Emerging Linguistic Functions in Early Infancy

Francisco Lacerda*, Eeva Klintfors, Lisa Gustavsson
Ellen Marklund and Ulla Sundberg

Department of Linguistics
Stockholm University
SE-106 91 Stockholm
Sweden
* frasse@ling.su.se

Abstract

This paper presents results from experimental studies on early language acquisition in infants and attempts to interpret the experimental results within the framework of the Ecological Theory of Language Acquisition (ETLA) recently proposed by (Lacerda et al., 2004a). From this perspective, the infant’s first steps in the acquisition of the ambient language are seen as a consequence of the infant’s general capacity to represent sensory input and the infant’s interaction with other actors in its immediate ecological environment. On the basis of available experimental evidence, it will be argued that ETLA offers a productive alternative to traditional descriptive views of the language acquisition process by presenting an operative model of how early linguistic function may emerge through interaction.

1. Introduction

Previous studies of the young infant’s ability to learn names of objects presented under controlled naturalistic settings have demonstrated that by about 7 to 8 months of age, infants are capable of interpreting arbitrary words as names of visual objects, provided the words and the objects co-occur consistently. For instance, a study by Gogate and Bahrick (Gogate & Bahrick, 1998) indicates that 7 month-old infants are able to explore audio-visual co-occurrences to establish arbitrary word-like associations between isolated speech sounds and objects. In addition, a more general assessment of the impact of audio-visual synchrony (Prince et al., 2004) strongly suggests that synchronic events may expose linguistically relevant audio-visual relationships. But while the young infant’s ability to establish sound-object links offers good support to the notion that association processes are likely to underlie early language acquisition, accounting for the language acquisition process in terms of relatively simple associative processes involving isolated words is problematic because it may lack general ecological relevance. Indeed, as often pointed out by scientists criticizing the emergentist views of the language acquisition process (Lidz, Gleitman, & Gleitman, 2003), words representing the names of objects available to the young infant tend to be embedded in utterances rather than uttered in isolation, an aspect that necessarily reduces the ecological relevance of experimental studies reporting referential learning from words presented in isolation. Thus, to further investigate the extent to which general association processes might underlie early language acquisition in ecologically relevant adult-infant interaction settings, a series of experiments were set up in which the target words were integrated in natural sentences (as those typically heard by infants) and arbitrarily combined with visual objects simultaneously accessible to the infants.

The present paper will argue that early language acquisition can indeed be seen as the result of an interactive process between the infant and its environment, through which the infant picks up linguistic regularities afforded in the ambient language. In the following we try to provide an empirical basis for our emergentist views of the early language acquisition process by reviewing some of our experimental studies addressing different aspects of early language acquisition in infants and examining the characteristics of the infant-directed. We will first review an experiment designed to test how different linguistic factors may influence the infant’s ability to derive word-object relationships from exposure to naturalistic audio-visual contingencies. Thereafter we will examine the characteristics of infant-directed speech from the perspective of the ETLA. Finally, we will address the issue of necessity of general-purpose versus language-specific processes underlying the infant’s ability to link visual and auditory information and form productive linguistic representations.

2. Emerging word-object associations

To investigate the generality of the word-object association process, a series of experiments were carried out to investigate the infant’s ability to derive the names of objects from experience with audio-visual stimuli, where natural sentences conveying implicit referential information are presented simultaneously with the visual images of the objects they refer to. One setup of these studies was already described in (Gustavsson et al., 2004) and will be briefly reviewed here.

This study used a Visual Preference procedure similar to the procedures used by Fernald and her colleagues (Fernald, Swingley, & Pinto, 2001; Swingley, Pinto, & Fernald, 1999). In general terms, the procedure can be described as inducing the infant’s response from its looking time towards alternative pictures displayed simultaneously and where one of the pictures is associated with the expected response.
2.1 Speech materials

The speech materials were Swedish sentences recorded by a female native speaker of Swedish. The utterances introduced non-words as names of the objects being displayed on the screen. Nine films were created to include all the possible combinations of position of the target word (initial, medial or final position in the utterance) and the utterances focal accent (falling on the utterances initial, medial or final words). The syntactic structure of the utterances was different from film to film but within each film the position of the target word and the part of the utterance receiving focal accent was kept constant. Furthermore, although the utterances within each were structurally equal, the non-target words were different from utterance to utterance in an attempt to mimic the variation typically observed in natural utterances. Examples of the utterances presented in two of the nine films are shown in table 1, where the focal accent is indicated by boldface and the position of the target word by XXX. For the placement of the target word and focal accent, the utterances were divided in three regions – initial, medial and final. The initial and final positions were defined by the first and the last word in the utterance. The medial position was defined as the remaining part of the utterance.

<table>
<thead>
<tr>
<th>Film 1 target word: final focal accent: final</th>
<th>Film 2 target word: final focal accent: medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titta här är söta XXX</td>
<td>Det är den söta XXX</td>
</tr>
<tr>
<td>Se på den lilla XXX</td>
<td>Se på den lilla rara XXX</td>
</tr>
<tr>
<td>Titta på fina XXX</td>
<td>Titta på fina glada XXX</td>
</tr>
<tr>
<td>Kolla in den glada XXX</td>
<td>Kolla glada XXX</td>
</tr>
</tbody>
</table>

Table 1. Example of the Swedish utterances presenting the target words. The target word is represented by XXX, standing for the non-words “Kucka” and “Dappa”. Focal accent is represented by boldface.

Each of the nine films was organized in three phases – baseline, exposure and test.

In the baseline phase, still images of two puppets were displayed side by side in a split-screen. The duration of the baseline phase was 30 s. During the baseline phase an especially composed short instrumental lullaby (Anna Ericsson, 2004) was playing, starting approximately 2 s after the onset of the visual display and finishing about 2 s before the end of the baseline phase. The infant’s looking towards each of the puppets during this phase was used as a measure of the subject’s preferential bias towards the puppets.

During the exposure phase, two short 20 s video sequences were played to show each of the puppets per se, introduced by the sentences referring to the particular puppet being displayed (see table 1). The sentences were evenly distributed throughout the duration of each video sequence. The first sentence started about 1 second after the onset of the visual display and the last sentence finished about 1 s before switching to the next video sequence. These video sequences were presented after each other, switching from one puppet to the other. The total duration of the exposure phase was 120 s, during which each of the individual video sequences was presented 3 times. The infants’ looking time towards the each of the puppets was taken as a measure of attention during the exposure phase.

In the test phase the two puppets were again displayed in a split-screen similar to that of the base-line but now the audio track played questions like “Where is XXX?” or “Can you see XXX?”, where XXX was the name of one of the puppets, implicitly introduced in the descriptions presented during the exposure phase. The test phase was 30 s long, just as the baseline phase.

2.2 Subjects

A total of 49 infants participated in the study. Some of the infants participated in more than one session, adding up to 78 sessions. The results presented here come from a total of 75 sessions, distributed as indicated in the table below.

The ages of the subjects at the time of their participation in the sessions ranged from 201 to 278 days (mean age was 239 days, s.d.=15 days). The age distribution for this sample was nearly Gaussian (skewness=0.180; kurtosis=0.503).

<table>
<thead>
<tr>
<th>Target word position</th>
<th>Focal accent</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>medial</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>final</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>22</td>
<td>25</td>
<td>75</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Number of data points used for specific combinations of target word and focal accent positions.

The selection criterion for the sessions above required that the looking bias towards one of the puppets did not exceed 35% of the baseline time.

2.3 Procedure

The subjects were video recorded during the experiments, using a camera placed just above the display they were looking at. To register the actual images that the infant was looking at and give the possibility of re-analysis, the film being displayed was mixed onto the upper left corner of the image of the subject’s face. This overlapping image used about 1/16 of the screen area and did not interfere with the image of the face of the infant. A time stamp placed at 40 ms intervals was also recorded on the lower right corner of the screen. This time stamp was subsequently used to compute a session-relative time, allowing the line up the start of the video films from different subjects.

In this experiment, the looking times towards each of the puppets were measured manually, frame by frame, a very time consuming procedure. The separation between the target images was about 30°, which was enough to allow clear decisions on which side of the screen the subject was momentarily looking at. Three levels of looking were coded – left, right and off.

On the basis of these codes, a “pre-to-post exposure gain” variable was defined as the net increment in looking time towards the puppet used as target.
Gain = Tgt − TgtB

where Tgt is the total looking time towards target puppet and TgtB is the total looking time during baseline towards the puppet that would become the target in the test phase.

3. Results

The results from the first sessions in which the 36 selected subjects participated are shown in figure 1, grouped according to the placement of the focal accent and the position of the target word in the utterances.

Given the reduced number of subjects each condition and the typical variance observed in this type of experiments, it is perhaps not surprising that no significant main effects for the target word or placement of the focal accent could be observed. As stated above, this analysis was carried out on a selection of all the sessions in which the infant’s initial bias towards any of the puppets was less than 35% of the total baseline time. Further analyses using all the available data from the 78 sessions did not change appreciably the pattern displayed on figure 1. The main difference was a broadening of the confidence interval for medial target word position with medial focus, due to an extreme negative gain outlier resulting from a strong bias towards the puppet that would function as target. A non-parametric display of the same data is shown in figure 2. The dependence of the median values on the target word and focal accent position is in good agreement with the pattern displayed in figure 1. There were no significant main effects or interactions for word position and placement of the focal accent. However a tendency for longer looking times was observed for the target word in focal position ($F(1,73)=2.957, p<0.090$). If the case of the target word in final position, with a focal accent in the initial position of the sentence is excluded, then a significant effect of the placement of the target word in focus is obtained ($F(1,65)=4.075, p<0.048$). Furthermore, there was a significant difference between the mean looking times for the group of sentences with the target word in focal position plus the sentences with the target word in final position and focal accent in initial position ($F(1,73)=5.579, p<0.021$).

4. Discussion

While there were no overall significant differences when considering all the data at once, the response pattern displayed in figures 1 and 2 strongly suggested that target words in focal position might have been easier to associate with the corresponding puppets than when focal accent did not fall on the target word. This means that 8-month-old infants seem to be able to pick up relevant linguistic information by listening to the word that is placed in focal position. An unexpected effect was however observed when the target word was in non-focal final position but the utterance had initial focal accent. It appears that the initial focal accent may have primed the infants to attend to the utterance, prompting the subjects to retrieve the less prominent target word delivered in sentence final position.

In summary, the results of this experiment seem to indicate a general ability to link recurrent target words with visual objects that are simultaneously available to 8-month-old infants, providing the ground for the linguistically relevant referential function. The fact that the strength of the responses varied significantly for different combinations of focal accent target word placement further suggests that the infants’ ability to pick up the linguistic referential function was modulated by prosodic patterns and primarily contingent on the coherence in the placement of the focal accent and the target word. An implication of this is that by 8 months of
age, deriving a linguistic referential function on the basis of exposure to running speech may not be simply a matter of co-occurrence of recurrent sound strings (representing the target word) and visual objects (the puppets to which the target words refer), rather a process in which the acoustic salience of the sound strings plays a decisive role. This notion is reinforced by the fact that focal accent on the initial part of the utterance seems to have enhanced the response to the target word that occurred in utterance final position.

Figure 3. Waveforms (-1 to +1), intensity curves (-60 dB to 0 dB) and $f_0$ contours (100 Hz to 600 Hz) for the first sentences in the conditions. Time scale in seconds.

Top – Focal accent and target word in final position
Mid – Focal accent initial, target word final
Bottom – Focal accent and target word medial.

The target word is uttered during the interval 0.9-1.6 s, approximately. The high and flat $f_0$ contour at about 1.5 s in this sentence is not due to saturation.

Three examples of sentences introduction the target word “kucka” are shown in figure 3 to illustrate the dramatic $f_0$ excursions associated with the focal accent. Such pitch variations are typical of infant-directed speech. They introduce over one octave increase in pitch and are likely to be salient enough to capture the infant’s attention towards the sound string being uttered. Our previous studies addressing younger (under 6 months of age) infants’ to discriminate between short utterances exposing the interaction between the target words and focal accent placement suggested however that those younger infants might not able to pay attention to the target word when strong focal information was added (Lacerda & Sundberg, 1996). To be sure Lacerda and Sundberg’s (1996) study concerned discrimination and the speech materials were much simpler than those used in the experiment above, but the results may probably be taken as an indication that the younger infants might not have succeeded in the current task.

5. Linking audio-visual information

The emergence of the linguistic referential function suggested by the study reported in the previous section may be seen as a consequence of a general multi-sensory representation process through which synchronic multi-sensory information is spontaneously associated, thereby exposing implicit cross-modal regularities (Lacerda et al., 2004a; Lacerda et al., 2004b; Lacerda, 2003). Because the efficient use of spoken language is based on the ability to relate sound symbols (however variable) to objects perceived (primarily but not exclusively) by other senses, a systematic (or at least predictable) link between the sound code and the objects it refers to must exist at some level of representation (Minsky, 1985). Note that in line with ETLa (Lacerda et al., 2004a), such a sound code is a generic reference to the concrete auditory impression of a word or a lexical phrase as a whole, not to the word’s representation in terms of linguistic concepts like phonemes or syllables nor to the sequence of words that may build up the lexical phrase. In this perspective, words, syllables and phonemes are an emergent consequence of the combinatorial pressure imposed by increasing representation needs (Nowak, Plotkin, & Jansen, 2000; Lacerda, 2003).

To address the issue of the generality of cross-modal links in infancy, we carried out a study to investigate the infant’s ability to use temporal synchrony to relate ecologically relevant auditory and visual speech information, the infant’s ability to relate ecologically relevant synchronous non-speech audio and visual information and the infant’s ability to relate synchronic non-speech audio with speech (articulatory) visual information.

The background for the present experiments is an early Kuhl and Meltzoff’s study (Kuhl & Meltzoff, 1982) showing that 18 to 20 weeks-old infants can pick up the correlation between acoustic and articulatory characteristics of speech sounds. In their study the infants were exposed to a split-screen displaying two faces, one articulating [a] and the other articulating /i/, while an audio signal consisting of either one of those vowels was played. Their results indicated significantly longer looking times towards the face whose articulation was consistent with the audio signal.

Also (Bahrick, 2004) carried out a study in which 5 months-old infants were tested on their ability to discriminate between different phenomena involving changes in rhythm or tempo. The tests were organized in three situations: 1) a multimodal situation, where a plastic hammer was seen while the sound of the hammer hitting a
surface was heard, (2) a unimodal situation where only the sound of the hammer was heard, and (3) a unimodal situation where the hammer was only seen but not heard. The rhythm of the events was subsequently manipulated in each of these three situations in order to create novel situations that the infants might discriminate. The results indicated that only the 5 month-old infants who received the bimodal redundant stimulation could detect the rhythm changes. According to other studies by Bahrick, infants tend to be less dependent on redundant information the older they get.

Our study attempted to expand the findings of Kuhl and Meltzoff (Kuhl et al., 1982) and Bahrick’s by investigating the ability of 6 to 8 months-old Swedish infants to perceive synchronous visual and auditory input, for both speech and non-speech events. We also introduced a methodological improvement by using a high resolution eye-tracking system, with a maximum resolution of about 0.5°, which allowed the presentation of four images on a single screen during the test phase instead of the two alternatives used by Kuhl and Meltzoff, thereby reducing to 25% the spontaneous chance level of looking at one of the images. Just as in Kuhl and Meltzoff’s case, we hypothesized that the infants would look significantly longer towards the images displaying motor activities coherent with the heard speech or non-speech signals.

6. Method

After a short calibration of the eye-tracking system, the infants were exposed to a short video film while their eye-movements were registered throughout the session. For this paper, only the infants’ average looking times towards the different quadrants of the split-screen will be considered for statistical analysis. However, the eye-tracking data was collected with high enough temporal and spatial resolution to allow a detailed study of the infants’ visual strategies but those results will be reported in a future paper.

6.1 Subjects

Of the forty infants who participated in this study four had to be excluded due to calibration errors. The resulting in 36 subjects (13 boys and 13 girls) aged 25-33 weeks (mean age 28.5 weeks). The subjects were randomly selected from the National Swedish address database (SPAR) targeting 6 to 8 months-old infants whose parents lived in the Stockholm metropolitan area.

6.2 Stimuli

The infants were exposed to a film showing a female actress against a blue background. The film consisted of four sequences: (1) a baseline for speech articulations, (2) a test phase for audio-visual coherence in speech stimuli, (3) a baseline for non-speech gestures and (4) a test phase for audio-visual coherence in non-speech stimuli.

In the speech part or the experiment the baseline consisted of four identical still images on a split-screen showing the actress’s face. The baseline of the non-speech part of the experiment was an animated video sequence showing four different tempos of hand clapping, one in each quadrant. This baseline sequence was identical to the one to be used in the test phase, but with a silent sound track.

In the test phase for the speech stimuli (figure 4), the actress was again shown on a four quadrant split-screen articulating the vowels [a] and [y] and the syllables [ba] and [by]. For periods of 20 seconds, the speech signal was synchronized with the film shown in one of the quadrants. In the first 20 seconds the speech signal consisted of repetitions of the syllable [by] and the organization of the four video tracks was (target position in boldface)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ba</td>
<td>by</td>
<td>y</td>
<td>a</td>
</tr>
</tbody>
</table>

Figure 4. Example of the speech sound part of the experiment. The actress is articulating [ba] (UL), [by] (UR), [y] (LL) and [a] (LR). The audio played was the syllable [by], i.e. the target image was UR.

Directly after the vowel [a] was presented in the next 20 seconds and the position of the visual targets was

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

After this the [by] syllable was repeated as target but the visual target was placed in another quadrant

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

Finally the [a] was presented again and the visual target once more relocated

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

The utterances were produced with rise-fall $f_0$ contours and the target images were placed in different quadrants for each of the four 20 seconds sequences, as shown in the tables above.

For the test phase with non-speech stimuli (figure 5), the actress was shown on a split-screen clapping hands in different tempos. The tempos were 157%, 101%, 63% and 49% of the original recording tempo. The audio was manipulated to 101% of the original recording tempo, thus synchronized with one of the images shown. The location of the videos on the split screen is given in the tables below.
Baseline for rhythm (clapping movements with no sound)

<table>
<thead>
<tr>
<th>Speed</th>
<th>101%</th>
<th>49%</th>
<th>63%</th>
<th>157%</th>
</tr>
</thead>
</table>

Test phase for clapping rhythm. The spatial organization of the videos is the same as during baseline except that now the sound track corresponding to the 101% speed plays the clapping sounds.

Baseline for coupling of clapping sounds with visual displays of [by] utterances. The display shows four images of the actress uttering [by] at different speeds. During the baseline the soundtrack is silent.

The video films for this test phase were identical to those of the baseline but now the sound track played a clapping sound synchronized with the articulatory movements shown on the lower left quadrant.

Half of the subjects were exposed to the speech sound part of the experiment first, followed by the hand-clapping part. The rest of the subjects were exposed to the two parts in reversed order.

6.3 Material

The equipment used for tracking the infant’s eye movements was Tobii 1750 eye-tracker integrated with a 17” TFT monitor. The system uses low intensity infrared light to create a static reference frame on the spherical surface of the eye and derives a gaze vector from the relative position of the pupil within that frame. The system performs gaze measurements 50 times per second and with a nominal accuracy of 0.5°. The eye-tracking data generated by the ClearView 2.2.0 software package that comes with the system was subsequently analyzed using Matematica 5.1 and SPSS 13.0.

6.4 Procedure

The experiments were carried out in a dimly lit studio where most of the light came from the screen connected to the eye-tracking system. The brightness of the display on this screen was enough to draw the infant’s attention towards the stimuli being presented. The infant sat in front of the screen at a distance of approximately 60 cm. The parent sat in the studio slightly behind and outside the infant’s visual field and listened to music played through sound-isolating head-phones equipped with active noise reduction. Before recording the gaze the system was calibrated using the infant’s fixations on special purpose calibration points that were displayed on an otherwise empty screen. The calibration procedure was typically carried out in less than one minute.

7. Results

An example of the infants’ responses during one of the speech sound conditions is displayed in figure 6. Each panel corresponds to a quadrant on the test screen. The condition illustrated in figure 6 refers to the 20 s video sequence during which the infants head the vowel [a] for the first time in the session. The video corresponding to the sound track was displayed on the lower left corner (LL) of the screen. The curves in each of the panels show the percentage of infants who, at a given time throughout the 20 s of that test phase, were looking at the quadrant represented by the panel. The curves indicate a looking preference towards the upper quadrants, with a slight dominance for the upper right quadrant, displaying the articulation of [ba] syllables. The upper left quadrant, receiving the next highest percentage of looking time through this test phase, displayed the articulation of [by]. The correct visual target was in fact displayed on the lower left quadrant in this test phase and appears to have in fact received the lowest average percentage of looking time.

Figure 5. Hand-clapping in four different tempos: 101% (UL), 49% (UR), 63% (LL), and 157% (LR) of the original tempo. The sound of hands clapping was synchronized with the target image (UL).

Figure 6. Percentage of infants looking towards each of the quadrants on the screen as a function of time. A running time-window of 200 ms was used. First presentation of [a]. The target image was placed on the lower left (LL) quadrant in this case.
To obtain the net individual gains in looking time towards each of the quadrants shown during the test phases, a repeated measures analysis of variance was performed using the looking times towards a given quadrant during the baseline and the test phase. The results, using film order as between subject’s order, indicated a significant gain for the upper left quadrant displaying [ba], when [by] was heard ($F(1,34)=5.243$, $p<0.028$). There was no significant interaction with film order but film order was a significant between subjects effect ($F(1,34)=4.303$, $p=0.046$). These results suggest thus that the infants matched the visual image of [ba] with the sound of [by], although the group of infants who started the session seeing the clapping sequences performed not as well as the group seeing first the speech stimuli. Another significant gain in looking behaviour was observed for seeing [ba] and when listening to [by] ($F(1,34)=6.196$, $p<0.018$). No significant interaction with film order or significant effect of film order was observed in this case.

The analysis of the looking behaviour during the baseline phases indicates that the infants tended to look most of the time towards the upper quadrants. To compensate for this bias, the results from were sorted in terms of video materials being displayed, rather than the quadrants on which they appeared. Thus, the total looking time towards visual [ba] while listening to [by], for instance, was computed by adding the gain in looking time towards the upper left quadrant during the first test phase where [by] was heard, with the gain in looking time towards the lower left quadrant during the other test phase during which [by] was played. The results from this type of analysis are shown in figures 7 and 8. ANOVA models using the individual subject’s looking times towards each of the visual displays shown in figure 7 revealed a within-subjects significant linear trend in looking behaviour towards [ba], [by], [a] and [y] ($F(1,35)=7.235$, $p<0.011$). A very significant within-subjects linear trend was also found for the pattern displayed in figure 8 ($F(1,35)=27.507$, $p<0.0005$).

### 7.1 Non-speech sounds

The same type of analysis was carried out for the video films involving clapping sounds. In this case all the four quadrants displayed the same type of action but the action was performed with different repetition rates. The results from the infant’s matching between clapping sounds and the videos showing the actress clapping at different rates are shown in figure 9. The gain in looking time is greatest for the upper left quadrant, which also is the quadrant showing the clapping movements in synchrony with the sound.

In the other situation involving clapping sounds the videos displayed the actress rhythmically uttering [by]. The results in this case did not show maximum looking time gain towards the lower left quadrant containing the utterances synchronized with the clapping sounds, as illustrated in figure 10. Instead, the maximum looking time was towards the lower right quadrant. However,
when the looking behaviour is organized as a function of the repetition tempo, a pattern of increasing looking times for increasing frequency in the repetition of the articulatory movements emerges. This is significant linear trend ($F(1,35)=9.365, p<0.004$).

Figure 10. Average looking time gains (in ms) and 95% confidence intervals for looking towards [by] while hearing clapping sounds.

8. Discussion

The results of these experiments do not support the strong notion that 8-month-old infants might be able to establish phonetically relevant correspondences between speech sounds and their underlying articulatory movements. Indeed, rather than looking at the quadrants displaying the articulatory movements associated with the speech sounds, the infants’ preferences seemed to follow the salience of the articulatory displays on the quadrants (ba>by>a>y). But this was not because they were unable to detect synchrony in general terms. As demonstrated by the tests with clapping sounds, the infants were able to pick up the correct audio-visual synchrony when clap sounds and images were present but they appear to treat speech sounds (or the articulatory movements associated with speech sounds) in a different way than non-speech sounds. In fact, the infants failed to detect synchrony between the non-speech sound and the synchronic articulatory movements. They looked instead longer towards the video film showing the most rapid alternations between closed and open lips, a response that is in line with the results from the “speech sound part” of the experiment.

9. Conclusion

Taken together, the two experiments reported here suggest that 8 month-old infants may be using unspecific associative functions to pick up relevant linguistic information on the basis of multi-sensory regularities available in their immediate linguistic environment. If this is true, the infant’s success in acquiring the relevant linguistic functions is in line with ETLA and may be more dependent on the structure of its linguistic ambient than on the unfolding of a language acquisition program. In addition, from the point of view of epigenetic robotics this may be a general productive approach, worth to pursue (Dominey & Boucher, 2004). Indeed, given the repetitive characteristics of speech directed to infants (IDS) about 3 months of age, ETLA suggests that it may be possible to derive meaning from the recurrent co-occurrences of auditory and other sensory information representing the infant’s immediate linguistic environment. In our MILLE-project, we are currently making efforts to model the early stages of language acquisition exploring the acoustic regularities available in repetitive IDS.

Acknowledgements

Research carried out with grants from the Swedish Research Council and the Bank of Sweden Tercentenary Foundation.

References


