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 hominin 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. 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/.
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.

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.

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

We didn't test discrimination of the natural sentences played back-forward when played forward.

We didn't test discrimination of the natural sentences played back-forward when played forward.

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 tamarins were considered habituated if they 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.

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.

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).

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 (PS1RR00168-37) for providing the tamarins; our research was approved by the Animal Care and Use Committee at Harvard University (92-16) and the CCPRPB Paris-Cochin.

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].