



# Speech categorization context effects in seven- to nine-month-old infants

Ellen Marklund, Francisco Lacerda and Anna Ericsson

Department of Linguistics, Stockholm University, Sweden  
 ellen@ling.su.se, frasse@ling.su.se, annaer@ling.su.se

## Abstract

Adults have been shown to categorize an ambiguous syllable differently depending on which sound precedes it. The present paper reports preliminary results from an on-going experiment investigating seven- to nine-month-olds sensitivity to non-speech contexts when perceiving an ambiguous syllable. The results suggest that the context effect is present already in infancy. Additional data is currently collected and results will be presented in full at the conference.

**Index Terms:** speech perception, context effects, infants

## 1. Introduction

The perception of speech sounds is influenced by the context in which they are heard. Experiments with ambiguous syllables following different kinds of context sounds have shown that the sound immediately preceding the speech sound determines which category it is perceived to belong to [1, 2]. These shifts in the perception of speech sounds do not appear to be a manifestation of peripheral auditory phenomena, such as forward masking [3], but rather of a more central phenomenon associated with phonetic and phonological processes. For instance, an ambiguous syllable, between a /da/ and a /ga/ in an acoustic continuum, is interpreted as /da/ when following the syllable /a/, but as /ga/ when following the syllable /ar/ [4]. Similar results have also been reported for non-speech contexts [5, 6, 7] and for situations where the context and the ambiguous syllable are separated by a silence or a neutral sound, suggesting that the effect is caused by spectral contrast. If the spectrum of the low intensity non-speech context sound is dominated by high frequencies, the following syllable is perceived as if it had more energy in low frequencies (/ga/ in the above mentioned case), and vice versa [8].

The purpose of the current study is to shed some light on the nature of this type of context phenomenon. In particular it investigates whether this context effect is present already in infancy. If young infants are affected by context effects in the same way as adults are, there is the possibility that the phenomenon might be linked to general auditory processing mechanisms rather than to the listener's phonetic and linguistic experience. At this stage in the study, seven- to nine-month-old infants are being investigated on their categorization behavior when presented with an ambiguous syllable preceded by non-speech contexts in different frequency ranges. Although infants in this age range already have a considerable experience of processing speech sounds in their ambient language [9, 10, 11] they still may not be affected by the type of context effects that are observed for adults. Data collection is on-going and the current paper reports the results obtained so far.

## 2. Method

The infants are shown a film in which they are trained to learn a correspondence between the sound of specific syllables and

the positions in which images appear on the screen. When presented with an ambiguous syllable in different contexts, their looking behavior is taken as a response as to how the ambiguous syllable is perceived.

### 2.1. Subjects

Participants so far are 38 infants between seven and nine months of age (mean age = 8.1 months) randomly selected from the Swedish demographic register based on date of birth and geographical criteria. The total number of participants in the current age group will be approximately 60, and a reference group of 20 adult subjects will also be tested.

### 2.2. Stimuli

The stimuli are 25 film sequences of 5 s, presented as a 125 s film. The film consists of 20 training sequences (randomized within the film), four test sequences (with locked positions) and one control sequence (also with locked position). The control sequence occurs after eight training sequences. The four test sequences, each preceded by three random training sequences, are then presented in fixed order (see table 1).

Table 1. Order and selection of the different sequence types within the film stimuli.

Number of sequences	Sequence type	Selection
8	Training	Random
1	Control	Fixed
3	Training	Random
1	Test	Fixed
3	Training	Random
1	Test	Fixed
3	Training	Random
1	Test	Fixed
3	Training	Random
1	Test	Fixed

There are four versions of the film, balanced for correspondence between syllable and image position on the screen, and the order of the test sequences (see table 2). Each subject is randomly assigned to one of the four film versions.

Table 2. Structure of the four film versions showing the syllable-box pairing, the syllable in the control sequence and the context frequency range in the four tests.

Film	Syllable-box	Control	Test 1	Test 2	Test 3	Test 4
1	/da/ - Left /ga/ - Right	/da/	High	Low	High	Low
2	/da/ - Left /ga/ - Right	/ga/	Low	High	Low	High
3	/ga/ - Left /da/ - Right	/da/	High	Low	High	Low
4	/ga/ - Left /da/ - Right	/ga/	Low	High	Low	High

Each of the film's 25 sequences starts with a context sound that is played while two boxes rotate on the screen. Next, a syllable is presented and an image either appears or does not appear in one of the boxes, depending on which kind of sequence it is.

The context sounds are short series of random tones within specified frequency ranges (here referred to as high, low and middle), and the syllables are /da/ and /ga/ and an ambiguous syllable taken from the acoustic continuum between them<sup>1</sup>. A categorical perception experiment with adult Swedish subjects was performed using the present continuum of syllables, and the most ambiguous syllable was chosen for the current experiment.

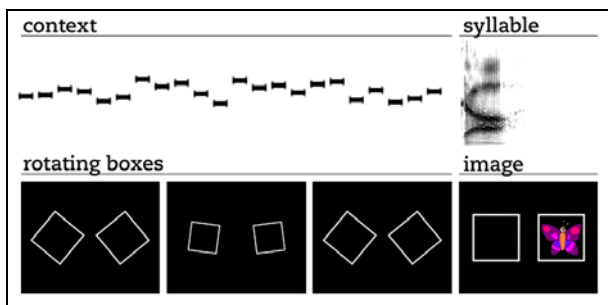


Figure 1. The general configuration of audio (top) and video (bottom) of a training sequence. During a context sound, two boxes rotate on the screen, and when the syllable is presented, an image appears in one of the boxes.

During the training sequences (see figure 1), the context sound is in the middle frequency range and assumed not to influence perception of the unambiguous syllable; it is present only to make the training as similar to the tests as possible. The syllable is either /da/ or /ga/. When it is played, an image appears in the box corresponding to the syllable (balanced across film versions, see table 2).

<sup>1</sup> The contexts were created by 22 tones, 700 ms long, followed by a 300 ms silent period. The frequencies of the tones were randomly selected from the high, medium or low frequency ranges, using a uniform distribution. The high range was 2.3 kHz to 3.3 kHz, the middle range was 1.8 kHz to 2.8 kHz and the low frequency range from 1.3 kHz to 2.3 kHz. The syllables were created as described in [8]. Sound files were provided by Lori Holt (MILLE-project).

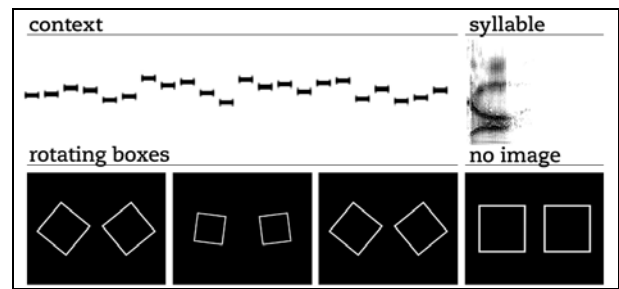


Figure 2. The general configuration of audio (top) and video (bottom) of the control and test sequences. During a context sound, two boxes rotate on the screen, and when the syllable is presented both image boxes are left empty.

The control sequence is identical to a training sequence except that no image appears while the syllable is presented (see figure 2). There are two versions of the control sequence; one with high frequency context and one with low frequency context, although only one is presented in each version of the film (see table 2). The control sequence is included in order to be able to confirm that the setup of the experiment is working; that the infants learn what they need to learn in order to give interpretable responses to the test stimuli.

In two of the four test sequences, the context sound is in the high range and the other two in the low range. The syllable presented is the ambiguous one, and no image is displayed in the boxes (see figure 2).

### 2.3. Procedure

Subjects are placed in their parents' lap in front of a Tobii XL Eye-tracker screen. After a short calibration, the film is presented and the subjects' eye movements are recorded, using Tobii Studio 2.0. During the experiment, the parents listen to music in headphones so they do not inadvertently influence their child's behavior regarding the audio of the film. If calibration fails or if any errors occur, the recording is excluded from data analysis. Of the 38 recordings, four were excluded due to system instability or failure to calibrate fully in the case of non-cooperative subjects.

### 2.4. Analysis

There are five time periods that are of interest for analysis. The first is the interval after syllable presentation in the control sequence until the end of the sequence. The other four are the corresponding time intervals in the four test sequences. It is during these periods that the looking behavior of the infants can be interpreted as a response; their looking behavior shows in which box they expect an image to appear, that is to say, how they have perceived the syllable.

For each of the relevant time intervals, the subject's accumulated looking time towards each of the boxes is calculated. If subjects do not look towards either box during this time interval, they must be considered not to have given any response, resulting in missing data for the sequence in question. Tobii Studio 2.0 is used for data preparation and statistical analysis is performed in SPSS 17.0.

## 3. Results and Discussion

Ideally, only participants with data present for all five relevant time intervals would be included in the analysis. However, given the relatively low number of subjects, the short response

interval, and the fact that infants typically cannot be expected to follow instructions of a test procedure, that is not a viable option at this stage. Instead, the five instances are so far considered as independent observations, rather than the repeated within-subject measures that they in fact are.

The number of possible responses for the control sequence is thus 34, while for the two types of test sequences it is 68 (since there are two instances of each test type in the film). All cases with missing data are excluded, leaving only those instances in which a response was given for analysis.

### 3.1. Control Sequence

The purpose of the control sequence is to make sure the participants pick up on the syllable-box correspondence and so are able to respond in an interpretable way. If the subjects have learned the connection, they are expected to look more towards the box corresponding to the syllable they hear (target) than to the other one. Fourteen participants' control sequences were excluded due to missing data, leaving 20 responses for analysis. A paired-samples 2-tailed t-test shows that while the mean looking time towards target is indeed slightly longer than towards non-target it is not significantly so ( $t(19)=1.195$ ,  $p<0.247$ ; see figure 3). However, the results still indicate that the subjects do in fact seem to learn the correlation between syllable and position on the screen, which is essential for the interpretation of the actual test data.

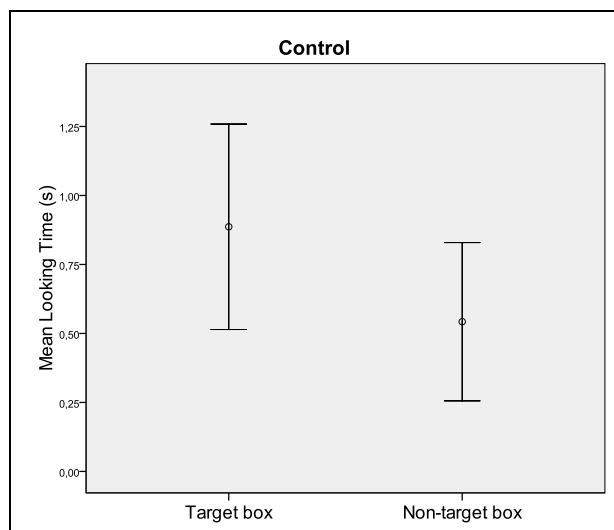


Figure 3. Mean looking time towards target (left) and non-target (right) based on 20 responses. The error-bars show the 95% confidence intervals.

### 3.2. Test Sequences

There were 37 non-responses in the high frequency context, resulting in 31 measured points of data. Based on the behavior of adults [8], infants were expected to perceive the syllable in the high frequency context as a /ga/. As seen in figure 4 the mean looking time is somewhat longer towards /ga/ ( $t(30)=-2.216$ ,  $p<0.04$ ).

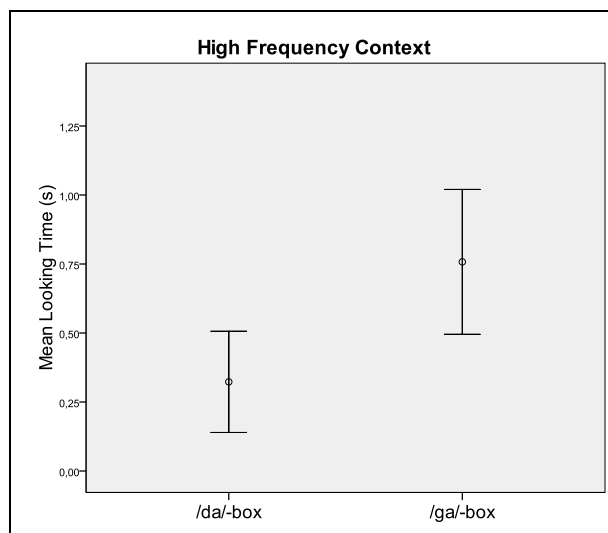


Figure 4. Mean looking time towards the box corresponding to /da/ (left) and /ga/ (right) based on 31 responses. The error-bars show the 95% confidence intervals.

In the low frequency context, there were 33 cases of missing data, leaving 35 responses. Again based on adults' behavior [8], the infants are expected to perceive the syllable as /da/. In figure 5 the looking times towards the two alternatives is shown, illustrating a slight tendency towards this behavior ( $t(34)=0.795$ ,  $p<0.432$ ).

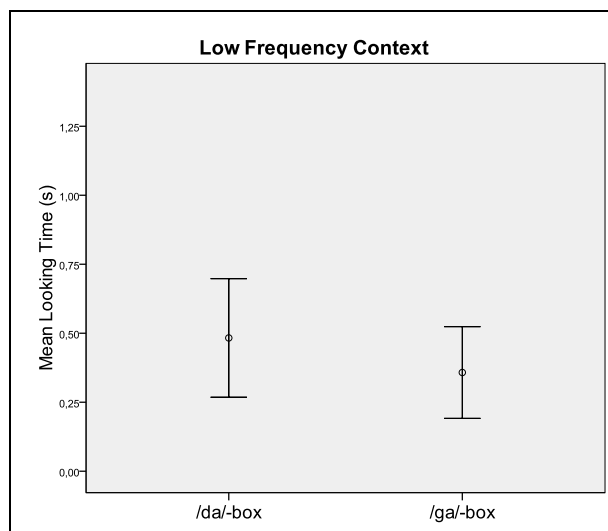


Figure 5. Mean looking time towards the box corresponding to /da/ (left) and /ga/ (right) based on 35 responses. The error-bars show the 95% confidence intervals.

An overall analysis of the pooled data indicates a weak tendency for a differentiated effect of the frequency range of the context ( $F(1,64)=2.653$ ,  $p<0.108$ ) and a significant interaction between the looking behavior and the contexts ( $F(1,64)=5.052$ ,  $p<0.028$ ), suggesting that the effect reported for adults is present already in infancy.

### 3.3. Conclusions

At this preliminary stage of the data collection it was not possible to set up a repeated measures' ANOVA. The infant's attention level in this kind of experiments tends to fluctuate and it is rather difficult to obtain enough complete data sets from a relatively small group of subjects, as required by a repeated measures analysis. Therefore the option of analyzing the pooled data is expected to provide a good first insight on the infants' general behavioral trend. By discarding the possible within-subjects' information, the current analysis runs the risk of missing any systematic changes in the infants' behavior during the session but it enables using enough group data to detect behavioral changes that can be attributed to different context conditions. With the addition of further subjects, it will become meaningful to carry out within-subjects' analyses that may disclose possible systematic fluctuations in the individual subject's behavior.

The preliminary analysis of the data suggests that non-speech contexts indeed influence perception of speech syllables in seven- to nine-month-olds, much as they do in adults. Data collection will be completed and the final results of the current experiment will be reported at the conference.

## 4. Acknowledgements

The present paper is Stockholm University's preliminary data report to the MILLE-project (K2003-0867), a project involving CMU, Pittsburgh, USA (Lori Holt) and KTH, Stockholm, Sweden (Rolf Carlsson and Björn Granström) and funded by the Bank of Sweden Tercentenary Foundation. The current study was also supported by Stockholm University's Faculty of Humanities (HumFak, ledande forskning). The authors would like to thank Kelly Smith for help with data collection, and Fredrik Myr and Ulrika Marklund for help with proof-reading the paper. Thanks also to Simon Carlgren, Mathilda Eriksson and Tove Jörgensen for collecting data in the adult categorical perception experiment.

## 5. References

- [1] Diehl, R. L. and Kluender, K. R., "On the categorization of speech sounds", in S. Harnad [Ed], *Categorical Perception: The Groundwork of Cognition*, 226-253, New York: Cambridge University Press, 1987.
- [2] Ladefoged, P. and Broadbent, D. E., "Information Conveyed by Vowels", *The Journal of the Acoustic Society of America*, 29:98-104, 1957.
- [3] Gelfand, S., "Hearing: An introduction to psychological and physiological acoustics", New York: Marcel Dekker, Inc., 1998.
- [4] Mann, V. A., "Influence of preceding liquid on stop-consonant perception", *Perception & Psychophysics*, 28(5):407-412, 1980.
- [5] Lotto, A. and Kluender, K. R., "General contrast effects in speech perception: Effect of preceding liquid on consonant identification", *Perception & Psychophysics*, 60(4):602-619, 1998.
- [6] Holt, L. L., "Auditory constraints on speech perception: an examination of spectral contrast", PhD thesis, Univ. Wis., 1999.
- [7] Lotto, A. J. and Holt, L. L., "Behavioral examinations of the level of auditory processing of speech context effects", *Hearing Research* 167:156-169, 2002.
- [8] Holt, L. L., "Temporally Nonadjacent Nonlinguistic Sound Affect Speech Categorization", *Psychological Science*, 4(16):305-312, 2005.
- [9] Kuhl, P., Williams, K., Lacerda, F., Stevens, K. N. and Lindblom, B., "Linguistic experience alters phonetic perception in infants by 6 months of age", *Science*, 255:606-608, 1992.
- [10] Best, C. T., McRoberts, G. W. and Goodell, E., "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system", *The Journal of the Acoustic Society of America*, 109:775-794, 2001.
- [11] Polka, L. and Werker, J. F., "Developmental changes in perception of nonnative vowel contrasts", *Journal of Experimental Psychology: Human Perception & Performance*, 20:421-435, 1994.