAN. Figure 2 shows an example of the gesture phases in the pointing-naming task and the approximate position of the pitch peak in the pointing gesture-speech combination; figure 3 shows an annotation example of the pointing-naming task in ELAN. In this figure, the top left panel displays the video image, the top right panel displays the controls, and the bottom panel displays the ties where prosodic and gestural cues are annotated, namely gesture form (i.e. pointing, shaking arms, etc.), gesture function (i.e. deictic,

iconic), gesture phases, transcription of target word, temporal position of the pitch accent, and temporal position of the pitch peak.

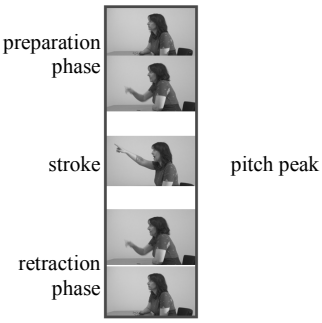

Figure 2. Example of the gesture phases and the position of the pitch peak with respect to the gesture phases.

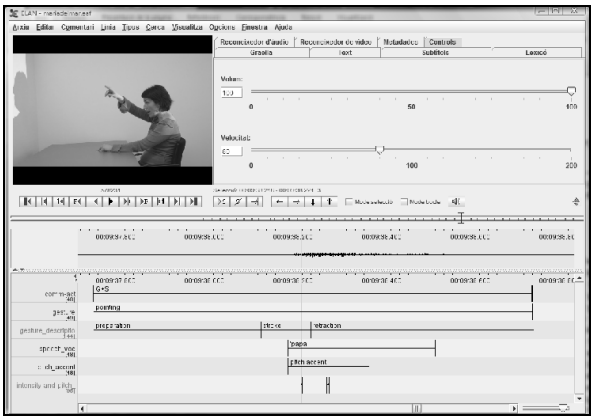

Figure 3. Example of annotation in ELAN in adult data.

2.2. Results

Results show that speakers temporally align the stroke of the pointing gesture with the pitch accent of the target word in 98% of the cases (*see* figure 4). No significant differences were found between monosyllabic and disyllabic words, or on the position of the pitch accent in disyllabic words.

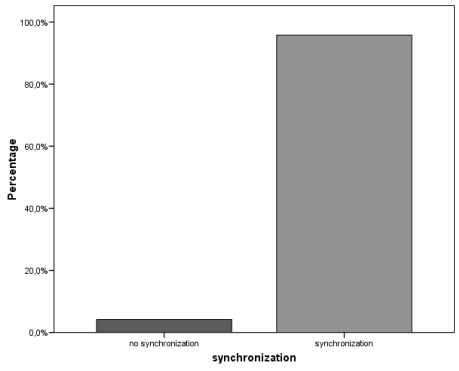

Figure 4. Percentage of synchronized and non-synchronized gesture-speech combinations.

When analyzing the position of the pitch peak in respect of the stroke of the gesture, results showed that the pitch peak tended to align with the beginning of the stroke (*see* figure 5). The precise acoustic analysis of the speech shown in figure 5 reveals that the pitch peak aligns with the beginning of the stroke, rather than the end of the stroke. This tendency is stronger in the case of the monosyllabic words.

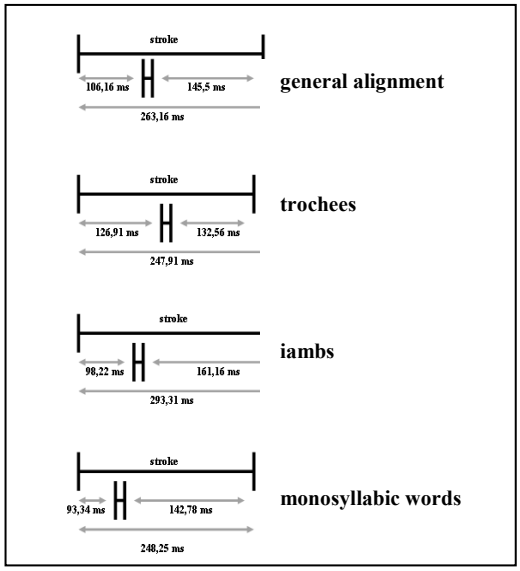

Figure 5. Alignment of the pitch peak in respect of the stroke of the pointing gesture depending on the number of syllables and the pitch accent position.

Results of the pilot study in adults confirmed the hypothesis of phonological synchronization between gesture and speech in adults, i.e. that the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech [9].

3. Study 2

3.1. Method

Participants. A longitudinal database of four Catalan-babbling infants who were recorded between ages 0;11 and 1;7 (a portion of the Esteve-Prieto corpus) was used for the analysis. All parents speak exclusively Catalan to their child and to each other, and they all live in a region where Catalan is regularly spoken by 80-90% of the population.

Procedure. All children were video-recorded at their homes during 30-minute weekly sessions using a SONY camera, model DCR-DVD202E PAL. Recordings were made by the first author of this study, who was previously acquainted with the families and the children. Children were always recorded in the same room, typically the living-room, during free play sessions. All children were recorded as they interacted with their mothers, except for one child who was recorded while interacting with both her father and her mother in most of the sessions. A tripod was used and placed as close to the child as possible with the camera pointing towards the child's face.

Recording sessions varied in length, lasting between 30 and 40 minutes, and took place weekly until children were 1;0, biweekly until children were 1;6, and monthly until children were 2;0.

Coding. All communicative acts produced by children at 0;11, 1;1, 1;3, 1;5, and 1;7 were classified as 'speech-only',

'gesture-only', or 'gesture-speech' combinations, for a total of 4,675 cases. For the purpose of this study, a communicative act is any act performed by gestural or speech means (or a combination between the two) that is used to perform some function in communication. Then, all 'gesture-speech' combinations involving a pointing gesture were analyzed in terms of prosody and gestures using ELAN [10], and Praat software package [2], for a total of 897 cases. Moreover, all meaningful words produced by children during the 30-minute recording session were also annotated to determine lexical developmental points.The goal was to investigate the temporal synchronization between the two, i.e. whether the gestural pointing strokes were aligned with the pitch peaks in vocalizations.

As for the gestural analysis, the gesture type in terms of form (shaking arms, pointing, saying bye-bye, etc.) and the gesture type in terms of function (emotive, gestural routine, deictic, or iconic) [14,16] were coded. The gesture phases of deictic gestures (preparation, pre-stroke hold, stroke, post-stroke hold, and retraction) were further coded in order to analyze the temporal synchronization. As for the prosodic analysis, the precise location of the pitch peak was annotated independently in Praat, i.e. the highest point in the fundamental frequency line. Figure 6 shows an example of annotation in ELAN. On the top left of the ELAN window, video images are displayed; on the top right, control buttons allow changes such as speed, volume, or navigation through the cases; on the bottom panel of the window, gestural and prosodic cues were annotated in different tiers as follows: gesture form (i.e. pointing, shaking arms, etc.), gesture function (i.e. deictic, iconic, gestural routing, etc.), gesture phases (preparation, stroke, retraction), utterance type (statement, request, vocative, etc.), and temporal position of the pitch peak. In addition, the waveform is displayed between the top and the bottom panels, to allow exporting the target audio portion to Praat and annotation of the prosodic information such as pitch peak position and pitch accent.

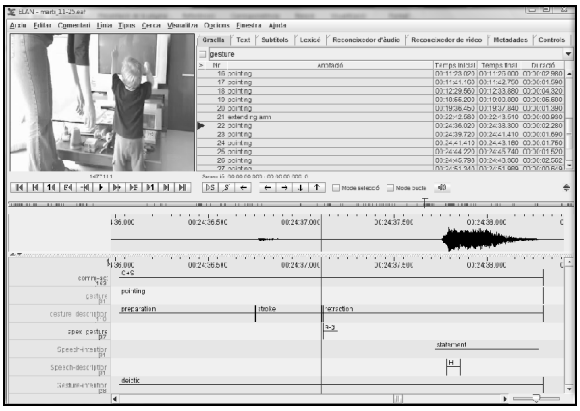

Figure 6. Example of annotation in ELAN of the children's data.

3.2. Results

Development of co-speech gestures. Results show that at 0;11, of all communicative acts containing gesture, only 43,26% were gesture-speech combinations. However, at 1;1 the gesture-speech combinations already represented 54,8% of communicative acts including gesture, and children produced an even higher proportion of gesture-speech combinations in the late babbling and one-word stages (*see* figure 7).

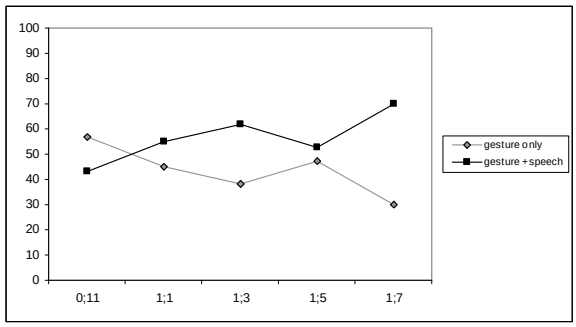

Figure 7. Development of 'gesture only' and 'gesture-speech' combinations from 0;11 to the one-word period.

Development of the temporal alignment of deictic 'gesture-speech' combinations. Figure 8 shows a summary of the temporal alignment of 'gesture-speech' combination and, specifically, the deictic gesture-speech combination. The results show that at 0;11, children do not synchronize gesture and speech yet, since the pitch peak is produced during the stroke of the gesture only in 30% of cases. At this early age, children produce most of the pitch peaks either before or after the gesture stroke. However, at the late babbling stage and early one-word periods, children already produce a higher proportion of synchronized 'gesture-speech' combinations than unsynchronized ones.

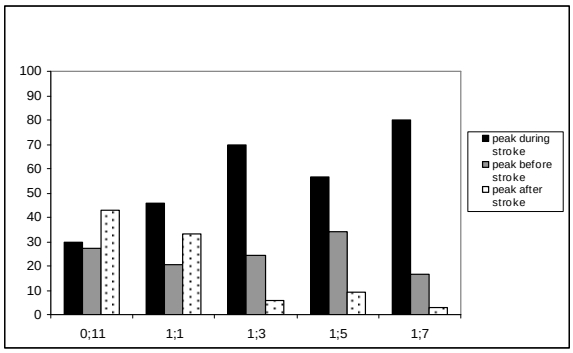

Figure 8. Evolution of the position of the pitch peak compared to the stroke. There is synchronization when the pitch peak is produced during the stroke.

Figure 9 (below) shows a more detailed analysis of the temporal alignment between the stroke of the pointing gesture and the pitch peak. The values show the mean position of the peak within stroke of the gestures and the distances (in ms) between the start and the end of the stroke and the peak pitch, at different ages of the four children. The results show that at the beginning of the babbling stage, the pitch peak tends to be aligned at the end of the stroke of the deictic gesture. However, at the late babbling stage and the one-word periods, the pitch peak has moved to the left and it is aligned at the beginning of the stroke.

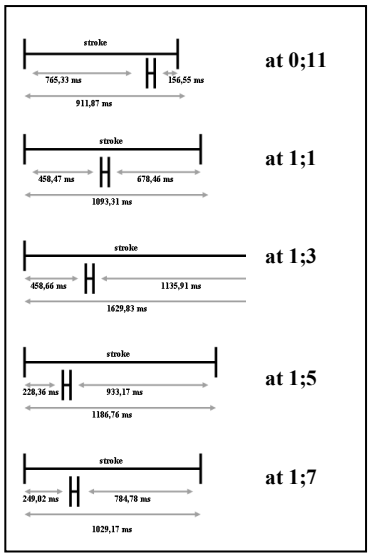

Figure 9: Longitudinal evolution of the temporal alignment of the pitch peak and the stroke. At the early babbling stage, the pitch peak is aligned at the end of the stroke, but at the one-word period it is aligned at the beginning of the stroke.

These results suggest that as early as the late babbling stage, communicative acts combining gesture and speech are produced more often than gesture-only acts. Interestingly, adequate alignment is shown to emerge in the late babbling and early one-word periods. Specifically, at the beginning of the babbling stage the pitch peak tends to be aligned at the end of the stroke, and as the child approaches the one-word period, the pitch peak is progressively aligned at the beginning of the stroke of deictic gestures.

4. Conclusions

The results of these studies suggest that children start behaving like adults in terms of gesture and speech integration earlier than predicted in other studies [4]. As for the development of co-speech gestures, it is shown that already at the late babbling stage gesture-speech combinations are produced more often than gesture-only acts. As for temporal alignment of deictic gesture-speech combinations, the adult-like alignment is shown to emerge in the late babbling and early one-word periods (from 1;1 to 1;5. Specifically, as the child approaches the one-word period, the *f*0 peak is progressively aligned at the beginning of the stroke. The alignment of the pitch peak with the beginning of the gesture stroke is also observed in the adult's results, suggesting that children adequately align gesture and speech as early as the late babbling period, and prior to successful communication by means of words.

5. Acknowledgements

We would like to thank the parents of the children recorded for agreeing to participate in this study: R. Pardo and J. Pérez, S. Cardó and T. Capitán, T. Coca and O. Tubella, and A. Gaudes and G. Parera. We would also like thank A. Bel, L. McNally, and J. I. Hualde for their valuable comments on this work. Finally, we would like to express our gratitude to J. Borràs-Comes and M. M. Vanrell for their help with the statistics. This research has been funded by projects FFI2009-07648/FILO (awarded by the Ministerio de Ciencia e Innovación) and 2009 SGR 701 (awarded by the Generalitat de Catalunya).

6. References

[1] Bates, E., Camaioni, L. & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly 21*, pp. 205-224.

[2] Boersma, P., Weenink, D. (2005). *Praat: doing phonetics by computer* (Version 4.3.01). University of Amsterdam 2005. [http://www.praat.org/].

[3] Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford.

[4] Butcher, C. & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In: D. McNeill (ed.). *Language and gesture*. New York: Cambridge University Press, pp. 235-258.

[5] Esteve-Gibert, N. & Prieto, P. (in progress). *Esteve-Prieto Corpus*. Universitat Pompeu Fabra, Barcelona.

[6] Goldin-Meadow, S. (1997). When gesture and words speak differently. *Current Directions in Psychological Science 6*, pp. 138-143.

[7] Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development 71*, pp. 231-139.

[8] Iverson, J. M. & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science 16*, pp. 367-371.

[9] Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In: M.R. Key (ed.). *The relationship of verbal and nonverbal communication*. The Hague: Mouton, pp. 207-227.

[10] Lausberg, H. & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. Behavior *Research Methods, Instruments, & Computers, 41(3)*, pp. 841-849.

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

[12] Loehr, D. (2004). Gesture and intonation. Doctoral Dissertation. Georgetown University, Washington, DC.

[13] McClave, E. (1991). Intonation and gesture. Doctoral Dissertation. Georgetown University, Washington, DC.

[14] McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

[15] Rochet-Capellan, A., Laboissière, R., Galván, A. & Schwartz, J.L. (2008). The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech and Language Hearing Research 51(6)*, pp. 1507-1521.

[16] Sansavini, B., Guarini, S. & Stefanini, C. (2010). Early development of gestures, object-related actions, word comprehension and word production, and their relationships in Italian infants. *Gesture 10(1)*, pp. 52-85.

[17] Tomasello, M., Carpenter, M. & Liskowsky, U. (2007). A New Look at Infant Pointing. *Child Development 78*, pp. 705-722.