

# Chapter 1

## Empirical Evidence for the Claim That the Vocal Theory of Language Origins and the Gestural Theory of Language Origins Are Not Incompatible with One Another



Nobuo Masataka

**Abstract** The author argues about the implications of the evolution of motherese for the emergence of language in the human history, and it occurred in both the vocal mode and the manual mode, the fact indicating that the gestural theory of and the vocal theory of language origins are not incompatible with one another. It is a commonplace observation that hearing adults tend to modify their speech in an unusual and characteristic fashion when they address infants and young children. The available data indicate that motherese or infant-directed speech is a prevalent form of language input to hearing infants and that its salience for preverbal infants results both from the infant's attentional responsiveness to certain sounds more readily than others and from the infant's affective responsiveness to certain attributes of the auditory signal. In the signing behavior of deaf mothers when communicating with their deaf infants, a phenomenon quite analogous to motherese in maternal speech is observed. Concerning the aspect of linguistic input, moreover, there is evidence for the presence of predispositional preparedness in human infants to detect motherese characteristics equally in the manual mode and in the vocal mode. Such cognitive preparedness, in fact, serves as a basis on which sign language learning proceeds in deaf infants. One can seek its evolutionary origins in the rudimentary form of teaching behavior that occurs in the adult–infant interaction in nonhuman primates as well as in humans, by which the cross-generational transmission of parenting is made possible, including that in the deaf community.

**Keywords** Motherese · Signed language · Language learning · Infant-directed signing · Infant-directed speech

---

N. Masataka (✉)  
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

## 1.1 Importance of Motherese when Considering Language Evolution

It is well known that there have been controversies about the origin of language and that some scientists, including Charles Darwin, sought it in animal vocalizations whereas others including Wilhelm Wundt considered manual gestures as the origin. Here, the author argues from the ontogenetic perspective the implications of the evolution of motherese for the emergence of language in the human history, and that it occurred in both the vocal mode and the manual mode, the fact indicating that the gestural theory of and the vocal theory of language origins are not incompatible with each other.

Developmentally, the earliest non-cry sounds produced by human infants do not transmit meanings, unlike words uttered by older individuals, but rather reveal fitness or express states. On the one hand, we make no assumption that very young infants intend to communicate with such sounds. On the other hand, adults who receive the sounds are provided with broad reception/interpretation capabilities that allow them to provide differentiated feedback to signals, as noted by Masataka (2003). This assumption is in accordance with results in animal communication indicating nonhuman primates are far more flexible in recognition/reaction to signals than in signal production or functional usage. Flexible responsiveness of receivers is a necessary condition for their being able to apply selection pressure on volubility and flexibility of infant vocal tendencies (Oller et al. 2016). Such reasoning could lead us to notice the importance of considering how selection pressures might engender increase in the tendency of an infant to produce spontaneous non-cry vocalization for arguing language evolution, and that in fact, evolution of increased infant spontaneous vocalizations began with developmental steps in individual infants under the selective pressure of their own caregivers.

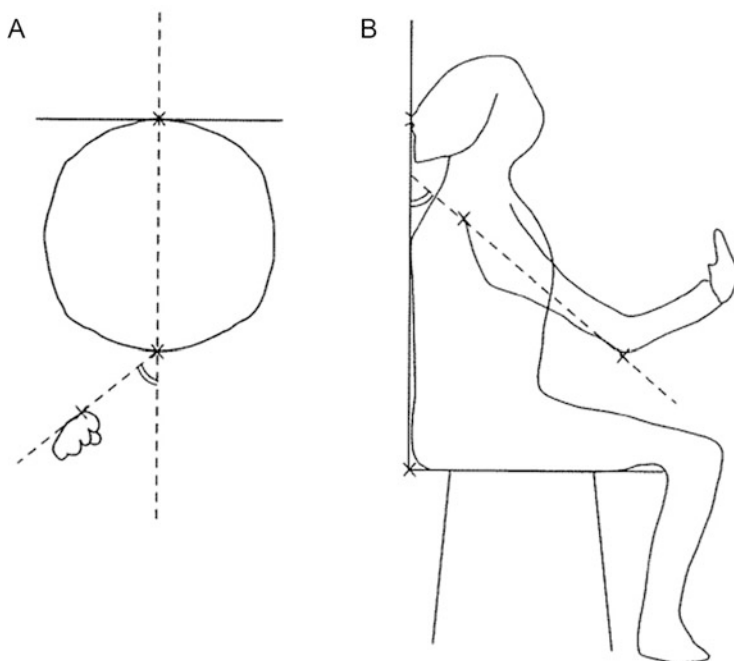
Indeed, recent researches have revealed the fact that during modern human development, infant vocal capabilities emerge at least partly in response to social interaction, where caregivers react to vocal capabilities of infants in accordance with a scaffolding principle requiring parental discernment and intuitive parenting to reinforce vocal exploration and learning. Both endogenous inclination of infants to explore the vocal space and interactive feedback from caregivers thus foster growth in vocal capability.

If this assumption about selection pressures for spontaneous vocalizations is valid, it follows that hominin parents would also have been selected to become aware of the fitness reflected in infant vocalizations and capable of responding to those indicators with selective care and reinforcement of vocalizations. And the author believes this to have been illustrated so far in the phenomenon that is called “motherese.”

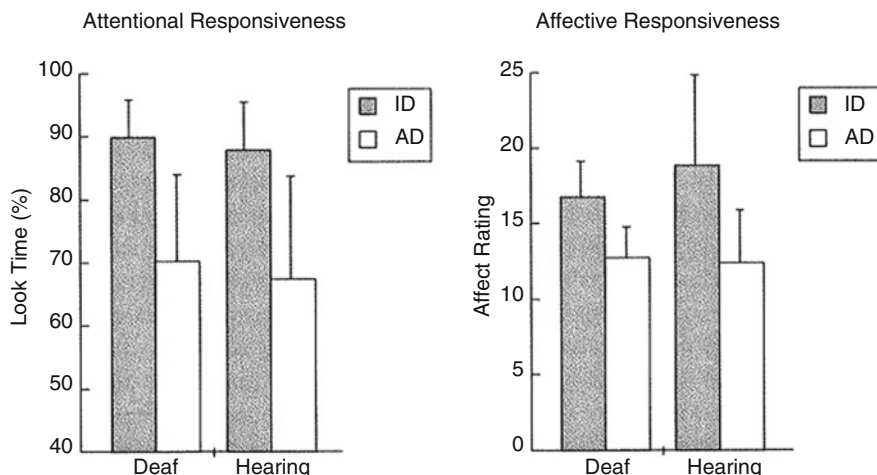
## 1.2 Implications of Infant-Directed Speech or Motherese in the Vocal Mode for Spoken Language Learning

Although the fact that hearing adults tend to speak to infants in an odd and characteristic fashion has been a commonplace observation, it was Ferguson (1964) who first offered a coherent description of the linguistic and the paralinguistic features of child-directed speech. In the languages investigated by Ferguson, which included English, Spanish, Arabic, Comanche, Giyak, and Marathi, the use of elevated pitch and exaggerated pitch excursions were the most prominent characteristics observed across cultures. Since then, this sort of speech style has been the focus of considerable research, and such speech is commonly referred to as “motherese” or infant-directed speech. Most studies included the claim that hearing adults speak to children in a high-pitched, even “squeaky” voice (Fig. 1.1).

The data that are available to date all indicate that prosodic properties characterizing infant-directed speech and infant-directed singing are a key component of language input to preverbal infants and that they serve as important social and attentional features in early development (Saint-Georges et al. 2013). This fact has led several researchers to explore the possibility that the effectiveness of exaggerated



**Fig. 1.1** Schematic representation of the planar view (a) and the side view (b) of the mothers as recorded on film. The points marked by a light pen on the digitizer are indicated by an X. (Cited from Masataka 1992)



**Fig. 1.2** Comparison of responsiveness to infant-directed (ID) signing versus adult-directed (AD) signing between hearing infants and deaf infants. Error bars represent standard deviations. (Cited from Masataka 1998)

properties in infant-directed speech for modulating infant attentional and affective responsiveness results from innate predispositions to respond selectively to those properties. Cooper and Aslin (1990) examined behavioral preferences for infant-directed speech over adult-directed speech in two groups of infants, one made up of 12 one-month-old infants and the other consisting of 16 two-day-olds. Infants of both groups were tested according to the same visual-fixation-based auditory-preference procedure. The results showed that both one-month-old infants and newborns preferred infant-directed speech over adult-directed speech. Although the absolute magnitude of the infant-directed speech preference was significantly greater in the older infants, who showed longer looking durations than the younger infants, subsequent analyses showed no significant difference in the relative magnitude of this effect, indicating that infants' preference for the exaggerated prosodic features of infant-directed speech is present from birth and may not depend on any specific postnatal experience. From these results, it appears that postnatal experience with language does not have to be extensive for the infants to learn about the prosodic features of their native language. However, it has been shown that newborns prefer the intonational contour and temporal patterning of a prenatally experienced melody (Panneton 1985). Thus, the possibility remained that both prenatal and postnatal auditory experience could affect the relative salience of prosodic cues for young infants (Fig. 1.2).

In a subsequent study, 4- and 9-month-old English-learning infants were reported to show a robust attentional and affective preference for infant-directed speech over adult-directed speech in Cantonese, though this language was completely "foreign" to them (Werker et al. 1994). Similarly, 4- to 7-month-old English-learning infants were reported to show as strong attentional and affective responsiveness to

infant-directed singing in foreign languages as to that in English (Trainor 1996). However, the infants included in these experiments had heard speech or songs before. In fact, all hearing infants of hearing parents are exposed to some form of speech and song as early as the prenatal period (DeCasper et al. 1994). Masataka (2003) investigated the attentional responsiveness to infant-directed speech and infant-directed song in two-day-old healthy infants, who were hearing and were born to congenitally deaf parents. A total of 15 infants participated in the experiment. All of the infants lived in typical nuclear families. The parents had acquired Japanese sign language as a first language and communicated with one another by the signed language exclusively. They lived in typical deaf communities. Since auditory experience for infants in utero mostly comes from their parents with regard to speech and song, the findings obtained from these infants were expected to provide a convincing answer to the question of whether preference for infant-directed speech and infant-directed song is predispositional or not.

For the stimuli presented, infant-directed and adult-directed speech, and infant-directed and adult-directed songs were prepared. Both speech samples and song samples were made up of a Japanese version and an English version, respectively. Speech samples and song samples were recorded from 10 adult females: five native speakers of Japanese and five of English were instructed to read or sing identical short scripts and play songs either to an infant or to an adult in Japanese and English, respectively. Acoustical comparisons of the speech samples and the song samples revealed that when directed to the infant, both speech and song showed such modification as reported to be typical of infant-directed speech or song; i.e., the pitch was elevated, the pitch contour was exaggerated, and the tempo became slower. When these infant-directed and adult-directed speech or song stimuli were presented, the infants were found to look longer at the infant-directed version as opposed to the adult-directed version whether the stimuli were those of speech samples or singing samples. Since intrauterine recordings taken near term have revealed that the maternal voice and heartbeat are audible in utero, but nonmaternal voices are rarely audible because of attenuation by maternal tissue and/or masking by intrauterine sounds (Querleu and Renard 1981; Querleu et al. 1981), these results strongly indicate a predispositional perceptual preference by human infants for exaggerated prosodic properties, whether they are in speech or song.

In addition to the fact that linguistic input to infants is modified prosodically, evidence has been presented that indicates the possibility of phonetic modification in motherese in a way that would enhance language learning. Kuhl et al. (1997) audiotaped 10 native-speaking women when speaking to their 2- to 5-month-old infants and when speaking to their adult friend in each of the US, Russia, and Sweden. Native-language words containing the vowels /i/, /a/, and /u/ were preselected for analysis in the three languages, and, with regard to each of the three vowels in each of the languages, the vowel triangle was compared between the speech sample recorded in the infant-directed condition and that recorded in the adult-directed condition.

The results revealed that when interacting with the infants, phonetic modifications occurred in the maternal speech as compared with when communicating with the

adults. Across all the three languages, mothers were found to produce acoustically more extreme vowels when addressing their infants, resulting in an expansion of the vowel triangle in the speech sample collected in the infant-directed condition. They did not simply raise all formant frequencies when speaking to their infants, as they might have done if they were mimicking child speech. Rather, formant frequencies were selectively modified to achieve an expansion of the acoustic space encompassing the vowel triangle. When vowel triangle areas were compared between the infant-directed sample and the adult-directed sample for each mother, the results were highly consistent across individuals. For each of the 30 mothers, the area of the vowel triangle was greater in the infant-directed condition than in the adult-directed condition. On average, they expanded the vowel triangle by 92% when addressing their infants (English, 91%, Russian, 94%, Swedish, 90%). The extent of the expansion did not differ across languages, suggesting that the phonetic modification took place essentially in the similar manner regardless of the difference of the languages.

An expanded vowel triangle is assumed to increase the acoustic distance between vowels, which eventually enables the infants to make distinctions among them more easily. Moreover, the possibility arises that by stretching the triangle, vowels would come to be produced that go beyond those produced in typical adult conversation. From both an acoustic and articulatory perspective, these vowels are “hyperarticulated,” and hyperarticulated vowels are perceived by adults as “better instances” of vowel categories. Indeed, a laboratory test showed that when listening to good instances of phonetic categories, infants show greater phonetic categorization ability (Iverson and Kuhl 1995). When prosodic properties of maternal speech are exaggerated, the infant listening to that is enabled to imitate the prosodic pattern more easily. Given the fact that the ability of vocal imitation of infants proceeds simultaneously in terms of prosody and in terms of phonetics during their first 6 months of life (Masataka 2003), it seems reasonable to hypothesize that the same would be the case for maternal speech with exaggerated phonetic properties: when the vowel triangle is expanded in maternal speech, the imitation of the vowel heard by the infant would be facilitated.

### **1.3 Characteristics of Signs of Deaf Mothers When Interacting with Their Deaf Infants**

Results of the first systematic attempt to analyze parental signing to infants were reported by Erting, Prezioso, and O’Grady Hynes (1990). Since 1985, they had been collecting videotapes of deaf mothers interacting with their deaf babies. Their preliminary observations supported previous researchers’ findings that the deaf mothers varied their signing regarding the type of movement, location on the infant’s body, the intensity and speed of movement, and rhythmic patterning so that they could get and maintain the infant’s attention effectively. These findings appeared to

be consistent with the claim that motherese in spoken languages is mainly concerned with prosodic modification. Based on the findings, they focused on the American sign for MOTHER produced by two deaf mothers who had acquired ASL as their first language during two different types of interaction: interaction with their deaf infants when the infants were between 5 and 23 weeks of age and interaction with their adult friends. When a total of 27 MOTHER signs directed to the infants were compared with the same number of the signs directed to the adults, the mothers were found to (1) place the sign closer to the infant, perhaps the optimal signing distance for visual processing, and to (2) orient the hand so that the full handshape was visible to the infants. Moreover, (3) the mothers' face was fully visible for the infants to see, (4) eye gaze was directed at the infants, and (5) the sign was lengthened by repeating the same movement. The results appear to support the claim that parents use special articulatory features when communicating with infants, including parents from a visual culture whose primary means of communication is visual-gestural instead of auditory-vocal.

Following that study, Masataka (1992, 1996, 1998) conducted a series of experiments on the sign motherese phenomenon in deaf infants and their deaf parents who acquired Japanese sign language (JSL) as their first language. When starting the series of experiments concerning sign motherese, Masataka (1992) attempted to replicate the finding with a larger sample of participants who lived in a different cultural environment from those studied by Erting et al. (1990) and with the use of more exhaustive methodology for the analysis of signs. In most signed languages that have been investigated so far, facial expressions are known to play a multifunctional role for transmitting the meaning embedded in each signing movement. As with spoken language, they are used to express affect. However, unlike in spoken languages, specific facial behaviors in the signed languages also constitute the required grammatical morphology for numerous linguistic structures (for instance, relative clauses, questions, and conditional sentences in the case of ASL). Such characteristics would pose tremendous difficulty in the quantitative analysis of signing behavior. In this regard, JSL has a practical advantage in that it relies on cues produced by head movement exclusively. Consequently, it appears relatively easy to separate linguistic dimensions from paralinguistic dimensions in sign production.

In all, 13 mothers participated in the recordings. Eight of them were observed when freely interacting with their deaf infants and when interacting with their deaf adult friends (Masataka 1992). The remaining five were instructed to recite seven prepared sentences either toward their infants or toward their adult friends. Recordings were made with each mother and her infant or friend seated in a chair in a face-to-face position. The height of each chair was adjusted such that the eyes of the mother, her infant, and her friend were about 95 cm above the floor. The infant's body was fastened to the seat by a seat belt. The mother's behavior was monitored with two movie cameras. One of the two provided the frontal plane view and the other provided the right-side view. At the beginning of each recording session, she was instructed to interact with her infant or friend as she normally might when they were alone together. She was also told not to move her head during the session.

For each recognized sign recorded on this tape, the following four measurements were performed: (a) duration (the number of frames); (b) average angle subtended by the right hand with respect to the sagittal plane of the mother; (c) average angle subtended by the right elbow with respect to the body axis of the mother; and (d) whether the mother repeated the same sign consecutively or not. For measuring the positions of the hand and the elbow, each of the frames was projected by a movie projector onto a digitizer, which was connected to a minicomputer. By plotting the position of the hand and the head or the position of the elbow and the body axis on the digitizer with a light pen, the computer measured the angle between them with an accuracy of 0.5 degrees. Subsequently, the measured values were averaged for each sign by the mother. These measurements were performed to analyze the degree of exaggeration of signing by the mother. As signed languages are processed by deaf infants in the visual mode whereas spoken languages are processed by hearing infants mainly in the auditory mode, it was hypothesized that exaggeration of signing behavior should appear in the pattern of movements of hands and arms that were making the signing gestures.

When the pattern of such movement was compared when the mother was interacting with her infant and when she was interacting with another adult, striking differences were found with regard to all of the four parameters measured. All of the differences were statistically significant. For the analysis of the duration of signs, values were averaged across participants for each condition. The duration was longer in the case of signs directed to a mother's infant than in the case of signs directed to her adult friend.

When the angle of the hand and elbow subtended to the sagittal plane or body axis was calculated across participants for each condition, the same tendency was apparent. Mean scores for maximum values of angles for each sign directed to infants significantly exceeded those for signs directed to adults with respect to both the hand and the elbow. Similarly, mean scores of averaged values of angles for each sign directed to infants significantly exceeded those for signs directed to adults. With regard to all three of these parameters, post-hoc comparisons revealed that each participant demonstrated a significant increase in these scores when she interacted with her infant. Comparison of the rates of repetition of signs for signs directed to her infant and for signs directed to her adult friend also showed that the rates for the 13 mothers when they interacted with their infant exceeded the rates when they interacted with their adult friend, and that each mother demonstrated a significant increase in repetitions when she interacted with her infant.

Overall, the results indicate the presence of striking differences between the signed language used by Japanese deaf mothers when they interact with their infants and when they interact with their adult friends. When interacting with their infants, they use signs at a relatively slower tempo and are more likely to repeat the same sign, and movements used to make the signs are exaggerated. The outcome of this experiment revealed a phenomenon that is analogous to motherese in maternal speech. At the time of that study, it was well known that signed languages are organized in an identical fashion to spoken languages with regard to most linguistic aspects (Klima and Bellugi 1979). When adults produce manual actions that are part



of the phonetic inventory of signed languages in communicating with young deaf children, who have only a rudimentary knowledge of language, one would predict that the social interaction should have as unusual a quality as when adults speak to young hearing children. That is, if the interaction is to be established and maintained, adults must be led to utilize special constraints in their activities and to produce signals, in either the signed or spoken modalities, in a somewhat modified manner according to the infant's level of attention and comprehension. This notion is strongly supported by the results of the experiments described above.

In particular, it is important to note that the mothers manipulated the duration, scope (angle), and word repetition rate of their sign production when signing to their infants as contrasted to their adult friends. This "sign motherese" is considered to be parallel to "speech motherese" in manipulating or varying the prosodic patterns of the signal because duration, scope, and repetition rate are all dimensions of prosody in sign, roughly analogous to duration, pitch, and repetition in speech. The fact that the mothers involved in the experiments clearly manipulated or changed these dimensions of sign production indicates that they were doing something in addition to and presumably apart from their manipulation of affective dimensions of sign production. In the case of speech motherese, actually, it is known to be frequently accompanied by pronounced modifications in facial expression (i.e., more frequent and more exaggerated smiling, arched eye-brows, and rhythmical head movement; Gusella et al. 1986; Sullivan and Horowitz 1983). Nevertheless, these components can all be regarded as paralinguistic in speech, and also in JSL, their occurrences, if they do occur, do not influence mother–infant communication from a grammatical perspective, as described earlier. Thus, it can be concluded that the phenomenon of motherese is a property of an amodal language capacity, at least with regard to prosodic dimensions.

## 1.4 Perception of Sign Motherese

A phenomenon quite analogous to speech motherese was thus considered to have been identified and designated sign motherese. However, although in the spoken modality the available data indicate motherese is a prevalent form of language input to infants, it is still unclear whether the same is the case for sign motherese. Certainly in the studies introduced above the experimenters were able to tell the difference between the form of infant-directed signing and that of adult-directed signing on behavioral and physical grounds. However, the question of whether deaf infants can perceive the difference remains unanswered. Continually monitoring the infants' degree of attention and understanding, mothers might modify various features of their signing so as to maintain the degree of the infants' responsiveness at an optimal level. Does it enhance the infants' acquisition of the basic units of signed language? In order to address this question, Masataka (1996, 1998) undertook the following study of the perception of sign motherese.

The experiment was made up of two parts; i.e., one involved congenitally deaf infants of deaf mothers while hearing infants of hearing mothers participated in the other. In the first experiment, seven deaf infants at 6 months of age were presented with a stimulus videotape that comprised excerpts of the infant- and adult-directed signs produced by the five deaf mothers described above. All of the deaf mothers of the infants, whose husbands were also deaf, were signers of Japanese sign language as their first language. In the stimulus tape, the following seven sentences were signed toward the infants of the mothers or toward their adult friends: “Good morning.” “How are you today?” “Get up.” “Come on now.” “If you get up right away, we have a whole hour to go for a walk.” “What do you want to do?” “Let’s go for a walk.”

Obviously, the question being asked in this study is whether deaf infants will attend to and prefer infant-directed signing over adult-directed signing. Therefore, the infant’s reactions to the stimulus presentations were videotaped and later scored in terms of attentional and affective responsiveness to different video presentations. The stimulus presentation and recording took place in a black booth. Throughout the experiment, each of the seven infants was presented the stimulus tape only once. During each session, a mother stood with her back to the video monitor and held the infant over her shoulder so as to allow the infant to face the video display at a distance of approximately 60 cm. The session was conducted only when the infant was quiet and alert. During the session, the mother was told to wear a music-delivering headphone and to direct her gaze to a picture on the wall, 90 cm to her right. Although deaf, the mother was asked to wear the headphone so that hearing mother–infant dyads with no exposure to a signed language could be comparatively investigated, which will be described below.

The infant’s reactions were evaluated by four raters, two for attentional responsiveness and the other two for affective responsiveness. As an indicator of attentional responsiveness, the percentage of the total video display time spent looking at the video screen, averaged between the two raters, was measured. For the measurement, the two watched the infant on screen independently and pushed a button whenever the infant fixated on the video display. These button presses were counted and timed by a computer. In order to measure affective responsiveness, each of the other two observers was independently told to attend to facial features as well as vocalizations of the infant and to rate the infant with a set of 9-point scales on the three dimensions that were originally used in the previous study of the perception of speech motherese (Werker and McLeod 1989). The three dimensions were treated as indexes of a single underlying factor, and one cumulative score of affective responsiveness was created for each infant by summing the ratings on the three dimensions by each rater and by averaging between the two raters. Thus, the maximum and the minimum cumulative scores would be 27 and 3, respectively, for each infant for each stimulus condition. The higher the scores received by the infants on the dimensions, the more positive emotions they were judged to be experiencing.

When the actual amount of time each infant fixated on the videotape during depictions of infant-directed signing versus adult-directed signing was analyzed statistically, a significant difference was found. The overall mean proportion of

time each infant fixated on the videotape was 90.4% (SD = 6.3) for infant-directed signing and 69.8% (SD = 14.7) for adult-directed signing. When the videotape segment of infant-directed signing was presented, the infants apparently looked at it longer than when the segment of adult-directed signing was shown. Concerning affective responsiveness, the response scores to the segment of infant-directed signing exceeded those to the segment of adult-directed signing in every infant participant. Overall mean scores were 16.9 (SD = 2.5) for infant-directed signing and 12.8 (SD = 2.1) for adult-directed signing. Thus, the infants were affectively more responsive to infant-directed signing than to adult-directed signing.

The results of this experiment revealed that sign motherese evoked more robust responsiveness than adult-directed signing did from the deaf infants. This offers the long-term prospect of identifying features of motherese that are not specific to a particular language modality. Certainly it is important to take account of the possibility that the attractiveness of sign motherese is due to something in how deaf mothers manipulated their affective behavior rather than characteristics of signs themselves. In this regard, it should be noted that the head movements of the subject mothers were limited in the initial videotaping. As already noted, this restriction does not influence mother–infant communication from a grammatical perspective. From the paralinguistic perspective, however, the communicative sample obtained from the mothers under those circumstances should be affected by this restriction. Yet, the infants showed greater responsiveness to infant-directed signing than to adult-directed signing, suggesting that human infants may have an equal capacity to attend to motherese characteristics in speech or sign. They could be predisposed to attend to appropriately modified input, regardless of the modality of the input.

No doubt, in order to answer the question of whether the particular modal capacities must be triggered by some amount of experience or not, one must examine the response to these patterns in infants who lack substantive experience (linguistic exposure) in the modality. Indeed, concerning speech motherese, several researchers have treated this issue. Among others, Cooper and Aslin (1990) showed that the infant's preference for exaggerated prosodic features is present when tested at the age of 2 days. From the results, it appears that postnatal experience with language does not have to be extensive for the infant to learn about the prosodic features of their native language. However, it has also been shown that newborns still show a preference for the intonational contour and/or temporal patterning of a prenatally experienced melody (Panneton 1985). Thus, both prenatal and postnatal auditory experiences affect the relative salience of prosodic cues for young infants. In a subsequent study, four- and nine-month-old English-learning infants were found to show a robust attentional and affective preference for infant-directed speech over adult-directed speech in Cantonese, though the language was completely “foreign” for them (Werker et al. 1994), but these infants had heard SPEECH before. No convincing evidence has ever been presented to prove the hypothesis. Rather, it can be said that as long as hearing infants are investigated, it is extremely difficult, if not possible, to address the question of whether exposure to any language is necessary for the infant to be able to react to the motherese form of the language.

Therefore, a subsequent study (Masataka 1998) was conducted to determine whether hearing infants with no exposure to signed language also prefer sign motherese. If such a preference exists, this could be a demonstration that human infants lock onto particular kinds of patterned input in a language modality completely independent of prior experience. The participants were 45 sets of hearing mothers and their first-born, full-term. Hearing infants with no exposure to a signed language (21 boys and 24 girls). All of the infants were 6 months old. Their mothers were all monolingual women who spoke only Japanese, were between 22 and 29 years old, and were middle class. The stimulus tape shown to each of the 45 infants was the same video tape used in the experiment on deaf infants. The protocol of stimulus presentations and that of scoring the infant's reactions to the stimulus presentations were also the same as in the previous experiment.

The analysis of the data obtained from the 45 infants revealed that there was a significant main effect for stimulus type with respect to their attentional responsiveness. They looked at the videotape filming infant-directed signing longer than at the tape filming adult-directed signing. The overall mean proportion of time that each of them fixated on the stimulus tape was 87.5% (SD = 8.2) for infant-directed signing and 66.0% (SD = 17.3) for adult-directed signing. Moreover, when comparing responsiveness for infant-directed signing versus adult-directed signing of the hearing infants with that of deaf infants involved in the above experiment, no significant difference was found in the scores between the two groups. Concerning affective responsiveness, too, a similar tendency was found. For the 45 hearing infants, overall mean scores were 18.9 (SD = 6.1) for infant-directed signing and 12.6 (SD = 3.4) for adult-directed signing. A statistically significant difference was found between the two stimulus conditions. Comparing these scores with those of the deaf infants, there was again not a significant difference between the two participant groups.

Clearly, hearing infants who have had no exposure to any form of signed language are attracted to motherese in JSL in a manner strikingly similar to that of deaf infants who have been exposed to the language from birth. Infants are attracted to sign motherese, regardless of whether they have ever seen sign language. Space and movement are known to be the means for transmitting morphological and syntactic information in signed languages. The continuous, analogue properties of space and movement are used in systematic, rule-governed ways in almost all signed languages that have been investigated so far (Armstrong et al. 1995; Newport 1981). JSL is not an exception for that. The abstract spatial and movement units are analogous in function to discrete morphemes found in spoken languages. Among those properties in spoken languages, it is hypothesized to be the grand sweeps in changes in fundamental frequency (dynamic peak-to-peak changes) that attract hearing infants to spoken motherese (Fernald 1985; Fernald and Simon 1984). As a visual analogue to such features, it is likely to be the sweeps of peak visual movement in sign motherese that attract deaf infants. These special properties evident in infant-directed communication may have universal attentional and affective significance, and human infants are predisposed to attend to the properties even if they have had no exposure to sign motherese before.

## 1.5 Emergence of Manual Babbling

The above findings would lead us to consider how the presence or absence of such linguistic input affects the linguistic behavior infants acquire. Concerning the issue of how the ontogenetic process of vocal behavior in deaf infants differs from that in hearing infants within the first year of their lives, little is known so far, partly because babbling has been anecdotally assumed to appear under maturational control. However, recent observations revealed that vocal ontogeny proceeds essentially similarly in deaf and hearing infants until the end of the marginal babbling stage (Masataka 2003). Thereafter, hearing infants come to produce canonical babbling whereas deaf infants are unable to do so. This was considered to be due to the necessity of auditory feedback for the acquisition of articulatory ability, an ability that makes the pronunciation of reduplicated consonant-vowel syllables possible.

Interestingly, the onset period of canonical babbling in hearing infants exactly coincides with the onset period of manual babbling which has been proposed to occur in deaf infants as well as in hearing infants by Petitto and Marentette. In the course of conducting research on signing infants' transition from pre-linguistic gesturing to first signs, they had been performing close analyses of the "articulatory" variables (analogous to "phonetic" level) of all manual activity produced by ASL deaf infants of ASL deaf parents from the age of 6 months. Actually, the presence of manual babbling had been noted by some researchers prior to the report of Petitto and Marentette. Prinz and Prinz (1979), on the basis of their observations between the age of 7 months and 21 months of a hearing infant born to a deaf mother, reported that the infant exhibited a group of manual behaviors that appeared to be produced as a result of imitation of signing of the mother. In another dyad of a hearing infant and its deaf mother, the infant was observed to clap and rub the hands in a circular motion by Griffith (1985). That behavior was interpreted as a prelexical form of a sign included in ASL. However, neither study had conducted such a systematic and quantitative analysis as Petitto and Marentette did. During the investigation, they noticed the presence of a class of manual activity that was unlike anything else that had been reported so far in previous literature.

These manual behaviors included linguistically relevant units but were produced in entirely meaningless ways, and they were wholly distinct in their pattern from all other manual activities, i.e., general motor activity, communicative gestures, and signs. They occurred between 9 and 12 months of age, the period corresponding to the canonical babbling stage for hearing infants acquiring speech. Indeed, subsequent analyses revealed that this class of manual activity was characterized by the identical timing, patterning, structure, and use of vocal behavior in hearing infants that is universally identified as babbling. As a result, it was termed "manual babbling." Further comparison of the ontogeny of manual action between hearing infants who were acquiring spoken language with no exposure to a signed language and deaf infants who were acquiring ASL as their first language revealed that both of them produced roughly equal proportions of communicative gestures (e.g., arms raised to request being picked up) at approximately 8 to 14 months of age.

Nevertheless, the deaf infants produced far more manual babbling forms (e.g., handshape-movement combinations exhibiting the phonological structure of formal signed language) than did the hearing infants, and manual babbling accounted for approximately 40% of manual activity in deaf infants and less than 10% of the manual activity of hearing infants.

Takei (2001) attempted to replicate the above findings in deaf infants growing up with exposure to JSL. He observed one male and one female infants, whose parents both acquired JSL as their first language. He made at least 1-h observations of each of the infants at their home when interacting with the mother freely with prearranged toys with the use of a video-recorder. Subsequently the video-recordings were transcribed in the laboratory. When the recorded manual activities were transcribed according to the same classification categories as employed by Pettito and Marentette (1991), manual babbling appeared first at the age of 6 months, and occurred most often when the infants were 10 months old. Immediately after that period, the first recognizable signs that were meaningful as JSL signs were recorded. After the onset of these first signs, the frequency of manual babbling tended to decline. Basically, the peak of manual babbling was found to be a reliable predictor of the onset of meaningful signs in the infant.

Taken together, these studies revealed that deaf infants with exposure to signed language showed manual babbling with much higher frequency than hearing infants. In addition, the phenomenon was most robust just prior to the onset of meaningful vocabulary in signed language in each of the deaf infants. In spite of this quantitative difference, however, even hearing infants, with no exposure to a signed language, were found to produce manual babbling. Given the fact that manual babbling precedes the onset of the first recognizable signs as vocal babbling precedes the onset of the first meaningful words, language capacity was taken by the authors of the article to be able to manifest itself equally as sign or speech in human infants; i.e., in hearing infants, the ability of vocal and gestural language can develop simultaneously and in a similar manner, at least during the preverbal period, though the degree to which they develop can vary, because some infants come to produce almost equal numbers of first words both in speech and sign around the identical period (Goodwyn and Acredolo 1993). In the case of deaf infants, the progression of acquiring spoken language is hindered earlier. This is to some extent due to the lack of auditory feedback that is crucial for the acquisition of articulatory ability in speech. As a consequence, they were considered to be obliged to exclusively rely on the manual mode to realize their linguistic capacity. This fact could partly account for a richer repertoire of manual babbling by the sign-exposed infants. However, it is not the only variable responsible for the difference.

Concerning other responsible variables, Masataka (2003) observed the progression of manual activity of 3 preverbal deaf infants of hearing parents with no exposure to a signed language and found that although the infants did perform manual babbling, the amount of the behavior was greater than that performed by hearing infants but much smaller than that performed by deaf infants with exposure to JSL. In that study, a comparison was made of the pattern of the development of manual activity between the age of 8 months and 12 months in three groups of

infants, i.e., three deaf infants of deaf parents with exposure to JSL, three deaf infants of hearing parents with no exposure to any signed language, and two hearing infants of hearing parents with no exposure to any signed language. When all of the infants' manual activities were transcribed in an identical manner, they were found to be classified into two types in every infant: syllabic manual babbling and gestures, and the infants of the three groups produced similar types and quantities of gestures during the study period. However, they differed in their production of manual babbling. Manual babbling behavior accounted for 25 to 56% of manual activity in deaf infants with exposure to JSL, but a mere 2 to 6% of the manual activity of hearing infants. Manual babbling as a percent of manual activity [manual babbling/(manual babbling + gesture)] showed a significant increase as the deaf infants developed, but did not exhibit such a change in the hearing infants. On the other hand, the value was intermediate between the values of these two groups for deaf infants with no exposure to any signed language (between 10% and 15%). Moreover, the value was greatest at the age of 8 months in all three groups of infants and showed a significant decrease as they developed.

Thus, the phenomenon of manual babbling does not necessarily imply that babbling is purely an expression of a "brain-based language capacity" that is amodal, or that babbling is not a phenomenon of motor development but is strongly associated with the abstract linguistic structure of language. Even vocal babbling arises with the aid of properties of motor organization that are shared by the motor system controlling the manual articulators, and thus the parallel progression of manual babbling and spoken babbling during the time course of their development per se should not be a very astonishing fact.

## 1.6 Antecedents of Sign Motherese

In all, sign motherese apparently acts as a behavioral adjustment to help children to learn languages using the manual mode. This fact should lead us to question what is predispositionally evolved in humans as the antecedents of such capability. In fact, Csibra and Gergely (2006) proposed that humans are provided with an evolutionary adaptation such that when presenting new information in general, adults create and children respond to a pedagogical context. Apparently, the enhancement of sign language learning that occurs in children caused by adults' sign motherese can fall into this context.

According to this account, children are sensitive to cues from adults that highlight information if the information is new and important for their learning of novel action that is important for their living in general. Namely, children are more likely to imitate acts that are marked as intentional and demonstrated for themselves. For example, if infants simply witness an adult bend over and touch a lamp with the hand in order to illuminate it, rather than having the adult specifically show them after making eye contact, infants are less likely to imitate it (Csibra and Gergely 2006). Given the fact that sign motherese is important for language learning in the

pedagogical context, one can expect similar levels of action learning in much broader contexts in infant-directed demonstrations relative to adult-directed demonstrations as long as both are presented in an intentional pedagogical fashion. Also, human children are extremely well known as voracious learners while adults are ubiquitous teachers!

In this regard, the findings reported by Brand, Baldwin, and Washburn (2002) should be noteworthy. In that study, teaching behavior about the usage of novel objects was video-recorded and subsequently analyzed in adults. When their performance was compared when it was directed toward 6- to 13-month-old infants and when it was directed toward another adult, a particular suite of embellished behavior was found to emerge when it was directed toward infants, which has been referred to as “motionese” or infant-directed action. Given young children’s fledging attentional control, the motionese system of behaviors incorporates some previously studied features, such as eye gaze and emotional expressiveness (Chong et al. 2003), but also involves a variety of other modifications. These include closer proximity to a child versus an adult partner, greater enthusiasm, a large range of motion, simplified action sequences, greater repetitiveness, and higher interactiveness, including more and longer gazes to infants’ faces and more turn-taking. Extensions of this work have also found evidence of longer pauses in infant-directed action compared with adult-directed action and a unique coordination of speech and action in demonstrations for children (Rohlfing et al. 2006).

On the basis of these findings, a question concerning the effectiveness of motionese as a teaching behavior would arise. Although the possibility remains that this medley of cues could be distracting or frustrating relative to the straightforward adult-directed action style, it should be more likely that these cues function to provide a richer learning experience than a standard demonstration. If so, then motionese would fit within a suite of behavioral adjustments for communication with infants and children, including the motherese in the vocal mode and in the manual mode that have been documented so far.

In fact, suggestive evidence for this possibility has been presented by Brand and Shallcross (2008), who reported that motionese preferentially attracted children’s attention, namely, that 6- to 13-month-old infants looked longer at infant-directed versus adult-directed demonstrations when both were available to view simultaneously. This was the case when the faces were digitally blurred, to obscure eye gaze and expressive information, suggesting a role for the entire suite of behaviors and not simply facial cues. Increased scrutiny or interest in the behaviors could lead children to encode and remember the actions better for subsequent imitation.

The results of the study also indicated the possibility that motionese may have functions beyond attracting children’s attention or preparing them for new information: motionese modifications may also help to make demonstrated acts easier to parse (e.g., by highlighting boundary points with repetition and eye gaze), stress which body parts and subtle physical motions are necessary (e.g., a horizontal twist before vertically pulling off a cup), and highlight the intentions behind the acts (e.g., by exaggerating facial expressions of surprise and satisfaction). If motionese



functions in this manner, children may learn more easily, and imitate more faithfully, when the suite of cues is shown throughout the demonstration.

Therefore, in order to directly test whether motionese influences young children's observational learning of novel acts, whether the special infant-directed action modifications parent use when teaching their children really improved 2-year-olds' imitation was investigated in a subsequent study (Williamson and Brand 2014). In that study, a total of 48 children saw an adult perform a series of acts on four novel objects using either an infant-directed style (e.g., larger range of motion and enhanced boundary marking with more repetition) or an adult-directed style. After each demonstration, the children received a test period for that object during which they were allowed to play for 25 s. When the videos recorded during that period were coded, children who saw any demonstration (either infant-directed or adult-directed) were found to show imitation in that they were more likely to produce the target acts than were children in a non-demonstration baseline group. Moreover, children who saw demonstrations augmented with motionese exhibited even higher levels of imitation than did children who saw adult-directed demonstrations, indicating that infant-directed demonstrations were particularly effective teaching behaviors.

In fact, a rudimentary form of motionese is known in nonhuman animals. Importantly, such behavior is assumed to serve as a basis on which social learning is enhanced in the animals.

The population-level instrumental use of objects present in the external world has been regularly reported among nonhuman animals. Perhaps the most famous case of such an instance is that of Imo, a Japanese macaque (*Macaca fuscata*) that learned that washing sand-covered sweet potatoes made them more palatable (Kawai 1963). Imo's exercise spread through her community and became standard practice within approximately 10 years. It is not clear, however, just how the transmission occurred, but it seems likely that observational learning played a part.

More recently, Masataka et al. (2009) saw evidence of the social transmission of flossing in a troop of free-ranging long-tail macaques (*Macaca fascicularis*) in Thailand. The researchers first observed nine adult monkeys routinely pull the hair from the head of women tourists and use the hair to floss their teeth. Flossing has since become widespread in the troop. Although as with the potato washing, there is no clear demonstration of how the behavior spread, Masataka et al. (2009) found evidence that the mother monkeys attempted to teach flossing to their offspring through modeling. When the behavior of seven female macaques was videotaped when they were flossing and their behavior was compared when their infant was nearby and when it was not, they paused more often, repeated flossing more frequently, and flossed for a longer time when their infants were present than when their youngsters were not around. Though there is no direct evidence that the infants imitated their mothers' behavior, the motionese-like behavior observed suggests that modeling played a part in the spread of flossing in the troop and that the spread was enhanced by such behavior modifications.

## 1.7 Role of Sign Motherese as a Part of Multimodal Motherese in the Development of Social Understanding in Infants

Taken together, many findings show that the so-called motherese phenomenon occurs either in the vocal mode (infant-directed speech, or speech motherese) or in the manual mode (infant-directed signing, or sign motherese). Motherese, whether in the vocal mode or in the manual mode, highlights mother–child adjustments during interactions. This is because these behavioral characteristics are a product that has evolved as the historical extension of motionese as a device of enhancing social learning that has merged during the evolution of nonhuman primates.

Mothers are able to adjust their motherese to children's age, cognitive abilities, and linguistic level. Therefore, motherese may arouse children's attention by signaling mothers' linguistic behavior that is addressed to the children. Mothers also adapt their motherese to children's reactivity and preferences. Mothers' continuous adjustments to their children result in the facilitation of exchanges and interactions, with positive consequences for sharing emotions and for learning and language acquisition. Children's reactivity is also important given that their presence increases motherese, and children's positive, contingent feedback makes them more affective, which in turn increases the quality of the motherese.

In all, motherese mediates and reflects an interactive loop between the child and the caregiver, such that each person's response may increase the initial stimulation of the other partner. At the behavioral level, the interactive loop is underpinned by the emotional charge of the affective level and the construction of intersubjective tools, such as joint attention and communicative skills. Based upon such findings, Gogate, Bahrick, and Watson (2000) claim that infant-directed communication, even if it occurs between hearing infants and their caregivers, is as seldom unimodal as it is in face-to-face communication between hearing adults, and that it is more complex and broader topic than has been assumed. They argue that the entire sensory system perceives information about the dialogue partner. Communication is so much more than speech, and manual and gestural movements. It encompasses more than simply focusing on visual and audible signals: variations in intonation, pitch, intensity, and gestures, facial expression, and eye gaze are recognizable. As reasoned by Ugur et al. (2015), humans need additional information to interpret ambiguous information in a dialogue, whose partners need to establish common ground in order to recognize each other's intentions and theory of mind. Infants may, however, have an advantage because child-directed communication is simpler, prosodic, and visual features are more exaggerated, which makes certain sensory input more salient, and intentions are therefore easier to recognize and patterns in the information stream can be more easily identified.

While there was a clear separation in child-directed research between that of different modality, mainly auditory and visual, the above reasoning has led researchers to orient toward the interplay of amodal input. Bahrick and Lickliter (2000) theorize that intersensory redundancy makes the given more salient in a

constant stream of arbitrary information, stating that young hearing infants prefer temporal synchronous and amodal stimuli over asynchronous stimuli. For example, when hearing somebody speak in the same rhythm, they are more likely to see her/his movements of lips in the same rhythm than to see her/his movements of lips in the different rhythm.

Temporal synchronous, redundant, and amodal information can direct attention to meaningful events, which may be beneficial for learning. In addition, naming an object and touching or moving or rather showing it are redundant information across the auditory and visual senses and therefore make the relation between the name and the object salient. Such reasoning has led researchers to no longer distinguish motherese in the vocal mode and that in the manual mode, but to include both in the category of multimodal motherese (Gogate et al. 2000). Coming from the perspective of motionese research, Brand and Tapscott (2007) explored the possibility that utterances enhanced the perception of action in 9- to 11-month-old hearing infants and found that co-occurring infant-directed speech and motionese facilitated the segmentation of the observed action into smaller units.

Reinforcing the notion of multimodal motherese, Nagai and Rohlfing (2009) found that when interacting with infants, as opposed to adults, parents attempt to employ a number of strategies in order to increase the saliency of the objects and their initial and final states. These strategies include suppressing their own body movements before starting to execute the task, and generating additional movements on the object, such as tapping it on the table. The resulting behavior is qualitatively different from adult-adult interaction in observable parameters, such as the pace or the smoothness of the movement. It is shown that infants also benefit from this behavior. Koterba and Iverson (2009) showed that hearing infants exposed to a higher repetition of demonstration exhibit longer bangs and shakes of objects whereas infants exposed to a lower repetition spent more time for turning and rotating objects. Ugur et al. (2015) developed a bottom-up architecture of visual saliency that can be used in an infant like learning robot. In this architecture, similar to an infant, the robot tends to find the salient regions in caregiver's action, when they are highlighted by multimodal motherese. Moreover, the saliency preference of the robot also motivates humans to use such motherese as if they were interacting with a human infant, thereby completing the loop. Due to the limited attention mechanism of the robot, human participants tried to teach the task (cup-stacking) by approaching toward the robot and introducing the object in the proximity of the robot. In general, participants amplified their movements and made pauses as if to give the robot a chance to understand the scene. Such delimiting pauses and exaggerations were those also observed in sign motherese as well as in spoken motherese.

The advantage infants enjoy with exposure to multimodal motherese could not be restricted to language learning, but be extended further to the other aspect of child development, e.g., to learning of social understanding that has apparently been a topic of great interest. While one hallmark of social understanding is the acquisition of a theory of mind between 3 and 5 years of age, which is the ability to predict and explain social behavior on the basis of mental state, children acquire certain social

cognitive abilities that are predictive of their later social understanding already in infancy (Rakoczy 2012). Such an early competence is the ability to encode human actions as goal-directed. Woodward (1998) showed 6-month-old infants events in which a person reached for one of two toys. After habituation to this event, infants looked significantly longer on test trials in which a person grasped the other toy (new goal trials) than when the person reached for the old toy, but the path of the grasping was changed (old goal trials). Because infants show this preference only for human goal-directed action, and not for the motion path of a mechanical claw, the looking-time patterns indicate that infants encode actions in terms of agent–goal relations. The ability to explain behavior by ascribing goals to an agent, which infants develop during the second half of their first year of life, has turned out to be an important early precursor of later mental state understanding. In fact, Aschersleben, Hofer, and Jovanovic (2008) found a positive correlation between 6-month-old infants' decrement of attention to goal-directed action and their false belief understanding at 4 years of age.

Licata et al. (2014) investigated the relationship between mother–child interaction quality and infants' ability to interpret actions as goal-directed at 7 months assessed by the testing described above. Interaction quality was assessed in a free play interaction using two distinct methods: one assessed the overall affective quality (emotional availability), and one focused on the mother's proclivity to treat her infant as an intentional agent with autonomous thoughts, intentions, emotions, and desires (mind-mindedness). There are good reasons to expect a positive association between both emotional availability and mind-mindedness, and infants' understanding of intentional action. Maternal emotional availability might be beneficial for infants' goal encoding, as learning especially is promoted in a warm and sensitive environment, in which a child can explore his/her environment while being supported by the mother. Maternal mind-mindedness, on the other hand, could be related to infant's goal encoding, as mind-minded mothers may have infants who are better at interpreting human actions as goal-directed because these mothers allow their infants to experience themselves as self-efficient by verbally commenting on their mental states and appropriately responding to them.

Analyses revealed that only maternal emotional availability assessed by the Emotional Availability Scales (EMS) (Biringen 2008), and not maternal mind-mindedness assessed according to the protocol developed by Meins and Fernyhough (2010), was related to infants' goal-encoding ability. The link remained stable even when controlling for child temperament, working memory, and maternal education.

EAS is a method that does not quantify distinct behaviors but analyzes the interactional style of the dyad. It is an emotion-focused measure that refers to the overall affective quality of the relationship. The construct of emotional availability is multidimensional, as it comprises different dimensions of caregiving. Licata et al. argue that the fostering of maternal emotional availability for infants' goal-encoding skills could be mediated through specific ways of the mother engaging in goal-directed behavior with her infant (such as multimodal motherese, and sign motherese), which in turn could support infants' goal sensitivity.

Infant-directed action represented by sign motherese could serve itself as a basis on which infants are enhanced to get access to the meaning of action by helping them to recognize goals and intentions that motivate action. An emotionally available mother could have several attributes that characterize such motherese: an emotionally available mother shows much positive affect and interacts with her child, showing turn-taking, eye contact, and joint attention. It is highly possible to assume that maternal emotional availability also enhances infants' attention and helps them to learn to recognize goals and intentions that motivate action. One reason for this relation could be that magnified emotions amplify information about intentionality and help infants recognize goal achievement (Brand et al. 2002), which is necessary to encode human actions as goal-directed.

More recently, in fact, the findings confirming such reasoning have been reported by a neuroendocrine study employing a computational method on video vignettes of human parent–infant interaction including 32 fathers that were administered oxytocin or a placebo in a crossover experimental design (Weisman et al. 2018). The role of such nine-amino acid peptide hormone in the initiation and maintenance of affiliative bonds and parental repertoire has been elucidated in humans as well as in nonhuman animals (Feldman 2012). Acute administration of oxytocin to a parent was found to enhance the care's, but at the same time also the infant's physiological and behavioral readiness for dyadic social engagement (Weisman et al. 2012). Yet, so far, the exact cues that are involved in this affiliative transmission process have remained unclear.

Weisman et al. revealed the fact that oxytocin administration markedly increased the maximum distance between the father and the infant but at the same time drove the father to reach their closest proximity to the infant earlier, as compared with the placebo condition. Moreover, a father's head tended to move faster and to reach its highest velocity sooner under oxytocin influence. Fathers showed greater acceleration which also tended to vary more. Finally, infant oxytocin increase following interaction with the father correlated with the father's maximum acceleration parameter.

Given that the behavioral variables assessed in the study obviously fall under the definition of motherese in the manual mode, overall implications of its results for deaf studies of cognition and learning should be profound. It implies that such modality is susceptible to intervention in key neuroendocrine pathway that is central to the initiation and maintenance of parental repertoire, namely, the oxytocin system. Moreover, the modality has manifested itself as a consequence of the evolution of the cross-generational transmission of parenting in humans. Future studies should examine whether oxytocin shapes proximity and motion among deaf mothers interacting with their infants, using a signed language, and whether the communication patterns shaped under the hormone affects the infants' subsequent cognitive development.

## 1.8 Conclusions

The reasoning that I have documented so far indicates that, developmentally, the early manual action performed by deaf infants, like non-cry sounds produced by hearing, infants do not transmit meanings, but rather reveal fitness or express states. We make no assumption that very young infants intend to communicate with such action. Moreover, this is also observed in hearing infants. On the other hand, adults who receive the sounds are provided with broad reception/interpretation capabilities that allow them to provide differentiated feedback to signals if they have acquired sign languages. Such reasoning could lead us to notice the importance of considering how selection pressures might engender increase in the tendency of an infant to produce spontaneous meaning less manual action for arguing language evolution at the manual mode, and that in fact, evolution of increased infant spontaneous sign action began with developmental steps in individual deaf infants under the selective pressure of their own deaf caregivers.

In all, during modern human development, infant linguistic capabilities emerge at least partly in response to social interaction either at the vocal mode or at the manual mode, where caregivers react to linguistic capabilities of infants in accordance with a scaffolding principle requiring parental discernment and intuitive parenting to reinforce linguistic exploration and learning. Both endogenous inclination of infants to explore the linguistic space and interactive feedback from caregivers thus foster growth in linguistic capability either at the vocal mode or at the manual mode.

If this assumption about selection pressures for spontaneous action either at the vocal mode or at the manual mode are valid, it follows that hominin parents would also have been selected to become aware of the fitness reflected in infant action and capable of responding to those indicators with selective care and reinforcement of the action. The author considers this to have been illustrated so far in the phenomenon that is called “motherese.” And the phenomenon of motionese implies the underlying evolutionarily antecedents for that, by which infants, whether hearing or deaf, are enabled to learn language independent of the auditory capability. In this sense, the gestural theory of language origins can be said to exert relevancy to exactly the same degree to that the vocal theory of language origins can exert relevancy.

## References

- Armstrong DF, Stokoe WC, Wilcox SE (1995) *Gesture and the nature of language*. Cambridge University Press, Cambridge
- Aschersleben G, Hofer T, Jovanovic B (2008) The link between infant attention to goal-directed action and later theory of mind abilities. *Dev Sci* 11:862–868
- Bahrick LE, Lickliter (2000) Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 35:190–201
- Biringer Z (2008) The emotional availability (EA) scales. Unpublished Coding Manual

- Brand RJ, Shallcross W (2008) Infants prefer motionese to adult-directed action. *Dev Sci* 11:853–861
- Brand RJ, Tapscott S (2007) Acoustic packaging of action sequences by infants. *Infancy* 11:321–332
- Brand RJ, Baldwin DA, Ashburn LA (2002) Evidence for “motionese”: modifications in mothers’ infant-directed action. *Dev Sci* 5:71–87
- Chong S, Werker J, Russell J, Carroll J (2003) Three facial expressions mothers direct to their infants. *Infant Child Dev* 12:211–232
- Cooper RP, Aslin RN (1990) Preference for infant-directed speech in the first month after birth. *Child Dev* 61:1584–1595
- Csibra G, Gergely G (2006) Social learning and social cognition: the case of pedagogy. In: *Processes of change in brain and cognitive development. Attention and performance*, vol 21. Oxford University Press, New York, pp 249–274
- DeCasper AJ, Lecanuet JP, Busnel MC, Garnier-Deferre C, Maugeais R (1994) Fetal reactions to recurrent maternal speech. *Infant Behav Dev* 17:159–164
- Erting CJ, Prezioso C, O’Grandy Hynes M (1990) The interactional context of deaf mother-infant communication. In: Voltera V, Erting CJ (eds) *From gesture to language in hearing and deaf children*. Springer, Berlin, pp 97–106
- Feldman R (2012) Oxytocin and social affiliation in humans. *Horm Behav* 61:380–391
- Ferguson CA (1964) Baby talk in six languages. *Am Anthropol* 66:103–114
- Fernald A (1985) Four-month-old infants prefer to listen to motherese. *Infant Behav Dev* 8:181–195
- Fernald A, Simon T (1984) Expanded intonation contours in mothers’ speech to newborns. *Dev Psychol* 20:104–113
- Gogate LJ, Bahrick LE, Watson JD (2000) A study of multimodal motherese; The role of temporal synchrony between verbal labels and gestures. *Child Dev* 71:878–894
- Goodwyn SW, Acredolo LP (1993) Symbolic gesture versus word: is there a modality advantage for onset of symbol use? *Child Dev* 64:688–701
- Griffith PL (1985) Mode-switching and mode-finding in a hearing child of deaf parents. *Sign Lang Stud* 48:195–222
- Gusella J, Roman M, Muir D (1986) Experimental manipulation of mother-infant actions. Paper presented at the international conference of infant studies, Los Angeles, September
- Iverson P, Kuhl PK (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J Acoust Soc Am* 97:553–562
- Kawai M (1963) On the newly-acquired behaviors of the natural troop of Japanese monkeys on Koshima Island. *Primates* 4:113–115
- Klima ES, Bellugi U (1979) *The signs of language*. Harvard University Press, Cambridge, MA
- Kotterba EA, Iverson JM (2009) Investigating motionese: the effect of infant-directed action on infants’ attention and object exploration. *Infant Behav Dev* 32:437–444
- Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, Stolyarova EI, Sundberg U, Lacerda F (1997) Cross-language analysis of phonetic units in language addressed to infants. *Science* 277:684–686
- Licata M, Paulus M, Thorermer C, Kristen S, Woodward AL, Socian B (2014) Mother-infant interaction quality and infants’ ability to encode actions as goal-directed. *Soc Dev* 23:340–356
- Masataka N (1992) Motherese in a signed language. *Infant Behav Dev* 15:453–460
- Masataka N (1996) Perception of motherese in a signed language by 6-month-old deaf infants. *Dev Psychol* 32:874–879
- Masataka N (1998) Perception of motherese in Japanese Sign Language by 6-month-old hearing infants. *Dev Psychol* 34:241–246
- Masataka N, Koda H, Urasopon N, Watanabe K (2009) Free-ranging macaque mothers exaggerate tool-using behavior when observed by offspring. *PLoS One* 4:e4768
- Masataka N (2003) *The onset of language*. Cambridge University Press, Cambridge
- Meins E, Fernyhough C (2010) *Mind-mindedness coding manual*. Unpublished Coding Manual

- Nagai Y, Rohlfing KJ (2009) Computational analysis of motionese toward scaffolding robot action learning. *IEEE Trans Auton Ment Disord* 1:44–54
- Newport E (1981) Constrains on structure: evidence from ASL and language learning. In: Collins W (ed) *Minnesota symposia on child psychology*. Erlbaum, Hillsdale, pp 65–128
- Oller DK, Griebel U, Warlaumont AS (2016) Vocal development as a guide to modeling the evolution of language. *Top Cogn Sci* 8:382–392
- Panneton RK (1985) Prenatal experience with melodies: effect on postnatal auditory preference in human newborns. Ph.D. dissertation, University of North Carolina, Greensboro
- Pettito LA, Marentette PF (1991) Babbling in the manual mode: evidence for the ontogeny of language. *Science* 251:1493–1496
- Prinz PM, Prinz EA (1979) Simultaneous acquisition of ASL and spoken English: Phase I: Early lexical development. *Sign Lang Stud* 25:283–296
- Querleu D, Renard K (1981) Les perceptions auditives du fœtus humain [Auditory perception of the human fetus]. *Journal de Gynecologie Obsterique et Biologie de la Reproduction* 10:307–314
- Querleu D, Renard K, Crepin G (1981) Perception auditive et reactive foetale aux stimulations sonores. *Journal de Gynecologie Obsterique et Biologie de la Reproduction* 10:307–314
- Rakoczy H (2012) Do infants have a theory of mind? *J Dev Psychol* 30:59–74
- Rohlfing KJ, Fritsch J, Wrede B, Jungmann T (2006) How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Adv Robot* 22:1183–1199
- Saint-Georges C, Chetouani M, Cassel R, Apicella F, Mahdhaoul A, Muratoni F, Laznik M-C, Cohen D (2013) Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review). *PLoS One* 8:e78103
- Sullivan JW, Horowitz FD (1983) Infant intermodal perception and maternal multimodal stimulation: implication for language development. In: Liositt LP, Rovee-Collier CK (eds) *Advances in infancy research*. Ablex, Norwood, pp 184–239
- Takei W (2001) How do deaf infants attain first signs? *Dev Sci* 4:71–78
- Trainor LJ (1996) Infant preferences for infant-directed versus noninfant-directed play songs and lullabies. *Infant Behav Dev* 19:83–92
- Ugur E, Nagai Y, Celikkanat H, Oztop E (2015) Parental scaffolding as a bootstrapping mechanism for learning grasp affordance and imitation skills. *Robotics* 33:1163–1180
- Weisman O, Zagoory-Sharon O, Feldman R (2012) Oxytocin administration to parent enhances infant physiological and behavioral readiness for social engagement. *Biol Psychiatry* 72:982–989
- Weisman O, Delaherche E, Rondeau M, Chetouani D, Feldman R (2018) Oxytocin shapes parental motion during father-infant interaction. *Biol Lett* 9:21030828
- Werker JF, McLeod PJ (1989) Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Can J Psychol* 43:320–346
- Werker JF, Pegg JE, McLeod PJ (1994) A cross-language investigation of infant preference for infant-directed communication. *Infant Behav Dev* 17:323–333
- Williamson RA, Brand RJ (2014) Child-directed action promotes 2-year-olds' imitation. *J Exp Child Psychol* 118:119–126
- Woodward AI (1998) Infants selectively encode the goal object of an actor's reach. *Cognition* 69:1–34