The coordination of vocalizations and communicative gestures in the transition to first words.

Eva Murillo (eva.murillo@pdi.ucm.es)
Departamento de Psicología Básica II. Universidad Complutense de Madrid
Facultad de Psicología. Campus de Somosaguas
28223 Pozuelo de Alarcón. Madrid. Spain

Almudena Capilla (almudena.capilla@uam.es)
Departamento de Psicología Biológica y de la Salud. Universidad Autónoma de Madrid
Facultad de Psicología. Campus de Cantoblanco
28049 Madrid. Spain.

Abstract

This work addresses the developmental changes in vocalizations acoustic features in relation to its coordination with communicative gestures in the transition to first words. Our hypothesis is that gestural-vocal coordination facilitates early lexical development, so that the acoustic features of vocalizations will be more similar to those of words when they are accompanied by gestures, specifically pointing gesture, and with a declarative function. Preliminary findings show differences in duration, fundamental frequency, and syllabicity parameters, related with gesture coordination and declarative function.

Keywords: Communicative development; gestures; vocalizations; multimodal communication; acoustic analysis, first words.

Introduction

Vocal and motor components are linked from the first months of life, and several motor and vocal milestones, appear to be related in the first year (see Iverson, 2010 for a review). Vocal and motor components seem to influence each other along early stages of language development (Bates & Dick, 2002). In the period of transition to first words, communicative gestures are very often produced with vocal components. Around the second year of life, near 70% of gestures are produced in coordination with vocalizations, and the coordination of gestures with deictic gestures, especially with pointing gesture, has a predictive value on subsequent lexical development (Murillo & Belinchón, 2012). It may be that adult's response to infant's communicative attempts is different if infant's communicative behavior includes gestural and vocal coordination, that is, has a multimodal character, as pointed out by Goldin-Meadow, Goodrich, Sauer and Iverson (2007). This way, adult's response can offer a linguistic model that the child can use to give "word form" to her vocalizations. Taking this into account, the aim of this study is to explore if vocal-gestural coordination in the transition to first words has an impact on vocal characteristics, expecting that vocalizations produced with gestures, especially with deictic gestures, will have more similar features to those of adult language.

Several studies have shown that there is formal continuity between infant preverbal vocalizations and their first words (Hsu, Fogel & Cooper, 2000; Oller, 1980; Stoel-Gammon & Cooper, 1984; Vihman, Ferguson & Elbert, 1986; Vihman, Macken, Miller, Simmons & Miller, 1985). In fact, the transition from preverbal vocalizations to words is not a sudden change, but a gradual process (i.e. Karousou & López-Ornat, in press; Majorano & D'Odorico, 2011).

Regarding gestures and motor actions, Fogel and Hannan (1985) observed the co-occurrence of index finger extension and vocalizations before the first 15 weeks, and Masataka (1995) found in three-month-old infants the association of index finger extension with syllabic vocalizations. In the period of transition to first words, communicative gestures such as pointing, reaching, giving or showing, become an usual way for infants to affect adult's intentions. The use of these gestures along the first year, especially deictic gestures, has a predictive value on subsequent lexical development (Bates, Benigni, Bretherton, Camaioni and Volterra,1979; Colonnesi, Stams, Koster & Noom, 2010), and gestural coordination with first words in the second year seems to predict different aspects of language development such as two-words combination (Morford and Goldin-Meadow, 1992) or overall vocal production (Capirci, Iverson, Pizzuto & Volterra, 1996). There is some evidence suggesting that there are differences in the vocalizations used depending on gestures that accompanies them and their communicative function (Murillo & Belinchón, in press), but these differences are based on acoustic analyses produced by trained judges. As pointed out by Nathani and Oller (2001), children vocal behavior is a very complex phenomenon, and the analysis of vocalizations is an extremely difficult process. To explore the nature of vocalizations that appear with communicative gestures in the period of transition to first words, we have chosen an acoustic analysis of objective parameters instead of acoustic judgment. Analysis of parameters such as duration, fundamental frequency, or intensity, has been a useful strategy to detect not only developmental changes in vocalizations related to language development (DePaolis, Vihman & Kunnari, 2008; Papaeliou & Trevarthen, 2006; Scheiner, Hammerschmidt, Jürgens & Zwirner, 2002; Snow, 2007), but also differences between typically developing children and children with developmental disorders (Bonneh, Levanon, Dean-Pardo & Adini, 2011; Diehl, Watson, Benetto, McDonough & Gunlogson, 2009; Oller et al., 2010). We based our analysis on different parameters defined by Oller et al. (2010), that have been sensitive to developmental changes and relevant in differentiating children with and without disorders such as autism or language delay.

Method

Participants

Eleven Spanish children (6 girls, 5 boys) were recorded every three months from 9 to 15 months of age. All of them came from biparental monolingual Spanish-speaking homes, and were born from full-term uncomplicated pregnancies and normal deliveries. No developmental or hearing problem was reported by parents, and all the infant participants were attending nursery school when data collecting started. Informed consent was obtained from parents who voluntarily agreed to participate in the study. Table 1 shows the mean age of participants at each recording session and session duration.

	Recording sessions					
	9 months		12 months		15 months	
	Mean	SD	Mean	SD	Mean	SD
Age (m;d)	9;3	0;8	12;0	0;8	15;2	0;8
Session duration	17’	5’44’’	16’	4’29’’	17’	8’28’’

Table 1: Mean age or participants an duration of the recording sessions.

Materials and procedure

During the observation sessions, the infant sat on a child-chair. A lapel microphone, continuously recording at a sampling rate of 44100 Hz, was placed in the child’s lapel in a way that they couldn’t see it to avoid distraction and recording problems. The primary caregiver (mother or father) was seated next to the child, and the experimenter showed a set of toys one at a time. The experimenter interacted with the child in a semi-structured play situation, responding to her communicative intents, but leaving the child to lead the interaction. All communicative behaviors (gestures, vocalizations or their coordinate use) addressed to the experimenter or to the caregiver were recorded and coded considering the type of gesture (pointing, reaching, give, conventional or symbolic gestures) and communicative functions (declarative, imperative or others). More details concerning recording setting characteristics, coding categories, and inter-observer agreement, can be found at Murillo and Belinchón (2012). For the present purpose, we excluded from the analysis all the communicative behaviors registered that did not include a vocalization.

Vocalization analysis. Infant’s vocalizations were extracted from audio recordings and segmented using the Praat program (http://www.fon.hum.uva.nl/praat/). We followed Bloom, Russel and Wassenberg, (1987) criterion to consider a vocal sound as a vocalization. A new vocalization was counted as beginning after any audible inspiration of after a period of a second or more of silence. Non-vocal sounds, cries and other vegetative sounds were excluded. 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 CVI was identified when the speech signal was followed by a noise signal by at least 50 ms (Oller et al. 2010). Given that in our sample vegetative sounds and cries were excluded, all the CVI obtained were *Speech-related verbal islands* (SVI) following Oller et al.’s (2010) classification. SVI definition includes utterances as babbling, prespeech vocalizations and real speech. Every SVI longer than 50ms was analyzed acoustically.

Firstly, we obtained the number of SVI per vocalization. Then, we obtained several acoustic parameters of SVIs, and classified then according to different criteria. We extracted SVI duration and classified them following Oller et al.’s (2010) criteria, with the exception of the addition of a new category “XS” (see Table 2).

Table 2: SVI duration categories

Category	Duration (ms)
XS	51-110
S	111-250
M	251-600
L	601-900
XL	900

Then, we obtained 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. We classified SVI according to their *f(0)* mean as shown in Table 3.

Table 3: Fundamental frequency classification

Category	*f(0)*
1	< 250
2	250-600
3	>600

Finally, we computed the canonical syllables (CS) parameter described by Oller et al. (2010), which provides a critical measure of the maturity of initial formant transitions. Formant frequencies (F1 and F2) were tracked based on the linear predictive coefficients (LPC). As in Oller et al.'s (2010), an SVI was categorized as a CS if (1) the SVI's category duration was either S or M, (2) the SVI was from Category 2 based on its *f(0)*, (3) the maximum slope change of F1-F2 was reached within 120 ms, and (4) to that point, either F1 or F2 slope were higher than 3 and 5 respectively.

Results

Vocalization rate and coordination with gestures

Children produced 1686 vocalizations, with a rate of 2.99 vocalizations per minute. Mean rate of vocalization was 2.24 at 9 months (*SD*=1.20; Min: 1.07; Max: 4.81), 3.48 at 12 months (*SD*=1.54; Min: 1.29; Max: 6.24) and 3.65 at 15 months (*SD*=2.22; Min: 0.26; Max: 7.17). To explore the vocal-gestural coordination changes with age, we conducted a repeated measures ANOVA, with vocalization+gesture (VG) rate as dependent variable and age (9, 12 and 15 months) as factor. Figure 1 shows the vocalization rate with and without gesture.

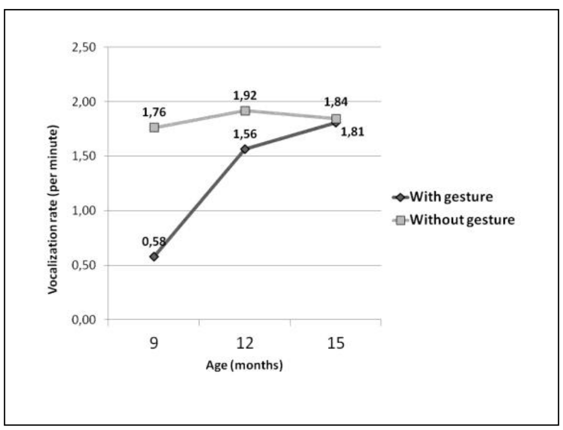

Figure 1: Vocalization rate with and without gesture at every age

We found a significant effect of age, $F(2,20)=7.583$; $p=.004$; $\eta^2=.431$, showing an increase of VG rate from 9 to 15 months (1.807 vs. 0.579; $p=.003$).

Child vocal islands analysis

We obtained 2427 SVIs from 1686 vocalizations. Several analyses including number of SVI per vocalization, duration of SVIs and fundamental frequency were conducted.

Number of vocal islands per vocalization. In order to examine the influence of gesture and communicative function on the number of SVI per vocalization, we conducted a repeated-measures ANOVA. We took as dependent variable the number of islands per vocalization, and as factors the gesture (pointing vs. reaching) and the communicative function (declarative vs. imperative). No significant effect of gesture was found. Nevertheless we found a function effect ($F(1,6)=6.063$; $p=.049$), showing that there are more SVIs when vocalization has an imperative function than when it has a declarative one.

Duration of SVI. For exploring developmental changes in SVI duration related with function of the communicative behavior, a repeated measures ANOVA was conducted. SVI duration was taken as dependent variable, and age (9, 12 and 15 months) and communicative function (declarative vs. imperative) as factors. There was no significant effect of age on SVI duration, but we found a communicative function effect ($F(1,9)=8.292$; $p=.018$, $\eta^2=.480$). Declarative SVI were longer than SVI with imperative function (356.65 vs. 294.9; $p=.018$).

Regarding gestural accompaniment, we examined if there were differences on SVI durations depending on the coordination of the vocalizations with a gesture. A repeated measures ANOVA was carried out, taking the communicative function (declarative vs. imperative) and the gestural coordination (with gesture vs. without gesture) as factors, and SVI duration as dependent variable. We found an effect of interaction between function and gestural coordination ($F(1,10)=8.108$; $p=.017$, $\eta^2=.448$). As can be seen in Figure 2, when there is vocal-gestural coordination, declarative SVIs are longer than imperative ones (428.35 vs. 315.76; $p=.029$).

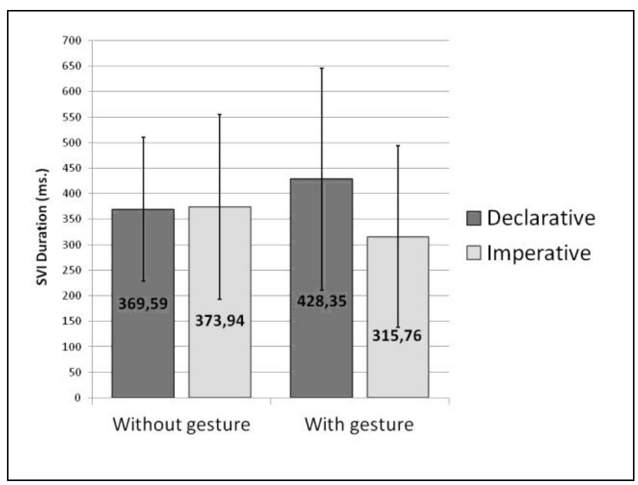

Figure 2: SVI duration with and without gesture according to the communicative function.

This difference is not observed when SVI is not coordinated with a communicative gesture.

Fundamental frequency analyses. The mean fundamental frequency was obtained for every SVI and classified in Category 1, 2 or 3. Category 1 includes SVI with pitch values that are below the range expected for a child's voice in speech-like utterances. Category 3 exceeds the maximum value expected for a child's voice in a speech-like utterance. According to this classification, the SVIs more similar to speech should be included in Category 2. In order to explore how SVIs with different *f(0)* values were associated with gestures and communicative functions, Spearman correlations were conducted.

Concerning communicative function, there was a positive and significant correlation between SVI's from Category 2 and declarative function (ρ=.46; p=.022). We also found a significant and negative correlation between SVIs from Category 3 and declarative function (ρ= -.059; p=.003).

Regarding gestures, we found a positive relation between reaching gesture and Category 3 of *f(0)* (ρ=.049; p=.015), and a negative correlation between Category 2 and reaching when the gesture has an imperative function (ρ=-0.70; p=.001).

When we included age in the analysis, we found that at 12 months, there was no significant correlation between gestures and frequency categories. However, at 9 months and at 15 month, we found the same pattern: a significant and negative correlation between Category 2 and reaching gesture (ρ =-.081 p=.048 at 9 months, and ρ =-.135 p <.001 at 15 months), and a significant and positive correlation between Category 3 and reaching gesture (ρ =.84; p=.039 at 9 months, and ρ= .157; p=.001 at 15 months).

Considering that Category 2 of *f(0)* is the one that includes vocalizations more similar to speech, we examine the proportion of gestures that appears with SVIs from this category. We calculated the proportion of pointing and reaching gestures of the total SVI in Category 2 at each age. Results showed no significant difference between gestures, but, as can be seen in Figure 3, there is a tendency of pointing gesture to increase along the period studied, not seen for reaching gesture. These results point out an interesting path for further research.

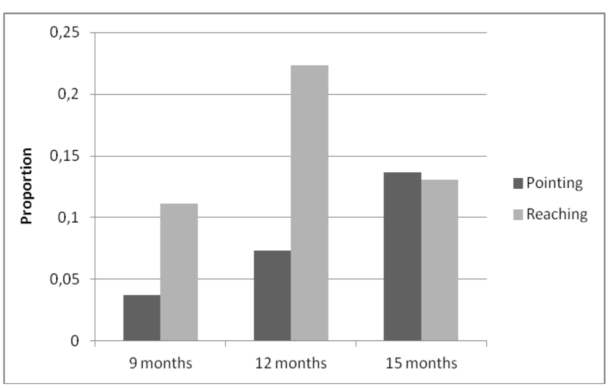

Figure 3: Proportion of SVIs from Category 2 that are accompanied by pointing and reaching gestures at each age.

Fundamental frequency and duration analysis. We analyzed how SVIs were distributed according to frequency and duration categories at the different ages, with the aim of identifying developmental changes in SVI related features. For the frequency Category 1, it was not possible to conduct the analysis due to the scarce number of observations.

For Category 2, at 12 months of age, we found that there were more SVIs with M and L durations than expected by chance ($\chi^2(4)$=64,183, p<.001), and less with XS duration. At 15 months, there were less SVIs with XS and XL duration, and more SVIs with M duration than expected by chance ($\chi^2(4)$=24,776; p<.001).

For Category 3, at 9 months there were less SVIs with XS duration and more with M, L and XL duration than expected by chance ($\chi^2(4)$=30,390; p<.001). At 12 months, less SVIs with XS and L duration and more with S duration than expected by chance were found ($\chi^2(4)$=38,347, p<.001). At 15 months, there were less SVIs with XS duration and more with L and XL durations than expected by chance ($\chi^2(4)$=40,049, p<.001). Results concerning frequency and duration distribution of SVIs at different ages are summarized in Table 3.

Table 3: Differences in SVI distribution according to age, *f(0)* category and duration.

Frequency 2	XS	S	M	L	XL
9 months	-		+	+	
12 months	-	+	+	+	
15 months	-	-	+		-
Frequency 3	XS	S	M	L	XL
9 months	-		+	+	+
12 months	-	+		-	
15 months	-	-		+	+

(+ = appearance above change, - = appearance below chance)

When we included gesture and communicative function in the analyses, without taking age into account, we found no differences in distribution according to *f(0)* and duration neither for declarative or pointing gestures, nor for declarative reaching gesture. We only found differences for reaching with imperative function, that appears with a frequency above chance when it is accompanied for a SVI with XS duration and from Category 1 of *f(0)(* $\chi^2(8)$=17,525, p=.025). That means that imperative reaching is associated to short and low vocal utterances.

Canonical Syllabes (CS). As described in the Vocalization analysis section, we classified every SVI as positive or negative according to CS parameter. In order to analyze the effect of gestural coordination and age, we calculated the proportion of CS islands accompanied by gesture at every age from the total of CS islands registered for each child. Mean proportions of CS with and without gesture at each age can be seen in Figure 4.

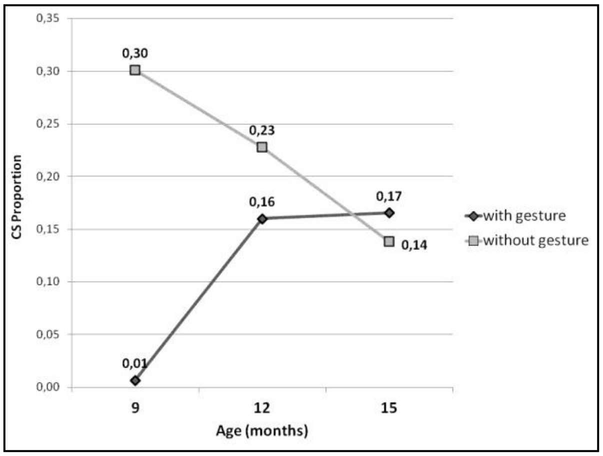

Figure 4: Mean proportion of CS islands with and without gesture at each age.

We conducted a repeated measures ANOVA taking as dependent variable the proportion of CS, and as factors age (9, 12 and 15 months) and the gestural coordination (with or without gesture). To assure the normality of the distribution, the arcsine square root transformation for proportion data was conducted.

Results show a main effect of gestural coordination ($F(1,10) = 6.339$; $p=.031$; $\eta^2=.388$) and an interaction effect of age and gestural coordination ($F(2,20) = 5.107$; $p=.016$; $\eta^2=.338$). Regarding the gestural coordination effect, we found that the proportion of CS islands is greater when the vocalizations appear without gesture vs. with gesture (.410 vs. .236; $p=.031$). The effect of interaction between age and gestural coordination shows that proportion of CS are greater without gesture only at 9 months (.527 vs. .025; $p=.006$), there are no differences between the proportion of CS islands with and without gestures at 12 or 15 months. We found that when the vocalizations are coordinated with a gesture, proportion of CS islands is greater at 12 months than at 9 months (.330 vs. .025; $p=.010$), and at 15 months than at 9 months (.352 vs. .025; $p=.007$). This means that the relative frequency of CS islands tends to increase with age when coordinated with gesture, especially from 9 to 12 months. There are no differences in CS islands proportion when the vocalizations are not coordinated with a gesture.

Conclusions

Considering the small size of our sample and the preliminary character of this study, results obtained must be interpreted with caution. However, there are some general ideas that can be extracted from our results and that can outline some clues for further research.

Gestural-vocal coordination becomes a very frequent phenomenon in the period of transition to first words. Meanwhile the rate of vocalizations alone remains stable, the rate of coordination of vocalization and gesture rises along the period studied. It is then necessary to consider the multimodal character of infant communication when studying precursors and predictors of language development.

The internal structure of vocalizations shows changes depending on communicative function. The number of SVIs is greater when the function of the vocalization is imperative, but the SVIs are shorter than the vocalizations with declarative function. Gestural accompaniment has also an impact on SVI duration, being longer when they are displayed with a gesture, but this is so only for SVIs with declarative function. Imperative behaviors require an understanding of others as agents, but declarative behavior requires an understanding of others as psychological beings (Liszkowski & Tomasello, 2011). Changes in formal aspects of communicative behavior as the internal structure of children utterances may be reflecting this nuances of other's understanding.

The category that includes the *f(0)* range more similar to speech, appears positively linked to declarative function and negatively related with imperative function or reaching gesture. On the contrary, the category that includes the vocalizations with the highest pitch, appears positively related to reaching and to the imperative function, but negatively related with declarative function.

We did not find a relation between the pointing gesture and the categories more similar to mature speech as clear as we expected. It is possible that the observation situation has favored the appearance of imperative behaviors and reaching gestures, to the detriment of declarative behaviors and pointing gesture. Pointing gesture is linked to declarative function, but it appears frequently with imperative function in our sample. Communicative behaviors with imperative function were much more frequent in our sample than with declarative function. This could affect the results obtained, especially in the analysis that implies narrower categories.

The use of vocal structures with CS coordinated with gestures increases with age, suggesting that maybe multimodality in children's communicative attempts can modify adult's response, as pointed out in previous studies (Goldin-Meadow et al., 2007). Adult's contingent response provides a linguistic model that enables the child to assimilate their own vocalizations to adult language. This can be especially relevant when child express a declarative intention, considering that the aim of declarative behavior is to produce a change not in the physical world, but in other's mental state. More research is needed to explore the effect of adult's contingent behavior on children's vocal and gestural characteristics.

References

Bates, E., Benigni, L., Bretherton I., Camaioni, L. & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. Nueva York: Academic Press.

Bates, E. & Dick, F. (2002). Language, gestures and the developing brain. *Developmental Psychobiology, 40*, 293-310.

Bloom, K., Russell, A. & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language, 14* (2), 211-227.

Bonneh, Y., 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), 1-7.

Capirci, O., Iverson, J., Pizzuto ,E. & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language, 23*(3), 645-673.

Colonnesi, C., Stams, G. J. J. M., Koster, I. & Noom, M. J. (2010). The relationship between pointing gesture and language: A meta-analysis. *Developmental Review, 30* (4), 352–366.

DePaolis, R., Vihman, M. & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics, 36,* 406-422.

Diehl, J., Watson, D., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high functioning autism. *Applied Psycholinguistics, 30*, 385-404.

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*(5), 1271-1279.

Goldin-Meadow, S., Goodrich, W., Sauer, E. & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science, 10* (6), 778-785.

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

Iverson, J. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language, 37*(2), 229-261.

Karousou, A. & López Ornat, S. (in press). Prespeech vocalizations and the emergence of speech: A study of 1005 Spanish children. *Spanish Journal of Psychology*.

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development, 26*(1), 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* (1), 47-66.

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* (2), 247-257.

Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language, 19*(3), 559-580.

Murillo, E. & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture, 12* (1),16-39.

Murillo, E. & Belinchón, M. (in press). Patrones comunicativos multimodales en la transición a las primeras palabras: Cambios en la coordinación de gestos y vocalizaciones.

Nathani, S. & Oller, D.K. (2001). Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations. *Behavior Research, Methods, Instruments & Computers, 33*(3), 321-330.

Papaeliou, C. & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing "communication" and "apprehension". *Journal of Child Language, 33*, 163-178.

Oller, D. K. (1980). The emergence of sounds of speech in infancy. In J. S. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, 1: Production*. New York: Academic Press.

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* (30), 13354-13359.

Snow, D. (2007). Polysyllabic units in the vocalizations of children from 0;6 to 1;11: Intonation groups, tones and rhythms. Journal of Child Language, 34(4), 765-797.

Stoel-Gammon, C. & Cooper, J.A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language, 11*(2), 247-271.

Talkin, D. (1995) A Robust Algorithm for Pitch Tracking (RAPT). In Kleijn, W.B.and Paliwal K. K. (Eds.) *Speech Coding & Synthesis*. Elsevier.

Vihman, M., Ferguson,C.A. & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics, 7*(1), 3-40.

Vihman, M., Macken, M.A., Miller, R., Simmons, H. & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61*, 397-445.