REPORT

The role of speech rhythm in language discrimination: further tests with a non-human primate

Ruth Tincoff,[1] Marc Hauser,[1,2] Fritz Tsao,[3] Geertrui Spaepen,[1] Franck Ramus[4] and Jacques Mehler[4,5]

1. *Department of Psychology, Harvard University, USA*
2. *Program in Neurosciences, Harvard University, USA*
3. *Department of Psychology, Hillsdale College, USA*
4. *Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS/CNRS/ENS, Paris, France*
5. *International School for Advanced Studies, Trieste, Italy*

Abstract

Human newborns discriminate languages from different rhythmic classes, fail to discriminate languages from the same rhythmic class, and fail to discriminate languages when the utterances are played backwards. Recent evidence showing that cotton-top tamarins discriminate Dutch from Japanese, but not when utterances are played backwards, is compatible with the hypothesis that rhythm discrimination is based on a general perceptual mechanism inherited from a primate ancestor. The present study further explores the rhythm hypothesis for language discrimination by testing languages from the same and different rhythmic class. We find that tamarins discriminate Polish from Japanese (different rhythmic classes), fail to discriminate English and Dutch (same rhythmic class), and fail to discriminate backwards utterances from different and same rhythmic classes. These results provide further evidence that language discrimination in tamarins is facilitated by rhythmic differences between languages, and suggest that, in humans, this mechanism is unlikely to have evolved specifically for language.

Processing a spoken language requires perceptual mechanisms that operate on the incoming signal and extract information relevant for understanding the linguistic content of the utterance. Human infants begin the language-learning process with sensitivities to many aspects of the speech signal, such as phonetic distinctions, statistical regularities and intonation patterns (for reviews see Jusczyk, 1997; Oller, 2000). Based on this prolific body of data, most researchers agree that human infants are tuned to spoken language from birth and possess perceptual mechanisms that guide the development of spoken language (e.g. Jusczyk & Bertoncini, 1987; Kuhl, 2000). This developmental conclusion raises important evolutionary questions about how our species acquired such mechanisms (Hauser, Chomsky & Fitch, 2002; Holt, Lotto & Kluender, 1998; Kuhl, 1986, 2000; Kuhl & Miller, 1975). Are these mechanisms unique to humans and specialized for speech processing? Are they shared with other species, but fine-tuned over the course of human evolution and further in individual human development? Or, are these mechanisms evolutionarily ancient, selected for general auditory processing functions and conserved over time? These are broad questions that can only be addressed by systematic comparative studies, using a wide variety of species (including closely related non-human primates and more distantly related animals) and tasks (training and spontaneous methods), testing an array of different perceptual and computational phenomenon (e.g. discrimination of phonetic and prosodic information, grammatical principles). The present study builds on earlier results to tackle one small portion of this very large comparative space. Specifically, we present an experiment designed to further understand the role of rhythmic cues in language discrimination by human infants and cotton-top tamarin monkeys.

Human newborns discriminate non-native languages from different rhythmic classes, but do not discriminate languages from the same rhythmic class (Mehler *et al.*, 1988; Nazzi, Bertoncini & Mehler, 1998; Ramus, 2002). These findings contribute to the more general proposal

Address for correspondence: Ruth Tincoff, Psychology Department, Harvard University, 33 Kirkland Street, 9th Floor, Cambridge, MA 02138, USA; e-mail: tincoff@wjh.harvard.edu

that infants build grammatical knowledge of the ambient language by means of 'prosodic bootstrapping', with rhythm playing one of the central roles (Jusczyk, 1997; Mehler, Dupoux, Nazzi & Dehaene-Lambertz, 1996; see chapters in Morgan & Demuth, 1996). The presence of rhythm discrimination within days of birth suggests that the perceptual mechanism is innately specified, triggered by the most minimal experience (Mehler *et al.*, 1988; Mehler *et al.*, 1996). As stated above, claims of innate systems in humans immediately raise questions about evolutionary ancestry. In particular, did the capacity to discriminate languages based on rhythmic cues evolve uniquely in humans and specifically for spoken language processing or is the mechanism shared with other animals, having evolved for more general auditory purposes?

The *rhythmic class hypothesis* suggests that natural languages can be sorted into a few major rhythmic classes based on linguistic timing units such as stress patterns or syllables (Abercrombie, 1967). Ramus, Nespor and Mehler (1999) proposed an empirical measure of rhythm as the proportion of time in utterances occupied by vowels (%V). Applying this measure of rhythm to a wide variety of languages provides an excellent fit with classical taxonomies of language using more intuitive notions of rhythmic class (e.g. Abercrombie, 1967; Cutler, 1994; Pike, 1945). Specifically, the rhythmic class hypothesis predicts that, in the absence of other cues or in the absence of attention to other cues, languages from different rhythmic classes will be discriminated; those from the same rhythmic class will not.

To determine if human newborns are uniquely endowed with the ability to discriminate languages according to rhythmic class, Ramus and colleagues (2000) compared their behavior with that of cotton-top tamarin monkeys, using the same auditory materials and same habituation–dishabituation testing techniques. When natural utterances were presented, tamarins discriminated Dutch and Japanese, two languages that belong to different rhythmic classes, but the human newborns did not. Ramus *et al.* suggested that the multiple levels of information present in the natural utterances, specifically language and speaker changes, hindered the human newborns' attention to rhythm. Consistent with that proposal, human newborns discriminated Dutch and Japanese when the utterances were resynthesized and the speaker changes eliminated. When the tamarins were tested in this condition, their discrimination was less robust than with the natural utterances, but nonetheless statistically significant. These differences suggest that human newborns and tamarins may not be responding to exactly the same cues in the utterances. In contrast, evidence for convergence in the underlying mechanism comes from the pattern of

results obtained with backwards utterances. Here, both tamarins and human newborns failed to discriminate Dutch and Japanese backwards utterances.

The Ramus *et al.* (2000) results raise questions about how tamarins and human newborns discriminate human languages. While similarities in closely related species suggest a commonly inherited mechanism, differences may provide important information about the how the mechanism has changed over evolutionary time or how it changes over the course of an individual's life. It may be that some components are available to one species and not to the other, or that tamarins and human newborns differentially attend to certain components of the speech signal (e.g. rhythm, stress, intonation, phonotactics). In addition, these species similarities and differences may change over the lifecourse as functions of maturation, the system's sensitivity to experience, or the amount of experience. The present study is a further step towards understanding the role of rhythmic cues in language discrimination and, in particular, the extent to which this mechanism is unique to humans or shared with other species. We tested three predictions that follow from the hypothesis that tamarins use rhythmic cues to discriminate human languages (Ramus *et al.*, 2000). First, the previous results showed that tamarins discriminate natural, forward utterances of Dutch and Japanese. If this finding indicates sensitivity to rhythmic cues, then tamarins should successfully discriminate a novel contrast of languages from different rhythmic classes. Second, if rhythmic cues are necessary for language discrimination in tamarins, then they should fail to discriminate two languages from the same rhythmic class. Third, the previous results showed that, like human newborns, tamarins failed to discriminate backwards utterances. For the language contrasts in the present study, we also predict no discrimination of backwards utterances from different and same rhythmic classes.

The habituation–dishabituation method used in the present study is modeled on that used in Ramus *et al.* (2000). Specifically, we tested whether tamarins can discriminate Polish and Japanese, languages from different rhythmic classes. We also tested whether tamarins can discriminate American English and Dutch, languages from the same rhythmic class. Although the human newborn studies cited above used British English, we use American English here in order to provide tamarins with the strongest test of discriminating same rhythmic class languages. The tamarins in our laboratory hear American English frequently.[1] If they are using cues other than

---

[1] The laboratory has had members who speak languages other than English. However, English is the common language and the one that tamarins would hear most frequently.

rhythm, then their increased exposure to American English may lead to discrimination of these same rhythmic class languages. This prediction is based on the findings of Nazzi, Jusczyk and Johnson (2000) that 5-month-old American English learning infants discriminate certain languages within their native stress-based rhythmic class. Specifically, among the stress-based languages of American English, Dutch and German, American 5-month-olds discriminated American English from Dutch, but failed to discriminate Dutch from German. In addition, the 5-month-olds discriminated American English from British English. While the infants still demonstrated between rhythmic class discrimination (i.e. Italian versus Japanese, British English versus Japanese) the rhythmic class hypothesis cannot account for the above findings. Nazzi *et al.* (2000) proposed the *native language acquisition* hypothesis as an alternative explanation. This hypothesis states that 5-month-olds have had sufficient exposure to learn the specific prosodic features of their native language but not those of the rhythmic class as a whole. Moreover, Nazzi *et al.* (2000) argue that within rhythmic class, discrimination is still based on prosodic level information and not necessarily segmental level, phonotactic or phonetic information. This conclusion is supported by the fact that the American English and British English sentences were identical in content and thus the phonotactic information was the same, and by acoustic analyses showing differences in prosodic measures between the dialects (e.g. pitch variation and syllable duration). Importantly for the current study, discrimination of languages from the same rhythmic class indicates that either the underlying rhythm mechanism has been changed by experience, or that another mechanism is operating. If tamarins discriminate American English from Dutch, then either the underlying mechanism is open to experience, or it is tuned to additional prosodic or segmental information. In contrast, if tamarins do not discriminate American English from Dutch, then this result is consistent with the rhythmic class hypothesis and suggests that either the mechanism is insensitive to experience or the experience was deficient in some way.

Perhaps the most intriguing result of Ramus *et al.* (2000) was that tamarins did not discriminate Dutch and Japanese utterances played backwards, a pattern that parallels those obtained with human newborns. Why human newborns have trouble with backwards speech has yet to be elucidated. When this finding was first discovered (Mehler *et al.*, 1988), it seemed natural to invoke an innate tuning to some specific properties of human speech, such as the temporal pattern of plosive consonants (e.g. [b]), or a prosodic universal, such as phrase-final lengthening. Regardless of these specifics, other primates should not possess such knowledge about human speech unless what appears to be speech-specific knowledge actually arises from more general characteristics of the primate (or vertebrate) perceptual system. We present the different (Polish versus Japanese) and same (American English versus Dutch) rhythmic class comparisons with backwards utterances to further test whether or not the failure to discriminate backwards speech is based on general properties of the primate perceptual system. If the general primate hypothesis is supported, then a major reinterpretation of infant speech perception data may be in order, which would have significant implications for theories of the evolution of language (Hauser *et al.*, 2002a; see chapters in Christiansen & Kirby, 2003 and Wray, 2002).

Method

Subjects

We tested 20 cotton-top tamarins (*Saguinus oedipus*), seven mated adult males, seven mated adult females, three sub-adult males, and three sub-adult females. All subjects were born in captivity at the New England Regional Primate Research Center, Southborough, Massachusetts, or the Primate Cognitive Neuroscience Lab, Harvard University, Cambridge, Massachusetts. Animals are housed in social groups consisting of a mated pair and, in some cases, their offspring.[2]

All mated adult subjects have experience in playback experiments (Ghazanfar, Flombaum, Miller & Hauser, 2001; Hauser, Newport & Aslin, 2001; Miller, Dibble & Hauser, 2001; Ramus *et al.*, 2000; Weiss, Hauser & Garibaldi, 2001). They readily move in and out of their home cages and into the test area.

Sub-adults were individuals between the ages of 10 and 24 months, were the offspring of a mated pair, and were housed with the mated pair. Tamarins are co-operative breeders and in the wild, offspring often remain with their family group and help with the care of subsequent offspring.[3] The sub-adult subjects were naïve to the

---

[2] For further information on the colony, see www.wjh.harvard.edu/~mnkylab

[3] We use the term 'sub-adult' because the usual criterion for adulthood – reproductive age – is complicated in tamarins due to characteristics of their breeding system. Significantly, tamarin females may reach physical maturity and undergo some hormonal signs of puberty (between 15 and 28 months of age), but the normal ovarian cycling necessary for conception is suppressed as long as a female remains with her family group and does not have access to a novel male (possibly up to 42 months, Ziegler, Savage, Scheffler & Snowdon, 1987). In contrast, males do not appear to be hormonally suppressed in their family group; male mating behavior may be constrained by the social environment, other physiological aspects of fertility, or interactions between the male and female reproductive states (Ginther, Carlson, Ziegler & Snowdon, 2002).

playback experiment, though they had participated in other experiments with different methods (Hauser, Pearson & Seelig, 2002). They also readily moved in and out of their home cage and into the test area. Before approximately 10 months of age, young tamarins tend not to leave their home cage and experience distress in the testing room.

Stimuli

We presented natural speech utterances extracted from a multi-language corpus previously used in experiments testing discrimination of language rhythm (Nazzi *et al.*, 1998; Nazzi *et al.*, 2000; Ramus *et al.*, 2000). The utterances in the corpus consisted of short (2–3 s) declarative statements that were recorded by multiple female native speakers. Further details regarding the acoustic characteristics of the utterances in the corpus can be found in Ramus *et al.* (1999).[4] American English and Dutch are drawn from the same, stress-based, rhythm class, Japanese and Polish from different rhythm classes.[5] For the present experiment, we selected a total of 52 utterances, 13 each from American English, Dutch, Japanese and Polish. For each language, 10 utterances served as habituation stimuli and three utterances served as test stimuli. The 10 habituation stimuli for a particular language consisted of five utterances recorded by one speaker of that language and five utterances recorded by a second speaker of that language. A third speaker of the language recorded the three test utterances. Utterances were selected to minimize differences in total sentence duration and average pitch between habituation and test speakers; the content of the utterances varied across speakers and across languages.

Procedure and design

The experiment was conducted in a soundproof booth (Industrial Acoustics Company, Model 400-A). A test box inside the booth was constructed out of opaque Plexiglas with a wire mesh front. The test box was designed explicitly to allow the tamarins to perch on the wire mesh front where they can easily be viewed by the experimenter. The tamarins can move freely in the test box, but they often cling to the wire mesh and cannot

cling to the smooth Plexiglas surfaces. A concealed loudspeaker (Alesis Monitor One) was located inside the soundproof booth, above and to the left of the subject's back. A video camera with a microphone recorded the experiment and was connected to a monitor outside of the booth. An experimenter seated outside the booth observed the tamarin on the monitor, presented stimuli and judged habituation trials. Stimuli were played from a Macintosh G4 computer via an Audiomedia soundcard and an Alesis amplifier. Subjects remained in the test box for an entire session without access to food or water; a session lasted approximately 30 min and upon completion, subjects were returned to their home cage with *ad libitum* water access.

The design of this experiment was modeled on those of Nazzi *et al.* (1998), Nazzi *et al.* (2000), and Ramus *et al.* (2000). An experimental session began with a habituation phase that presented utterances produced by two speakers of one language. A test phase immediately followed. During the test phase, all novel speakers were presented; three utterances were produced by a novel speaker of the habituation language (Same) and three utterances were produced by a novel speaker of the new language (New).[6] The session began when the experimenter judged the tamarin to be relatively stationary and looking 180 degrees down and away from the concealed loudspeaker; this position was spontaneously achieved, with no food reinforcement. Once the subject was in the ready position, an experimenter presented the first habituation sentence and made a judgment of the subject's response ('online coding'). A 'Yes' response was scored if the subject turned its head in the direction of the concealed speaker during the playback or within 2 seconds of its offset. A 'No' response was scored if there was no head movement at all, or if head movement occurred, it was directed away from the speaker. If the experimenter was unsure of the response, then 'Ambiguous' was recorded. The inter-trial interval was set to a minimum of 10 s and a maximum of 60 s. If the subject failed to achieve a ready position within this interval, then the experimenter presented the next trial regardless of the subject's position. If the subject was looking at

---

[4] Utterances from American English were first recorded for Experiment 5 of Nazzi *et al.* (2000) and therefore were not included in the Ramus *et al.* (1999) analyses.

[5] There has been debate over whether Polish is a stress- or syllable-timed language. Ramus, Dupoux and Mehler (2003) present perceptual evidence that Polish is neither stress-, syllable-, nor mora-timed, but might represent a new rhythmic class. Importantly, Polish has a very different rhythm compared to mora-timed Japanese.

[6] This design raises the possibility that discrimination was based on idiosyncratic differences between the New and Same Language test speakers. However, the backwards conditions serve as controls for that. Based on the findings of Sheffert, Pisoni, Fellowes and Remez (2002), talker information is likely preserved in backwards speech; human adults trained with backwards speech samples successfully discriminated talkers. This finding would predict that if some characteristic of the test speakers makes them inherently easy to discriminate, then this characteristic should also be present in the backwards utterances and should lead to discrimination of the backwards utterances. As shown in the results, this is not a concern for the present study.

the loudspeaker when the playback was initiated or was jumping or calling, then the experimenter recorded the trial as 'Bad'.

During habituation, the experimenter presented a minimum of 10 trials (one of each of the five utterances from the two speakers) before evaluating if the subject had reached the habituation criterion. A subject was considered habituated if it unambiguously failed to respond (i.e. No response) on three consecutive trials. During online coding of habituation, Ambiguous and Bad responses were treated as Yes responses. Once a subject reached the habituation criterion, the experimenter then presented six test trials (three New speaker, Same language and three New speaker, New language) following the same procedure as described for habituation. The first test trial always presented a sentence from the New language. The remaining five test trials alternated, Same-New-Same-New-Same.

Two aspects of the experimental design differed from that used by Ramus *et al.* (2000) on tamarins. First, we used a habituation criterion of three consecutive No responses rather than two. We changed this criterion to increase our confidence in the extent of habituation, and thus, decrease the likelihood that a Yes response during the test phase was due to spontaneous recovery of attention. Second, we presented six utterances (three New speaker, New language, three New speaker, Same language) in the test phase with a fixed order (New language-Same language-New-Same-New-Same). In Ramus *et al.* (2000), each subject received a single test trial, either a test sentence produced by a new speaker of the habituated language (speaker change) or by a new speaker of the new language (language change). Therefore, we gained the statistical power of a repeated-measures comparison over the between-subjects comparison previously used. However, order effects were then a concern. We elected the fixed order in the present study as a conservative approach designed so that all tamarins were comparably habituated when presented with the first New speaker, New language test trial. If the first test trial was a New speaker, Same language sentence, then some subjects would likely not respond to the speaker change, essentially changing their habituation state, and possibly leading to a greater likelihood of response to the New speaker, New language trials (Ramus *et al.*, 2000). Finally, we eliminated the post-test playback of a tamarin call as this failed to be diagnostic of the tamarins' overall level of habituation.

Of the 20 subjects, 17 participated in four successful experimental sessions, one for each of the conditions: Forward Polish/Japanese, Forward American English/ Dutch, Backwards Polish/Japanese, and Backwards American English/Dutch. For each animal, at least four

days elapsed between the experimental sessions for each condition. The order of conditions and habituation language were counterbalanced across animals. One adult subject participated in three successful experimental sessions; the experimenter aborted two attempts on the Forward Polish/Japanese session due to the subject's inability to achieve a ready position. One sub-adult subject participated in only one successful session (Backwards American English/Dutch) and was excluded from further testing due to an overall lack of response to the playback procedure. We aborted one adult on his first experimental session because he rarely maintained the ready position. Therefore, the 20 original subjects yielded 72 experimental sessions out of a possible 80. The final coding identified 18 of these 72 experimental sessions as containing at least one Yes response on the last three habituation trials; these sessions were excluded from further analysis. The final data set included 13 Forward Polish/Japanese experimental sessions, 13 Forward American English/Dutch, 14 Backwards Polish/Japanese, and 14 Backwards American English/Dutch.

Offline coding, reliability and dependent measure

The last three habituation trials and the six test trials were coded offline using digitized video clips. Offline observers judged the trials following the criteria stated above for online coding. Importantly, if an offline observer judged any of the last three habituation trials to contain a Yes response, then the entire experimental session was excluded from further analyses. The trials were digitized (30 frames/s) from the videotapes and saved as individual video clips identified by random numbers. Using the audio track, we marked the onset and offset of the auditory stimulus and a 2-s window following the offset. During coding, the observer muted the sound, used the visual markers to identify the coding window, and judged the subject's response from the stimulus onset to the offset of the 2-s window. Clips were coded in random number order, as opposed to the actual experimental session order. This procedure ensured that coders were blind to the experimental condition, trial type, trial order and stimulus. Two offline observers judged 70% of the digitized video clips. One offline observer judged the remaining video clips. The two observers agreed on 91% of the trials with $\kappa = .82$, indicating excellent agreement (Cohen, 1960; Fleiss, 1981). A third offline observer judged the remaining trials containing disagreements. The response agreed upon by two of the three observers was entered as the final response. When there was disagreement by all three observers, the trials were excluded from the final analysis (five of the 41 trials judged by three observers).

The dependent measure calculated for each animal in each condition was a difference score comparing mean percentage of Yes responses on New language test trials with mean percentage Yes responses on Same, habituated language test trials. A Wilcoxon signed-ranks test was then run on these difference scores using a two-tailed α-level of .05. Greater responding on New test trials than on Same test trials indicates discrimination.

Results

The rhythmic class hypothesis predicts that tamarins should discriminate natural utterances drawn from languages from different rhythmic classes (e.g. Polish and Japanese), but should fail to discriminate natural utterances drawn from languages from the same rhythmic class (e.g. American English and Dutch). The results support these predictions. Tamarins in the Forward Polish/Japanese condition initiated significantly more orienting responses on the New language test trials than on the Same language test trials ($M_{New} = 69\%$, SD = 31.96, $M_{Same} = 41\%$, SD = 33.88; $n_{New>Same} = 11$, $n_{New<Same} = 1$, $T(-) = 10$, $z(13) = -2.28$, $p = .03$). In contrast, when tamarins were tested in the Forward American English/Dutch condition they showed no discrimination, initiating an equal percentage of orienting responses on New and Same language test trials ($M_{New} = 51\%$, SD = 35.11, $M_{Same} = 51\%$, SD = 32.28; $n_{New>Same} = 4$, $n_{New<Same} = 5$, $T(+) = 21$, $z(13) = -0.18$, $p = .86$). These results are summarized in the top panel of Figure 1.

We also tested the hypothesis that backwards speech would disrupt rhythm processing and lead to no discrimination. In contrast to the tamarins' successful discrimination of Forward Polish and Japanese, they did not initiate significantly more orienting responses to New than to Same language test items in the Backwards Polish/Japanese condition [$M_{New} = 41\%$, SD = 29.93, $M_{Same} = 55\%$, SD = 25.05; $n_{New>Same} = 3$, $n_{New<Same} = 6$, $T(+) = 16.50$, $z(14) = -0.71$, $p = .48$]. They also failed to respond more to New language test items in the Backwards American English/Dutch condition [$M_{New} = 56\%$, SD = 33.18, $M_{Same} = 52\%$, SD = 36.41; $n_{New>Same} = 4$, $n_{New<Same} = 4$, $T(-) = 17$, $z(14) = -0.14$, $p = .89$]. These results are illustrated in the bottom panel of Figure 1.

The above pattern of results is also supported by a chi-squared analysis on the subjects' New-Same difference scores from all conditions (Forward Polish/Japanese, Forward American English/Dutch, Backward Polish/Japanese, Backward American English/Dutch). Significantly more subjects in the Forward Polish/Japanese condition oriented to New test items than to Same test items compared to subjects in the other conditions [$\chi^2(3) = 8.714$, $p = .03$].
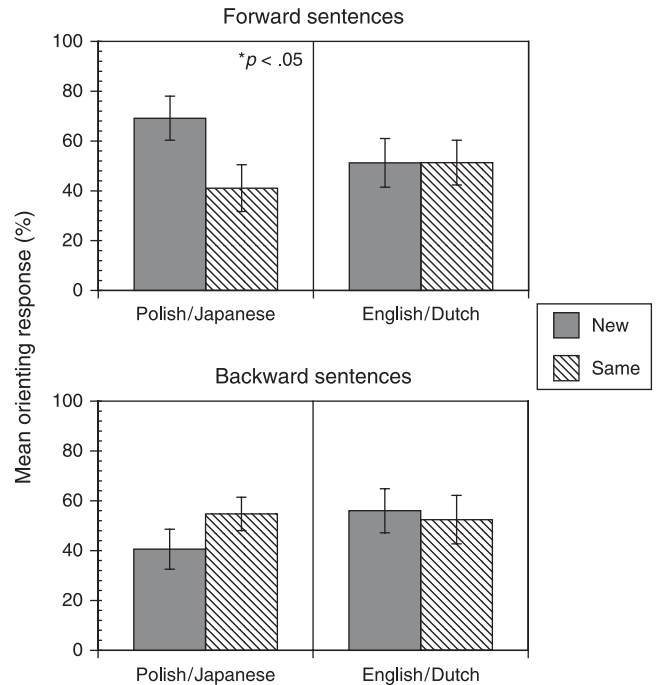
Figure 1   *Mean orienting response (%) to the New language (solid bars) and Same language (striped bars) test trials. The top graph presents responses to Forward utterances and the bottom graph to Backwards utterances. The Polish/Japanese comparison represents two languages from two different rhythmic classes; the English/Dutch comparison represents two languages from the same rhythmic class.*

Finally, we conducted two exploratory analyses to test for effects due to the age of the subject and the number of habituation trials the subject experienced. One prediction is that mated adult tamarins may respond differently from sub-adults due to possible effects of familiarity with the testing procedure, possible effects of hormonal state on attention or perception, maturation of the rhythm perception mechanism, amount of experience hearing human language, or the mechanism's sensitivity to experience-induced changes. No significant differences were found between the New–Same difference scores for mated adults and sub-adults across all four test conditions [Forward Polish/Japanese: mated adults, $n = 8$, $M_{Diff} = 25\%$, SD = 23.74, sub-adults, $n = 5$, $M_{Diff} = 33\%$, SD = 23.69, $U(13) = 15.50$, $p = .45$; Forward American English/Dutch: mated adults, $n = 9$, $M_{Diff} = 11\%$, SD = 40.90, sub-adults, $n = 4$, $M_{Diff} = -25\%$, SD = 32.12, $U(13) = 8.5$, $p = .14$; Backwards Polish/Japanese: mated adults, $n = 10$, $M_{Diff} = -3\%$, SD = 33.50, sub-adults, $n = 4$, $M_{Diff} = -41.5\%$, SD = 41.99, $U(14) = 10.00$, $p = .16$; Backwards American English/Dutch: mated adults, $n = 10$, $M_{Diff} = 2\%$, SD = 32.05, sub-adults, $n = 4$, M_{Diff}
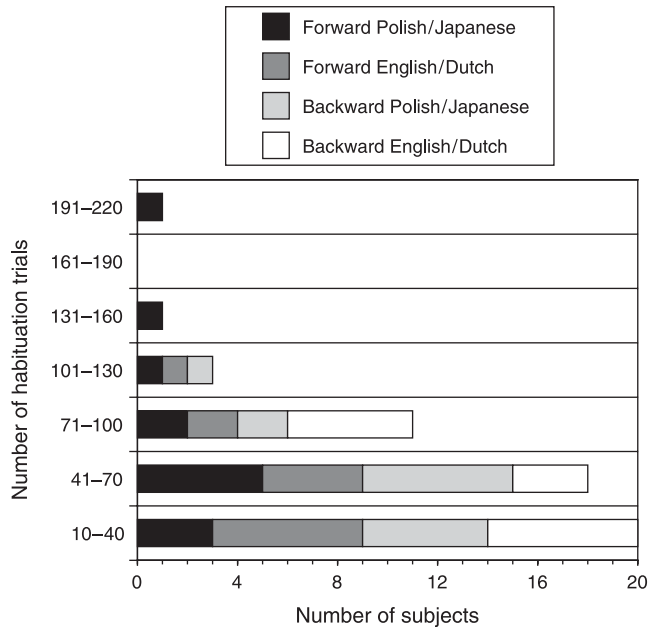
Figure 2 *Histogram of the number of subjects in each of the experimental conditions as a function of the total number of habituation trials to which they were exposed.*
Note: The *x* and *y* axes are reversed for illustration purposes.

= 8.5%, SD = 41.99, $U(14) = 18.00$, $p = .77$]. The Appendix presents the individual difference scores and further illustrates that the sub-adults' scores are equally distributed among the mated adults' scores in each of the test conditions. The exploratory analysis of the habituation trials revealed that subjects differed greatly in the number of habituation trials they experienced before reaching the criterion (range = 13–197). In particular, the most extreme number of habituation trials, 151 and 197, occurred in the Forward Polish/Japanese condition (see Figure 2). This observation raised the possibility that the successful discrimination of the Forward Polish/Japanese test items was due to subjects in that condition receiving more habituation trials, and thus more experience, before they reached the criterion than the subjects in the other conditions. However, two analyses rule out the possibility that successful discrimination in the Forward Polish/Japanese condition was based on a greater amount of experience with the habituation sentences. First, there were no significant differences among the experimental conditions in the mean number of trials to habituation [see Figure 2; Forward Polish/Japanese: $n = 13$, $M = 74.39$, SD = 53.69; Forward American English/Dutch: $n = 13$, $M = 51.08$, SD = 28.96; Backward Polish/Japanese: $n = 14$, $M = 55.36$, SD = 29.23; Backward American English/Dutch: $n = 14$, $M = 51.71$, SD = 23.42; $H(3) = 1.69$, $p = .64$]. Second, when the two subjects with the

extreme number of habituation trials were removed from the Forward Polish/Japanese analysis, the significant effect remained [$M_{New} = 70\%$, SD = 31.61, $M_{Same} = 36\%$, SD = 34.91; $n_{New>Same} = 10$, $n_{New<Same} = 0$, $n_{New=Same} = 1$, $T(-) = 0$, $z(11) = 2.81$, $p = .005$].

Discussion

Ramus *et al.* (2000) concluded that in parallel with human newborns, tamarins discriminate forward utterances of Dutch from forward utterances of Japanese. Furthermore, they concluded that both human newborns and tamarins failed to discriminate these same languages when the utterances were played backwards. The present study extends the results of Ramus *et al.* (2000) using a novel comparison of two different rhythmic class languages (Polish and Japanese) and more stringent methodological criteria. The results show that tamarins respond more to utterances from the new, different rhythmic class language than they do to utterances from the same, habituated language. Second, we extended the previous findings by establishing that, like human newborns, tamarins do not discriminate utterances when the languages are from the same rhythmic class, specifically, American English and Dutch. Importantly, the tamarins' lack of discrimination mirrors that of the less experienced newborns and not that of the more experienced 5-month-olds. This pattern of results suggests that, unlike human newborns, the tamarins' capacity to discriminate languages based on rhythmic cues cannot be altered by experience. However, it is possible that the youngest members of both species could be more sensitive to experience-induced changes. If so, then the tamarins would be at a disadvantage as they were not immersed in natural spoken language during their development as human infants are. An important issue for future research will be to determine whether or not the tamarins' auditory system can be tuned with early exposure to language, and if so, how this experience modifies their discrimination abilities. One caveat to these findings is that the discrimination test did use a fixed trial order. The purpose of the fixed trial order was to simplify the discrimination test for the tamarins, compared to a randomized test order, and to ensure the habituation criterion was equally met across animals, compared to a counterbalanced order. A second important issue for future studies with tamarins, using a within-subjects measure and explicit habituation to a criterion, is to determine the effects of a fixed test order.

Together these data are consistent with the conclusion that tamarins are attending to rhythmic cues in the language discrimination task and are not attending to other

acoustic cues that could lead to more fine-grained discriminations within a rhythmic class. The shared pattern of discrimination and failures for tamarins and human newborns lends support to the proposal that the same mechanism is present in both species. This proposal represents the current working hypothesis. Further research is necessary to identify further similarities and, importantly, differences in the spontaneous behavioral responses of human newborns and tamarins, and in other species. Recent training experiments show that rats can also discriminate forward, but not backwards, sentences of Dutch and Japanese, raising the possibility that the mechanism underlying our capacity to discriminate languages based on rhythmic cues may have evolved as early as the origin of mammals (Toro, Trobalon & Sebastián-Gallés, 2003). Further research will need to address possible distinctions in the mechanisms underlying spontaneous and trained behaviors, as well as determining if differences exist in the underlying mechanisms even when the behavioral responses are the same. A nice illustration of this distinction comes from work on number representation, where symbol-trained chimpanzees perform numerical discrimination tasks as do human children and adults, but the mechanism underlying this ability is completely different in these two species (Biro & Matsuzawa, 1999; Dehaene, 1997; Gallistel & Gelman, 2000; Hauser, 2000). Distinguishing between these hypotheses will require tests involving other species, tasks and stimuli, including the use of neurophysiological and neuroimaging preparations. Such tests are critical for they will enable us to address the more general and fundamental question lying at the core of research on language evolution: are the perceptual and computational mechanisms that we recruit for language comprehension and production specific to language or, did these mechanisms evolve for more general, language-independent functions? (Hauser *et al.*, 2002a; Kuhl, 1986, 2000; Trout, 2003; chapters in Christiansen & Kirby, 2003; Wray, 2002). More specifically, experiments with stimuli that isolate information such as rhythmic, phonotactic or speaker cues, will help identify differences that point to species-specific mechanisms.

The present results also replicate the earlier finding that tamarins fail to discriminate languages when the utterances are played backwards, regardless of rhythmic class. The specific cause that renders backwards speech so difficult to process needs to be determined. Earlier proposals, based on evidence from human newborns, suggested that failures with backwards utterances indicated an innate tuning to specific properties of human speech (Mehler *et al.*, 1988). For example, one speculation was that backwards speech disrupts the spectral patterns contained in natural speech utterances. This spectral information reflects the articulatory dynamics of the speaker's vocal tract and aids the listener in recovering the discrete linguistic units from the continuous acoustic signal. If the mechanism guiding language discrimination is tuned to these speech-specific spectral patterns, then distorting that information would disrupt the mechanism and could eliminate discrimination. Evidence that tamarins fail to discriminate backwards utterances argues against such proposals that rely on speech-specific constructs. Alternatively, the tamarins' and human newborns' inability to discriminate backwards utterances may suggest that the mechanism guiding language discrimination is constrained by sensitivity to general properties of the primate vocal production system, such as vibrating vocal folds or resonant vocal tracts. One further possibility, consistent with rats' inability to discriminate backwards sentences (Toro *et al.*, 2003), is that backwards speech violates more fundamental properties of all natural sounds, suggesting that the perceivers' processing mechanisms are tuned to very basic distinctions that characterize naturally produced sounds.

What are the implications of these results for explanations of language acquisition? Clearly, tamarins did not acquire perceptual sensitivity to rhythmic cues in order to bootstrap grammar. Yet, according to many of the theoretical papers mentioned earlier, a sensitivity to rhythmic cues allows human infants to begin assembling the information relevant for building grammatical structures. If our interpretation of the tamarin data is correct, then the related inference is that the language acquisition device has recruited a general perceptual ability of the primate auditory system for language-internal purposes.

Appendix Difference scores (New% – Same%) for mated adult tamarins and sub-adults

Test condition	Subject	Age	Difference score (%)
Forward Polish/Japanese	1	adult	−33
	2	*sub-adult*	0
	3	adult	33
	4	adult	33
	5	adult	33
	6	adult	33

Appendix *Continued*

Test condition	Subject	Age	Difference score (%)
	7	adult	33
	8	adult	33
	9	*sub-adult*	33
	10	*sub-adult*	33
	11	adult	33
	12	*sub-adult*	33
	13	*sub-adult*	67
Forward American English/Dutch	1	*sub-adult*	−67
	2	adult	−33
	3	*sub-adult*	−33
	4	adult	−33
	5	adult	−33
	6	adult	0
	7	adult	0
	8	*sub-adult*	0
	9	*sub-adult*	0
	10	adult	33
	11	adult	33
	12	adult	67
	13	adult	67
Backwards Polish/Japanese	1	*sub-adult*	−100
	2	adult	−67
	3	adult	−33
	4	adult	−33
	5	*sub-adult*	−33
	6	*sub-adult*	−33
	7	adult	0
	8	adult	0
	9	adult	0
	10	adult	0
	11	*sub-adult*	0
	12	adult	33
	13	adult	33
	14	adult	33
Backwards American English/Dutch	1	adult	−33
	2	adult	−33
	3	adult	−33
	4	*sub-adult*	−33
	5	adult	0
	6	adult	0
	7	adult	0
	8	adult	0
	9	*sub-adult*	0
	10	*sub-adult*	0
	11	adult	17[a]
	12	adult	33
	13	adult	67
	14	*sub-adult*	67

[a] One trial removed based on offline coding as Bad or Ambiguous. See Method for details.

Acknowledgements

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.

Biro, D., & Matsuzawa, T. (1999). Numerical ordering in a chimpanzee (*Pan troglodytes*): planning, executing, and monitoring. *Journal of Comparative Psychology*, **113**, 178–195.

Christiansen, M.H., & Kirby, S. (Eds.) (2003). *Language evolution*. Oxford: Oxford University Press.

Cohen, J.A. (1960). Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Cutler, A. (1994). The perception of rhythm in language. *Cognition*, **50** (1–3), 79–81.

Dehaene, S. (1997). *The number sense*. Oxford: Oxford University Press.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

Gallistel, C.R., & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences*, **4**, 59–65.

Ghazanfar, A.A., Flombaum, J.T., Miller, C.T., & Hauser, M.D. (2001). The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saginus oedipus*): playback experiments with long calls. *Journal of Comparative Physiology, A*, **187**, 27–35.

Ginther, A.J., Carlson, A.A., Ziegler, T.E., & Snowdon, C.T. (2002). Neonatal and pubertal development in males of a cooperatively breeding primate, the cotton-top tamarin (*Saguinus oedipus oedipus*). *Biology of Reproduction*, **66**, 282–290.

Hauser, M.D. (2000). *Wild minds: What animals really think*. New York: Henry Holt.

Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition*, **78**, B53–B64.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002a). The faculty of language: what is it, who has it, and how did it evolve? *Science*, **298**, 1569–1579.

Hauser, M.D., Pearson, H., & Seelig, D. (2002b). Ontogeny of tool use in cotton-top tamarins (*Saguinus oedipus*): recognition of functionally relevant features in the absence of experience. *Animal Behaviour*, **64**, 299–311.

Holt, L.L., Lotto, A.J., & Kluender, K.R. (1998). Incorporating principles of general learning in theories of language acquisition. In M. Gruber, C. Derrick Higgins, K.S. Olson & T. Wysocki (Eds.), *Chicago Linguistic Society, Vol. 34: The Panels* (pp. 253–268). Chicago: Chicago Linguistics Society.

Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P.W., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Language and Speech*, **31** (3), 217–238.

Kuhl, P.K. (1986). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology*, **45**, 233–265.

Kuhl, P.K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, **97** (22), 11850–11857.

Kuhl, P.K., & Miller, J.D. (1975). Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants, *Science*, **190**, 69–72.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, **29**, 143–178.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant's viewpoint. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116). Mahwah: Lawrence Erlbaum.

Miller, C.T., Dibble, E., & Hauser, M.D. (2001). Amodal completion of acoustic signals in a nonhuman primate. *Nature Neuroscience*, **4**, 783–784.

Morgan, J.L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah: Lawrence Erlbaum.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24** (3), 756–766.

Nazzi, T., Jusczyk, P.W., & Johnson, E.K. (2000). Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. *Journal of Memory and Language*, **43**, 1–19.

Oller, D.K. (2000). *The emergence of the speech capacity*. Mahwah: Lawrence Erlbaum.

Pike, K.L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Ramus, F. (2002). Language discrimination by newborns: yeasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, **2**, 85–115.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265–292.

Ramus, F., Hauser, M.D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and cotton-top tamarin monkeys. *Science*, **288**, 349–351.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: perceptual studies. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain.

Sheffert, S.M., Pisoni, D.B., Fellowes, J.M., & Remez, R.E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance*, **28** (6), 1447–1469.

Toro, J.M., Trobalon, J.B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats, *Animal Cognition*, **6** (2), 131–136.

Trout, J.D. (2003). Biological specializations for speech: what can the animals tell us? *Current Directions in Psychological Science*, **12** (5), 155–159.

Weiss, D., Hauser, M.D., & Garibaldi, B. (2001). Cotton-top tamarin long calls: production and perception of calls that encode information about individual, sex, and group identity. *Journal of Comparative Psychology*, **115**, 258–271.

Wray, A. (Ed.) (2002). *The transition to language*. Oxford: Oxford University Press.

Ziegler, T.E., Savage, A., Scheffler, G., & Snowdon, C.T. (1987). The endocrinology of puberty and reproductive function in female cotton-top tamarins (*Saguinus oedipus*) under varying social conditions. *Biology of Reproduction*, **37**, 618–627.