REGULAR ARTICLE

The *Artistic* Infant Directed Performance: A Mycroanalysis of the Adult's Movements and Sounds

Silvia Español[1] · Favio Shifres[2]

Abstract Intersubjectivity experiences established between adults and infants are partially determined by the particular ways in which adults are active in front of babies. An important amount of research focuses on the "musicality" of infant-directed speech (defined melodic contours, tonal and rhythm variations, etc.) and its role in linguistic enculturation. However, researchers have recently suggested that adults also bring a multimodal performance to infants. According to this, some scholars seem to find indicators of the genesis of the performing arts (mainly music and dance) in such a multimodal stimulation. We analyze the adult performance using analytical categories and methodologies of analysis broadly validated in the fields of music performance and movement analysis in contemporary dance. We present microanalyses of an adult-7 month old infant interaction scene that evidenced structural aspects of infant directed multimodal performance compatible with music and dance structures, and suggest functions of adult performance similar to performing arts functions or related to them.

Keywords Performing arts · Music · Dance · Infant directed speech · Early interaction

Introduction

The Musicality of Infant Directed Speech and the Multimodal Infant Directed Performance

The earliest experiences of intersubjectivity between adult and baby on which the infant's further communicative development will be based (Trevarthen 1998) are settled

---

✉ Silvia Español
    silvia.ana.es@gmail.com

[1] National Scientific and Technical Research Council, Buenos Aires, Argentina

[2] Laboratory for the Study of Musical Experience, Faculty of Fine Arts, Universidad Nacional de La Plata, Buenos Aires, Argentina

🌀 Springer

during the infant's first year. These experiences are sustained by intuitive parenting, a general and natural ability of adults to protect, feed, stimulate and teach culture features to their infants (Papoušek 1996a; Papoušek and Papoušek 2002). The Infant Directed Speech (IDS) is one of the best-known aspects of intuitive parenting.

Studies on IDS have shown (a) its functions concerning both regulation of the baby's emotional states and arousal, and favoring language acquisition, (b) its changes as an adaptation to infants' development throughout their first 2 years (Papoušek 1996a, b; Fernald 1989).

At the beginning, IDS enables to capture and keep the baby's attention, offering stimuli to respond to. Mother's speech to infants aged 3 to 5 days contains whispering and repetitive intonations contours. A few weeks after, IDS becomes a powerful device to regulate the baby's emotional state and behavior. Mothers tend to use a higher tessitura and sentences are short and repetitive with clear pauses. Melodicity (understood as the voice's pitch modulation) clearly appears as one of IDS features: melodic contours convey the speaker's intentions and babies perceive some basic categorical messages: approval, prohibition, attention, comfort. Therefore, as indicated by Fernald (1989), in IDS melody is the message. Repetition- variation is a general way in which IDS organizes. Parents tend to use a set of five or six melodic prototypes, which are often repeated, beyond the utterances' varied word content. However, they do not seem monotonous or boring. On the contrary, just like a long set of variations on a musical theme, melodic contours change, increasing and decreasing both, arousal and tension. This form is crucial when trying to extend early adult-baby social play throughout time, where an atmosphere of mutual enchantment prevails. The repetition-variation form is an ideal stimulus for infants because: (1) if the stimulus were always the same they would habituate and loose interest; and (2) repetition generates a certain regularity that allows to anticipate the course of time, to predict what will come next (Rivière 1986/ 2003; Stern 1985). Towards the 5th month, maternal speech changes to a mutual, more animated game: voice is still high, but expression pauses between them are longer and there is a greater variety of melodic contours, rhythms, and dynamic contrasts. These features - especially high pitch, expanded pitch contours, rhythmicity, repetition, and reduced speaking rate - make IDS sound much more musical than adult directed speech (Fernald 1989; Fernald et al. 1989). Bergeson and Trehub (2007) found that mothers use 'signature tunes'—i.e., voice patterns of melodic contour and interval (distance between successive pitches) features, which are idiosyncratically used- in their speech to infants. They found that mothers repeatedly use a small set of individually distinctive tunes, varying the verbal content that accompanied those tunes. This has led some scholars to characterize IDS as a form of music (Brandt et al. 2012).

Besides talking to their babies, adults throughout all cultures spontaneously sing to them (Trehub et al. 1993a). Lullabies are sung in a particular way, moreover, adults from different cultures can identify them thanks to this pecularity, even if they do not specifically know the musical culture to which they belong (Trehub and Schellemberg 1995), and can recognize when the lullaby is being performed in the presence of a baby as different to when this situation is being simulated (Unyk et al. 1992; Trehub et al. 1993b). In this way, the authors highlight the particularly expressive modality characteristic of IDS, that goes beyond the particular structure of the repertoire of lullabies itself. Despite highlighting the difference between singing lullabies with the expressiveness triggered by the baby's presence, such expressiveness is not described in the terms

musical studies use to analyze features in expressive performance - mainly by using timing, dynamics, articulations and timbre (or spectrum quality).

Multimodality is another prototypical trait of IDS. Since the earliest works in the 70's, it's been generally considered that IDS' vocal elements are embedded in multi-modal patterns of preverbal communication that include tactile, kinesthetic, and vestibular forms of stimulation (Papoušek 1996b). In turn, different modes of temporal organization of the dyad's earliest exchanges were detected: turn taking interactive synchrony and other co - active patterns. Bateson (1979) observed a turn taking pattern during the baby's first months of life which he called protoconversation. Although this pattern is basically produced between vocalizations, he also observed movement, look, and facial expression. Condon and Sanders (1974) noticed that from the earliest days the baby moves in precise and sustained movement segments synchronic with the adult's articulated speech structure, generating an "interactive dance". Stern and colleagues (Stern 1974; Stern et al. 1977) observed the temporal organization of multimodal patterns that adults direct to babies finding kinesic and vocal phrases—either performed independently, jointly, or alternatively- that are repeated creating larger units. Discrete smaller units or maternal acts are repeated conforming more extended ones according to the baby's response. If the mother manages to gain the baby's attention, he will produce smiles, head movements, or vocalizations that will work as a stimulus for the mother to continue developing more concatenated maternal acts. This kind of temporal organization of multimodal maternal performance emerges during the first months of life to decrease during the second half of the first year, when the baby starts to become interested in events and objects in the world. Technological advances, which allowed audio computer analysis of adult- infant vocalizations, led these pioneering works to be promptly abandoned by studies restricted to the temporal structure of dialogue. Interest in the kinetic component and the idea of multimodal interaction continued to be present in later research, but the initial extremely laborious coding analysis of body movements, after watching videos frame by frame, were left behind (Jaffe et al. 2001).

Nevertheless, Stern (1985) went further with the idea that adults offer babies a multimodal patterned performance noticing that under this situation they resemble music listeners or spectators of an abstract dance. In this line, Dissanayake (2000b, 2001, 2014) suggested there are antecedents of the temporal arts in early interactions linked to our phylogenetic past: The extended period of vulnerability in human breeding created a selective pressure for psychological proximity and cognitive mechanisms, ensuring an extensive and better maternal care. This was the reason for the origin of "*making special*" in affiliative behaviors, present in other primates - such as facial expressions, gestures, movements and sounds - through dynamic, rhythmic and crossmodal modeling. To attract their baby's attention parents elaborate or make special ordinary sounds and movements: they repeat (phrases, movements, sounds, facial expressions) and exaggerate (its melodic contours, amplitude, duration and pauses between them). This elaboration -which entails saliency, novelty, expectation, and therefore emotion- of what already where affiliative guidelines, has the power to directly lead to a state of inherently pleasurable mutuality. According to Dissanayake, the speech modifications of IDS are only one aspect of the dynamic, rhythmic and crossmodal modeling of the whole parental performance that includes sounds and movements in the same level. Maternal behavior uses repetition in any possible way:

vocalization, movement, and tactile stimulation. But what a mother does in all possible modalities is not repeated in an identical manner: she makes subtle variations in speed, suspense, vocal accompaniment, in such a way that babies can predict what comes next, avoiding habituation or boredom. Dissanayake (2008) suggests that "artifying" - or making special by means of aesthetic operations- is the ancestral activity or behavior that gave rise to, and continues to, characterize or imbue all instances of what today we call art. Miall and Dissanayake (2003) have called *Babytalk* to the infant directed vocal and kinetic performances that typically appear between the baby's 6th week and 6th month. They were pioneers in studying infant directed performances with poetic resources in *Babytalk*, founding parallelisms, hyperboles, alliterations, assonance and rhymes, and episode organization with clear beginning or introduction, ending, and, sometimes, refrains or coda. They consider *Babytalk* as a privileged resource for unfolding the first dyadic psychological connection, namely experiences of primary intersubjectivity (Trevarthen 1998).

In recent years, with the help of technological advances related to the capture and analysis of movement, a renewed interest in the kinetic component of multimodal performance awoke. Brand et al. (2002), Koterba and Iverson (2009) and Licata et al. (2014) analyzed adult movement from the perspective of action to goal, such as grasping an object, finding that, in their interactions with babies 8 months and older, mothers modify their movements. They named this modification and simplification of gestures, actions, or signs that adults use while interacting with infants or toddlers "motionese", noticing it might assist infants in processing human action, specifically to attend to objects and explore around them. The attention to movement in speech directed to infants parallels the interest the role of body and movement has received in music studies. In this context, studies on embodied music cognition pay particular attention to body dynamics in order to comprehend processes involved in the practice of musical meaning, typically in social contexts (Godøy and Leman 2010). Advances in the study of embodied music cognition, along with the methodological developments involved, particularly those related to the capture and analysis of movements, are emerging promising for directed to infant performance studies. However to this date, we do not have a hermeneutical development to interpret the enormous wealth of data offered in terms of movement quality variables (such as momentum, position, ground speed, etc.) as considered in the context of aesthetic elaboration (along the lines of Dissannayake's making special or "*artifying*" (2000a, 2008). In that sense, two theoretical constructs emerging at the intersection between the study of early interaction and performing arts bring new tools that help understand early interactions highlighting the role of movement. The first one is *Communicative Musicality* (Malloch 1999/2000; Malloch and Trevarthen 2009): interaction modes where the participants' rhythms, sound and motor gestures mate and coordinate and a particular way of "being together" is experienced from birth and throughout life, in communal activities and temporal or performing arts. The second one is *Forms of Vitality*. This is the term proposed by Stern (2010) to capture the way in which the human mind deals with dynamic experiences, crucial both to interpersonal encounters and performing art experiences. Forms of vitality can be grasped from experiences and can be directly observed in the behavior of others. They concern the "how", the manner in which things are done. A form of vitality is a *Gestalt,* a spontaneous integration emerging from holistic experiences of movement, time, force, space and directionality/intentionality that come with it. Both

phenomena have been particularly observed as occurring during the first year: the first manifestations of *communicative musicality* tend to appear towards the 2nd month, with a peak of increasing ability showing at about 6–7 months (Trevarthen and Reddy 2007), likewise, in early social play, parents play with movement and sound modulating *forms of vitality* to enjoy themselves in intersubjective encounters (Stern 2010). Both theories establish links between early interactions and the temporal arts, emphasizing the continuity between the intuitive behavior at stake in early intersubjective encounters and elaborated behaviors involved in social phenomena we recognize as art forms. According to both theories the *repetition-variation* form is a core feature of IDS, music, dance, and poetry. Simultaneously, in the field of communicative musicality, quantitative studies on movement involved in communicative expression of gesture in music performance are carried on (Lee and Shögler 2009), and the study of the development of infant vocal and bodily rhythms is back on track assisted by techniques that facilitate sound and movement microanalysis (Mazokopaki and Kugiumutzakis 2009).

The features of sound and movement displayed during the multimodal 7–month-old-infant directed performance will be the focus of this paper. Our analysis sets on the adult's behavior, instead of spontaneous reactions of the communicating infant. This distinguishes our approach from those used in previous studies (e.g., Brand et al. 2002), setting it closer to Miall and Dissanayake's (2003) pioneer work on the poetics of babytalk. Our research is not about 'motionese'. We are interested in adult-to-baby sound and movement performances that babies reciprocate by smiling, moving and vocalizing. This approximates our work to the pioneering research of Stern (1974) and Stern et al. (1977). If as suggested by Dissanayake (2008), "*artifying*" is the ancestral activity that gave rise to, and continues to characterize the arts, watching adult directed to baby performances provided with art tools that have not yet been used can help us understand and offer new empirical evidence about their organization. Some of these tools come from studies in music performance and movement analysis in contemporary dance.

Performing Arts and Multimodal Infant Directed Performance

Music Performance Studies Relevant to Analysis of Multimodal Infant Directed Performance

Music performance studies began to develop in the 80s on the basis of research in psychology, systematic musicology, and musical praxis, focusing on the expressive aspects of music performance—as compared to the normative aspects indicated by the musical score or the corresponding musical standard (See Shifres 2014 on problems about relationship between expressive and nominal values in music performance). This field of study emphasizes on music interpretation. Of particular interest to this study is that, almost naturally, this field was extended towards addressing communication between musicians and listeners. The aim of these works is not only descriptive (of musical performance's expressive features) but also explanatory of the intersubjective relations elaborated from it. The perspective of Music Performance Studies links descriptions of musical phenomena with the characteristics that make each performance a unique and unrepeatable event, mainly focusing on the identification of behavioral patterns that could account for idiosyncratic and expressive aspects of the performance not captured by musical notation.

Studies in music performance extend the object of study from the systematization of observations of issues beyond the scope of musical notation. Thus, for example, it is not only important to recognize how the durations of sounds can be identified with durational discrete categories (proportional) but also how the actual sound duration systematically differs from those categories. Therefore, defining "systematic" differences is crucial. Particularly, the key in the study of expressive variables (the most important are timing and dynamics) is to see how the regulatory framework (the metrical framework and dynamic range respectively) is set, wherein that regulatory framework deviates (for example, depending on what does the meter lengthen or shorten, or dynamics increase or reduce) and depending on what structural, emotional, communication issues, etc. it deviates from. A conspicuous example of this paradigm is the pioneering work of Sundberg and Verrillo (1981) in which the authors describe and typify the final phase *ritardando* as a typical case of systematic expressive deviation. Under it, the musical phrases are expressively demarcated through notes progressively longer as the end of the phrase comes closer. Subsequently Todd (1985) explained the systematic nature of such lengthening relative to hierarchies of phrases. According to this idea, the boundaries of phrases embedded in larger units are marked with smaller retardandi than the limits of more comprehensive musical utterance. Thus music performance studies sought to find behavioral patterns related to expressive timing according to structural aspects of the composition. Because the practice of musical performance in the nineteenth century was characterized by the use of durational deviations—e.g., lengthening and shortening of sounds—in regard to the nominal durations indicated by the score through the use of musical figures (quarters, eightnotes, etc.), with an expressive purpose, and that use was, back then, called *Rubato* (in Italian: stolen, or stolen time). Similarly, other aspects of expressive characterization of sound in music, such as dynamics (or loudness modulation), and articulations (the expressive characteristic derived from the length of the pauses between two consecutive sounds, combined with type of the sound's attack) are relevant. In this context increases or decreases in sonority regarding a level media related to structural aspects are studied. In this line, for example the use of *diminuendi* (decreasing intensity) relative to the end of phrases was described. Finally, it is worth noticing that music performance studies progressively covered other areas beyond structural explanations. Thus, since the 90s, more than a few studies have inquired on the expressive deviation systematizing related to the intended emotional content (Gabrielsson and Lindström 2010) or to kinetic content (Kendall and Carterette 1990).

In general, studies in IDS have adopted a musicological canon for describing its musical features, based on structural analyses according to notational categories. In that way, IDS is usually described in terms of rhythmic patterns defined by discrete categories of proportional durations, vocal sound pitch patterns in terms of melodic contour (ascending-descending pitch), and regular metrical structures, among others. Often these descriptions are accompanied by sound data's transcriptions in music notation. Beyond such, the structural descriptions in musical terms that abound in the literature on IDS, primarily coming from the work of the Papoušeks, the performative perspective becomes the natural environment for the study of communication in early childhood. To study IDS as an expressive musical discourse requires detecting patterns of systematic deviation of expressive parameters related to some normative value. After

this, some speculations about the meaning of those deviation patterns in the context of dyadic interaction are required.

Contemporary Dance Studies Relevant to Analysis of Multimodal Infant Directed Performance

With the emergence of modern dance, movement's dynamic experience stood in the spotlight. Choreographers looked to non-linear, random, and non-narrative forms, and the expression of movement came to the fore (Stern 2010). According to Laban (1971), each phrase of movement, every little weight transfer, a simple gesture of a body part, reveals some trait of feeling. Any move can be performed with multiple expressive variations: an apple can be eaten in a hungry and nervous manner, with focused movements triggered rapidly and suddenly or in a languid carefree way with scattered soft and slow movements that are hardly complete. Laban-Bartenieff's Movement Analysis System describes dance movements from their *Shape* and *Effort* (Laban 1971; Davies 2006; Newlove 2007). Shape and Effort permit capturing the dynamics and expression of movement. Shape comprises (i) *Body*: concerns body parts used in the observed movement; (ii) *Dimension or space*: concerns the relation of the body with the surrounding space, being the kinesphere the space limited by the points reached by the limbs without changing the body position, and defining three possible 'attitudes' of the body in this space: vertical (up-down), horizontal (right-left) and sagittal (front-back); (iii) *Form*: understood in terms of the basic opening/closing opposition of breathing, on vertical, horizontal and sagittal planes, in such a way that three forms are obtained: rising-descending, spreading-enclosing and advancing-retiring. In turn, Effort indicates the movement quality according to the attitude of dedication or fighting against space (direct vs. flexible), time (sudden vs. sustained) and weight (strong vs. light). While combining these factors eight basic types of Effort are obtained: (1) *pressing* (direct - sustained - strong), (2) *gliding* (direct - sustained - light), (3) *punching* (direct - sudden - strong), (4) *dabbing* (direct - sudden - light), (5) *wringing* (flexible - sustained - strong), (6) *floating* (flexible - sustained - light), (7) *flicking* (flexible - sudden - light), (8) *slashing* (flexible – sudden – strong). Each of the eight basic types of Effort can be performed in a continuum Flow: from a free flow (difficult to stop) to a bound flow (easy to stop). The theory highlights the feelings running along these efforts. According to Newlove (2007) every basic type of effort is accompanied by a movement sensation: pressing is accompanied by a heavy-long-threadlike sensation or sinking, gliding is accompanied by a light-long-threadlike sensation of elatedness, punching is accompanied by a heavy, short-threadlike sensation of dropping, dabbing is accompanied by a light-short-threadlike sensation of stimulation, wringing is accompanied by a heavy-long-pliant sensation of relaxation, floating is accompanied by a light -long-pliant sensation of suspension, flicking is accompanied by a light-short-pliant sensation of excitement, slashing is accompanied by a heavy-short-pliant sensation of collapsing. The movement sensation of flow relates to fluency: free flow with a "flowing on" sensation, bound flow with the feeling of pausing or being withheld Laban-Berteneieff.

This paper presents an empirical study on Infant-directed multimodal (sound and movement) performance (IDP). It recognizes antecedents in the previously mentioned works by Stern and colleagues. Nevertheless, unlike those studies, the present one is

based on performing arts methods. Its starting point is the idea—previously developed by Trevarthen, Stern, Dissanayake, among others—of the continuity between the intuitive behavior at stake in early intersubjective encounters and the elaborated forms of behavior involved in social phenomena we recognize as art forms. It hypothesizes that, given this ontogenetic link, performing arts methods may shed light on some features of adult-to-infant directed performances that have not been yet identified.

Aims

This work aims to analyze infant-directed performance using some analytical methods coming from performing art studies in order to uncover some aspects of infant directed performances that remain unknown. To reach such a goal an adult-7-month-old infant interaction scene was analyzed according to two methods of performing arts: (i) Laban's Movement Analysis System and (ii) and expression analysis in musical performance.

Method

One adult-infant interaction scene was selected from the audiovisual register of a previous longitudinal case study research about preverbal communication between the infant's 7th month to the 24th month. The mentioned study was carried on by the first author of this paper who interacted with the first son of a Spanish middle class family (Español 2004). The native language of all the participants is Spanish. The selected scene belongs to the first session of interaction between researcher and infant; it was recorded at the baby's home when he was 7 months old. It is characterized by spontaneity, playfulness and naturality, and some everyday life contingencies can be observed.

Choosing a researcher-infant dyad (a particular case of the classic adult-child dyad) permits to simultaneously take two fundamental perspectives on knowledge generation in reference to interactive systems: emic and ethic—correspondingly internal and external to the analyzed system—according the use of this term in ethnography and semiotics (Riba 1990). The researcher can register her own experiences and reactions on-line, and thanks to the availability of recorded audiovisual resources she can observe them remotely, retrieving her own previous experience. The selection of the dyad researcher/baby is consistent with Reddy's (2008) psychology of intersubjective phenomena's methodology proposal. Consequently, the scene selection was made on the basis of (a) the researcher's judgment on the quality of intersubjective involvement experienced at the time the scene occurred, and (b) the presence of clear patterns of the baby's social involvement (warbling, smiles, kicking, and intense eye contact) in the recorded material.

The analysis focused on the adult's performance, which was observed considering its sound and movement. Although we observed the baby's performance, our audiovisual analysis is not on the spontaneous reactions of the communicating infant itself. Separated analysis of sound and movement belonging to this scene were presented elsewhere (Español 2007; Shifres 2007). In this paper we present a conjoint analysis of the performance's sound and movement.

To analyze the scene's sound component, the audio was detached from the video and processed with Adobe Audition 6®. Besides certain sound analysis functions, the

software allows identifying the onset of each sound with millisecond precision. This makes it possible to accurately determine the duration of each sound and establish patterns of behavior over time, often assisted with graphic devices. This sound editor also allows separating each sound in order to isolate them and to analyze its attributes individually. Another software, PRAAT (widely used in research for analyzing spoken and sung voice), was used for such analyses, mainly intensity, pitch and sound formats. The software provides measurements of such attributes at different scales allowing versatile descriptions and interpretations according to the purpose the given explanation.

Movement was analyzed using Laban-Bartenieff's Movement Analysis System. Movement categorizations of *Shape* and *Effort* were made using Anvil 4,0 -a video-annotation software developed by Michael Kipp (2004). The software allows (i) to draw each category on a time line and watch the video image at the same time, and (ii) to manipulate the video's speed reproduction. The five categories of Shape and Effort belonging to Laban's system were included, and several video observations were run to analyze them separately (i.e., first, the video was observed to analyze body, then, dimension, etc.), keeping track of the corresponding observation on the annotation bar (that displays a time bar according to the video frames). This analysis was run without listening to the sound track aiming that sound and linguistic components did not affect the observers' judgments. Later, iterative observations of the video with sound were carried out, and the analysis of sound and movement was studied jointly.

Results

General Description of Scene

In the scene, the baby is sitting on his baby-seat and the adult is on the floor beside him. The 1.20 min scene is initiated by a movement made by the adult in order to attract the baby's attention, and ends because the baby turns his face around, probably in an attempt to reduce his level of arousal. It is a clear case of what Stern (1985) calls mutual behavior regulation. In between we find a sequence of sound and movements units, well organized and less organized moments. For a better analysis understanding, the scene was divided in three sections. The boundaries between sections are marked by the reestablishment of calm after an external stimulus suddenly breaks in.

General Features of the Analysis

Sound and movement microanalyses of the complete scene were run. Due to clarity and space reasons, only the sound and/or movement microanalyses corresponding to some relevant points are shown here. These microanalyses reveal the alternating predominance of sound and movement components in the configuration of the performance (when movement was leading, movement analysis was privileged and vice versa), and show the theoretically most interesting details of each section. Likewise, a multimodal analysis of each section (as an observational description) and an interpretation of the data are presented with the purpose of relating the results of all sound and movement microanalyses into the frame of the complete section. For each section a general description of it is introduced at the beginning.
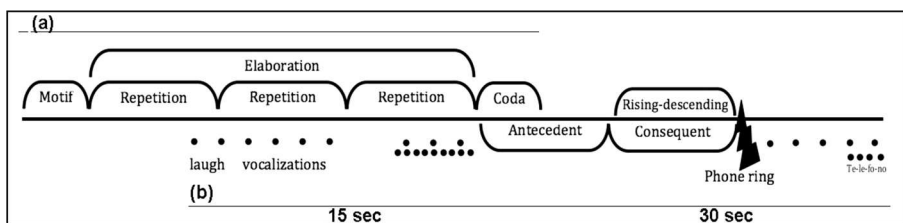
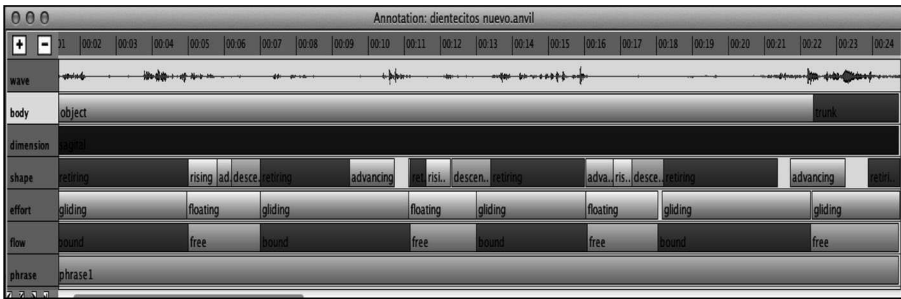## First Section

General Description

The first section comprehends the action of covering and uncovering the baby's face with a shawl. Initially the baby (H) is looking the other way while the adult (S) contemplates him and makes small movements to catch his eye, but as this turns unsuccessful she appeals to a way of calling the baby by slowly covering and uncovering him with the shawl. Immediately H starts to follow the direction of the shawl turning his head until his gaze crosses that of S. From there, what began as an almost casual movement becomes an ostensibly elaborated one, initiating a pleasant and joyful sharing play. The movement with the shawl is of paramount significance concerning the following kinetic-sound structure: firstly, it constitutes the *motif* that will be repeated with variations, determining the *thematic content* of the performance as a whole, and secondly, it defines the basic time span for the units that will structure the entire section. After the silent start, the sound comes just after S laughs (which takes place towards the end of the first repetition). This is also a remarkable fact, since it seems that only when the sound inrushes through an almost involuntary act such as laughter, the sound dimension is understood as a forum for exchange. By this means, the inclusion of sounds does not alter the atmosphere of intimacy achieved by movement and eye contact. On the contrary, sounds are added to this *aesthetic of encounter* contributing to the exchange. Towards the end, before the phone sound breaks out the atmosphere, the performance is kept organized by the sound component produced by S while movements show a lesser organization. The section ends when S quietly returns to the center of her kinesphere, after the calm interrupted by the phone ring is restored. Figure 1 displays a kind of section's script and highlights the involvement and relative importance of the sound and movement components of the multimodal performatic structure, which will be described below. Therefore, before presenting the multimodal analysis of the section, a microanalysis of movement, for the first part (a), and sound, for the second one (b) will be introduced.

Movement Microanalysis of Part (a)

Figure 2 shows show the varied repetition of the movement of removing the shawl brings about a movement phrase (23 s.). Repeated labeled rectangles allow observing

Fig. 1 Outlining of sound components (*below the horizontal axis*) and movement (*above the horizontal axis*) of the IDP scene's first section. The *horizontal axis* represents time. The *arcs* represent units of thematic sense (phrasing), and *points* represent the beat pulses of different hierarchical levels (tactus, subdivision). The sound elements that favor these beats (laughs, vocalizations, words, etc.) are also indicated

Fig. 2 Movement description of the beginning of section 1 according to the categories of analysis Laban - Bertenieff, scored using ANVIL 4-0. The *horizontal numerical scale (above)* indicates the seconds. The *first row* shows the dynamic envelope of the sound signal. The following rows correspond to the variables of Laban system (body, dimension, shape, effort, and flow). Finally in the last row units that emerge from the analysis of these variables are noted

the categories of Laban's Movement Analysis System that characterize the motif and its varied repetitions (elaborations) of a movement phrase.

Thus, microanalysis shows: (i) the initial movement of removing the shawl (3 s.), with Effort (row 4) *gliding*, and Shape (row 3) *retiring*. (ii) The elaborated repetition of the previous movement: *gliding-floating* pattern, in terms of Effort, and *rising -advancing - descending - retiring* pattern, in terms of Shape, with Flow (row 5) alternating between *bound* and *free*. (iii) Other two repetitions of the motif with variations of the Effort pattern (*retiring - rising - descending* and *advancing - rising - descending*) and a variable extent of *retiring*. (iv) *Coda or* closure with simplified movements by (a) taking away the object - characteristic element of both the motif and its elaborated repetitions (row 1 alternating *object-trunk* labels), and (b) simplification of the *rising-descending* pattern into the simpler *advancing-retiring* (row 3). The last *advancing* movement is the phrases' widest, marking both the closest physical contact point between the dyad members, and the longest movement (until reaching the target). The recurrent *retiring* movement closes the phrase. The adult backs down and reaches the center of her kinesphere.

Sound Microanalysis of Part (b)

Figure 3 displays a timing analysis of the section's sound component. Each bar is a complete phrase, and the horizontal axe represents time in msec. The vertical axe at the left represents the first "downbeat" of each phrase, while the dotted bars indicate the other downbeats of the phrases according to the onset of the accented syllables and an equal division of the silent time span. In that way it is possible to appreciate regularity of vocalizations and temporal relations among the phrases. This analysis shows that the sound component introduces a *beat level* of time regularity, and then a subordinate level (beat division), as it was suggested with the dots in Fig. 1. After time regularity is well established, the regular phrase sequence seems to finish with a lengthening of the last beats of phrase 5. After that, phrases 6 and 7 show no regular pulse. At the end of phrase 7 the phone rings. The last phrase can be seen, in that way, as an intent to reestablish regularity, firmly marked when S explains to H that the disrupting sound came from the telephone ("te-le-fo-no").

Figure 3 also displays the main data related to pitch and intensity details, where they are most relevant. Pitches are indicated in term of proximity of its F_0 (fundamental frequency) with a musical scale category (according to the software algorithm), and intensity in decibels (Db). Therefore, some relations of pitches (ascents and descents, registers, and so on) and intensity (crescendos, diminuendos, echoes, and so on) can be appreciated. Concerning pitch, two relations are interesting. Firstly, at phrase 3 an ascent is congruent with the interrogative character ("Do we get the shawl?"). This ascent raises the vocal register in such a way that the following "ti-ti-ti-ti-ti…" are noticeable high. Secondly, the end of phrase 6 is high while the end of the following phrase (7) is low, contributing to configure the antecedent-consequent idea (since an *antecedent* is understood as every phrase that creates tonal tensions, that are relaxed at the end of the following phrase, the *consequent* (Berry 1986); in virtue of the *gravity force* of the musical tonal system (Larson 2012) a high pitch will tend to proceed to a lower one in order to reach more stable conditions). Concerning intensity, it is worthy to note a diminuendo at phrase 5 ("I get it, I get it, I get it"), where each repetition of the text is progressively softer. In that way, this diminuendo is accompanying the end of the regular sequence.

Multimodal Analysis and Interpretation

Movement along the three repetitions-elaborations is temporarily regular. However, in the second repetition, the motif of retiring the shawl is accelerated and lasts a second less. This acceleration can be interpreted as expressive as long as (a) it is linked to H's demand to remove the shawl, and S's caring attitude to meet that demand; (b) it is the second repetition, whereby the interval expected time (the normative measurement of that interval) is already established, so that the change is shown as a *deviation* from the expected measurement. Furthermore, the acceleration is accompanied by vocal sounds. As anticipated, S laughs at the end of the first repetition of the motif, and the emergence
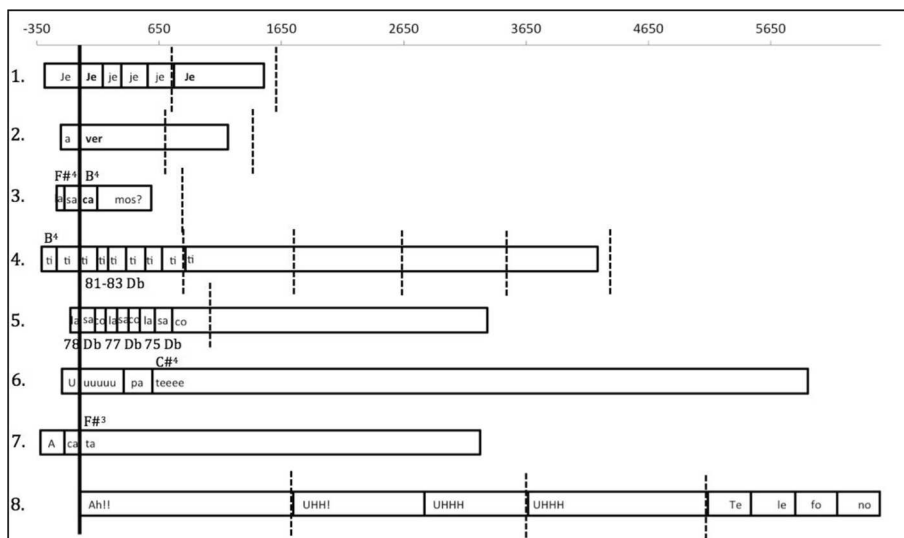
Fig. 3 Time analysis of sound component. Phrases are represented by *bars* (the text is transcribed one syllable in each box). See explanation in the text

of her laughter's sound enables the performance's sound dimension. During this repetition, sounds are progressively shorter and higher. As it was said, the repetition finishes with a change of voice register ("ti-ti-ti-ti…"). In that way, she accompanies the shawl's removal of the second elaboration (shortened), with sharp, regular, short sounds (stacattos): "ti -ti -ti -ti -ti -ti". In this way, the sound accompanies the expression shortening of the time unit.

The third motif repetition is again extended with the aim of closing the entire sequence (last elaboration), in the manner of a final *ritardando*. Again, this temporal deviation (this time shortening) is understood as such not only for its closing location in the sequence but also for the way in which sounds accompany it: a) first, the movement occurs in silence (the 4 m of silence in phrase 4, Fig. 3), which after the heard sounds increases the expectation for the shawl removal, and b) a breathed-in vocalization accompanies the last shawl withdrawal ("I take it out, I take it out, I take it out"*)* with an according decreasing of loudness, so that all the sequence ends with a lengthening (*ritardando*) and a decreasing of loudness (*diminuendo*).

Then, what appears as the ending sequence of the movement's elaborations -the simple advancing-retiring shape pattern- is accompanied with a vocalization ("uuuúpateee!"). The vocalization occurs precisely at the time of *advancing*. This is the longest *advancing* movement of all repetitions of the motif and has the same duration as the "u-uh-pate" sound. The recurrent *retiring* movement occurs in silence, and closes the phrase. The adult backs down and reaches the center of her kinesphere. The ("uuuúpateee!") vocalization is in a higher pitch range than the immediately previous vocalizations (similar to the one at "tititititi"). Returning to this tessitura rebuilds expectations. The expectation generated in the previous unit, suggests the achieving of a consistent unit that will satisfy expectations and close the loop. So although movement seems to close the previous episode, sound is telling the baby that the performance continues. Another important aspect that unifies the *antecedent-consequent* structure is the articulation contrasting the vocalizations of the first sequence of motif and elaborations. While there, vocalizations tended to emphasize each syllable in a *staccato* style ("ti-ti-ti-ti-ti-ti" and "la sa-co, la sa-co" in which each syllable was very short and with a very strong consonant), in this part vocal articulation binds sounds in a more *legato* style.

S puts the shawl back up covering her face and upper body once again; at that precise moment the phone rings. This sound interferes with the atmosphere of intimate contact previously accomplished, and with the performative plot continuity, mainly due to the pitch range and loudness that differ greatly from those that surrounded the scene, just at a time when what is being built seems to depend on the thread of sound level.

S lowers the shawl a little bit uncovering only her face, pronouncing a breathy "ahhh". She freezes her movement and talks to the baby establishing eye contact. Here, freezing is a major movement innovation. Then, she pulls down the shawl, looks towards the phone, looks back at the baby and sustaining eye contact she comes closer to the baby, with a soft movement at the sagital plane, and says te-le-fo-no (phone) returning to her kinesphere. Freezing occurs in response to the phone's sound and accompanies the expression of surprise, but then it breaks into the initial unit movement's repetition. The surprise and the abandonment of the movements' design appear to be offset by a systematic action in the sound level to *take control of the situation*. Thus, from the first sign of surprise, "Aaah!" a series of interjections follow one

another ("uy! ah!, uh!"), retaking the initial pulse's regularity. Similarly these interjections retake the upper register. Notably after 4 beats, control is regained. And already at a low register, to reduce tension, provide a safety message and end the interruption, S pronounces very clearly, in detail and with syllabic isochrony the word *phone*: "Te-lé-fo-no" (Fig. 3). In this way, the section ends with the confidence to continue with the pleasant and joyful exchange.

Second Section

General Description

The second section begins just as the first ends. It consists of two well distinguished parts. The first one is a typical *peek-a-boo* game, in which S says to de baby from behind a shawl, "Habibi… Habibi…(appearing) Here is Habibi! (behind the shawl again) Habibi, I don't see you… Habibi, where are you? (appearing) Here you are! … (then we might be able to hear the word) noooo (when H tries to suck the shawl)". Unlike the multimodal performance at the first section, this one has a more complex, hierarchical and embedded structure, probably due to its more tradition-bound nature. Generally, a *peek-a-boo* presents two contrasting elements: during the first one the dyad loses eye contact (usually by interposition of a blanket, or another object), and during the second visual contact is recovered while removing the object. In this case, the first element is elaborated by repetition. Likewise, the entire sequence is elaborated by repetition. However, as this elaboration relaxes some tonal tensions (as it will be seen next), this repetition can be understood as a resolution, and the completed cycle as an *antecedent-consequent* structure (Fig. 4).

The second one involves a new fortuitous event that alters the sharing atmosphere and the subsequent strategy to recover it. The top holding the baby-seat in the position in which it was set gives in, descending two positions. Unlike the first interruption (a sudden phone ring) this time the irruption is clearly multimodal: on one hand, H's position changes (descending abruptly) and that two-time-descending of the chair runs along two loud sounds coming from the seat's top when unhooking. The strategy to overcome the fact is also multimodal: both stimuli (kinetic and sound) give S the items to build the next unit of behaviors to handle the contingency that threatens the encounter's continuity. For these reasons there are relevant sound and movement data in both parts, therefore they will be analysed separately.
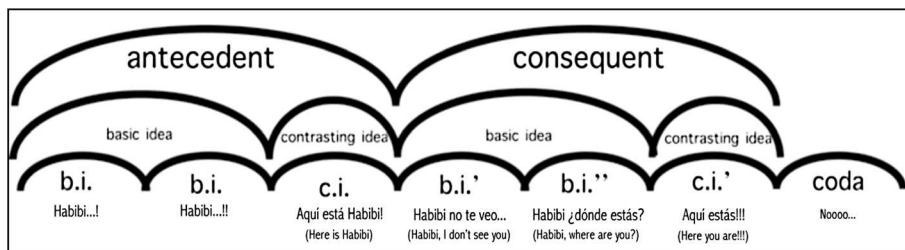
Fig. 4 Grouping structure (phrases) of the *peek-a-boo* play. The repetition of the basic and contrasting ideas in a hierarchical structure covering an overall ratio of *antecedent-consequent* is observed. Below, the original text version in Spanish pronounced while playing (*top*), and its English translation (*below*)

Sound Microanalyses of the Peek-a Boo Play

Figure 5 shows the duration of each syllable in milliseconds in order to display the duration of each sentence. Since pauses are relevant in this analysis, they are highlighted in black in the graphic. The figure allows appreciating that when there are more words, syllables are shortened in proportion with the same syllables in previous phrases with less text, preserving the organization of rhythmic pattern according to hierarchical categories. For example, the first syllables of unit 4 last on average half what they last in units 1 and 2, in a way that they can be represented by musical figures linked by the ratio 1:2.

Also, the relative duration of the seven units and, especially of the pauses separating them (in black), clearly indicate the demarcation of the hierarchical organization. Thus, longer pauses occur at the boundaries between units of superior hierarchical order, according to the embedded phrase structure (see Fig. 4), for example at the end of unit 3, and even more so at the end of the unit 6. This timing management is similar to the one that can be identified in the expressive musical performance of such structures (see Todd 1985). Note also how unit 4 is shortened returning to the time span similar to unit 1; that is setting the parallelism between both units. According to studies in expressive musical performance, the indicated microtiming variations would be intended to highlight the rhythmic relationships between units and the hierarchical organization of the whole unit's phrasing structure (see e.g., Todd 1985; Sundberg 1993; Palmer 1997). A *diminuendo* accompanies the expressive resource *rallentando* of the ending of period over the last unit, also a resource much used in musical performance to indicate formal unit closure (Todd 1992).

However, considering the section's timing in more detail, a tendency to use sound and rest lengthenings or shortenings in order to generate tension and relaxation is observed. For example, the relationship between units 1 and 2, based on the same text, does not account for a change of rhythmic pattern, but a lengthening of such. Thus, unit 2 stretches every sound of unit 1. Although this stretch helps to highlight the hierarchical upper level unit comprising both units (1 and 2) it undoubtedly contributes to increase tension and therefore expectation towards the end of unit 2, of which the final pause is also lengthened.
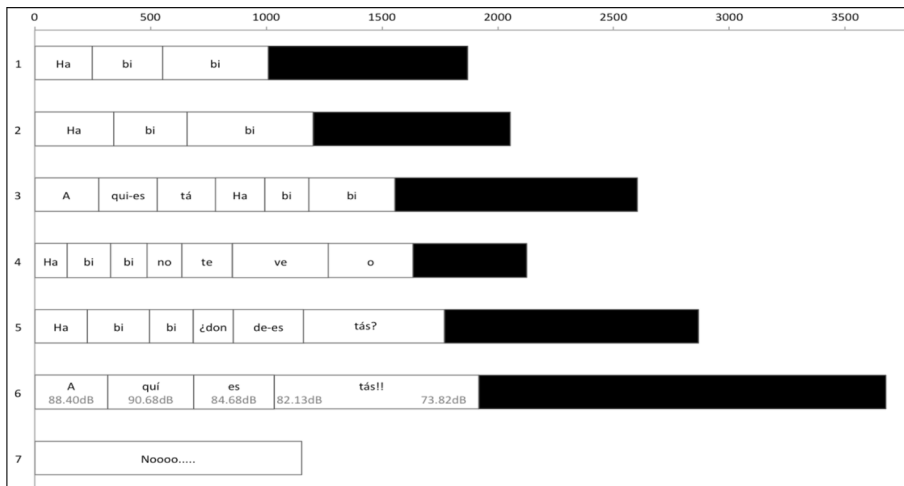

Fig. 5 Syllable duration (*in white*) and pauses (*in black*) in msec. of each of the 7 minimal units corresponding to Fig. 4's graphic

Similarly, if the lengthening of unit 2 is compared to that of unit 5, we can notice that the repetition of the basic idea in the consequent generates further expectation from the final pause lengthening, before tension resolution in unit 6. Thus, the consequent's relationship tension-relaxation is much more emphasized than the antecedent's. In this way, sound and pausing lengthening, besides clarifying the hierarchical structure, contributes to generating expectation, sustaining attention, and increasing tension, which is crucial to the sense of the game. In addition, the first time, the contrasting idea has a *feminine ending* (the last sound is unstressed), while the second time presents a *masculine ending* (on a stressed syllable). Therefore the closing is stronger the second time, hence the notion of *antecedent* (which rises tension) and *consequent* (which resolves tension) is derived.

Concerning the pitch component, the whole fragment is the scene's most tonic one. After both spectral and melodic (F_0) analysis, it was possible to clearly identify pitches (Fig. 5). Combining data of durations, prosodic accentuation of the used words, and pitch analyses, a transcription in music notation might be outlined (Fig. 6), in which it is possible to see some pitch relationships used as melodic motives. There is a melodic motif (basic idea) consisting of an interval (approximately a Major 6th.) ascending, with a second sequenced 6th, in the reiteration of the idea. The higher repetition of the basic idea ("Habibi!") contributes to increase the tension that relaxes with the contrasting idea, which, in contrast, displays a falling interval, compensating the previous melodic movement. In the consequent, the motif's elaboration of the basic idea is the reiteration of the highest tone, which finally resolves at the end of the contrasting idea (last tone). The *Coda* where S articulates the word "Nooo!" indicates she reproves H's action of taking the shawl to his mouth. This coda articulates one long syllable, and the lowest pitch of the entire section. S uses duration and pitch to express the sense of disapproval.

Multimodal Analysis of the Peek-a Boo Play and Interpretation

Movement analysis yields an organization of kinetic features that is highly consistent with the organization of sound. Figure 7 shows such movements organized into two major units -corresponding to the organization of antecedent-consequent- and a coda.

Fig. 6 Music notation transcript taken from the sound signal's duration and pitch measurements, interpreted according to expressive phrasing criteria. *Crosses* indicate that these sounds are articulated in a quasi sung manner (*Sprechgesang* style), and the *square notes* indicate the spoken voice. The *upwards* and *downwards peaks* indicate shifts in pitch with respect to the annotated pitch
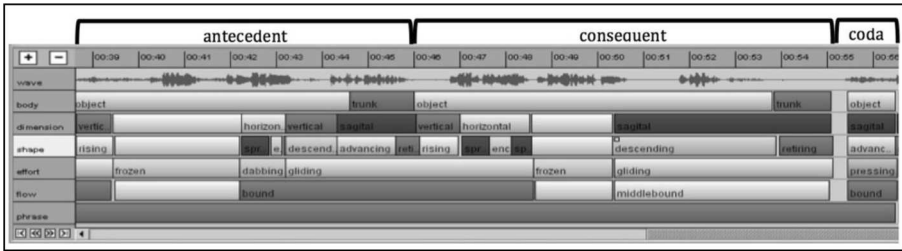
Fig. 7 Laban analysis of movement's Anvil notation (in order to compare with sound phrasing see upper row displaying the sound wave envelope; it is possible to appreciate the seven minor units)

The *retiring* shape form coincides with the end of the larger units. It is also noticeable how the *retiring* is longer as the end of the consequent is more lengthened. In that way the idea of *final cadence* as a closing gesture is reinforced.

The graphic analysis of Fig. 7 shows how movement precedes the spoken voice and frames the scene. Once movement initiates the scene, a sequence of movements and sounds with matched segmentation can be appreciated. The different hierarchic levels observed in the musical analysis (basic idea, contrasting idea, antecedent - consequent phrase, etc.) may be discerned from the movement (Fig. 7) although movement precedes the voice (see upper row displaying sound signal). The first unit –the antecedent- is made of: rising -the shawl-, freezing and small movements at the top on the horizontal plane, and descending and advancing -the torso- towards the baby and retiring-the shawl. The second unit—the consequent- is made of: rising -the shawl- again, movements at the top on the horizontal plane with the shawl, and freezing, and a more extended descending and advancing towards the baby. The sequence is similar to the previous one but inverted: freezing-movement at the top; movement at the top-freezing. In these major phrases: (i) "*rising*" movements anticipate speech (functioning like an anticipation of the call "Habibi….."), just when there is no sound (pauses); (ii) the instances of "*freezing* at the top" and "horizontal movements at the top" match the basic idea and its repetition, when the vocal pitch is ascending; (notice that the voice remains in the high register for a longer period of time during the consequent); and (iii) "*descending* and *advancing*" towards the baby match the contrasting idea and the pitch's descending, by jump and glissando (the ending *ritardando* matches the greatest extent and duration of the *descending* and *advancing* movement; notice as well that the melodic contour reaches its highest pitch and widest tessitura at that point, matching the most ostensible movement); (iv) the "*retiring*" shape form coincides with the end of the larger units. It is also noticeable how *retiring* is longer as the end of the consequent is more lengthened. In that way the idea of *final cadence* as a closing gesture is reinforced. When the consequent ends, the baby takes the initiative by catching the shawl and bringing it to the mouth. The adult initiates the final *retiring* movement with a different Effort (*pressing*), and gives rise to the *coda*: the game is over.

Sound Microanalyses of A Fortuitous Event and Resolution

After the peek-a-boo game's coda ("Noooo…!"), when the baby-seat falls descending two positions, the sound component acquires a clear rhythmic-metrical structure that differs markedly from everything that was happening before. Consequently, the most noticeable feature of this fragment is its clear rhythmic structure. The speech becomes more ono-matopoeic, with more marked consonants and shorter syllable durations. Syllables in Fig. 8
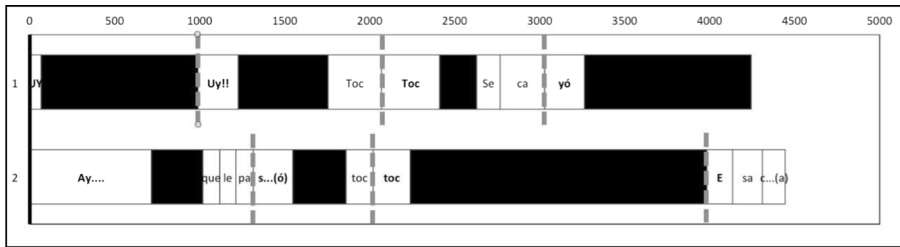
Fig. 8 Relative durations of sounds (*white*) and pauses (*black*) in the two units of *A fortuitous event and resolution*. Stressed syllables are marked with *dotted lines* showing the pulse's regularity in the first part. The pauses account for the staccato or legato articulation of sounds and phrasing (way of articulating phrases using timing resources)

show durations of each syllable and pauses between them. Here the *staccato* nature of the first sentence can be observed stressing accents over and therefore the pulse's regularity as well. By contrast, in the second sentence S lengthens the syllables, bringing them closer together. Notice how she lengthens the syllable "Ay!" to a point where the emotional content of what she was saying changes. This point represents the division of the unit into two subunits. In the first one, the content is surprise, in the second one is the manifested adult empathy with the disturbing situation suffered by the baby. As part of that empathy, she concludes "esa ca … (rita)" (that little face…) moved by the baby's face of affliction. The idea of "the little face" is what will give rise to the next section, it is important to highlight this because following the pulse set, these three syllables complete the rhythmic pattern in order to reach a state of balance to complete the more stable eight-time rhythmic unit. The prolonged pause, with which the first unit ends, separates them with a caesura. Finally, within the regularity framework of the whole fragment, the lengthening of the stressed syllables (strong beat) stands out. This shows that despite being a rhythmic sequence, regulating the change of character, and articulating speech in meaning units, they are reinforced by the phrases ending's lengthening and the metrically stronger sounds.

Movement Microanalyses of A Fortuitous Event and Resolution

The baby-seat's fall happened on the vertical plane, with a *direct*, *sudden* and *strong* Effort (*punching*) and accentuated at the bottom. As Figs. 9 and 10

Fig. 9 Movement description of *A fortuitous event and resolution* according to the categories of analysis Laban - Bertenieff, scored using ANVIL 4-0
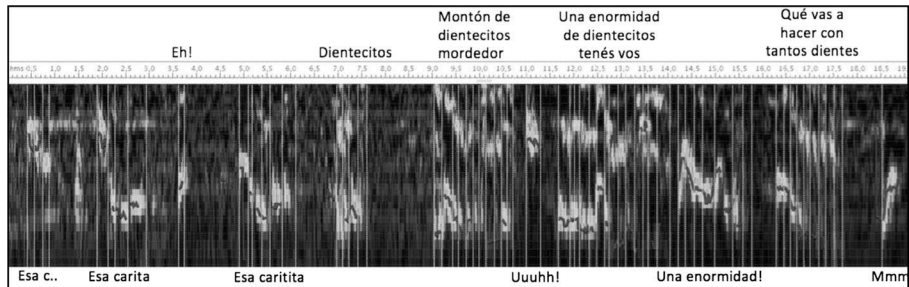
Fig. 10 Spectrogram of the third section. Timeline in seconds is at the top. The text of more tonic or *sung* phrases is indicated under the graphic (underlining the *more melodic* fundamental frequency contour), while the text of recited (spoken) units is indicated in the graphic's *upper panel* (highlighting also the clear first overtone)

displays, S interrupted her movement just at this instant, quickly responding to the baby, imitating the *baby-seat's* fall, with a similar movement but elaborating it by changing one element: moving the shawl twice with a *light*, *direct* and *sudden* Effort (*dubbing*), conserving the vertical plane. Then, she rapidly elaborated the unit by drawing her arms back: with a *gliding* and *floating* Effort, moving them forward with two new and very smooth *dubbing* (like a smooth rocking) movements. The phrase closes *frozen*, in a point of calm reaching the goal of the previous movement, sustaining eye-contact, probably to check if the baby was restored to a calming state.

Multimodal Analysis of A Fortuitous Event, Resolution and Interpretation

Meanwhile, motion analysis reveals that unlike previous phrases in which an initial motif resulted in a development according to different elaboration principles, here a fortuitous sound and kinetic event give S the material for her performance. The sudden fall of the chair, which the baby experiences in its own body, is a *vertical, descending* and *punching* (direct – sudden - strong) movement conveying a *dropping* feeling. In light of this event, S responds with a very similar—but subtly different—movement: using the complementary *rising* form, and a *dabbing* (light-direct-sudden) effort. Notice that *dabbing* differs from *punching* only in terms of weight; also the conveyed feeling is now *stimulated*. Moreover, S repeats this movement, emulating the double fall of the seat. Vocalizations also emulate the seat's fall and the soft timbre of her voice unites with the movement's qualities. Firstly, the direct and sudden-effort movements with the shawl join the "toc toc" vocalization that keeps a similar duration. Secondly, other very smooth (like a smooth rocking) movements are accompanied by the second "toc toc" vocalization. Thus, the adult gives the baby a changed and elaborated repetition of the disturbing stimulus, interpreting in a non-verbal manner that the stimulus is now friendlier. The phrase closes *frozen and silent* and upholding the eye-contact, probably to be sure that H has regained his calm state. Sound and movement microanalysis allows us to observe once again the hermeneutic intention of adult performance facing disturbing stimuli. Expressive resources

are in the service of this goal also configuring a sort of a *non-verbal hermeneutic.*

Third Section

General Description

The third section begins when the atmosphere of intersubjective encounter after the seat fall is restored, and ends when the baby turns his head and interrupts the visual contact with the adult thus concluding the exchange. Here, S cuts loose the shawl for the first time. This section consists of two differentiated parts. In the first one, a sweet and tender melody (quasi sung expressions) is raised by H's expression of grief after the seat's fall. The second begins when, after the baby smiles, S observes his recently out little teeth and composes a rhythmic game from the word *dientecitos* (little teeth).

Sound Microanalysis

The first phrases of this section present more tonic, quasi sung, sounds, while the units of the second part are more the recited type. The graph shows lighter bands corresponding to more energetic frequencies (they have a higher sound level). That band indicates the frequency corresponding to the pitch of the sound heard. In recited units, frequencies are more uniform and generally lower than in sung units. Besides in them is noticeable not only the fundamental frequency but also the first overtone.[1] In sung units frequencies are much more variable. The first three units show a downward melodic contour (the second and third, first descend and then maintain their pitch). The second part is spoken, so the melodic contour is much more monotonous, and in a markedly lower register.

Concerning rhythm, consider the difference between the second and third phrases involving the addition of one syllable ("carita" and "caritita": the second word is a diminutive of the former; Fig. 11). In this change, the word's prosodic stress is modified (in "carita", the stress falls on "ri"; in "caritita", on "ti"). The graph shows the timing parallelism accuracy of both stresses. This means that the above syllables are arranged in the same time lapse, so that the pulse remains highly regular. Finally, the durations of the syllables of the second part are more homogeneous.

In relation to the timing expressive features, Fig. 11 shows the use of syllable lengthening in a *ritardando* manner in the sung phrases (2, 3 and 7). In the spoken

---

[1] The phonation act does not result in a simple tone, but in a set of tones or spectrum. The lowest tone of this set is called fundamental, which generally corresponds with what we hear as sound's pitch, while others are overtones. Ideally the intensity of each overtone decreases as these turn higher. So the first overtone will be stronger than the second and this in turn stronger than the third. This idealization gets more fulfilled in spoken voice than in sung voice given that in the latter there are higher frequencies that often reach higher sound levels than the inferior ones. It is then expected that the spoken voice shows more intensity in the lowest overtone, while the singing voice depends on the tessitura, the emission type, etc.

phrases progressive increase of syllables it is also noticed from the phrase 4 increasing the chronometrical density (number of sounds per unit time), and elaborating 4 rhythmic phrases form the *motif* of *dientecitos* (little teeth). At the same time, these phrases show great regularity (particularly phrase 8).

Multimodal Analysis and Interpretation

The initial sweet and tender melody (phrases 1–3) unfolded downward melodic contours and seems to be used to attune H's face of affliction. Such three units are re-editing the first section strategy: a short motif with two repetitions. The first part, in which S says "esa ca …" alluding to H's expression of grief, is accompanied by an upwards/downwards vertical head movement.. In the second part, five phrases of sound and movement are sequenced. As shown in Fig. 11 sound chronometrical density (with a relative decrease of silences between phrases, or breaths) increases. It matches an increasing kinetic activity until in the last phrase S comes closer to the baby giving soft, light touches on his chest with her finger. This is the moment of greatest direct body contact of the whole scene, with lots of quick and ecxititing finger movements. Finally the baby turns his face away and the scene ends when the adult touches the baby. It seems that the stimulation intensity overloads, until the baby puts a final end to the scene. Further on, as it can be appreciated, throughout this whole section the adult's on-line creation focalizes primarily in speech, in such a way that we estimate that an analysis of its poetry, following the resources used by Miall and Dissanayake (2003), might enrich the section's artistic analysis as a verbal-sound-kinetic whole.
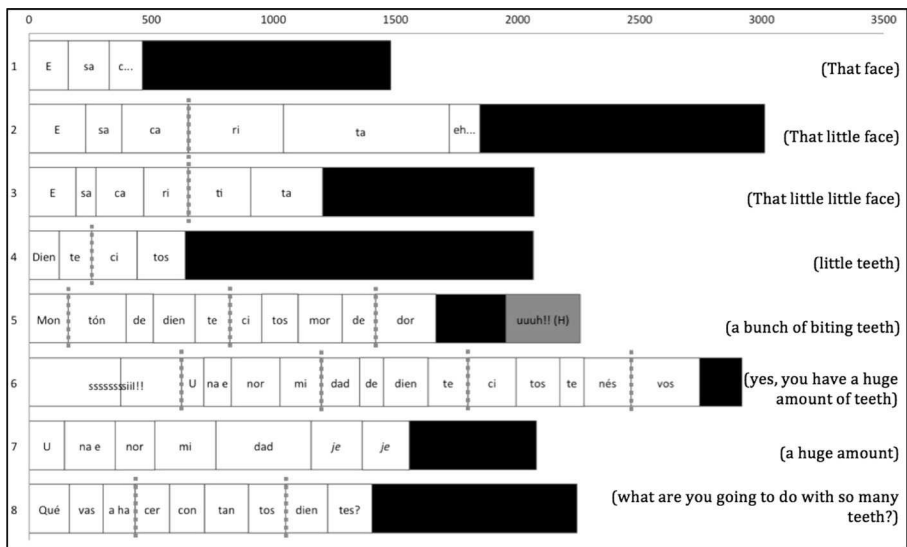
Fig. 11 Duration of syllables and pauses of the third section (*Dotted lines* preceding the accented syllables; English translation of the text at the right)

Conclusions

We have analyzed the features of an infant directed performance belonging to a social interaction between an adult/researcher and a baby that from both, an emic and an ethic point of view, is considered of high intersubjective involvement. The analyses were made using methods coming from the performing arts—especially Laban's Movement Analysis System and expression analysis in musical performance -under the assumption that such methods could shed light on hitherto unobserved aspects of adult behaviors that leads to sensitive ways of being together in early infant-adult encounters. In this way, we obtained a description of the adult's behavior when facing the baby, in a "body on body" interaction, that highlights its expressive features and aesthetic qualities displaying them as crucial for the strengthening of the intersubjective bond. The obtained description brings out the genetic link, and emphasizes the continuity, between the intuitive behavior at stake in early intersubjective encounters and elaborated forms of behaviors involved in social phenomena we recognize as art forms—highlighted by Dissanayake (2000a, b), Stern (1985, 2010), Imberty (2002), Cross (2010), Trevarthen (1999/2000), among others. This link has been established through "*artifying*" behaviors, the "making special" by means of aesthetic operations (Dissannayake 2000b, 2008), and elaboration of *dynamic forms of vitality* (Stern 2010), on the basis of *communicative musicality* (Malloch and Trevarthen 2009) as a cognitive/embodiment supporting skill.

The use of analytical tools and concepts developed for the exegesis of artistic expressions result in a plausible and consistent description of the infant directed performance. Overall, this methodological approach showed that:

(i)   As indicated by Stern et al. (1977), the temporal organization of multimodal patterns that adults direct to babies has kinetic and vowel phrases that can be run separately, jointly, or alternately.

(ii) Adults not only provide movement and sound units, but also holistic performing sound-movement attractive units: the phrases of sound and movement; this unity is achieved through compositional elaboration and the use of expressive resources of the culture's performing arts;

(iii) There is an expressive mode of operation on the temporal patterns that structure adult-to-baby directed behavior, making each exchange a one-of-a-kind performance despite settling in ritualized structures and/or culture dependents.

(iv) In that expressive context, kinetic and sound phrases modify their impact on the significant configuration of the performances as it develops, sometimes predominating one modality or the other, or complementing both in the expression of a common direction: for example indicating the end of a phrase.

(v)   Such expressive modality is carried out in order to prolong the attention support and strengthen the intersubjective engagement. Studies on IDS have indicated that the use of the repetition-variation structure reinforces the infant's attention. Our analyses show that, furthermore, adults intuitively use *expressive resources* to attract and sustain the baby's attention. The performance's timing seems to be the crucial aspect: durations' shortenings and lengthenings (of movements and sounds) are the main resource to increase expectation and sustain attention.

Specifically, the performed microanalyses granted us the possibility to detect structural and expressive aspects in the adult's performance directed to a 7-month infant that are compatible with music and dance structures and their expressive resources. These are:

(i) *The use of composition structures and organization modes proper to temporal arts.* The adult creates the entire performance—of speech, sound and movement—mainly using the basic *repetition-variation* strategy. The use of this strategy has already been observed in IDS studies. Our work shows that the adult's movements, and not only the speech, are organized in the form repetition-variation. Interestingly, our analyses also reveal that the adult arranges multimodal phrases, in similar ways to music and dance. For performances with original sources (adults spontaneous creations), composition seems to obey to a linear elaboration of motives. In this case, like in non-choreographic contemporary dance, for example *Contact Improvisation*, adults create an on-line performance, with no pre-designed plan, by elaborating motifs throughout the repetition-variation form (see First Section and Third Section). On the contrary, for performances based on traditional sources, the sequence of movements and sounds seems to be governed by a pre-designed plan showing a more sophisticated hierarchically elaborated structure corresponding to the structure of the repertoire piece taken as source (see the *peek-a boo* play in Second Section).

(ii) *The use of the timing, dynamics, and quality voice as expressive resources.* In performances based on traditional sources, it is precisely on the basis of these pre-established guidelines that the government of microtiming, dynamics, and other expressive attributes of sound, allow a modeling of the sound result contingent with the dynamics of exchange that is taking place. The use of pauses and *ritardandi* gradients according to the structural hierarchy of phrases, the handling of length and accelerations to contribute to tension-relaxation relationships and, consequently, to the discourse's cohesion, are some of them. Moreover, *rubati* matches variations in timing and dynamics of movements (more extended and marked movements). Handling of loudness, which is also a frequent resource for expressive contrasts in the interaction, creating intimate or expansive atmospheres, exciting or calming mood characters, is also contributing to the discourse's coherence. As in expressive music performance, calm is often identified with closure (see First Section and Second Section). Changes in sound quality are also used to attune to the baby's mood. For example, as descending melodic contour has been associated with sadness for centuries, it is easy to associate the *descending sung phrases* with the intention to attune with H's face of affliction in Third Section.

(iii) *The use of expressive resources for suggesting unit boundaries and accents, and pointing tension-relaxation relationships.* Much as when the adult adopts *repertoire pieces* as when he creates on-line performances, sound and movement also concur concerning the use of expressive resources for suggesting unit boundaries and accents, and pointing tension-relaxation relationships. For example, the longest *ritardando* sound matches the amplest movement, and the vocal pitch also matches the movements' height in space (see the *peek-a boo* play in Second Section).

(iv) *The adjustment of movements and sounds as a non-verbal hermeneutic to facilitate encounters and* regulate the atmosphere's intimacy and everyday life. Dance and musical performance analysis allows us to observe a hermeneutic intention of adult performance facing disruptive stimuli. In Second Section, when *the seat* suddenly fell, adult sounds and movements emulated the seat's fall but in a softer way, giving the baby a changed and elaborated repetition of the disruptive stimulus, interpreting in a non-verbal manner that the stimulus is now friendlier. In First Section, regular pronunciation of the word *telephone*, adjusting *to a pulse* also contributed to tension reduction, providing a safety message and recovering the peaceful atmosphere. Adjusting the performance's events to a regular pace seems to contribute to both, initiate and recover contact when at risk by the impact of external stimuli. On the contrary, when the intersubjective interchange occurs without external interference, and the infant's attention is clearly focused on the adult's performance, the pulse becomes more flexible (See First Section *antecedent-consequent*). Thus, it is possible that adults resort to holding on to the underlying pulse in order to capture the infant's attention and to take control of a situation threatened by external factors, calming the infant by showing control over any contingency. Pulse regularity can soon become more flexible, and timing deviations may emerge providing new meanings to the interchange. In this way, pulse may be understood as an expressive trait in a given context, and its systematic relaxation—or the patterns of systematic deviation regarding that pulse—, may also be considered as an expressive characteristic in a different context.

Although the baby's behavior has not been the focus of our work, the findings of this study and the non-systematic observations of infant behavior we performed made it possible to suggest that *Infant Directed Performances initiate an aesthetic enculturation.* The expressive resources observed in the adult performance are similar to the expressive resources of artistic events in her culture (i.e., timing and dynamics) used in a very similar way. The adult lengthens the phrases' endings according to a hierarchical grouping structure of the sequence in a very familiar manner, because this manner is the most common in tonal music expressive performances where that type of structure is strongly rooted. Thus it is possible to think, that as adults deliberately transmit characteristics of their culture by using games and musical repertoire of their cultural tradition, other characteristics of the culture are implicitly communicated through the use of certain expressive resources. *Rubato*, for example, is based on the infant*s' prim*ary ability to recognize, join to, and predict isochronous patterns, and detect micro-deviations from them (Trevarthen 1999/2000). Regarding this, the *work* parents do during early interactions may be understood as a way of "teaching" the expressive value that the culture has assigned to such micro-deviations. Each culture has its own repertoire of expressive resources for its performing arts. Studies on IDS have shown the regulative function of the baby's arousal and emotional states and its role in language acquisition. From this study it is possible to suggest that at least one cultural feature—the expressive manners of music performance—as part of intuitive parenting, is opening the infant a door to understand, enjoy and contemplate the arts in his culture. This hypothesis about the functional value of the adult-to-infant directed performance, requires further research focused on the baby's participation.

Concisely, the carried out microanalyses exhibit that infant directed performances may be characterized in terms of the categories used in the performing arts realm. They also suggest the functional resemblance between infant directed performances, music, and dance (they allow to form holistic meaningful units, can modify the baby's mood, and search to extend the attention and maintain contact) and the aesthetic enculturation function of infant directed performances. These results support the hypothesis of adult performance artisticity when interacting with infants. Intersubjectivity between adults and infants is partially determined by the particular ways in which adults are active in front of babies. Therefore, this study indicates that intersubjetcivity is initially an aesthetic phenomenon.

We have focused on interactions with 7-month-old infants. The first half of the first year of life is a crucial time span in development, dyadic adult-baby interactions that support primary intersubjectivity experiences have peaked, soon after we will find the emergence of triadic interactions where objects are incorporated. There is much yet to be done. It is necessary (i) to investigate whether structural aspects detected in adult-to-infant directed performance and their functional value vary or remain stable throughout the first year of life; and (ii) to keep exploring other structural and functional aspects of adult-infant interaction that are compatible with music and dance structures and functions. Microanalysis bring to light information that naturalistic observation in real time cannot give. The wealth of information provided by microanalysis offsets the problem of the generalization of the results, but certainly the study of the subject should be supplemented by other methodological approaches.

Conflict of Interest   The authors declare that they have no conflict of interest.

References

Bateson, M. C. (1979). The epigenesis of conversational interaction: A personal account of research development. In M. Bulowa (Ed.), *Before speech. The beginning of interpersonal communication* (pp. 63–78). Cambridge: Cambridge University Press.

Bergeson, T. R., & Trehub, S. E. (2007). Signature tunes in mothers' speech to infants. *Infant Behavior & Development, 30*, 648–654. doi:10.1016/j.infbeh.2007.03.003.

Berry, W. (1986). *Form in music* (2nd ed.). New Jersey: Prentice-Hall.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science, 5*, 72–83. doi:10.1111/1467-7687.00211.

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology, 3*, 327. doi:10.3389/fpsyg.2012.00327.

Condon, W. S., & Sanders, L. W. (1974). Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science, 183*(4120), 99–101.

Cross, I. (2010). La música en la cultura y la evolución. *Epistemus, 1*(1), 9–19.

Davies, E. (2006). *Beyond dance. Laban's legacy of movement analysis*. New York: Routledge.

Dissanayake, E. (2000a). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music*. Cambridge: The MIT Press.

Dissanayake, E. (2000b). *Art and intimacy: How the arts began*. Seattle: University of Washington Press.

Dissanayake, E. (2001). Becoming *Homo Aestheticus*: sources of aesthetic imagination in mother-Infant interactions. *Substance, 30*(1/2), 85–103.

Dissanayake, E. (2008). The arts after Darwin: Does art have an origin and adaptive function? In K. Zijlmans & W. van Damme (Eds.), *World art studies: Exploring concepts and approaches* (pp. 241–263). Amsterdam: Valiz.

Dissanayake, E. (2014). Homo musicus: ¿Estamos biológicamente predispuestos para ser musicales? In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaría sobre la musicalidad humana* (pp. 195–215). Buenos Aires: Paidós.

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana*. Madrid: Antonio Machado.

Español, S. (2007). La elaboración del movimiento entre el bebé y el adulto. In M. de la P Jacquier & A. Pereira Ghiena (Eds.), *Música y Bienestar Humano. Actas de la VI Reunión de la Sociedad Argentina para las Ciencias Cognitivas de la Música (SACCoM)* (pp. 3–13). Buenos Aires: SACCoM.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Development, 60*, 1497–5100.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501. doi:10.1017/S0305000900010679.

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. Juslin & J. A. Soloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford: Oxford University Press.

Godøy, R. I., & Leman, M. (Eds.). (2010). *Musical gestures: Sound, movement, and meaning* (p. 307). New York: Routledge.

Imberty, M. (2002). La musica e il bambino. In J.-J. Nattiez (Dir.) *Enciclopedia della musica* (pp. 477–495). Torino: Giulio Einaudi Editore.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C., & Jasnow, M. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in Child Development, 66*(2), 1–132.

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129–164.

Kipp, M. (2004). Anvil – a video annotation research tool. http://www.dfki.de/~kipp/anvil. June 2008.

Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: the effect of infant-directed action on infants' attention and object exploration. *Infant Behavior & Development, 32*(4), 437–444.

Laban, R. (1971). *The mastery of movement*. Boston: Plays.

Larson, S. (2012). *Musical forces. Motion, metaphor and meaning in music* (p. 369). Bloomington: Indiana University Press.

Lee, D. N., & Shögler, B. (2009). Tau in musical expression. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 83–104). Nueva York: Oxford University.

Licata, M., Paulus, M., Thoermer, C., Kristen, S., Woodward, A., & Sodian, B. (2014). Mother–infant interaction quality and infants' ability to encode actions as goal-directed. *Social Development, 23*(2), 340–356.

Malloch, S. (1999/2000). Mothers and infants and communicative musicality. *Musicæ Scientiæ, Special Issue*, 29–57.

Malloch, S., & Trevarthen, C. (Eds.). (2009). *Communicative musicality: Exploring the basis of human companionship*. Nueva York: Oxford University.

Mazokopaki, K., & Kugiumutzakis, G. (2009). Infant rhythms: Expressions of musical companionship. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality. Exploring the basis of human companionship* (pp. 185–208). Oxford: Oxford University Press.

Miall, D., & Dissanayake, E. (2003). The poetics of *Babytalk*. *Human Nature, 14*(4), 337–364.

Newlove, J. (2007). *Laban for actor and dancers. Putting Laban's movement theory into practice: A step-by-step guide*. New York: Routledge.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48*(1), 115–138.

Papoušek, H. (1996a). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliege & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 37–55). Oxford: Oxford University Press.

Papoušek, M. (1996b). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 88–112). Oxford: Oxford University Press.

Papoušek, M., & Papoušek, H. (2002). Intuitive parenting. In M. H. Bornstein (Ed.), *Handbook of parenting. Volumen 2. Biology and ecology of parenting* (pp. 183–203). Mahwah: Lawrence Erlbaum.

Reddy, V. (2008). *How infants know minds*. London: Harvard University Press.

Riba, C. (1990). *La comunicación animal. Un enfoque zoosemiótico*. Barcelona: Anthropos.

Rivière, A. (1986/2003). Interacción precoz. Una perspectiva vygotskiana a patir de los esquemas de Piaget. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comp.) *Ángel Rivière. Obras Escogidas* (pp. 109–142). Madrid: Panamericana Vol II.

Shifres, F. (2007). La Ejecución Parental. Los componentes performativos de las interacciones tempranas. (Parental Performance. Performing components in early interactions). In M. de la P Jacquier & A. Pereira Ghiena (Eds.), *Música y Bienestar Humano. Actas de la VI Reunión de la Sociedad Argentina para las Ciencias Cognitivas de la Música (SACCoM)* (pp. 13–24). Buenos Aires: SACCoM.

Shifres, F. (2014). Algo más sobre el enlace entre la infancia temprana y la música: El poder expresivo del rubato. In S. Español (Ed.), *Psicología de la Música y del Desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* (pp. 21–70). Buenos Aires: Paidós.

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187–213). Oxford: Wiley-Interscience.

Stern, D. N. (1985). *The interpersonal world of the infant. A view from psychoanalysis and developmental psychology*. Nueva York: Basic Books.

Stern, D. N. (2010). *Forms of vitality. Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. New York: Oxford University Press.

Stern, D. N., Beebe, B., Jaffe, J., & Bennet, S. L. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 177–202). London: Academic.

Sundberg, J. (1993). How can music be expressive?. *Speech Communication, 13,* 239–253.

Sundberg, J., & Verrillo, V. (1981). On the anatomy of the retard: a study of timing in music. *The Journal of the Acoustical Society of America, 68*(3), 772–779.

Todd, N. P. (1985). A model of expressive timing in tonal music. *Music Perception, 3*(1), 33–57.

Todd, N. P. (1992). The dynamics of dynamics: a model of musical expression. *Journal of the Acoustical Society of America, 91*(6), 3540–3550.

Trehub, S. E. & Schellemberg, E. G. (1995). Music: Its relevance to infants. En R. Vasta (Ed.) *Annals of Child Development, 11,* 1–24. New York: Jessica Kingsley Publisher.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993a). Adults identify infant-directed music across cultures. *Infant Behavior and Development, 16,* 193–211.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993b). Maternal singin in cross-cultural perspective. *Infant Behavior and Development, 16,* 285–295.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). Cambridge: Cambridge University Press.

Trevarthen, C. (1999/2000). Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicæ Scientiæ, Special Issue,* 155–215.

Trevarthen, C., & Reddy, V. (2007). Consciousness in infants. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 41–57). MA: Blackwell Publishing.

Unyk, A. M., Trehub, S. E., Trainor, L. J., & Schellemberg, E. G. (1992). Lullabies and simplicity: a crosscultual perspective. *Psychology of Music, 20,* 15–28.

Silvia Español Doctor in psychology, is a developmental researcher at CONICET (The National Scientific and Technical Research Council, Argentina). Her field of interest is socio cognitive development in infancy. Her work is located on the border between cognitive developmental psychology, psychology of music and the study of human movement.

Favio Shifres (PhD) is a music psychology researcher at UNLP (National University of La Plata, Argentina). His field of interest is communication in music performance and aural skill development of musicians in the framework of the embodied music cognition. His perspective of these topics follows research on pos-colonial epistemologies.