



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:
This is an **author produced version** of a paper published in:

Journal of Child Language 43.4 (2016): 890-913

DOI: <http://dx.doi.org/10.1017/S0305000915000343>

Copyright: © Cambridge University Press 2015

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription

Properties of vocalization- and gesture-combinations in the transition to first words

Eva Murillo

Universidad Complutense de Madrid

Almudena Capilla

Universidad Autónoma de Madrid

Author note

Eva Murillo, Departamento de Psicología Básica II, Universidad Complutense de Madrid; Almudena Capilla, Departamento de Psicología Biológica y de la Salud, Universidad Autónoma de Madrid.

Acknowledgments: We would like to thank the staff and families from E.I. Caperucita, E.I. Los Cedros and E.I. Soletes for their invaluable cooperation. We would also like to thank Victoria Marrero for her helpful comments on a previous version of this paper. A preliminary version of this study was presented at the Tilburg Gesture Research Meeting in July 2013.

Correspondence concerning this article should be addressed to Eva Murillo, Departamento de Psicología Básica II, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 .Pozuelo de Alarcón. Madrid. Spain.

E-mail: eva.murillo@pdi.ucm.es

Abstract

Gestures and vocal elements interact from the early stages of language development, but the role of this interaction in the language learning process is not yet completely understood. The aim of this study is to explore gestural accompaniment's influence on the acoustic properties of vocalizations in the transition to first words. Eleven Spanish children aged 0;9 to 1;3 were observed longitudinally in a semi-structured play situation with an adult. Vocalizations were analyzed using several acoustic parameters based on those described by Oller et al. (2010). Results indicate that declarative vocalizations have less protosyllables than imperative ones, but only when they are produced with a gesture. Protosyllables duration and $f(0)$ are more similar to those of mature speech when produced with pointing and declarative function than when produced with reaching gestures and imperative purposes. The proportion of canonical syllables produced increases with age, but only when combined with a gesture.

Key words: gestures, language development, vocalizations, multimodality, acoustic analysis.

Introduction

Social and language development are highly influenced by both vocal and gestural components of communication. On the one hand, infants are able to produce different types of vocalizations from the second week of life (Keller & Schölmerich, 1987). From three months of age, and throughout the first year, vocalization rate increases (Camp, Burgess, Morgan & Zerbe, 1987), and qualitative changes occur contingent upon social stimulation (Bloom, Russell & Wassenberg, 1987; Masataka, 1993a,b). Subsequently, the transition from these preverbal vocalizations to words occurs in a continuous, gradual way (Hsu, Fogel & Cooper, 2000; Karousou & López-Ornat, 2013; Majorano & D'Odorico, 2011; Vihman, Ferguson & Elbert, 1986). On the other hand, the use of gestures throughout the first year, especially the pointing gesture, has predictive value for subsequent lexical development (Bates, Benigni, Bretherton, Camaioni & Volterra, 1979; Camaioni, Castelli, Longobardi & Volterra, 1991; Rowe, Özçaliskan & Goldin-Meadow, 2008), though it should be noted that this predictive value for language development only applies when the pointing gesture has a declarative or general function, and not when it has an imperative function (Colonnaesi, Stams, Koster & Noom, 2010).

Vocal and motor components not only co-develop, but also interact from the first stages of language development. This interaction is supported by the tight link between the vocal and motor components throughout the first years of life (see Iverson, 2010, for a review). According to McNeill (1992), gesture and speech can be considered as parts of a single communication system, and are linked to the same underlying thought processes. From this point of view, the developmental linkages between vocal and

motor components can be seen as the ontogenetic basis of speech and gesture coordination (Iverson & Thelen, 1999).

Coordination between vocalizations and manual configurations is present as early as 2-3 months of age: infants employ the index finger extension more frequently with syllabic vocalizations (perceived by adults as more speech-like) than with vocalic ones or without vocalizations (Fogel & Hannan, 1985; Masataka, 1995). Critically, this association is constrained to index finger extension, and does not apply to other manual actions, such as grasping. At the end of the first year, manual rhythmic movements are related to the emergence of canonical babbling, and the vocalizations accompanied by rhythmic activity have different acoustic properties from those produced without it (Ejiri & Masataka, 2001).

By the end of the first year, children start using gestures to convey meanings to others (e.g., Bates et al., 1979; Carpenter, Nagel & Tomasello, 1998). Communicative gestures like pointing or reaching become increasingly frequent in the first few months of the second year, and tend to be accompanied by vocalizations: around 70% of communicative gestures are produced with vocalizations during this stage of development (e.g., Cochet & Vauclair, 2010a; Iverson & Goldin-Meadow, 2005; Leung & Rheingold, 1981; Rowe et al., 2008). Caselli, Rinaldi, Stefanini and Volterra (2012), studying children aged 0;8 to 1;6, found that it is after 1;4 that word production exceeds gesture production in children's communicative repertoires. Gestures and speech remain strongly associated throughout the language development process, and the gestural-vocal system evolves, fulfilling new specific functions in later language learning (Colletta, Guidetti, Capirci, Cristilli, Demir, Kunene-Nicolas & Levine, 2014).

In the last decade, research has gone a step further, focusing not only on the co-occurrence of speech and gestures, but also on their interactive effect on language development and its predictive value for linguistic. In this regard, Iverson and Goldin-Meadow (2005) observed that children aged 0;10 to 1;2 relied heavily on gestures to refer to objects. At a lexical level, items appeared initially in children's gestural repertoires, emerging subsequently in their verbal lexicons. Shortly afterwards, and before they produce sentences combining words, children produce constructions based on similar structures coordinating gesture and speech (Özçaliskan & Goldin-Meadow, 2005, 2009). Thus, the age at which children start producing pointing + noun combinations predicts the onset age for determiner + noun constructions (Cartmill, Hunsicker & Goldin-Meadow, 2014). Moreover, the onset of gesture-plus-word coordination conveying two elements of a proposition (supplementary coordination) predicts the onset of two-word combinations (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005). By means of supplementary coordination, children are able to convey two different semantic pieces, that is, sentence-like meanings. Also, the employment of gesture-plus-word coordination at 1;10 predicts sentence complexity at 3;6 (Rowe & Goldin-Meadow, 2009).

Importantly, vocal and gestural coordination predicts subsequent linguistic development even when the vocal component is not yet a word. Murillo and Belinchón (2012) found that the coordinated use of gesture (specifically, pointing), vocalization and social gaze at 1;0 is a strong predictor of lexical development three months later. Wu and Gros-Louis (2014) also found that gesture-vocal coordination was related to infants' linguistic skills at 1;3. In the same line, infants who were able to integrate pointing and speech at

1;0 showed better vocabulary abilities at the age of 1;6 (Igalada, Bosch & Prieto, 2015).

Butcher and Goldin Meadow (2000) found that children temporally coordinate gestures and speech, and the two modalities are also semantically integrated, but they claimed that this synchrony or coordination begins in the transition from the one-word to the two-word stage. Similar results are reported by Capirci, Contaldo, Caselli and Volterra (2005). However, Esteve-Gibert and Prieto (2014) have expanded these results to the period of transition to first words: children temporally coordinate gestures with vocalizations already at the babbling stage. Children combine gestural and vocal elements with an adult-like pattern: gesture onset precedes speech onset, gesture stroke onset co-occurs with speech onset and stroke onset precedes the beginning of the accented syllable. It seems, thus, that before speaking their first words, children are able to synchronize gesture and prosodic cues.

A plausible explanation for the role of gestural-vocal coordination in language development has to do with adults' reaction to multimodal communication. It seems that adults' response to infants' communicative attempts is different if the infant's behavior includes gestural and vocal coordination. The coordination of gesture and intonation contour can facilitate the comprehension of the infant's intention by adults (Balog & Brentari, 2008). In fact, maternal responses to infants' pointing are associated with improvements in language skills (Wu & Gros-Louis, 2014). Children combine gestures and vocalizations according to the caregiver's attentional state and the adult's response to their communicative attempt (Gros-Louis & Wu, 2012). At the same time, caregivers respond differentially to children's communicative behavior depending on its multimodal character and on how verbal and gestural elements are combined. In this regard, Fasolo and D'Odorico (2012) showed that mothers tend to label the gesture's

referent and produce the associated function words when the child produces an isolated gesture or a gesture combined with preverbal production. However, when the child produces a complementary gesture-plus-word combination, mothers reply by producing function words followed by the imitation of the word uttered by the child. If the child's communicative behavior is a supplementary gesture-plus-word combination, mothers tend to respond with an utterance completed syntactically, expanding the child's verbal utterance with one more argument. It seems, thus, that mothers expand children's utterances by augmenting their complexity and adding predicates or new arguments. Begus, Gliga and Southgate (2014) found that infant's learning is affected by adults' information about the object they had pointed to. Similarly, mothers' sensitive response to socially-directed vocalizations from the infant contributes to the emergence of vocal usage and the shaping of vocal development (Gros-Louis, West & King 2014). Considering this, the adult's response to multimodal communicative attempts may offer a linguistic model that the child can use to give "word form" to his or her vocalizations. Only recently has the relationship between the features of vocalizations and communicative gestures been addressed. Acoustic analysis carried out by trained judges revealed that vocalizations differ depending on the gestures that accompany them and their communicative function (Murillo & Belinchón, 2013). Grünloh and Lizskowski (2014) found that children aged 1;2 vocalized differently when pointing to request than when pointing to inform, regardless of the distance to the target object or of the hand shape adopted for pointing (whole-hand vs. index finger). Regarding prosody, they found that rising intonation was linked to informative and expressive pointing, whereas requestive pointing was associated with rising and flat intonation patterns. In addition, acoustic parameters such as pitch range and duration can differentiate between communicative and investigative vocalizations from the age of 0;9. Infants

produce shorter vocalizations with a wider pitch range when they are interacting with their parents than when playing alone. In addition, infants can use particular prosodic cues to express different communicative intentions. For example, they use an expanded pitch range and longer duration when expressing discontent, a wide pitch range but short duration when expressing satisfaction, and a narrow pitch range and short duration when producing responses or statements (Esteve-Gibert & Prieto, 2013).

The analysis of acoustic parameters (e.g., duration, fundamental frequency or intensity) has been very useful for detecting not only changes in vocalizations related to language development (DePaolis, Vihman & Kunnari, 2008; Papaeliou & Trevarthen, 2006), but also differences between typically developing children and children with developmental disorders (Bonneh, Levanon, Dean-Pardo & Adini, 2011; Oller, Niyogi, Gray, Richards, Gilkerson, Xu, Yapanel & Warren, 2010). In a similar line, we believe that detailed examination of the features of vocalizations could add valuable information to existing knowledge about vocal-gesture interaction and its role in early language development.

From an embodied and developmental perspective of language learning, and considering gesture and speech as parts of the same communicative system, the interaction between vocal and gestural elements should be already present before children are able to use words. With this in mind, the aim of this study is to investigate the interaction between vocalizations and gestures in the transition to first words. Unlike the case of previous studies, we performed a thorough analysis of the acoustic parameters of vocalizations, so as to investigate whether vocal-gestural combination has an impact on vocal characteristics. Our analysis strategy was based on the parameters and categories defined by Oller et al. (2010), which have been sensitive to developmental changes and relevant in differentiating children with and without disorders such as autism or language delay. Critically, these parameters allow us to

detect “speech-related vocal islands” (SVIs), which can be considered as precursors of mature syllables. We hypothesize that vocalizations will have more similar features to mature speech (in terms of proto-syllable structure, duration and fundamental frequency) when combined with a gesture than when produced alone, and that these features will depend on the type of gesture (pointing vs. others) and pragmatic function (declarative vs. imperative).

Method

Participants

Eleven Spanish children (6 girls) were recorded every three months from 0;9 to 1;3. All of them came from monolingual Spanish-speaking homes. All but one were first born, and all came from two-parent families. They were all born from full-term uncomplicated pregnancies with normal deliveries. No hearing or developmental problems or concerns were reported by parents, and the children were achieving developmental milestones within the typical range. All the infants were attending nursery school when the data collection began. Informed consent was obtained from parents who voluntarily agreed to participate. At the end of the study a DVD with the recordings of their son or daughter was provided to parents. Observation sessions were programmed within the week of each infant’s birthday, and when this was not possible, the criterion was extended by a week. Mean age of participants and session duration are shown in Table 1. A total of 563 minutes of video was recorded.

INSERT TABLE 1 HERE

Materials and procedure

The infant was seated in a baby chair, and the primary caregiver (mother or father) was seated on the right next to the infant (see Figure 1). The experimenter positioned herself in front of the child. A lapel microphone, continuously recording at a sampling rate of 44100 Hz, was placed on the child's lapel in such a way that they couldn't see it, to avoid distraction and recording problems. Although this observation setting did not recreate a natural setting for play, it effectively served the purpose of recording vocalizations and gestures. Particularly, as the microphone was wired to the camera, it was critical to keep children seated in the chair to prevent them from walking around the room.

INSERT FIGURE 1 HERE

The experimenter showed the child a set of toys, one at a time. The child was allowed to play with the toys while s/he showed interest. The set included balloons, bubbles, a picture book, a symbolic play set with plates, glasses and spoons, a spinning top, toy cars, and a wind-up toy. The same set was used for all participants and sessions. All the toys were presented to all participants in each session, though the order of presentation was not previously established. The experimenter interacted with the children, leaving them to lead the interaction and responding to their communication attempts. The primary caregiver was asked not to elicit or provoke communicative behaviors, but was encouraged to respond in a natural way to communicative attempts from the child.

Data analysis

All the communicative behaviors addressed to the experimenter or to the caregiver were coded according to the categories showed in Table 2. We considered as communicative those behaviors that were triadic, that is, that referred to some external entity, and that included gesture, vocalization and/or look directed at the adult. Behaviors with an unclear referent or without any sign of being adult-directed (orientation towards the adult or gaze use) were not considered in our sample.

INSERT TABLE 2 HERE

Two trained observers coded samples of 6 observation sessions including different children at different ages (18% of the total recordings). Agreement between coders was 92% for gesture ($k = .90$, $N = 155$) and 87% for communicative function ($k = .79$, $N = 388$). For the present purposes, all communicative behaviors recorded which did not contain vocalization were excluded from the analyses.

Gesture analysis

We only considered manual gestures in our coding system. As can be seen in Table 2, we adapted our coding categories from previous studies on communication and language development. Gesture categories included deictic gestures (pointing and reaching), which typically appear at the end of the first year and play a crucial role in language development (e.g., Bates et al., 1979; Carpenter et al., 1998; Colonna et al., 2010).

Besides deictic gestures, there are other gestures which also seem to play a role in language development: symbolic and conventional gestures. By means of these gestures, children start using an action to represent an object. Symbolic and conventional gestures typically develop along with early words, and were therefore

included in our coding system, following Acredolo and Goodwyn's description (Acredolo & Goodwyn, 1988).

Finally, we established an "Other" category which included gestures not fitting any of the previous categories and unclear gestures.

Pragmatic function analysis

We coded the pragmatic function of each communicative behavior, whether it was a vocalization, a gesture or a multimodal (gesture + vocalization) attempt. When the goal of the behavior was to share attention with the adult about an object or event, we considered it as a declarative behavior (for example, when playing with bubbles, the child looks at his/her father and vocalizes).

When the purpose of the behavior was to obtain a change in the physical world (for example, to obtain an object, or to get the adult to carry out an action) we considered it as an imperative behavior.

We coded in the category "Other", those behaviors with an expressive or rejection function, together with behaviors that occurred after a question or request from the adult. We also coded in this category gestures and vocalizations performed as part of play routines, and those behaviors whose function was unclear. Criteria were applied in a highly conservative way, with the aim of including in declarative and imperative categories only those behaviors that clearly had these pragmatic functions.

Vocalization analysis

Infants' vocalizations were extracted from audio recordings and segmented using the Praat program (Boersma & Weenink, 2014). To consider a vocal sound as a vocalization, we followed the Bloom et al. (1987) criterion. A new vocalization was

counted as beginning after any audible inspiration or after a second or more of silence. We excluded non-voiced sounds, (sounds which do not produce a visible trace on the Praat spectrogram), cries and other vegetative sounds (such as burps or hiccups). We obtained a total of 1686 vocalizations.

Once the vocalizations were segmented, we extracted the child vocal islands (CVI) based on the Robust Algorithm for Pitch Tracking (RAPT) (Talkin, 1995), as implemented in the Voicebox toolbox for Matlab. A child vocal island was identified when the acoustic energy level rose to 90% above baseline for at least 50 ms and ended when it fell to less than 10% above baseline for at least 50 ms, but not more than 300 ms.. In general CVIs correspond to syllables with very strong differentiations of acoustic energy level between nuclei (or vowels) and margins (or consonants) (Oller et al. 2010). Given that in our sample vegetative sounds and cries were excluded, all the CVIs obtained were speech-related vocal islands (SVI), following Oller et al.'s (2010) classification. Analysis of SVIs focused on acoustic effects of rhythmic “movements” of jaw, tongue, and lips (i.e., articulation), which underlie syllabic organization, and on acoustic effects of vocal quality or “voice” (Oller et al., 2010). The SVI definition includes utterances such as babbling, pre-speech vocalizations and real speech. We obtained 2427 SVIs from 1686 vocalizations. An example of SVIs in a vocalization is shown in Figure 2. Every SVI longer than 50 ms was acoustically analyzed, extracting the parameters detailed below.

INSERT FIGURE 2 HERE

First, we obtained the number of SVIs per vocalization. Then we extracted SVI duration, which was classified following Oller et al.'s (2010) criteria, with the exception of a newly-added category size “Extra-small” (see Table 3). We also computed the

fundamental frequency $f(0)$ using the autocorrelation method implemented in the Colea toolbox for Matlab (<http://www.utdallas.edu/~loizou/speech>). We employed a 30 ms long Hamming window, updated every 20 ms. SVIs were classified according to their $f(0)$, as also shown in Table 3.

INSERT TABLE 3 HERE

Next, we computed the canonical syllables (CS) parameter described by Oller et al. (2010), which provides a critical measure of the well-formedness of the initial formant transitions of each SVI with respect to initial transitions of syllables in mature speech. Formant frequencies (F1 and F2) were tracked based on the linear predictive coefficients (LPC). As in Oller et al. (2010), an SVI was categorized as a canonical syllable if (1) the SVI's category duration was either Small or Medium, (2) the SVI was of Medium Category based on its $f(0)$, (3) the maximum slope change of F1-F2 was reached within 120 ms, and (4) up to that point, either F1 or F2 slope was higher than 3 and 5 respectively. These criteria represent an approximation to the traditional acoustic specifications for canonical syllables in the infant vocalization literature (Oller et al., 2010).

Finally, to obtain a measure of pitch direction, we took each vocalization and removed the silence between SVIs. We then computed the fundamental frequency $f(0)$ for each vocalization and detected 3 reference points on it (p1: first time point; p2: intermediate point; p3: last time point). If the maximum/minimum $f(0)$ was located in either the first or the last time point, p2 was defined as the middle time point. Otherwise, p2 was defined as the local maximum/minimum. Only vocalizations with at

least three well identified $f(0)$ points were analyzed ($N=1315$). Based on the relationship between the three reference points, pitch direction was first categorized as flat or non-flat. A vocalization was defined as flat if there were less than two semitones of difference between the minimum and the maximum $f(0)$ values. Non-flat vocalizations were further classified based on pitch direction as (1) rising ($p1 < p2 < p3$), (2) u-shaped (falling – rising; $p1 > p2 < p3$), (3) falling ($p1 > p2 > p3$), (4) inverted u-shaped (rising – falling; $p1 < p2 > p3$).

Results

Children produced 1686 vocalizations, at a rate of 2.99 vocalizations per minute. Mean vocalization rate was 2.24 at 0;9 ($SD=1.20$; Min: 1.07; Max: 4.81), 3.48 at the age of 1;0 ($SD=1.54$; Min: 1.29; Max: 6.24) and 3.65 at the age of 1;3 ($SD=2.22$; Min: 0.26; Max: 7.17).

In the next two sections we present the results of the analysis of speech-related vocal islands (SVIs) and the different parameters derived from them (i.e., number, duration, fundamental frequency, and canonical syllables parameter), together with the analysis of the vocalizations' pitch direction.

Speech-related vocal island analysis

From the 1686 vocalizations uttered by the children we obtained 2427 SVIs, of which 603 were produced at 0;9, 965 at 1;0 and 859 at 1;3. Out of all the SVIs, 30% ($n=715$) were accompanied by a gesture.

Number of vocal islands per vocalization

In order to explore the influence of gesture accompaniment and pragmatic function on the number of SVIs per vocalization, we conducted a Linear Mixed Model (LMM)

analysis. The number of vocal islands per vocalization was the dependent variable. The pragmatic function (declarative vs. imperative) and the presence of gesture (with gesture vs. without gesture) were the fixed factors; subject was a random factor.

We found a main effect of gestural accompaniment on the number of islands per vocalization ($F(1, 1443.949) = 4.63; p=.032$): vocalizations produced with a gesture had more vocal islands than vocalizations produced without one (1.484 vs. 1.354). We also found a main effect of pragmatic function ($F(1, 1435.548) = 9.762; p=.002$), whereby imperative vocalizations had more vocal islands than declarative ones (1.323 vs. 1.515). Interestingly, the results also indicate an interaction effect between gestural accompaniment and pragmatic function ($F(1, 1441.581) = 10.473; p=.001$). Pairwise comparisons using the Bonferroni correction showed that when they were produced with a gesture, declarative vocalizations had less vocal islands than imperative ones, but not when produced alone (1.291 vs. 1.677; $p<.001$) (see Figure 3).

INSERT FIGURE 3 HERE

Conducting the same analysis, these differences were not found to be present at 0;9: we found no main effect of gesture accompaniment ($F(1, 340.120) = .036; p=.849$), of the pragmatic function ($F(1, 341.998) = .508; p=.477$) or of the interaction between them ($F(1, 338.934) = .023; p=.879$). At 1;0, we found an interaction effect between gesture accompaniment and pragmatic function, ($F(1, 567.430) = 7.276; p=.007$), indicating that declarative vocalizations have fewer vocal islands than imperative ones when produced with a gesture, but not when produced alone (1.285 vs. 1.708; $p=.018$); imperative vocalizations also have more vocal islands when produced with a gesture than when produced without verbal accompaniment (1.708 vs. 1.27; $p<.001$). The results showed no main effect of gesture accompaniment ($F(1, 568.389) = 2.231; p=.136$) or pragmatic

function ($F(1,568.992)=1.801$; $p=.180$) on the number of vocal islands per vocalization. By contrast, at 1;3, we found a main effect of gesture accompaniment ($F(1,487.227)=5.149$; $p=.024$) and pragmatic function ($F(1,293,620)=6.051$; $p=.014$): vocalizations had more vocal islands when produced with a gesture (1.542 vs. 1.345; $p=.024$) and when they had an imperative function (1.553 vs. 1.335; $p=.014$). We also found an interaction effect ($F(1,524,277)=4.055$; $p=.045$), showing that imperative vocalizations had more vocal islands when produced with a gesture than when produced alone (1.738 vs. 1.367; $p<.001$). On the other hand, declarative vocalizations had fewer vocal islands than imperative ones only when they were produced with a gesture (1.346 vs. 1.738; $p=.007$).

Duration analyses

We examined whether SVI duration depends on the coordination of the vocalization with a gesture and on the pragmatic function of the communicative behavior. For this purpose, we carried out an LMM analysis, with SVI duration as dependent variable and pragmatic function (declarative, imperative) \times gestural coordination (with, without gesture) as fixed factors; subject was the random factor.

We found no main effect of gesture accompaniment ($F(1,2118.259)=.029$; $p=.865$) or of pragmatic function ($F(1,13.82)=2.093$; $p=.865$) on SVI duration. Gesture accompaniment and pragmatic function interaction was only marginally significant ($F(1,2112.914)=3.782$; $p=.052$). Declarative vocalizations were longer than imperative ones, but only when they were produced with a gesture, though these differences did not reach statistical significance.

As expected considering previous literature (e.g., Cochet & Vauclair, 2010b; Liszkowski & Tomasello, 2011), in our sample the declarative function was linked to

pointing gestures, whereas the imperative function appeared to be associated with reaching gestures ($\chi^2(4, N= 715)=401.8; p<.001$). We only found 59 conventional and symbolic gestures that are included in the category “other”. Table 4 shows gesture type distribution according to its communicative function.

INSERT TABLE 4 HERE

Considering this link between gestures and communicative functions, we explored whether gestures were associated with specific duration patterns of SVIs. In order to do so, a Chi-square test with the duration categories and gesture type was conducted ($\chi^2(8, N= 713)=30.98; p<.001$). As can be seen in Table 5, pointing gestures appeared to be associated with SVIs of Medium duration. By contrast, reaching gestures were produced with Extra-large vocalizations more often than would be expected by chance.

INSERT TABLE 5 HERE

In order to explore developmental changes in the association of pointing and reaching with verbal island duration, we grouped short (Extra-small, Small, and Medium) and long categories (Large and Extra-large). We conducted a Chi-square test with type of gesture (pointing and reaching) and SVI duration categories (short and long) at every age (0;9, 1;0, 1;3). Results showed no differences in SVI duration distributions according to type of gesture at the age of 0;9 ($\chi^2(1, N= 58) =.189; p=.664$). However, at 1;0 we found a clear link between short categories and pointing, and between long categories and reaching gesture ($\chi^2(1, N= 333) =6.310; p=.012$). This association is also found at 1;3 ($\chi^2(1, N= 263) =4.117; p=.042$).

Fundamental frequency analyses

To investigate the effect of gesture accompaniment and pragmatic function on the fundamental frequency of vocalizations, we carried out a LMM analysis, with SVI mean $f(0)$ as dependent variable and pragmatic function (declarative, imperative) \times gestural coordination (with, without gesture) as fixed factors. Subject was again the random factor.

Figure 4 shows mean fundamental frequency according to pragmatic functions and gestural accompaniment.

INSERT FIGURE 4 HERE

We found no main pragmatic function effect on $f(0)$ ($F(1, 19.282) = .103; p=.751$).

Regarding gestural accompaniment, SVIs produced with a gesture had higher $f(0)$ means than when produced alone ($F(1, 1907.171) = 11.732; p=.001$). We also found an interaction effect between pragmatic function and gestural accompaniment ($F(1, 1870.969) = 4.383; p=.036$). Declarative vocalizations have higher $f(0)$ values when produced with a gesture than when produced alone (705.567 vs. 607.862; $p=.001$). This difference is not found when vocalizations have an imperative purpose.

In order to define which ranges of fundamental frequency were positively associated with specific pragmatic functions, every SVI was classified as Low ($N=45$), Medium ($N=1179$) or High ($N=1009$). According to Oller et al. (2010), The Low category includes SVIs with pitch values below the range expected for a child's voice in speech-like utterances. On the other hand, the High category exceeds the maximum value expected for a child's voice in a speech-like utterance. Following this classification, the SVIs more similar to speech should fall into the Medium category. In order to explore how SVIs with different $f(0)$ were associated with communicative functions and gestures, we conducted Spearman correlations.

As regards communicative function, there was a positive and significant correlation between SVIs from the Medium category and declarative function ($\rho=.46$; $p=.022$). Regarding gestures, we found a positive relation between reaching gesture and High category of $f(0)$ ($\rho=.049$; $p=.015$). It is important to take into account that these effects were age-dependent, appearing only at 0;9 and, more markedly, at 1;3 (positive correlation between High category and reaching gesture: $\rho=.84$; $p=.039$ at 0;9, and $\rho=.157$; $p=.001$ at 1;3).

Canonical Syllables (CS) analyses

As described in the vocalization analysis section, we classified every SVI as positive or negative according to the canonical syllable parameter. In order to analyze the effect of gestural coordination and age, we calculated the proportion of canonical syllable islands accompanied by gesture at every age from the total of canonical syllable islands produced per child. Mean proportion of vocal islands including canonical syllables produced with and without gesture at each age is shown in Figure 5.

INSERT FIGURE 5 HERE

We then performed a repeated-measures ANOVA: age (0;9, 1;0 and 1;3) \times gestural coordination (with or without gesture), taking as dependent variable the proportion of vocal islands including canonical syllables. To assure the normality of the distribution, the arcsine square root transformation for proportional data was conducted.

Results showed a main effect of gestural coordination ($F(1,10) = 6.34$; $p=.031$; $\eta^2=.388$) and an interaction effect between age and gestural coordination ($F(2,20) = 5.11$; $p=.016$; $\eta^2=.338$). This interaction is explained by the lower proportion of canonical syllables at 0;9 coordinated with a gesture compared to those produced without a gesture (.025 vs.

.527; $p=.006$) (see Figure 6). In addition, for vocalizations coordinated with a gesture, the proportion of canonical syllable islands was greater at 1;0 than at 0;9 (.330 vs. .025; $p=.010$), and at 1;3 compared with 0;9 (.352 vs. .025; $p=.007$). This means that the relative frequency of canonical syllable islands tends to increase with age when coordinated with a gesture, especially from 0;9 to 1;0. On the contrary, there were no differences in the canonical syllable islands proportion when the vocalizations were not coordinated with a gesture. As can be seen in Figure 6, the canonical syllable islands tend to increase only when they are produced with a gesture, but not when produced alone.

Pitch direction of vocalizations

Firstly we explored the differences in intonation patterns according to gestural accompaniment and pragmatic function. We found no differences in the intonation patterns distribution depending on the pragmatic function (nor on the vocalizations accompanied by gestures ($\chi^2(4, N= 331) = 5.328; p=.255$) or the vocalizations produced alone ($\chi^2(4, N= 810) = 5.153; p=.272$). However, on including age in the analysis, the differences found in intonation patterns when vocalizations were produced without a gesture did not reach statistical significance ($\chi^2(8, N= 952) = 14.931; p=.06$); but there were differences in intonation pattern distribution when vocalizations were produced with a gesture at different age points ($\chi^2(8, N= 360) = 16.155; p=.04$). At age 0;9, there were fewer rising and more inverted u-shape vocalizations than would be expected by chance, in contrast, at 1;3 there were fewer inverted u-shape vocalizations than we would expect by chance.

Regardless of gestural accompaniment, we found differences in intonation pattern distribution depending on age and pragmatic function (see Table 6).

INSERT TABLE 6 HERE

These differences did not appear at 0;9 ($\chi^2(4, N=279)=4.494; p=.368$) or at 1;0 ($\chi^2(4, N=440)=5.301; p=.258$), but only at 1;3 ($\chi^2(4, N=422)=13.274; p=.01$). At this age, rising vocalizations appeared with imperative more often than with declarative function, and flat vocalizations appeared with declarative more often than with imperative function. We did not find differences on intonation patterns depending on the type of gesture (pointing, reaching and other) ($\chi^2(8, N=360)=9.172; p=.328$).

Discussion

The main goal of this study was to explore some acoustic features of vocalizations produced by children with and without communicative gestures in the transition to first words. We hypothesized that the multimodal character of communication would have an impact on the acoustic properties of vocalizations produced in this period.

We found that gestural accompaniment has an impact on certain acoustic features of vocalizations. When produced with a gesture, vocalizations have more vocal islands or “protosyllables” with higher $f(0)$ than when produced alone. In addition, the number of canonical protosyllables accompanied by a gesture tends to increase with age, especially between 0;9 and 1;0. This increase is not found when the canonical syllables are produced without a gesture. These findings suggest that vocalizations are progressively assimilating speech features, especially when they are accompanied by a gesture. Other more complex properties, such as specific intonation patterns, seem to be more linked to pragmatic functions than to its combination with gestures.

Besides this gestural impact on vocalization features, we also found specific patterns on vocal properties depending on the interaction between the gestural accompaniment and the pragmatic function of the communicative behavior.

First, as mentioned earlier, we found that vocalizations produced with a gesture have more vocal islands or “protosyllables” than when produced alone. In addition, when coordinated with a gesture, pragmatic function has an influence on the number of vocal islands, with imperative vocalizations showing more vocal islands than declarative ones. We did not find this influence when vocalizations are produced without a gesture. This finding is consistent with previous research showing the predictive character of pointing with vocalization (Igalada et al., 2015; Murillo & Belinchón, 2012; Wu & Gros-Louis, 2014) on later language abilities, and the relevance of declarative function on this pointing predictive role (Colonnese et al., 2010).

We found no differences on duration depending on gestural accompaniment.

Declarative islands tended to be longer than imperative ones when produced with a gesture, but these differences did not reach statistical significance.

Nevertheless, with specific gestures and vocal island duration, we found a pattern of association between the Medium duration category and pointing gestures. By contrast, the reaching gesture is found more frequently with Extra-large vocal islands. As reported by Oller et al. (2010), SVIs from the Small and Medium categories suggest speech-like rhythmic organization because the durational values indicated are typical of syllables in speech. Vocal islands from the Large and Extra-large categories suggest the opposite, because the corresponding ranges are beyond the durations of typical syllables (see Oller et al., 2010, Appendix).

We also found developmental changes in this respect. From 1;0 onwards, proto-syllables from the Small and Medium categories were linked with pointing, and Large and Extra-large SVIs linked with reaching gesture.

This suggests that from 1;0, proto-syllables coordinated with pointing gestures tend to have a more similar duration to speech sounds than proto-syllables coordinated with other gestures such as reaching. It could be that SVI duration varies according to the synchrony of the vocalization with the gesture, as previous research has shown (Esteve-Gibert & Prieto, 2014).

In our fundamental frequency analysis, we found that mean $f(0)$ was higher when protosyllables were produced with a gesture than when produced alone specifically with declarative attempts. This pragmatic function appeared to be associated with SVI Medium duration. We also found a positive correlation between the reaching gesture and the high $f(0)$ category, although, surprisingly we did not find a direct relation between the imperative function and the high category. Similarly, we did not find a relation as clear as expected between the pointing gesture and the $f(0)$ categories more similar to mature speech. It might be that our observation situation favored the appearance of imperative behaviors to the detriment of declarative attempts. The pointing gesture is linked to the declarative function, but it frequently appeared with an imperative purpose in our sample. This could have affected the results obtained in this section, especially in the analysis of narrower categories.

As for the canonical syllable parameter, the proportion of SVIs categorized as positive increases with age when the vocalizations are coordinated with gestures, but this change is not observed when the vocalizations are produced alone. The canonical syllable parameter can be interpreted as indicating that syllable-like units or “proto-syllables”

are organized in the same way as syllables in mature speech. The increase of vocal islands with these canonical characteristics is particularly obvious from 1;0 onwards when they are produced with a gesture. Interestingly, at this age we also found differences between the number of vocal islands per vocalization depending on the presence of a gesture and the pragmatic function of the communicative attempt.

Likewise, it was not until 1;0 that we found a clear link between the SVIs durations more similar to speech and pointing gestures. These findings are in line with those of previous studies suggesting that gestural and vocal elements become integrated in the transition to first words (Esteve-Gibert & Prieto, 2014). At this stage, vocalizations accompanying gestures gradually acquire patterns of duration and syllabic structure more similar to those of adult speech.

Finally, to explore changes in pitch direction related to the use of communicative gestures, we classified the vocalization's pitch contour as flat, rising, falling, u-shaped or inverted u-shaped. We found that intonation patterns seem to be more linked to pragmatic functions than to gestural accompaniment, though we did not find a clear relation between intonation patterns and pragmatic functions until the age of 1;3. At this age, rising contours appear related to imperative functions, whereas declarative functions appears to be related to flat contours. Beyond this association seen at 1;3, we did not find a clear correspondence of intonation patterns with pragmatic functions. It might be that our pragmatic categorization was too wide to capture subtle differences in communicative purposes that can be related to specific intonation patterns. For example, we did not find any difference between informative or expressive pointing as proposed by Grünloh and Lizskowski (2014). Unlike us, they found a relationship between rising intonation and declarative pointing, whereas requestive behaviors were linked to rising and flat contours. There is enormous controversy and divergent findings regarding the

use of intonation patterns by children in the early stages of language development (see Snow & Balog, 2002, for a review). Despite this controversy, some studies have reported that falling contours precede rising contours in children's development. By contrast, rising contours are used more frequently than falling contours in utterances directed to the mother (Snow & Balog, 2002). Rising contours have also been associated with situations requiring a response vs. situations not requiring a response from the adult, and this seems to be the case in our sample. However, these situations include requests for objects or actions (categorized as "imperative function" in our study), and can also include the request for information generated in a labeling situation, that is, when the child is asking for the name of something. This protointerrogative function, categorized as declarative in our study, is linked to the pointing gesture (Rodriguez, 2009). Infant intonation patterns are highly complex, and it may be difficult to find a direct correspondence between specific intonation contours and pragmatic functions (see, for example, Thorson, Borrás-Comes, Crespo-Sendra, Vanrell & Prieto, 2014). Further research is needed to define communicative intentions in a more specific way in order to link subtle changes in pragmatic functions to different pitch contours.

The multimodal communication attempts from the child are progressively acquiring the formal properties (in terms of pitch, syllabic structure and duration) seen in mature speech. The fact that this change is not observed in vocalizations produced without a gesture offers support to the idea of the differential feedback provided by adults to multimodal communicative behaviors from the child. Adults are more likely to provide a verbal response when the infant's communicative attempts include gestures (Olson & Masur, 2013). Wu and Gros-Louis (2014) have shown that mothers provide more verbal responses to pointing with vocalizations than to object-directed vocalizations. The probability of the mother labeling a toy was higher if the child pointed with vocalization

than if s/he only vocalized. In addition, mothers respond differentially depending on the characteristics of infant vocalization. They tend to imitate and expand the child's utterance after consonant-vowel vocalizations, but not after vowel-like utterances (Gros-Louis, West, Goldstein & King, 2006).

Hence, it appears that gestural-vocal coordination can enhance the communicative response obtained from caregivers. This way, infants can gradually adjust their utterances to the feedback provided by adults in terms of pitch, duration and syllable-like structure, as our results suggest. Further research in naturalistic interaction settings is needed to disentangle these mechanisms.

Conclusions

In conclusion, this study shows that children vocalizations show differences on certain acoustic characteristics when they are produced with or without a gesture. Vocalizations have more SVIs when produced with a gesture, and their mean $f(0)$ is higher than when produced alone. Results also show that the proportion of canonical protosyllables tends to increase when vocalizations are produced with gestures, but not when they are produced without it. In addition, some specific acoustic patterns are associated with specific gestures: short duration categories are more linked to pointing gesture whereas long categories are linked to reaching ones.

Pragmatic function also interacts with gestural accompaniment: we found that declarative vocalizations have less SVIs than imperative ones only when they are produced with a gesture. Declarative vocalizations have also higher $f(0)$ when produced with a gesture than when produced alone, but this difference is not found for imperative vocalizations. By contrast, intonation patterns seem to be linked to pragmatic function and less influenced by gestural accompaniment.

The analysis of acoustic parameters represents a fruitful contribution to the understanding of vocal-gestural interaction in early language development. It serves to reassert, by means of objective measures, the changes in children's vocalization features depending on the use of gestures previously observed using acoustic judgments.

As Vigliocco, Perniss and Vinson (2014) point out, language is learnt as a multimodal process in a multimodal context, and should therefore be studied from a multimodal perspective. Both gestural and vocal components are part of a single communication system that reorganizes itself and evolves, by means of social interaction, throughout the process of language development.

References

- Acredolo, L. & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development* **59**, 450-66.
- Balog, H. L. & Brentari, D. (2008). The relationship between early gestures and intonation. *First Language* **28**, 141-63.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). The emergence of symbols: Cognition and communication in infancy. New York: Academic Press.
- Begus, K., Gliga, T. & Southgate, V. (2014). Infants learn what they want to learn: responding to infant pointing leads to superior learning. *PloS one*, **9**, e108817.
- Blake, J., McConnell, S., Horton, G. & Benson, N. (1992). The gestural repertoire and its evolution over the second year. *Early Development and Parenting* **1**, 127-36.
- Bloom, K., Russell, A. & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language* **14**, 211-27.
- Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.76, retrieved 8 May 2014 from <http://www.praat.org/>.
- Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L. & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience* **4**, 237.
- 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*, 235–58. Cambridge: Cambridge University Press.
- Camaioni, L., Castelli, M. C., Longobardi, E. & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language* **11**, 345-59.

- Camp, B. W., Burgess, D., Morgan, L. J. & Zerbe, G. (1987). A longitudinal study of infant vocalization in the first year. *Journal of Pediatric Psychology* **12**, 321-31.
- Capirci, O., Contaldo, A., Caselli, M. C. & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture* **5**, 155-77.
- Carpenter, M., Nagell, K. & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development* **63**, 1-174.
- Cartmill, E. A., Hunsicker, D. & Goldin-Meadow, S. (2014). Pointing and naming are not redundant: children use gesture to modify nouns before they modify nouns in speech. *Developmental Psychology* **50**, 1660.
- Caselli, M. C., Rinaldi, P., Stefanini, S. & Volterra, V. (2012). Early action and gesture “vocabulary” and its relation with word comprehension and production. *Child Development* **83**, 526-42.
- Cochet, H. & Vauclair, J. (2010a). Features of spontaneous pointing gestures in toddlers. *Gesture* **10**, 86-107.
- Cochet, H. & Vauclair, J. (2010b). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development* **33**, 432-42.
- Colletta, J., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N. & Levine, S. (2014). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *Journal of Child Language* **42**, 122-45.
- Colonesi, C., Stams, J. M., Koster, I. & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review* **30**, 352-66.

- DePaolis, R. A., Vihman, M. M. & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics* **36**, 406-22.
- Ejiri, K. & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science* **4**, 40-8.
- Esteve-Gibert, N. & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *Journal of Child Language* **40**, 919-44.
- Esteve-Gibert, N. & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication* **57**, 301-16.
- Fasolo, M. & D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture* **12**, 1-15.
- Fogel, A. & Hannan, T. E. (1985). Manual actions of nine- to fifteen-week-old human infants during face-to-face interaction with their mothers. *Child Development* **56**, 1271-9.
- Goldin-Meadow, S. & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 85-107. New Jersey: LEA.
- Gros-Louis, J., West, M. J., Goldstein, M. H. & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development* **30**, 509-16.
- Gros-Louis, J., West, M. J. & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy* **19**, 385-408.

Gros-Louis, J. & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant Behavior and Development* **35**, 773-8.

Grünloh, T. & Lizskowski, U. -Prelinguistic vocalizations distinguish pointing acts. *Journal of Child Language*, available on CJO2015.
doi:10.1017/S0305000914000816.

Hsu, H., Fogel, A. & Cooper, R. B. (2000). Infant vocal development during the first 6 months: Speech quality and melodic complexity. *Infant and Child Development* **9**, 1-16.

Igualada, A., Bosch, L. & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development* **39**, 42-52.

Iverson, J. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language* **37**, 229-61.

Iverson, J. & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science* **16**, 367-71.

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

Karousou, A. & López-Ornat, S. (2013). Prespeech vocalizations and the emergence of speech: a study of 1005 Spanish children. *The Spanish Journal of Psychology* **16**, 1-21.

Keller, H. & Schölmerich, A. (1987). Infant vocalizations and parental reactions during the first 4 months of life. *Developmental Psychology* **23**, 62-7.

Leung, E. H. & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology* **17**, 215-20.

Liszkowski, U. & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development* **26**, 16-29.

Majorano, M. & D'Odorico, L. (2011). The transition into ambient language: A longitudinal study of babbling and first word production of Italian children. *First Language* **31**, 47-66.

Masataka, N. (1993a). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *Journal of Child Language* **20**, 303-12.

Masataka, N. (1993b). Relation between pitch contour of prelinguistic vocalizations and communicative functions in Japanese infants. *Infant Behavior & Development* **16**, 397-401.

Masataka, N. (1995). The relation between index-finger extension and the acoustic quality of cooing in three-month-old infants. *Journal of Child Language* **22**, 247-57.

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

Murillo, E. & Belinchón, M. (2012). Gestural-vocal coordination:- Longitudinal changes and predictive value on early lexical development. *Gesture* **12**, 16-39.

Murillo, E. & Belinchón, M. (2013). Patrones comunicativos multimodales en la transición a las primeras palabras: cambios en la coordinación de gestos y vocalizaciones. [Multimodal communicative patterns on the transition to first words: Changes in the coordination of gesture and vocalization]. *Infancia y Aprendizaje* **36**, 473-87.

- Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U. & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences* **107**, 13354-59.
- Olson, J. & Masur, E. F. (2013). Mothers respond differently to infants' gestural versus nongestural communicative bids. *First Language* **33**, 372-87.
- Özçalışkan, Ş. & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition* **96**, 101-13.
- Özçalışkan, S. & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes* **24**, 190-217.
- Papaeliou, C. F. & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing 'communication' and 'apprehension'. *Journal of Child Language* **33**, 163-78.
- Rodríguez, C. (2009). The 'circumstances' of gestures: Proto-interrogatives and private gestures. *New Ideas in Psychology* **27**, 288-303.
- Rowe, M. L. & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science* **12**, 182-7.
- Rowe, M., Özçalışkan, Ş. & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language* **28**, 182-99.
- Snow, D. & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua* **112**, 1025-58.
- Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (eds.), *Speech Coding & Synthesis*, 497-518. New York: Elsevier.

Thorson, J., Borrás-Comes, J., Crespo-Sendra, V., Vanrell, M. & Prieto, P. (2014). The acquisition of melodic form and meaning in yes-no interrogatives by Catalan and Spanish speaking children. *Probus* **4**, 27.

Vigliocco, G., Perniss, P. & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**(1651), 20130292.

Vihman, M. M., Ferguson, C. A. & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics* **7**, 3-40.

Wu, Z. & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language* **34**, 72-90.

Table 1

Mean age of participants and duration of the recording sessions.

	Recording sessions					
	0;9		1;0		1;3	
	Mean	SD	Mean	SD	Mean	SD
Age (y;m;d)	0;9;3	0;0;8	1;0;0	0;0;8	1;3;2	0;0;8
Session duration	17'	5'44''	16'	4'29''	17'	8'28''

Table 2

Coding categories

Coding categories	
Multimodal combinations	
Gesture only	Gesture is produced without vocalization
Vocalization only	Vocalization is produced without gesture
Gesture + vocalization	There is an overlap of at least one second between gesture and vocal sound production, or one element is produced within the previous or subsequent second of the other element production
Gesture	
Point	Extension of the arm and index finger
Reach	Arm is extended, palm usually down, hand open and fingers straight, as defined by Blake, McConnell, Horton and Benson (1992)
Symbolic	Gestures that symbolically represent objects, events, desires and conditions. Adapted from Acredolo & Goodwyn (1988)
Conventional	Say hello or bye with the hands, clapping, etc.
Other	Any gesture observed not included in the previous categories
Communicative function	
Declarative	The goal of the behavior is to share attention with the adult about an object or event
Imperative	The purpose of the behavior is to obtain a change in the physical world: to obtain an object, or to get the adult to carry out an action
Other	Behaviors with an expressive or reject function, behaviors that occur after a question or request from the adult, the play routines and the behaviors whose function is not clear, are included in this category

Table 3

SVI classification according to their duration and fundamental frequency

Category	
Duration (ms)	
51-110	Extra-Small (XS)
111-250	Small (S)
251-600	Medium (M)
601-900	Large (L)
> 900	Extra-Large (XL)
$f(0)$ (Hz)	
< 250	Low
250-600	Medium
> 600	High

Table 4.- Gesture type frequencies according to its pragmatic function

Pragmatic function	Gesture type		
	Pointing	Reaching	Other
Declarative	98	18	24
Imperative	150	376	3
Other	6	8	32
Total	254	402	59

Table 5

Gestures distribution depending on SVI duration: frequency and adjusted residuals.

Gesture	SVI Duration										Total
	Extra-Small		Small		Medium		Large		Extra-Large		
	N	AR	N	AR	N	AR	N	AR	N	AR	
Pointing	37	-1.6	92	.5	102**	3.5	11*	-2.5	11*	-2.2	253
Reaching	75	.8	144	.4	109**	-3.2	34	.9	39**	2.8	401
Other	14	1.3	15	-1.6	18	-.3	10**	2.8	2	-1.2	59
Total	126		251		229		55		52		713

* $p < .05$; ** $p < .01$

Table 6

Intonation patterns depending on age and pragmatic function. Frequencies and adjusted residuals.

		Age					
		0;9		1;0		1;3	
Intonation		N	AR	N	AR	N	AR
Rising	Declarative	4	.6	9	-2.3	9	-2.2
	Imperative	2	-.6	25	2.3	27*	2.2
U-shaped	Declarative	27	1.9	37	.1	32	-1.4
	Imperative	19	-1.9	44	-.1	57	1.4
Falling	Declarative	5	-.1	11	.1	13	-.4
	Imperative	6	.1	13	-.1	20	.4
U-inverted	Declarative	71	-.7	115	.9	90	.7
	Imperative	89	.7	129	-.9	113	-.7
Flat	Declarative	22	-1	27	.3	36*	2.8
	Imperative	33	1	30	-.3	25	-2.8

* $p < .05$

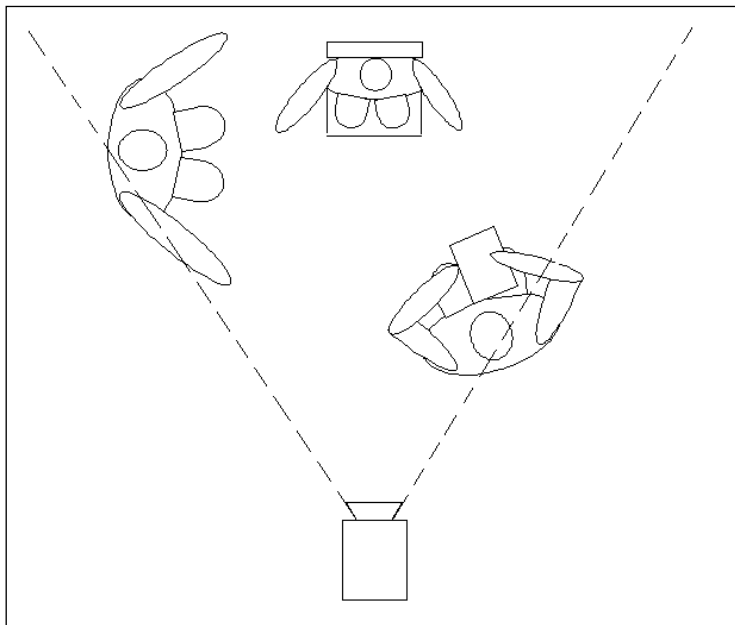


Fig.1.- Sketch of the observation setting

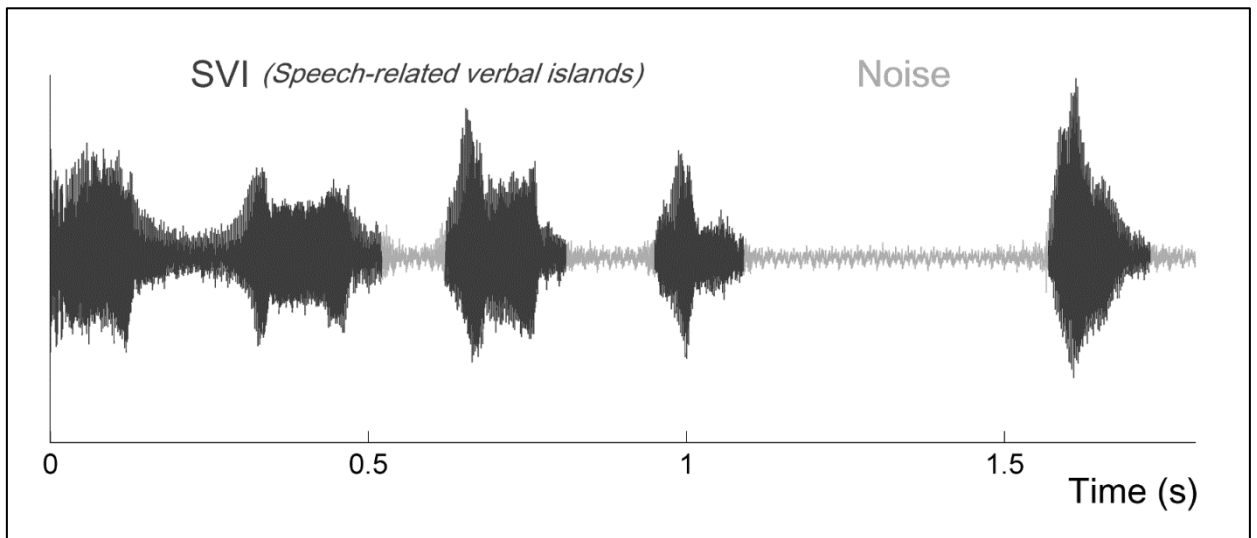


Figure 2

Example of the SVIs in a vocalization.

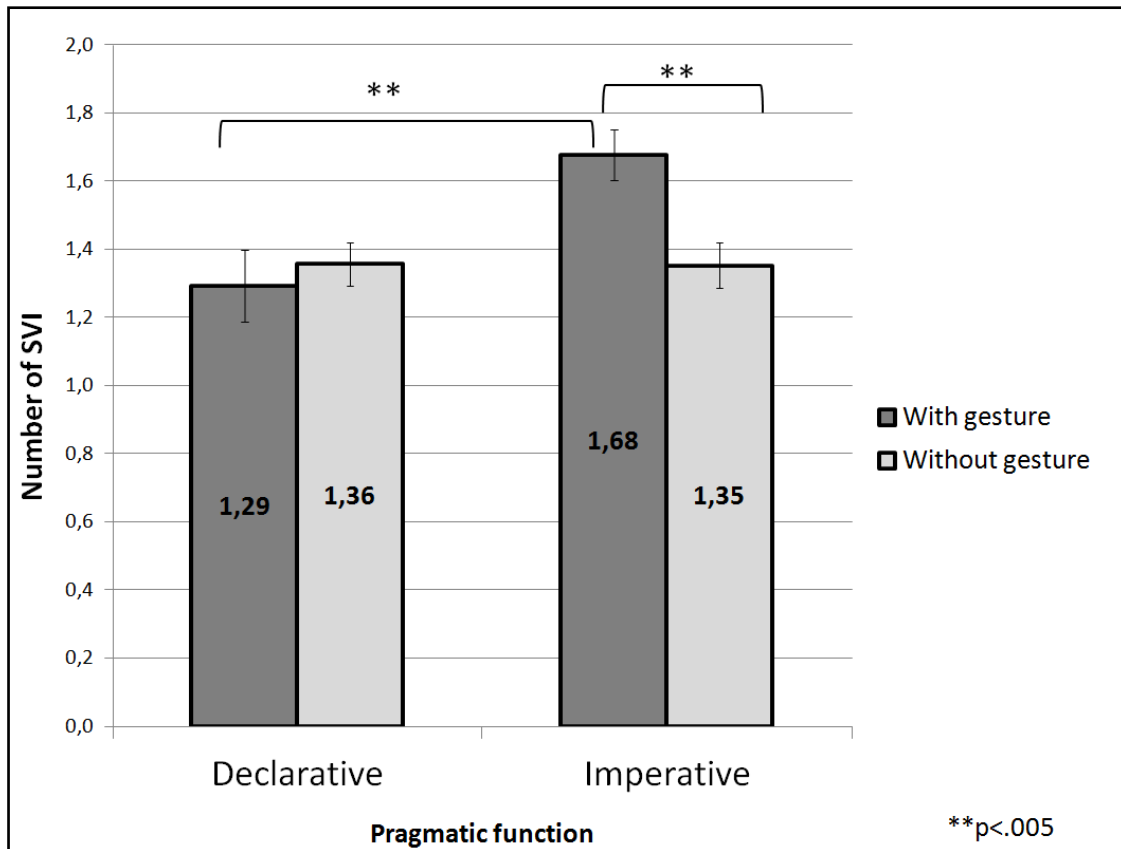


Figure 3

Mean number of SVIs per vocalization with and without gesture according to its pragmatic function

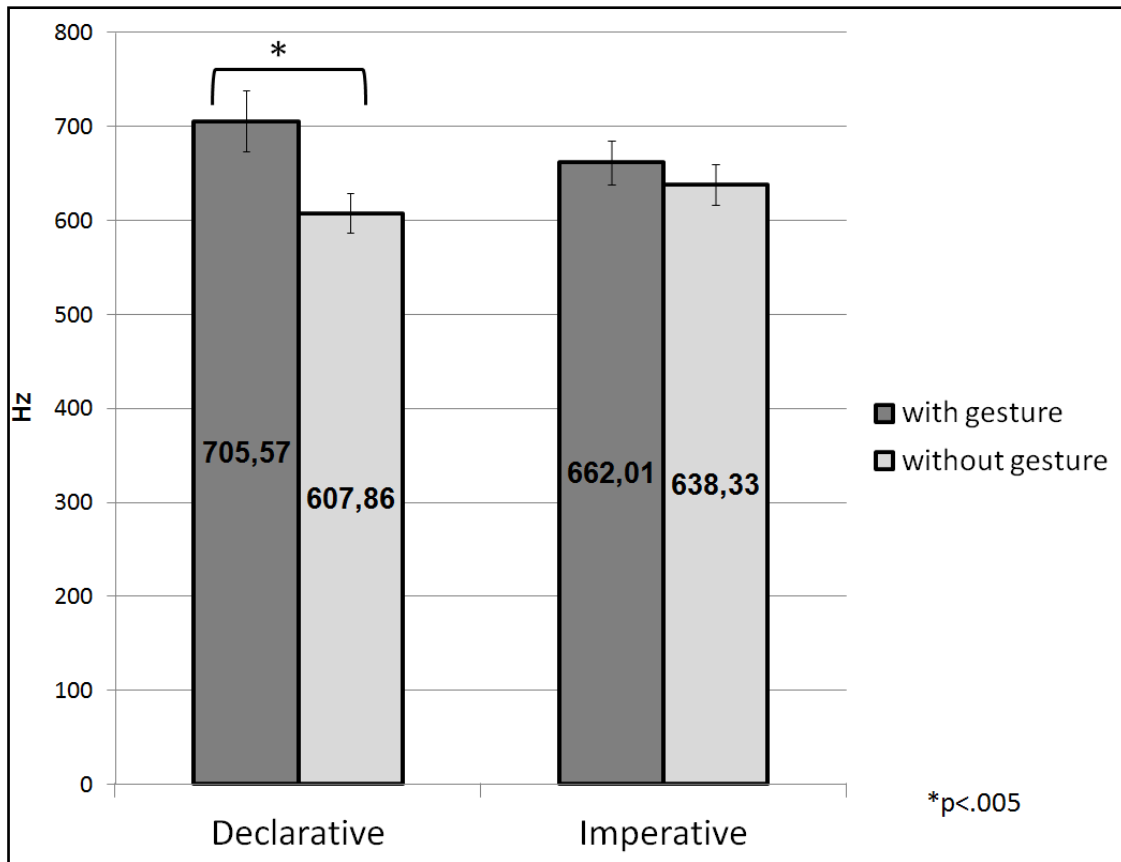


Figure 4

Mean $f(0)$ according to pragmatic function and gestural accompaniment.

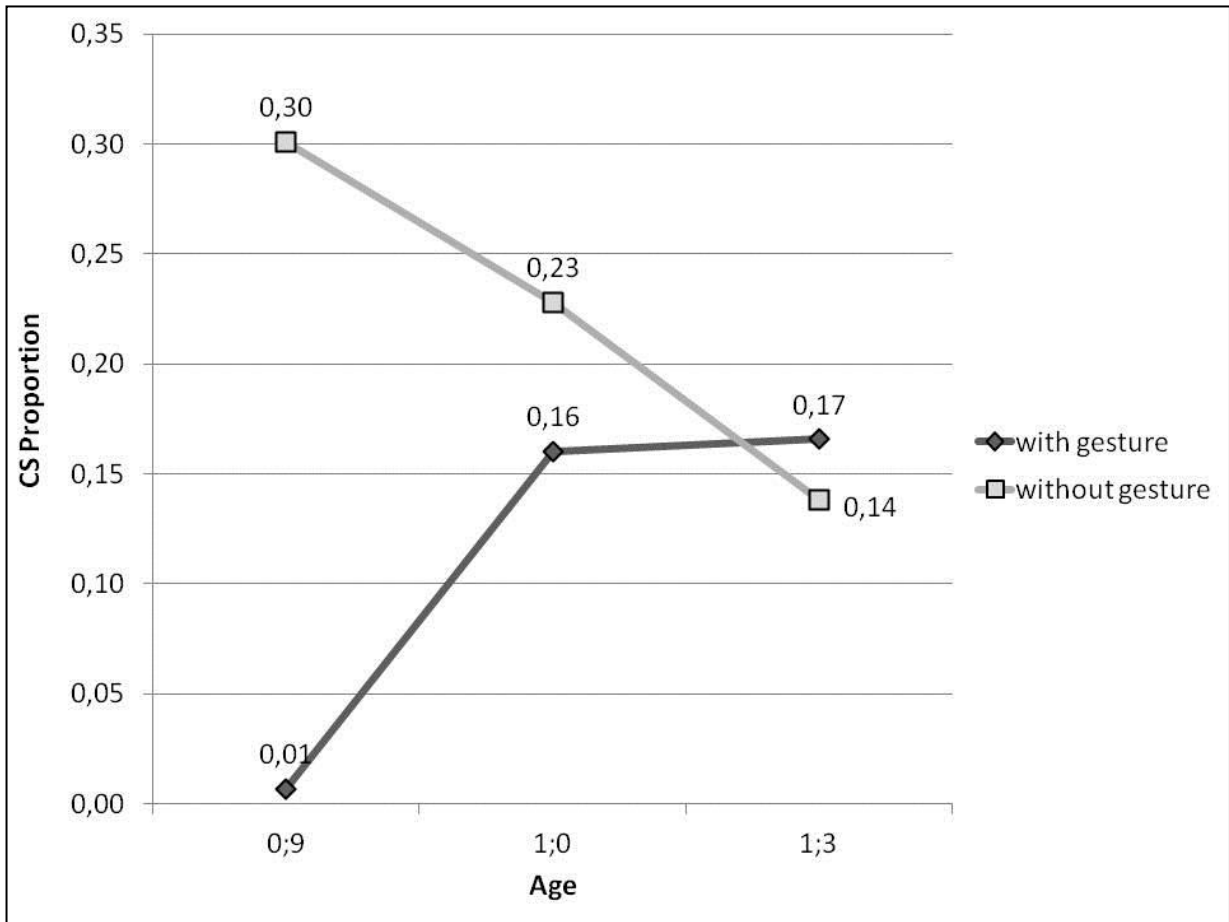


Figure 5

Mean proportion of CS verbal islands with and without gesture at each age from the total of CS verbal islands produced by each child.