Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 57 (2014) 301–316

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Infants temporally coordinate gesture-speech combinations before they produce their first words

Núria Esteve-Gibert [a,*,1], Pilar Prieto [b,a]

[a] Dpt. of Translation and Language Sciences, Universitat Pompeu Fabra, Spain
[b] Institució Catalana de Recerca i Estudis Avançats, ICREA, Spain

Available online 14 June 2013

**Abstract**

This study explores the patterns of gesture and speech combinations from the babbling period to the one-word stage and the temporal alignment between the two modalities. The communicative acts of four Catalan children at 0;11, 1;1, 1;3, 1;5, and 1;7 were gesturally and acoustically analyzed. Results from the analysis of a total of 4,507 communicative acts extracted from approximately 24 h of at-home recordings showed that (1) from the early single-word period onwards gesture starts being produced mainly in combination with speech rather than as a gesture-only act; (2) in these early gesture-speech combinations most of the gestures are deictic gestures (pointing and reaching gestures) with a declarative communicative purpose; and (3) there is evidence of temporal coordination between gesture and speech already at the babbling stage because gestures start before the vocalizations associated with them, the stroke onset coincides with the onset of the prominent syllable in speech, and the gesture apex is produced before the end of the accented syllable. These results suggest that during the transition between the babbling stage and single-word period infants start combining deictic gestures and speech and, when combined, the two modalities are temporally coordinated.


*Keywords:* Early gestures; Early acquisition of multimodality; Early gesture-speech temporal coordination

## 1. Introduction

There is a broad consensus in the literature on the tight relationship and mutual influence between gesture and speech. Many researchers have stated that gesture and speech form an integrated system in communication (De Ruiter, 2000; Kendon, 1980; Kita, 2000; McNeill, 1992). Important features that back up the speech-gesture integration analysis in adults are that most of the gestures are produced together with speech, and that the two modalities are (a) semantically and pragmatically coherent, and (b) temporally synchronized, i.e., the most prominent part of the gesture is temporally integrated with speech (McNeill, 1992).

Studies on the temporal coordination of gesture and speech provide strong evidence for the claim that gesture and speech form an integrated system in adults. It has been shown that the most prominent part of the gesture typically co-occurs with the most prominent part of the speech (Kendon, 1980). But different anchoring regions in speech have been proposed to serve as coordination sites for gesture prominence locations: speech onset (Bergmann et al., 2011; Butterworth and Beattie, 1978; Ferré, 2010; Levelt et al., 1985; Roustan and Dohen, 2010), prosodically prominent syllables (Krahmer and Swerts, 2007; Loehr, 2007; Leonard and Cummins, 2010), or prosodically prominent syllables with intonation peaks (De Ruiter, 1998; Esteve-Gibert and Prieto, 2013; Nobe, 1996). Taking together these findings, there is general agreement in the literature that (a) prominences in gesture and speech are temporally coordinated, (b) the prominence in gesture is represented by the gesture stroke (in the case of a deictic gesture, the interval of time during which the arm is extended) or the

* Corresponding author. Tel.: +34 935422409; fax: +34 935421617.
*E-mail address:* nuria.esteve@upf.edu (N. Esteve-Gibert).
[1] Postal address: Roc Boronat, 138, 08018 Barcelona, Spain.

0167-6393/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2013.06.006

gesture apex (the specific point within the stroke interval at which the finger is maximally extended), and (c) the prominence in speech is represented by the prosodically prominent syllable. In our study, these measures will be taken into account in order to investigate the development of the temporal coordination of gesture with speech in the early stages in language development.

But are gestures coordinated with speech in young infants to form an integrated system the way that they are in adults? Iverson and Thelen (1999) and Masataka (2003) state that speech and gesture coordinations have their developmental origins in early hand-mouth linkages. Based on the dynamic systems theory of motor control, they propose that systems activating mouth and arms can influence and entrain one another, and these entrainments are dynamic and flexible such that activation in one system can affect the other in the form of a looser or tighter temporal synchrony. However, a given behavior must be strong and stable (with low threshold and high activation) to pull in and entrain the activity of the complementary system. Iverson and Thelen (1999) propose four developmental stages of dynamic progression, namely, (1) in newborns, an early oral-manual system in which instances of hand-mouth contact and co-occurrences of hand movements with vocalizations are frequent; (2) from 6 to 8 months, rhythmical movements with the hands and mouth showing an increasing control over the manual and oral articulators, and possibly indicating the transition into the speech-gesture system; (3) from 9 to 14 months, a more articulated control over the two modalities, which are then more directed to communication, with the gesture modality predominating but with entrainment also occurring between the two, and showing a tight relation between early gesture production and a later language development, and (4) from 16 to 18 months, a tighter control over both modalities, leading to the emergence of synchronous gesture and speech combinations.

In order to investigate deeply the temporal overlap between the occurrence of vocalizations and rhythmic activities of the limbs in infants, Ejiri and Masataka (2001) investigated the vocal and motor behavior of 4 Japanese infants from 0;6 to 0;11. In a first study, they examined the temporal overlap between vocalizations and rhythmic activities during the babbling stage. The authors found that vocalizations very frequently co-occurred with rhythmic actions, and interestingly that these coordinated behaviors increased immediately before and during the month in which canonical babbling was initiated. In a second study, they compared vocalizations co-occurring with rhythmic actions to vocalizations not co-occurring with rhythmic actions, and they found that syllable and formant frequency durations were shorter in vocalizations co-occurring with rhythmic actions than in non-co-occurring ones.

Similarly, Iverson and Fagan (2004) described early infants' production of vocal-rhythmic movement coordination by testing 47 infants between the ages of 0;6 and 0;9. Results showed that at 0;7 vocal-motor coordination was a stable component of infants' behavioral repertoires, and that these early coordinations were a developmental precursor to the gesture-speech system. The authors based this statement on three observations: (1) infants at all ages coordinated vocalizations with single-arm rhythmic movements more often than with both-arm movements; (2) at all ages the proportion of coordinated right-arm movements was higher than that of left-arm movements, paralleling adult-like behaviors; and (3) most of the coordinations followed the temporal patterns of organizing gesture-speech productions, since motor activities were synchronous with or slightly anticipated vocalization onsets.

The abovementioned studies focusing on rhythmic movements revealed that vocal and motor rhythmic movements are precursors of the coordination between the gesture and speech modalities. However, very few studies have investigated the patterns of that early coordination itself, i.e., the early synchronization between communicative gestures and vocalizations. To our knowledge, only Butcher and Goldin-Meadow (2000) have performed such a study. The authors analyzed six children longitudinally in spontaneous play situations during the transition from one- to two-word speech in order to investigate whether (1) at that age children produce gestures with or without speech, (2) children temporally coordinate gesture and speech, and (3) children semantically integrate the two modalities. First, they found that the production of utterances containing gesture remained stable across the ages analyzed, but with a difference between age groups: at the beginning of the single-word period gestures were generally not accompanied by speech, and at the end of the single-word period children mainly combined them with speech. Second, they found that it was not until the beginning of the two-word period that children produced gesture-speech combinations in which the speech co-occurred with the most prominent part of the gesture (defined by them as the stroke or peak of the gesture, i.e., the farthest extension before the hand began to retract). Finally, the study showed that the proportion of gestures produced in combination with meaningful speech (as opposed to *meaningless* speech) increased across the ages analyzed. In conclusion, Butcher and Goldin-Meadow (2000) suggested that it is not until the beginning of the two-word period that children integrate gesture and speech as a single system to communicate intentionally.

The present paper aims at contributing to the research on early coordination between communicative gesture and speech by describing the emergence of gesture-speech combinations and their temporal coordination. Following up on Butcher and Goldin-Meadow (2000), we incorporate two innovative aspects in our study. The first innovative aspect is an analysis of the emergence of communicative gesture-speech combinations starting from the babbling period. The babbling period emerges in the middle of the first year of life and it is a crucial stage in language development because it provides the raw material for the production of early words (Oller et al., 1976; Vihman et al.,