Electrophysiological evidence for automatic phonetic processing in neonates

G. Dehaene-Lambertz[1,2,CA] and M. Pena[1]

Laboratoire de Sciences Cognitives et Psycholinguistique (CNRS UMR 8554 and EHESS), 54 boulevard Raspail, 75270 Paris cedex 06; [2]Service de Neuropédiatrie, Centre Hospitalier Universitaire Bicêtre (AP-HP), France

[CA,1]Corresponding Author and Address

Received 28 June 2001; accepted 1 August 2001

At least two fundamental properties should be present in a network computing a phonetic representation: categorical perception and normalization across different utterances. Normalization processes were studied at birth by recording high density evoked potentials to strings of syllables in sleeping neonates. We compared the response to a change of phoneme when irrelevant speaker variation was present or absent. A mismatch response was recorded at the same latency in both cases, suggesting that relevant phonetic information was extracted from the irrelevant variation. Combined with our previous work showing that the mismatch response is sensitive to categorical perception in infants, this result suggests that a phonetic network like that of adults, is already present in the infant brain. Furthermore, efficient phonetic processing does not require attention. *NeuroReport* 12:3155–3158 © 2001 Lippincott Williams & Wilkins.

Key words: Auditory perception; Brain; Event-related potentials; Infant; Language; Phoneme

INTRODUCTION

The degree of similarity between infants' and adults' representations of speech is critical to our understanding of infants' general predisposition to acquire language. Do these representations share the same functional properties and neural bases in adults and in infants? If we examine phoneme perception for example, two main functional properties are present in adults: categorical perception, that is the capacity to perceive differences along one acoustic dimension categorically, and perceptual constancy, that is the capacity to perceive similarity between sounds when irrelevant acoustical variations are introduced. The later property is essential to speech comprehension. It allows listeners to perceive the appropriate syllable while its acoustic characteristics undergo considerable variation due to speaker's vocal tract size and shape, speech register, speech rate, environmental noise, etc.

Electrophysiological recordings allow us to explore the neural bases of behaviors and their similarity in adults and infants. In adults, a phonetic network dependent on the individual's native language, whose generators are left-lateralized and primarily involve the left planum temporale, has been isolated [1]. In 3-month-old infants, we have described a network sensitive to phonetic boundaries, reacting only to a change that crosses the /ba/ /da/ boundary and not to a change of similar acoustical magnitude within the phonetic category [2]. If this network computes a phonetic representation from the speech input, it should also present perceptual constancy. Thus, the goal of our experiment is to study the mismatch response to a change of phoneme when speaker variability is or is not present. The comparison of the timing of the response will determine if normalization is immediate or requires a second step in the computation of stimulus representation. Furthermore we will study the automaticity of this process by testing sleeping neonates.

MATERIALS AND METHODS

Subjects: Sixteen French full-term (> 39 weeks' gestation) neonates (11 girls, 5 boys) were tested during the first week of life (2–6 days, mean 3.5 days). Pregnancy and delivery were normal. Mean birth weight was 3753 g (3150–4380 g). Our procedure was approved by the local French ethical committee (CCPPRB Paris Cochin) and all parents provided their written informed consent for the participation of their babies in the experiment.

Stimuli: Two syllables /pa/ and /ta/ produced by four female speakers were used. The average root mean square of the syllable waveforms was equalized and the syllables were presented at an intensity of 78 dB (WC). The fundamental frequency of each speaker was 205, 183, 177 et 184 kHz (Praat 3.9.3 software). Syllable intonation varied across speakers as did the duration of the syllables (219, 263, 250, 229 ms for /ta/ and 227, 242, 267 and 257 for /pa/). In order to control that there was enough acoustical variation among the different utterances for them to be discriminable, the syllables were presented in pairs to five naive adults, who had to detect all perceptible changes. They were able to perceive a change of utterances in 92%

of the similar phonetic pairs, indicating that the acoustical variations between the different utterances were indeed perceptible. The percentage of detection of a change of phoneme was 98% in the pairs of syllables produced by the same speaker, and 94% in the pairs of syllables produced by two different speakers. The percentage of false alarms was 1%.

Procedure: Each trial was made up of four stimuli (stimulus onset asynchrony 600 ms, inter-trial interval 4 s). The last syllable of the trials was always produced by the same speaker (speaker 1) and is thereafter called the test syllable; the three first syllables constitute the context. In standard trials, the test syllable belonged to the same phonetic category as the context. In deviant trials, it came from the other phonetic category. For half the trials, the context was /pa/ and for the other half /ta/. In same speaker trials, the same physical syllable was repeated in the context and in the standard test. The deviant test syllable was produced by the same speaker (speaker 1) but by definition belonged to the other phonetic category. In different speaker trials, each syllable was produced by a different speaker. Order of speakers was randomly constituted for each trial with no repetition of speaker within the same trial, the last syllable always produced by speaker 1. The 8 types of trials: context (/pa/ or /ta/) × condition (standard *vs* deviant) × speaker (same *vs* different speakers) were randomly presented for a total of 200 trials (25 trials by condition) with the constraint that all conditions be presented five times for every 40 trials. Stimuli randomization, presentation and synchronisation with the ERP recording system were carried out using the EXPE software package [4] on a PC compatible with a Proaudio Spectrum 16 D/A Board.

Evoked brain responses were collected using a 64-channel geodesic electrode net referred to the vertex. The net was placed on the sleeping neonate who was seated in the lap of one of the experimenters, facing a speaker. Only one baby awoke during the first minutes of the experiment and thus was not included. All other babies remained asleep during the entire experiment and received 200 trials. Sleep stages were not taken into account. It is probable that sleep stages varied during the run. Because trials were randomly presented within blocks of 40 trials, this factor should affect all conditions similarly. EEG activity was digitized at 125 Hz over a 3144 ms epoch including a 150 ms baseline. Channels contaminated by eye or motion artifacts were automatically rejected and trials with >25 contaminated channels were rejected. The remaining trials were averaged for each subject and each of the eight different conditions, baseline corrected, digitally filtered between 0.5 and 20 Hz. An average reference transformation was applied to obtain the absolute potential. Four conditions were considered collapsing both context syllables (/pa/ and /ta/): standard, deviant, same speaker or different speakers. An average of 141 trials (117–177) have been kept for each neonate, that is 34.5, 34.4, 36.1 and 35.7 respectively for the standard and deviant same speaker and standard and deviant different speakers conditions. Two-dimensional reconstructions of scalp voltage at each time step were computed using spherical spline interpolation.

RESULTS

Analyses of the stimulus repetition (S1–S3): Contrary to the two peaks evoked responses described in older babies, the response to speech sound in neonates has only one peak whose maximum is at 292 ms post syllable onset. Its positivity is medial, around and in front of the vertex. Negativity recovers both temporal regions and occipital region. As in older, awake babies, the evoked response decreases in amplitude with syllable repetition in sleeping neonates. As shown in Fig. 1, the decrease is greater for the first repetition than for the subsequent ones and is present for both speaker conditions.

To analyze the habituation of the evoked response, a cluster of five electrodes in front of the vertex, that is at the maximum of the positivity, was selected. A repeated measures ANOVA was done on the voltage averaged across a 160 ms temporal window centered on the peak maximum after each of the first three syllables, with stimulus number (1, 2 and 3), speaker (same or different speakers), condition (standard and deviant) as within-subjects variables. There was a main effect of stimulus number ($F(1.5,30) = 5.47$ $p = 0.018$, Greenhouse-Geisser correction for repeated measures). No other main effect or interaction was significant. The decrease of amplitude after the first repetition of the syllable was significant (S1 *vs* S2: $F(1,15) = 5.85$, $p = 0.029$) and did not interact with speaker ($F(1,15) = 3.47$, $p = 0.08$). When *post-hoc* analyses were calculated for each type of trial, the effect of stimulus number and the decrease between S1 and S2 were significant only for different speaker trials (respectively $F(1.42,30) = 7.67$, $p = 0.006$, $F(1,15) = 9.11$, $p = 0.009$).

Electrophysiological response to a change of phoneme: The change of phoneme induces an increase in amplitude of the peak relative to the standard condition. As shown in Fig. 2, the response to the change is very similar in the two types of speaker trials. It is similar in timing, beginning around 190 ms post-onset in both cases, and in topography although the mismatch response for different speaker trials is wider on the scalp than the response recorded in same speaker trials with a more anterior positivity maximum in the different speaker condition.

To analyze the mismatch response in each type of trial, two clusters of five electrodes each were chosen located at the maximum of the positivity of the mismatch response for each type: frontal (maximum for different speaker) and central (maximum for same speaker). For each location, an ANOVA was performed on the voltage averaged across a 160 ms temporal window centered on the peak maximum with the same variables as above (stimulus number (1–4 now), speaker and condition) plus hemisphere (left and right). At the central location, there was a significant interaction of stimulus number (S1 S2 S3 *vs* S4) and condition ($F(1,15) = 5.05$, $p = 0.040$) due to an effect of condition for S4 only ($F(1,15) = 6.49$, $p = 0.022$). These two effects did not interact with speaker (both $F(1,15) < 1$). *Post-hoc* analyses revealed significant effects only for the same speaker condition (same speaker: stimulus number × condition $F(1,15) = 4.79$, $p = 0.045$, main effect of condition restricted to S4 $F(1,15) = 5.30$, $p = 0.036$; different speakers: stimulus number × condition $F(1,15) = 1.79$, $p = 0.201$, main effect of condition restricted to S4 $F(1,15) =$

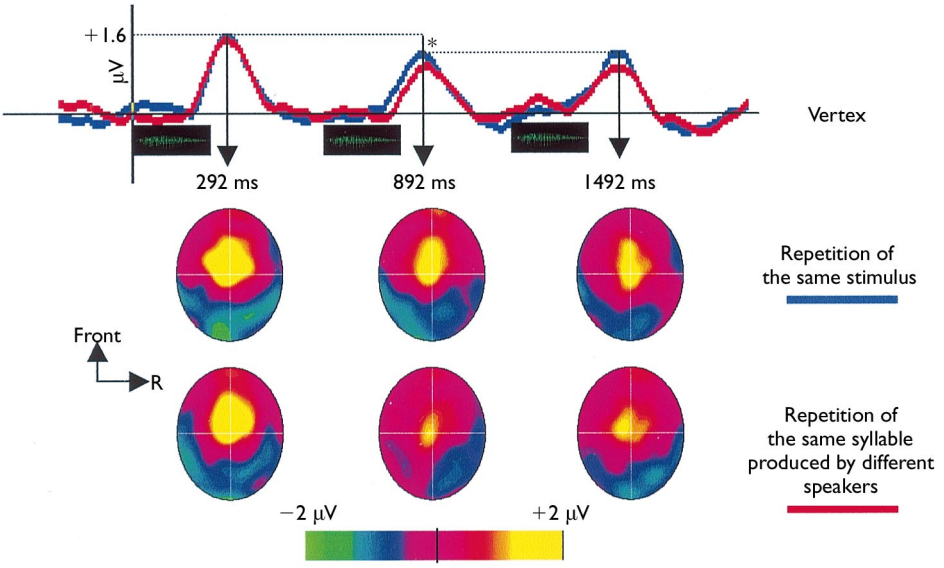

Fig. 1. Grand-averaged responses to the context stimuli, showing the decrease of the amplitude of the evoked response in same and different speakers trials. Top: waveform recorded at the vertex. Second and third lines: maps of evoked responses at the maxima of the peaks for both speaker conditions.

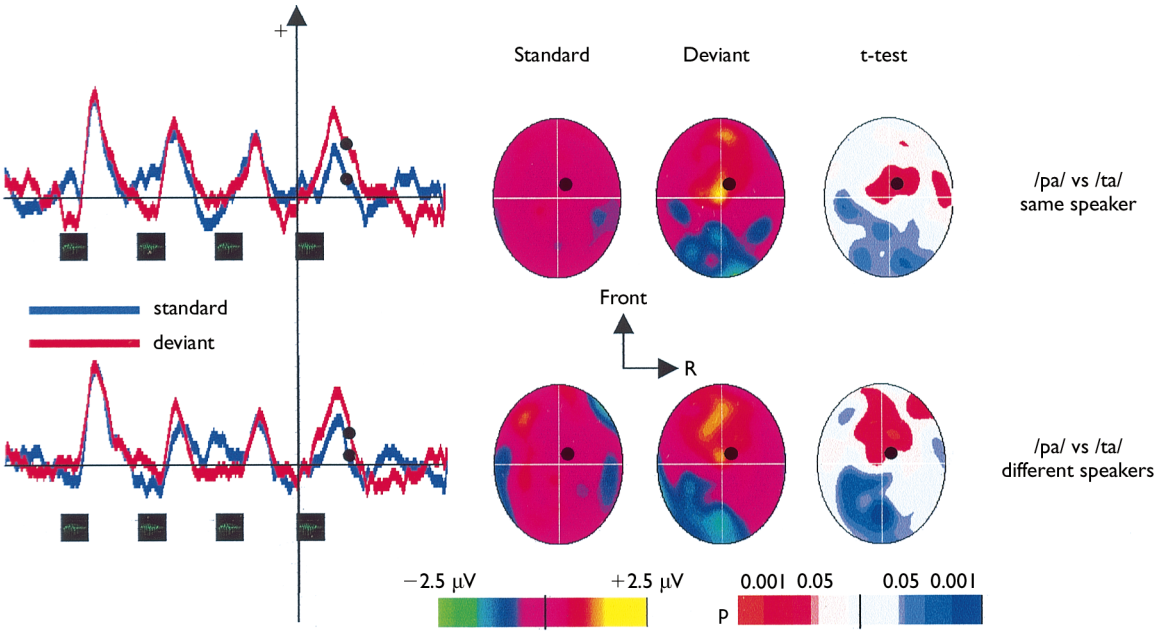

Fig. 2. Grand-averaged responses to the last syllable of the trials (S4) for same speaker and different speakers conditions. Left: ERP from a right paravertex electrode (filled circle on maps). Right: maps of evoked responses to standard, deviant syllables at 332 ms following stimulus onset (filled circle on ERP). Right-most column, maps of statistical significance (*t*-test) of deviant *vs* standard stimulus at the same time.

3.16, $p = 0.096$). At the frontal location, the stimulus number × condition interaction like the triple interaction condition × stimulus number × speaker was marginally significant (respectively $F(1,15) = 4.14$, $p = 0.060$ and $F(1,15) = 3.40$, $p = 0.085$), confirming that the mismatch response was broader in different speaker trials than in same speaker trials. Indeed, the stimulus number × condition interaction was significant only for different speaker trials ($F(1,15) = 8.05$, $p = 0.012$ for different speaker; $F(1,15) < 1$ for same speaker trials). This interaction in different speakers trials was due to an effect of condition for S4 only ($F(1,15) = 5.70$, $p = 0.031$). No interaction with Hemisphere was significant in these analyses.

DISCUSSION

Our main result is the presence of a significant difference between standard and deviant trials in the "different speaker" trials suggesting that even in the presence of irrelevant acoustical variability, a common feature, the phoneme identity, has been identified in the context syllables and

that a change of phoneme has been detected in the deviant trials. It is indeed important to note that syllable intonation, duration and syllable formants due to the speakers' voice characteristics varied within a phonetic category and that these variations were easily discernible by adults. Neonates are able to perceive such variations: they can discriminate their mother's voice from another [5,6] or between two strangers' voices [7] and electrophysiological response to a voice change has been recorded [8]. Thus, only a network able to extract the phonetic category notwithstanding perceptible irrelevant acoustical variations can elicit such a mismatch response in the different speaker trials. This result confirms the few behavioral experiments that have examined perceptual normalization in babies [9,10] and in older infants [1,11]. Moreover, we show here that normalization is present from birth and is not the consequence of the establishment of phonetic prototypes following extensive exposure to speech.

If we compare trials with or without irrelevant acoustical variability, it becomes apparent that the evoked responses are very similar, both during the context and the test periods. In particular, the timing of the mismatch response is similar in both speaker conditions, demonstrating that at the stage at which deviant syllables are detected, this is done using a normalized representation. We had previously shown that the mismatch response to a within-category change was also at the same latency as the mismatch response to a between category change [2]. Together these two experiments show that the phonetic representation is not computed after the acoustical representation but in parallel with it. This computation also appears to be automatic and attention-free since it was obtained in sleeping neonates.

Finally, electrophysiological studies have shown that a phonological representation, that is dependent on the native language is maintained in sensory memory and is computed by a dedicated network in adults. Indeed, no mismatch response, or just a weak one, has been recorded to a foreign phonetic contrast in several experiments, while the same contrast in native subjects induces a strong mismatch response [3,12–15]. In infants, Cheour *et al*. [16] have shown that the mismatch responses in 1-year-old babies is also influenced by the native language. Given that the mismatch response is similar in infants and adults and that both populations show categorical perception and normalization, two main properties of the adult's phonological network, we would like to suggest that there is a continuity between neonates and adults and that this network, present in neonates, is perfectly set to process the relevant properties of speech stimuli and is thereafter shaped by the linguistic environment.

REFERENCES

1. Kuhl PK. *Infant Behav Dev* **6**, 263–285 (1983).
2. Dehaene-Lambertz G and Baillet S. *Neuroreport* **9**, 1885–1888 (1998).
3. Dehaene-Lambertz G. *Neuroreport* **8**, 919–924 (1997).
4. Pallier C and Dupoux E. *Behav Res Methods Instrum Comput* **29**, 322–327 (1997).
5. DeCasper AJ and Fifer WP. *Science* **208**, 1174–1176 (1980).
6. Ockleford EM, Vince MA, Layton C *et al*. *Early Hum Dev* **18**, 27–36 (1988).
7. Floccia C, Nazzi T and Bertoncini J. *Dev Sci* **3**, 333–343 (2000).
8. Dehaene-Lambertz G. *J Cogn Neurosci* **12**, 449–460 (2000).
9. Kuhl PK and Miller JD. *Percept Psychophys* **31**, 279–292 (1982).
10. Jusczyk PW, Pisoni DB and Mullennix J. *Cognition* **43**, 253–291 (1992).
11. Kuhl PK. *J Acoust Soc Amer* **66**, 1668–1679 (1979).
12. Dehaene-Lambertz G, Dupoux E and Gout A. *J Cogn Neurosci* **12**, 635–647 (2000).
13. Näätänen R, Lehtokovski A, Lennes M *et al*. *Nature* **385**, 432–434 (1997).
14. Sharma A and Dorman MF. *JASA* **105**, 2697–2703 (2000).
15. Winkler I, Kujala T, Tiitinen H *et al*. *Psychophysiology* **36**, 638–642 (1999).
16. Cheour M, Ceponiene R, Lehtokoski A *et al*. *Nature Neurosci* **1**, 351–353 (1998).

Acknowledgements: This study was supported by Ministère français de la Santé et de la Recherche PHRC 1995 N° AOM95011, Groupement d'Intérêt Scientifique Sciences de la Cognition N° PO 9004 and the McDonnell Foundation.