

Language discrimination by human newborns and by cotton-top tamarin monkeys

Franck Ramus

Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS/CNRS),
54 Boulevard Raspail, 75006 Paris, France.

Marc D. Hauser, Cory Miller and Dylan Morris

Department of Psychology, Harvard University, Cambridge,
MA, 02138, USA.

Jacques Mehler

Laboratoire de Sciences Cognitives et Psycholinguistique
(EHESS/CNRS), 54 Boulevard Raspail, 75006 Paris,
France.

Humans, but no other animal, make meaningful use of spoken language. What is unclear, however, is whether this capacity depends on a unique constellation of perceptual and neurobiological mechanisms, or whether a subset of such mechanisms are shared with other organisms. To explore this problem, we conducted parallel experiments on human newborns and cotton-top tamarin monkeys to assess their ability to discriminate unfamiliar languages. Using a habituation-dishabituation procedure, we show that human newborns and tamarins can discriminate sentences from Dutch and Japanese, but not if the sentences are played backwards. Moreover, the cues for discrimination are not present in backward speech. This suggests that the human newborns' tuning to certain properties of speech relies on general processes of the primate auditory system.

A fundamental question in the study of language evolution and acquisition is the extent to which humans are innately endowed with specialized capacities to comprehend and produce speech. Theoretical arguments have been used to argue that language acquisition must be based on an innately specified language faculty (1, 2), but the precise nature and extent of this "language organ" is mainly an empirical matter, which notably requires studies of human newborns as well as non-human animals (3-5). With respect to studies of humans, we already know that newborns as young as four days old have the capacity to discriminate phonemes categorically (6) and perceive well-formed syllables as units (7-9); they are sensitive to the rhythm of speech, as shown in experiments where newborns distinguish sentences from languages that have different rhythmic properties, but not from languages that share the same rhythmic structure (10, 11); however newborns don't discriminate languages when speech is played backwards (10), and neurophysiological studies suggest that both infants and adults process natural speech differently from backwards speech (12, 13). All of these studies indicate that humans are born with capacities that facilitate language acquisition, and that seem well attuned to the properties of speech. Studies of non-human animals, however, show that some of these capacities may predate our hominid origins. For example, insects, birds, non-primate mammals, and primates process their own, species-typical

sounds in a categorical manner, and some of these species perceive speech categorically (14-18).

Our aim in this paper is to extend the comparative study of speech perception in three directions. First, we have conducted joint experiments on human newborns and on monkeys using the same design and the same material. Second, whereas most studies of non-human animal speech perception involve extensive training prior to testing on a generalization task, our experimental approach - the habituation-dishabituation paradigm - involves no training, and parallels the method used in studies of infant speech perception. Thus, conditions are met to appropriately compare the two populations. Third, most studies of speech processing in animals involve tests of phonemic perception. Here, we extend the analysis to sentence perception, thereby setting up a much broader range of perceptual problems.

Our experiments were run on human newborns and cotton-top tamarin monkeys (*Saguinus oedipus oedipus*). The stimuli consisted of 20 Japanese and 20 Dutch sentences, uttered by 4 female native speakers of each language. Conditions where the two languages are pitted against one another were compared with conditions where speakers of the same language are contrasted. As an additional factor, sentences within a session were either played forward or backwards. To more readily control for prosodic features of the signal, all conditions were rerun using synthesized exemplars of the original sentences. Synthesized sentences were created with the MBROLA diphone synthesizer (19). Phoneme duration and fundamental frequency were preserved, whereas the phonetic inventory was narrowed to only one phoneme per manner of articulation: all fricatives were synthesized as /s/,

Correspondence should be addressed to Franck Ramus. E-mail: f.ramus@ucl.ac.uk. Present address: Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, GB.

vowels as /a/, liquids as /l/, plosives as /t/, nasals as /n/, and glides as /j/ (20). Thus, each synthesized sentence preserved only the prosodic characteristics of its natural counterpart, while eliminating lexical and phonetic information.

Newborns were tested using the high amplitude sucking procedure and a habituation/dishabituation design. Sentences were elicited by the newborns' sucking on a pacifier. In the language change condition, newborns were habituated to 10 sentences uttered by 2 speakers in one language and then switched to 10 sentences uttered by 2 new speakers in the other language. In the speaker change condition, newborns were habituated to 10 sentences uttered by 2 speakers from one language and then switched to 2 new speakers in the same language. A significant increase in sucking following the language change, as compared to the speaker change, is taken as evidence that newborns perceive a significant difference between the two languages (21).

Thirty-two newborns were tested (22) on the natural language-forward experiment, 16 in the language change condition and 16 in the speaker change condition. Figure 1A shows that the two groups did not differ significantly, and thus that newborns failed to discriminate the two languages [$F(1,29) < 1$] (23). This result appears to conflict with previous experimental work showing that newborns discriminate English and Japanese. However, our experiment exposes newborns to great speaker variability (4 voices) (24), and this factor has previously been shown to impair infants' discrimination abilities (25). If speaker variability is responsible for the absence of discrimination, then we would predict successful discrimination with fewer speakers. To test for this possibility, we ran a second experiment using synthesized speech, thereby reducing the number of voices to one, that of the speech synthesizer (26).

Thirty-two additional newborns were tested (27) on the forward language and speaker discrimination, using the synthesized versions of the original sentences. Figure 1B shows that newborns in the language change condition increased their sucking significantly more during the 2 minutes following the switch than newborns in the speaker change condition [$F(1,29) = 6.3$, $p = 0.018$]. This indicates that newborns discriminate sentences of Dutch from sentences of Japanese, relying exclusively on prosodic cues. Moreover, this result shows that the newborns' failure to discriminate in Experiment 1A was probably due to speaker variability.

To determine the specificity of the newborns' capacity to discriminate languages, we tested 32 more newborns with the same synthesized sentences, but played backwards (28). Figure 1C shows that newborns fail to discriminate languages played backwards [$F(1,29) < 1$] (29). Moreover, the interaction between Experiments 1B and 1C (forward vs. backwards) is marginally significant [$F(1,59) = 3.6$, $p = 0.06$]. The finding that newborns discriminate two non-native languages played forward but not backwards suggests that the newborns' language discrimination capacity may depend upon specific properties of speech that are eliminated when the signal is played backwards. However, before drawing such a conclusion, it is important to directly assess the speech-specificity of this capacity, by testing it on another species.

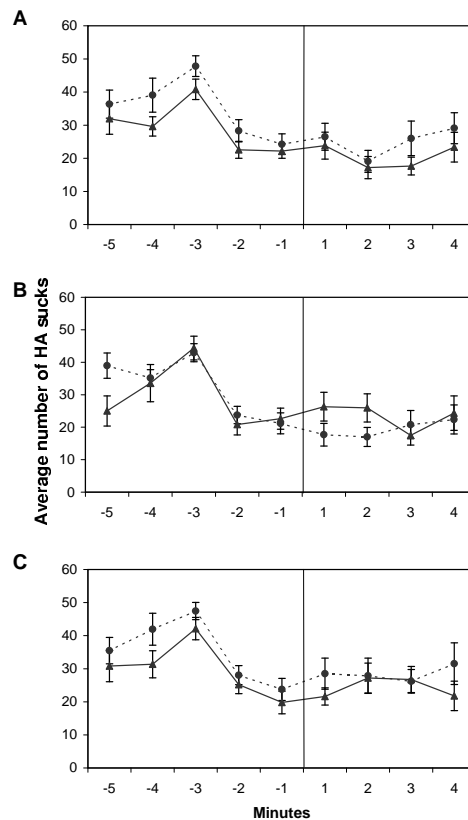


Figure 1. Average number of high amplitude sucks per minute for babies in the control (speaker change, dotted lines) and in the experimental group (speaker and language change, solid lines). Minutes are numbered from the time of change, marked with a vertical line. Error bars represent ± 1 standard error of the mean. 1A: natural sentences played forward. 1B: same sentences synthesized. 1C: same sentences synthesized and played backwards.

Cotton-top tamarins ($n = 13$) were tested with the same stimulus set as the newborns. Instead of sucking rate, however, we used a head orientation response toward the loudspeaker. During the habituation phase, a tamarin was presented with sentences uttered by 2 speakers in one language and then tested with a sentence uttered by a new speaker, either in the same language (speaker change condition), or in the other language (language change condition). Recovery of orientation toward the loudspeaker was interpreted as an indication that the tamarin perceived a difference between the habituation and test stimuli (30).

Experiment 2A involved natural sentences of Dutch and Japanese, played either forward or backward (31). Figure 2A shows that 10 out of 13 tamarins ($p < 0.05$, binomial test) dishabituated in the language change condition, whereas only 5 out of 13 dishabituated to the speaker change ($p = 0.87$). The difference between language and speaker

change is significant ($p < 0.05$, chi-square test). This result suggests that the tamarins discriminated Dutch from Japanese, irrespective of speaker variation. Surprisingly, such a pattern is not observed when the sentences are played backwards: only 5 out of 13 tamarins dishabituated to the backwards language change ($p = 0.87$), and this pattern is not significantly different from the speaker change condition ($p > 0.2$). These results parallel those obtained with newborns on the synthetic stimuli.

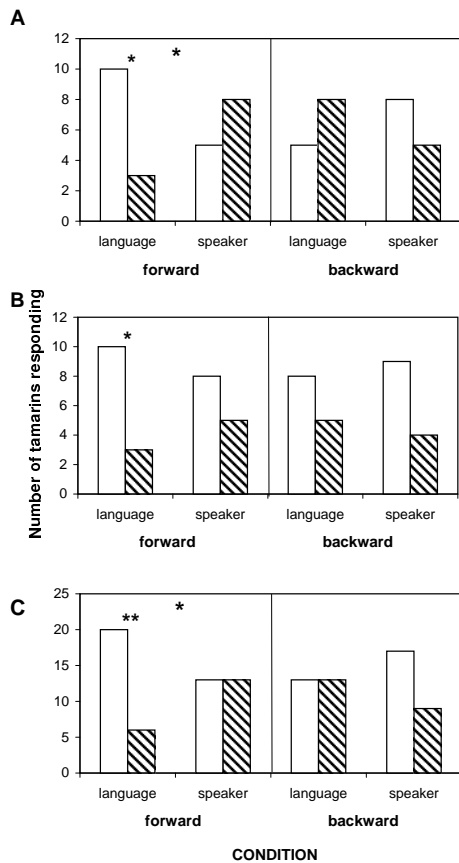


Figure 2. Number of tamarins responding positively (white bars) and negatively (hatched bars) to the test sentence depending on the condition: language or speaker change, sentences played forward or backwards. 2A: natural sentences. 2B: synthesized sentences. 2C: data from experiments 2A and 2B pooled together. * $p < 0.05$. ** $p < 0.01$.

In Experiment 2B, the same tamarins were tested on both the speaker and language conditions, but with synthesized sentences. Figure 2B shows that 10 out of 13 tamarins dishabituated to the forward language change ($p < 0.05$). Although the number of subjects dishabituating to the speaker change failed to reach statistical significance ($p = 0.29$), the increased numbers in this condition led to a non-significant difference between language and speaker change for the synthesized sentences ($p > 0.3$). For backward sentences, subjects failed

to show a statistically significant level of dishabituation to either the language or speaker change ($p = 0.29$ and $p = 0.13$). Experiment 2B suggests that tamarins' ability to discriminate Dutch and Japanese is diminished when only prosodic cues are available.

When the data from Experiments 2A and 2B are pooled (Figure 2C), the overall result is clear: when sentences are played forward, tamarins significantly dishabituate to the language change ($p = 0.005$) but not to the speaker change ($p = 0.58$), and the difference between language and speaker change is significant ($p < 0.05$). When sentences are played backwards, no such effect is observed. This overall result parallels that obtained with human newborns: both species discriminate sentences of Dutch and Japanese played forward, but not backwards.

The pattern of our results suggests both striking similarities and differences between the monkey and the human auditory systems. First, we have shown that tamarins, like human newborns, are able to process not just isolated syllables, but also whole strings of continuous speech, and to extract enough information thereof to discriminate between Dutch and Japanese. Second, their ability to do so above and beyond speaker variability suggests that they are able to extract auditory equivalence classes, that is to extract abstract linguistic invariants despite highly variable acoustic shapes (17, 32). Third, the fact that, like newborns, tamarins fail to discriminate when speech is played backwards, suggests that their language discrimination capacity does not rely on trivial low-level cues, but rather on quite specific properties of speech. Since tamarins have not evolved to process speech, we in turn infer that at least some aspects of human speech perception may have built upon pre-existing sensitivities of the primate auditory system. Finally, unlike newborns, tamarins fail to discriminate the language change more than the speaker change when speech is resynthesized. This leaves open the possibility that human newborns and tamarins may not be responding to exactly the same cues in the sentences: tamarins might be more sensitive to phonetic than to prosodic contrasts.

References and Notes

1. N. Chomsky, *Language and Problems of Knowledge*. (MIT Press, Cambridge, MA, 1988).
2. S. Pinker, *The Language Instinct*. (William Morrow and Company, Inc., New York, 1994).
3. A. Doupe, P. Kuhl, *Ann. Rev. Neurosci.* 22, 567 (1999)
4. A. Ghazanfar, M. D. Hauser, *Trends in Cognitive Science* 3, 377 (1999).
5. E. S. Spelke, E. L. Newport, in *Handbook of Child Psychology, Volume 1: Theoretical Models of Human Development* R. M. Lerner, Ed. (Wiley, New York, 1998) pp. 275-340.
6. P. D. Eimas, E. R. Siqueland, P. W. Jusczyk, J. Vigorito, *Science* 171, 303 (1971).
7. J. Bertoncini, J. Mehler, *Infant Behavior and Development* 4, 247 (1981).
8. J. Bertoncini, C. Floccia, T. Nazzi, J. Mehler, *Language and Speech* 38, 311 (1995).
9. R. Bijeljac-Babic, J. Bertoncini, J. Mehler, *Developmental Psychology* 29, 711 (1993).

10. J. Mehler, et al., *Cognition* 29, 143 (1988).
11. T. Nazzi, J. Bertoncini, J. Mehler, *Journal of Experimental Psychology: Human Perception and Performance* 24, 756 (1998).
12. H. J. Neville, D. L. Mills, *Mental retardation and developmental disabilities research reviews* 3, 282 (1997).
13. D. Perani, et al., *NeuroReport* 7, 2439 (1996).
14. G. Ehret, B. Haack, *Naturwissenschaften*, 68, 208 (1981).
15. R. A. Wyttenbach, M. L. May, R. R. Hoy, *Science* 273, 1542 (1996).
16. P. K. Kuhl, J. D. Miller, *Science* 190, 69 (1975).
17. P. Kuhl, in *Categorical perception: The groundwork of cognition* S. Harnad, Ed. (Cambridge University Press, 1987) pp. 355-386.
18. M. D. Hauser, *The evolution of communication* (Bradford Books/MIT Press, Cambridge, MA, 1996).
19. T. Dutoit, V. Pagel, N. Pierret, F. Bataille, O. van der Vrecken, *The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes*, IC-SLP'96, Philadelphia (1996). This software is freely available from <http://tcts.fpms.ac.be/synthesis/mbrola.html>.
20. The production of this kind of stimuli is described in great detail in F. Ramus, J. Mehler, *Journal of the Acoustical Society of America* 105, 512 (1999).
21. The experiment takes place in a sound-attenuated booth, with only the baby and the experimenter inside. The experimenter is blind to the experimental condition and listens to masking noise during the test. Newborns are randomly assigned to the control or to the experimental group. Order of presentation of the languages is counterbalanced across subjects. During a given phase, sentences corresponding to the condition are played in a random order. The habituation phase lasts at least 5 minutes. The habituation criterion is a 25
22. Subjects are full-term healthy newborns, aged between 2 and 5 days, recruited at the Port-Royal maternity hospital in Paris. Forty-two additional babies were tested, and their results were discarded due to: rejection of the pacifier (1), sleeping or insufficient sucking before the switch (12), crying or agitated (9), failure to meet the habituation criterion (9), sleeping or insufficient sucking after the switch (6), loss of the pacifier after the switch (4), computer failure (1).
23. The slight tendency for babies in the control group to suck more overall than those in the experimental group, though visible on Figure 1A, is not significant [$F(1,31)=2.6$, $p=0.12$].
24. In the English-Japanese discrimination experiments by Nazzi et al. (11), the variability due to the 4 voices was much reduced by low-pass filtering the stimuli. In other experiments (10), a single bilingual speaker was used.
25. P. W. Juszyk, D. B. Pisoni, J. Mullenix, *Cognition* 43, 253 (1992).
26. While the resynthesis process reduces all voices to one, other (prosodic) characteristics of the different speakers are preserved. We didn't test discrimination of the natural sentences played backwards, due to newborns' failure to discriminate these sentences even when played forward.
27. Twenty additional babies were tested, and their results were discarded due to: sleeping or insufficient sucking before the switch (6), crying or agitated (4), failure to meet the habituation criterion (1), sleeping or insufficient sucking after the switch (3), loss of the pacifier after the switch (6).
28. Seven additional babies were tested, and their results were discarded due to: crying or agitated (1), sleeping or insufficient sucking after the switch (4), loss of the pacifier after the switch (2).
29. The tendency for babies in the control group to suck more than those in the experimental group during the habituation phase is not significant [$F(1,31)=2.7$, $p=0.11$].
30. Before running these experiments, all tamarins had participated in a habituation-dishabituation experiment involving their own, species-typical vocalizations. Thus, all subjects were familiar with the general test set up. Subjects sit in the test cage within the acoustic chamber, and do so without stress. They also sit or hang on the front panel during testing, thereby allowing relatively unambiguous observations of head turning behavior. Experiments were run by transporting a tamarin from their home room to a test room, acoustically and visually isolated from all other tamarins. Two observers watched the session from a monitor outside of the test room. Stimuli were played back from a concealed speaker only when the subject's head and body were oriented approximately 180 degrees away from the speaker. A positive response was scored if the subject turned and oriented toward the speaker within the playback period. If the response was ambiguous, we ran the trial again, but with a different exemplar from the habituation series. A subject was considered habituated if he or she failed to respond on two consecutive trials. Following habituation, the test stimulus was played. The final trial of the session was a post-test playback, presenting the long call of a tamarin. Given the salience of the long call, we expected the tamarins to respond. If they failed to do so, we excluded the entire session, under the assumption that failure to respond to all stimuli represents general habituation to the test set-up. Inter-trial interval within a session was set at a minimum of 15 sec and a maximum of 60 sec. All trials were videotaped. After running a session, trials were digitized onto a computer and subsequently scored blind with respect to test condition by stepping through the experiment frame-by-frame. Two observers scored each test trial; inter-observer reliability was 0.92.
31. Each tamarin was tested in the four conditions, with the order of presentation of languages counterbalanced across subjects. Inter-session interval was no less than 4 days, with a median of 7 days. When a session's data were excluded due to an ambiguous response to the test sentence, or failure to respond to the post-test trial, the tamarin was tested again in the same condition after 1-3 weeks, in order to complete all the conditions. The significance of the proportion of monkeys reacting to a given change is assessed through a binomial test, and the difference between two conditions is assessed through a 2X2 chi-square test.
32. The fact that newborns fail under this very condition in Experiment 1A is likely due to their immature auditory system, since susceptibility to speaker variability seems to resolve a few months after birth (25).
33. Work supported by the Délégation Générale pour l'Armement and a National Science Foundation Young Investigator Award to MDH. We thank the Port-Royal maternity hospital for providing access to the newborns, and the parents for their participation. All parents gave informed consent. We thank the New England Regional Primate Research Center (P51RR00168-37) for providing the tamarins; our research was approved by the Animal Care and Use Committee at Harvard University (92-16) and the CCPRB Paris-Cochin.