JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2003, **80,** 205–215 NUMBER 2 (SEPTEMBER)

SPEECH PERCEPTION IN RATS: USE OF DURATION AND RISE TIME CUES IN LABELING OF AFFRICATE/FRICATIVE SOUNDS

PHIL REED, PETER HOWELL, STEVIE SACKIN, LISA PIZZIMENTI, AND STUART ROSEN

UNIVERSITY COLLEGE LONDON

The voiceless affricate/fricative contrast has played an important role in developing auditory theories of speech perception. This type of theory draws some of its support from experimental data on animals. However, nothing is known about differential responding of affricate/fricative continua by animals. In the current study, the ability of hooded rats to ''label'' an affricate/fricative continuum was tested. Transfer (without retraining) to analogous nonspeech continua was also tested. The nonspeech continua were chosen so that if transfer occurred, it would indicate whether the animals had learned to use rise time or duration cues to differentiate affricates from fricatives. The data from 9 of 10 rats indicated that rats can discriminate between these cues and do so in a similar manner to human subjects. The data from 9 of 10 rats also demonstrated that the rise time of the stimulus was the basis of the discrimination; the remaining rat appeared to use duration.

Key words: speech perception, phonemes, rise time, duration, discrimination learning, rats, humans

Several theorists have suggested that the production and understanding of speech is somehow ''special'' (see Kelley, 1986; Liberman, 1996). That is, speech sounds are thought by such theorists to be discriminated from one another, and labeled, differently from the way in which nonspeech sounds are discriminated from one another. Such a view has been termed a *motor theory* of speech perception (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Although there have been several reformulations of motor theory (e.g., Liberman & Studdert-Kennedy, 1978; Repp & Liberman, 1987), they all share several features of the original, and none is as explicit as the view put forth by Liberman et al. (1967). Motor theory suggests that there are invariant relations between particular speech sounds and their motor analogues (Kerzel & Bekkering, 2000; but see Lindblom, 1991; Paliwal, Lindsay, & Ainsworth, 1983), and that a brain module causes perception of these phonemic structures directly by comparison of the speech sound input and motor analogue required for its production (Liberman & Mattingly, 1985). That is, the perception and labeling of speech sounds depend on the listener being able to compare the speech sound input to the motor programs that would be required to produce that sound. Once an appropriate match has been found, the sound can be labeled as a particular phoneme.

Such theorists argue further that, as humans are unique in being the only species possessing speech and language, the processes allowing the discrimination and labeling of speech sounds may be specific to humans. Obviously, such a theory would predict that species such as rats, which are not capable of producing speech, should not perceive and label speech sounds in a manner similar to humans. Although a demonstration of human-like speech perception does not rule out the possibility of a special speech-perception mechanism in humans, it would show that the existence of such a mechanism is not a necessity, and that some other aspect of the situation must support the discrimination of the sounds.

In speech perception, theories such as the above motor theory have implied little explanatory role for auditory discrimination processes that involve the cues contained in the stimuli themselves. That is, the auditory differences among the stimuli that potentially could support the learning of a discrimination between those cues are not suggested by motor theories as critical in the specialized

These data were first presented at the Experimental Analysis of Behaviour Group Meeting, London, and at the Fourth European Meeting for the Experimental Analysis of Behavior, Amiens, France. Thanks are due to Lisa A. Osborne for her support. This research was supported by a grant from the Medical Research Council of the United Kingdom.

Correspondence concerning this article should be addressed to Phil Reed, Department of Psychology, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP, Wales, United Kingdom (e-mail: p.reed@swan.ac.uk).

processes that allow speech perception and the subsequent labeling of speech sounds. There is a class of theories, however, that does assume a central role for the properties of the auditory cues in explaining many phenomena in speech perception. Hence, a crucial issue for such theories of speech perception is the extent to which perceptual processes are based upon the properties of the speech signal. That is, whether the dimensions of the speech sound stimuli (phonemes) are sufficiently different from one another to allow differential responding (labeling) to those stimuli based on the differences in the dimensions of the sound stimuli themselves. Thus the aims of an auditory theory of speech perception are to specify the acoustic/auditory features that are responsible for the perception/labeling of the phonemic distinction, and to provide a mechanism through which various acoustic/auditory factors can interact to determine the phonetic percept.

If such a mechanism is responsible for speech perception, then all animals with the appropriate auditory sensitivities should be able to perceive speech stimuli in a similar manner to that found in humans. That is, the stimuli belonging to the same speech sound continua should be categorized similarly across these species. In fact, several studies on speech categorization have been conducted using nonhuman species. In general, these studies have shown that various nonhuman species, such as chinchillas (e.g., Kuhl & Miller, 1975), quail (e.g., Kluender, Diehl, & Killeen, 1987; Lotto, Kluender, & Holt, 1997), and monkeys (e.g., Sinnott, Brown, & Borneman, 1998; but see Sinnott, Brown, Malik, Kressley, 1997), do categorize speech sounds in a similar manner to humans. This suggests that the mechanisms that allow perception and categorization of speech sounds are not unique to humans and are auditory in nature.

The voiceless affricate/fricative contrast (distinguishing "***ch***ip" from "***sh***ip") has played a particularly important role in supporting such auditory theories when human subjects have been employed. Such a contrast can be made on the basis of a number of interacting acoustic features. Important for the current purposes are the duration and the rise time of the sound stimulus. Compared to the fricative (***sh***), the affricate sound stimulus (***ch***) is typically shorter in duration and has a faster rise time (Gerstman, 1957; Howell & Rosen, 1983). Typical audiograms illustrating the difference between a "***ch****a*" and a "***sh****a*" sound are displayed in Figure 1.

The first empirical evidence of an acoustic feature responsible for phonetic categorization came from a study reported by Cutting and Rosner (1974; see also Delgutte, 1982). They reported a close correspondence between labeling and discrimination of speech sounds (heard as voiceless affricate/fricative sounds) and nonspeech sounds that varied in rise time (heard as plucked–bowed sounds as they resemble a stringed instrument being either plucked or bowed, respectively). On such continua, a rise time continuum was created in which sounds go immediately (from their start) to their maximum amplitude, or take some time (the rise time) to get to their maximum amplitude. The "pluck/bow" is the stimulus classification imposed on stimuli from such a rise time continuum, being a verbal label for the stimuli derived from the different sounds produced when plucking (short rise time) or bowing (long rise time) a string of a violin, guitar, and so forth. Cutting and Rosner also claimed that the discrimination function for a nonspeech, pluck–bow continuum had a peak at the category boundary obtained from a labeling function, just as the speech continuum did. This claim was particularly important as the discrimination peak had hitherto been thought to be unique to speech sounds. Although it is now clear that the discrimination function for pluck–bow continua of this specification does not have a peak (cf. Rosen & Howell, 1981), labeling performance *is* similar for both the speech continuum and the nonspeech continuum.

That speech and nonspeech sounds are perceptually characterized as similar to one another is one problem for a view that suggests that speech sounds are processed by a dedicated specialized mechanism. Such a motor theory view might predict differences in the manner of processing of such speech and nonspeech stimuli. If this finding could be combined with a demonstration of similarities between human and nonhuman discrimination of such speech and nonspeech sounds, these results would be problematic

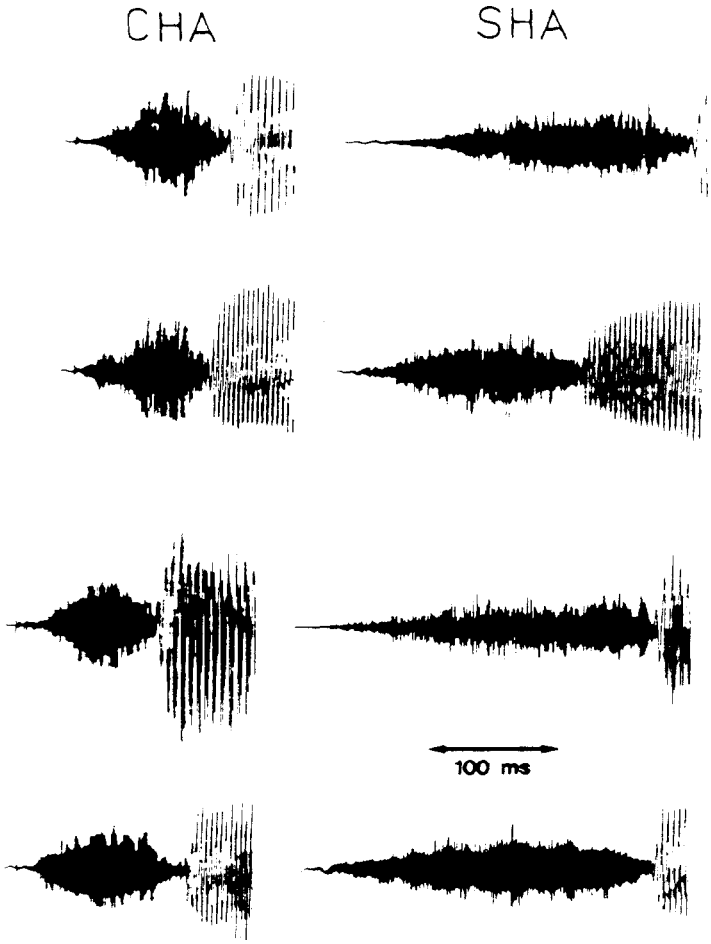

Fig. 1. Audiogram of "*cha*" and "*sha*" sounds showing rise time and duration differences between the stimuli.

for motor theories of speech perception. Unfortunately, nothing is known about the way in which mammals (apart from humans) respond to the affricate/fricative continuum (i.e., nonhuman animals have not been tested on a ***ch***a–***sh***a continuum).

To allow documentation of the cross-species similarities in responding to this continuum, the present experiment employed rat subjects and tested their categorization of speech sounds. Correspondences between human and rat performance on speech continua would provide strong support to an auditory point of view (i.e., a view suggesting that a discrimination can be based on the properties of the stimulus, rather than in the perceptual system of the organism) because motor theory would suggest the systems necessary for speech perception would be absent in the rat and other nonverbal creatures (see Kuhl, 1987, for a review). Such correspondences between rat and human speech perception would also tend to challenge theories that claim that the perception of speech depends on specialized mechanisms that are unique to humans. Similarly, correspondences in perceptual behavior across speech and analogous nonspeech continua (both in humans and rats) would also tend to support auditory theories of speech perception.

The nonspeech analogues of the ***ch***a–***sh***a syllables used in the present study were constructed so that they involved rise time only, or both rise time and frication duration, as cues to discriminate between the stimuli. The selective involvement of the duration cue was achieved by constructing signal-correlated noise (SCN) versions of the original speech

sound (Schroeder, 1968). The SCN transformation maintains the envelope information when it is applied to a signal. Specifically, the transformation maintains the rise time of stimuli. The other effect to note is that the speech sound is transformed to white noise. This effect allows selective involvement of frication duration as a cue.

On the one hand, applying the transformation to the fricated (*sh* or *ch*) portion alone (i.e., to create a noise–speech stimulus continuum) broadens the noise bandwidth, but the duration of the noise section can still be easily determined because the vowel onset (i.e., *a*) is readily apparent. On the other hand, applying the transformation to the vocalic (*a*) portion (i.e., speech–noise) leads to the percept of a continuous noise, in which duration of the fricated portion is not discernable (this same situation applies when the transformation is made on the whole of the signal [i.e., noise–noise]). Thus duration is neutralized as a cue with SCN versions of all the syllables (noise–noise), or when applied to the vowel portion alone (speech–noise), but not with an SCN version applied to the fricated portion alone (noise–speech) or with full speech syllables (speech–speech).

If the animals' generalization from the speech to analogous nonspeech versions depends on whether the vocalic portion is an SCN version, then duration would be implicated as a cue underlying perception of this phonemic distinction. Alternatively, if animals generalize irrespective of whether the vowel is SCN, then rise time would appear to be the critical cue. One further aspect to note about the current experiment is that generalization testing was assessed without any specific training on the alternative versions. Finally, because the auditory sensitivity of rats differs from the auditory sensitivity of humans, comparison of human performance with that of the rats' performance was based on filtered sounds, so that these sounds for the humans approximate those that were heard by the rats.

METHOD

Subjects

Ten experimentally-naive, male Lister Hooded rats were used in the present experiment. One rat became ill and died during the course of the experiment, and its data were not reported in this study. The rats were 4 to 7 months old at the start of the experiment and had initial free-feeding body weight ranges of 350 to 560 g. The rats were initially deprived to 85% of their free-feeding body weights and thereafter were maintained at these weights for 6 weeks. After every 6-week period, 15 g was added to the rats' target weight, and they were maintained at this new weight for a further 6 weeks. The rats were housed in groups of three and four and were kept on a 12:12 hr light/dark cycle. Testing took place in the light part of the cycle.

Six humans acted as subjects, 2 males and 4 females. They were aged between 18 and 25 years and had no history of hearing difficulty.

Apparatus

Training for the rats was conducted in two identical operant-conditioning chambers 23.5 cm by 23.5 cm by 20.5 cm. Reinforcers (45 mg Noyes food pellets) could be delivered to a centrally located food tray that was covered by a hinged, clear plastic flap. There was a response lever on each side of the hinged flap (referred to as left and right response levers). A force of approximately 0.30 N was required to operate the levers. A speaker with flat frequency response within 3 dB, between 50 Hz and 15 kHz, was located on the ceiling of the chamber through which the auditory stimuli could be presented. This speaker was positioned directly above the position of the rat's ears when the rat was in the center of the conditioning box between the two levers. The chamber was not illuminated during the course of the experiment.

Stimuli

The word *sha,* spoken by a female, was recorded into a computer at 30 kHz. Twelve bits of quantization were recorded into the computer at 30 kHz, which was the maximum obtainable on the sound card employed (Data Translation® Model dt 2811). A section of constant amplitude frication preceded the vowel. Successive stimuli were created by reducing the duration in 10-ms steps, from a maximum of 85 ms. Then the onsets were contoured to give rise times varying from 5 to 85 ms, in 10-ms steps, with the most abrupt

rise applied to the shortest stimulus, the next most abrupt rise applied to the next shortest sound, and so on. Thus these stimuli covary in frication duration and in rise time at onset, so that the shortest sounds like *cha*.

A nonspeech version of each stimulus was created by randomly flipping the polarity of each sample in the digitized speech waveforms, with a probability of .5 (SCN, cf. Schroeder, 1968). Such a transformation preserves many of the temporal properties of the signal but makes its spectrum white. There is, thus, relatively little effect on the properties of the aperiodic frication in the sounds (maintaining their rise time and frication, and the aperiodic nature of the sound). The vowels (i.e., *a*), however, sound very different after the transformation as they are transformed from quasi-periodic versions to aperiodic stimuli.

The speech–noise and noise–speech stimuli were constructed as follows. First, for both sets of stimuli, the end of the affricate portion was located. The SCN transformation was applied up to the start of the vowel for the noise–speech stimuli (converts only the fricated portion to white noise) or from the vowel to the end of the sound for the speech–noise stimuli (converts only the vowel portion to white noise).

Procedure

Rats. The rats were initially magazine trained in two 30-min sessions. During each of these sessions, the response levers were retracted from the chamber and reinforcement was delivered according to a random time 60-s schedule. For the first session, the flap covering the food tray was taped open to allow easy access to the pellets. For the second session, the flap was lowered to its standard resting position. Following magazine training, the rats were trained to press the levers in four sessions during which a continuous reinforcement (CRF) schedule was in effect. For each of these sessions, both of the levers were inserted into the chamber, and reinforcement followed a response to either lever. Each of these sessions lasted for 30 min.

The critical experimental manipulations began immediately following the last session of CRF training. For these sessions, both levers were inserted into the chamber and remained there throughout the session. Sessions consisted of sixty 15-s trials, during which reinforcement was available. Each trial was separated by a 45-s intertrial interval (ITI). During each of the trials, one of two sounds (*cha* or *sha*) was presented. For 5 rats (Geof, Euan, Tony, Timm and Ch), when the stimulus *cha* was presented, left-lever responses were reinforced. Responses to the right lever were not reinforced. For these rats, when the stimulus *sha* was presented, responses to the right lever were reinforced, but responses to the left lever were not. The opposite contingencies were operative for the other rats (Bo, Hen, Jo, and Ro).

Table 1

Number of trials to performance criteria during acquisition training.

Rat	Trials
Geof	43
Euan	118
Timm	37
Chris	107
Ro	108
Bo	116
Jo	56
Tony	45
Hen	46

Training consisted of two phases: A phase during which the rats were trained on the endpoints, and a generalization test phase. During the endpoint phase, for the first two sessions one response was required for reinforcement during the stimulus presentation. This requirement was increased during Sessions 3 and 4. In these sessions, reinforcement required either one or two responses. This number was determined randomly for each successive reward. For the remainder of the experiment, reinforcement was arranged according to a random-ratio 3 schedule. The number of required responses varied from reward to reward and were determined randomly. Each session consisted of 30 trials of each stimulus type and lasted for 30 min. This phase of training continued until the rats reached a criterion of six continuous sessions during which they reached 80% correct responses on each lever. The number of trials taken to reach this criterion are displayed in Table 1 for each rat. As soon as the rats reached this criterion, they were shifted to the generalization test phase.

There were four tests during the generalization phase of the study. Each session contained nine stimuli in the series. In the first test, the rats were given six sessions in which there were 56 trials. Twenty-eight of these trials were as described above (14 with *cha,* and 14 with *sha*). The other 28 trials corresponded to four nonreinforced trials for each of the seven midpoint stimuli (i.e., stimuli with 15- to 75-ms rise time, in 10-ms steps). Following six sessions of this testing, the rats were given four sessions of endpoint training as before, and were then tested with the SCN sounds. This test took the same form as the above, but with 28 trials for the speech endpoints (reinforced), and 36 nonreinforced trials, four each with the two endpoint and seven midpoint SCN sounds.

After six sessions of these contingencies, the rats were then returned to discrimination training with *cha–sha* for four sessions. The third test used the speech–noise stimuli. Again, there were 28 trials for the speech endpoints (reinforced), and 36 trials for the speech–noise stimuli—four for each of the endpoint stimuli and four for each of the seven midpoint sounds (nonreinforced). After four more sessions of baseline training, the subjects experienced the noise–speech test. This test took the same form as that described above except the stimuli from the noise–speech continuum were employed. The rats were then exposed to the same procedure a second time.

Humans. The humans labeled four continua. Two of the continua were the speech–speech and the noise–noise continua that were played to the rats. The other two continua were based on the above speech–speech and noise–noise continua. As the audible frequency range for rats and humans is different, a set of stimuli were created for the human subjects that would contain only the frequencies audible to the rats. The procedure to create these stimuli for the humans was based on the results of Kelly and Masterton (1977) for the albino rat (*rattus norvegicus*). The stimuli were created by filtering the speech and noise stimuli prior to digitizing them. High-pass filtering of the stimuli occurred at 8 kHz, 12 dB/octave rolloff. The stimuli were also low-pass filtered at the Nyquist frequency leaving a total bandwidth of 7 kHz (i.e., 8 to 15 kHz).

The humans were presented with the nine stimuli, varying in rise time as described for the rats. Each of the stimuli was presented 10 times to each of the human subjects. The order of presentation of the stimuli was randomized. The subjects were presented these stimuli through headphones and were seated in front of a response panel that comprised two keys. They were instructed to press the key on their left if they perceived the stimulus as *cha,* and to press the key on their right if they perceived the stimulus as *sha.*

RESULTS

Figure 2 shows the obtained labeling functions for the sound continua. The results for each of the four different continua are given: all speech (speech–speech), all noise (noise–noise), frication portion noise (noise–speech), and vocalic portion noise (speech–noise). These functions represent the percentage of responses emitted to the lever associated with reinforced responding when the endpoint sound *cha* was presented (i.e., the *cha* appropriate lever). The rat's scores for each stimulus were calculated by dividing the number of times that the rat made its first response during the stimulus presentation to the lever associated with reinforced responding in the presence of the endpoint speech–speech stimulus *cha* by the total number of trials associated with the stimulus in question during the experiment. If no responses were made during a trial, that trial was excluded from the calculation.

Inspection of Figure 2 shows the performances of the rats on the speech–speech continuum were highly similar to one another. Their performance with speech–speech stimuli generalized across the continuum. Rat Tony's function was shallow, and Rats Euan, Ch and Hen were somewhat noisy. Categorization curves were fitted by cumulative normal functions using maximum likelihood (Bock & Jones, 1968). This test, also known as Probit Analysis, weights the observations with respect to the reliability it is possible to achieve for that number of observations, and with respect to their position in the distribution. From these lines, the 50% point, or category boundary, and the standard error of the estimate (s.e.) were estimated. With these statistics, analysis proceeded by transforming

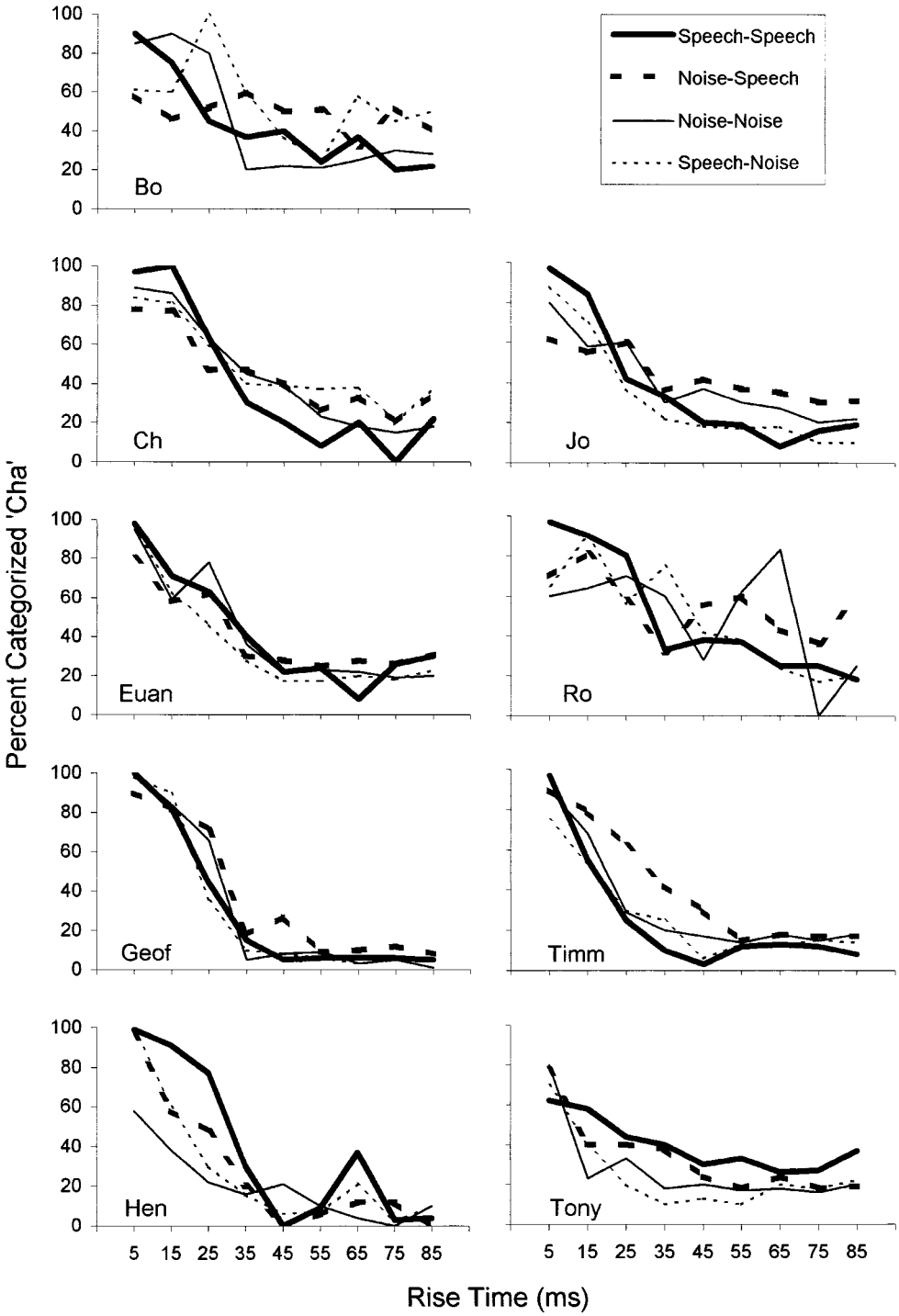

Fig. 2. Labeling functions for rat subjects depicting data from the all speech affricate/fricative continuum (speech–speech), SCN versions of each of the speech stimuli (noise–noise), SCN versions of the affricate portion alone for all stimuli (noise–speech), and SCN versions of the vowel portion alone for all stimuli (speech–noise).

the parameters to standard normal deviates. The category boundary for an individual was defined by the probit analysis as ± 1.96 standard error of measurement (SEM) around the 50% point, and all of these intervals overlapped the location of the 50% point for the group. Thus it appears that all animals place the speech category boundary at a similar position on the continuum.

Performance on the noise–speech continuum (which includes duration and rise time as cues, as does the speech–speech continuum) was similar to that observed on the speech–speech continuum. Rat Tony's performance on noise–speech was not as shallow as with speech sounds, indicating that the shallow function with speech appears to be attributable to variability in performance. Rat Euan's data were somewhat noisy, but there seemed to be clear evidence of categorization both on the noise–speech continuum and on the speech–speech continuum. Probit analysis showed that the 8 rats remaining in the analysis exhibited category boundaries within ± 1.96 s.e. of the pooled data of these same rats. Thus there is no evidence for category boundary differences between rats that can categorize the continuum. Further comparison showed that the category boundaries of the noise–speech continuum all fell within ± 1.96 s.e. of the pooled estimate on the speech continuum for these 8 rats.

The remaining two conditions converted the vowel sound (*a*) to noise, and either left the frication (*cha* or *sha*) as relatively narrowband frication noise (i.e., speech–noise), or converted it to white noise (i.e., noise–noise). As argued, performance on these continua cannot be based on duration, so any evidence of categorizing behavior along these continua indicates the use of the rise time cue in categorization. Inspection of Figure 2 shows that the rats categorized both of these continua of sounds in a similar way to the speech continuum on which they were initially trained. Analysis proceeded as before for the 7 rats (excluding Bo and Ro). For both the speech–noise and noise–noise continua, each rat's data fell within ± 1.96 s.e. of the pooled estimates of these 7 rats for the same test continuum (noise–noise and speech–noise, respectively). Moreover, when the individual estimates were compared with ± 1.96 s.e. estimated for the three other test continua, for these 7 rats all individual boundaries fell within the same confidence limits. Thus it appears that the rats can generalize their initial training on the speech–speech continuum to continua they have not heard before. Because rise time is the only cue that would permit the rats to do this, it seems that most of the rats relied on this cue.

Two rats (Rat Bo and Rat Ro) did not show the same pattern of performance as the other 7 rats shown in Figure 2. Rather, these rats displayed a different profile from one another, as well as from the other rats. Performance of Rat Bo was consistent with predictions based on the subject using the duration cue. That is, for Rat Bo, the two continua to which the initial categorization did not generalize (noise–noise and speech–noise) had the duration cue neutralized. This subject, thus, appeared to use duration as a cue. Rat Ro showed little evidence of categorization for any continuum other than that trained.

The results for the human subjects listening to speech and noise are presented (before and after filtering in each case) in Figure 3. These data show that there was good generalization between nonfiltered and filtered continua and between speech and noise for all subjects (with the exception of Subject H3). Probit analyses showed that there was no significant difference between the functions for speech and noise continua. Moreover, the boundaries for the filtered and nonfiltered continua for the humans did not differ from each other, suggesting that the properties underlying this discrimination were contained similarly with the speech and noise stimuli and were not restricted to speech stimuli.

Figure 4 presents the group-mean category boundaries obtained for the categorization of the speech–speech and noise–noise sounds for the 7 rat subjects who completed all generalizations successfully, and for the 6 human subjects employed. (These data reflect performance on the filtered versions of the sounds in the case of the humans.) Inspection of these data shows similar performance on the speech and noise continua within a species, all ts < 1. The category boundary for the humans was at a longer rise time for the rats on both discriminations, smallest $t(11) = 3.01$, $p < .01$.

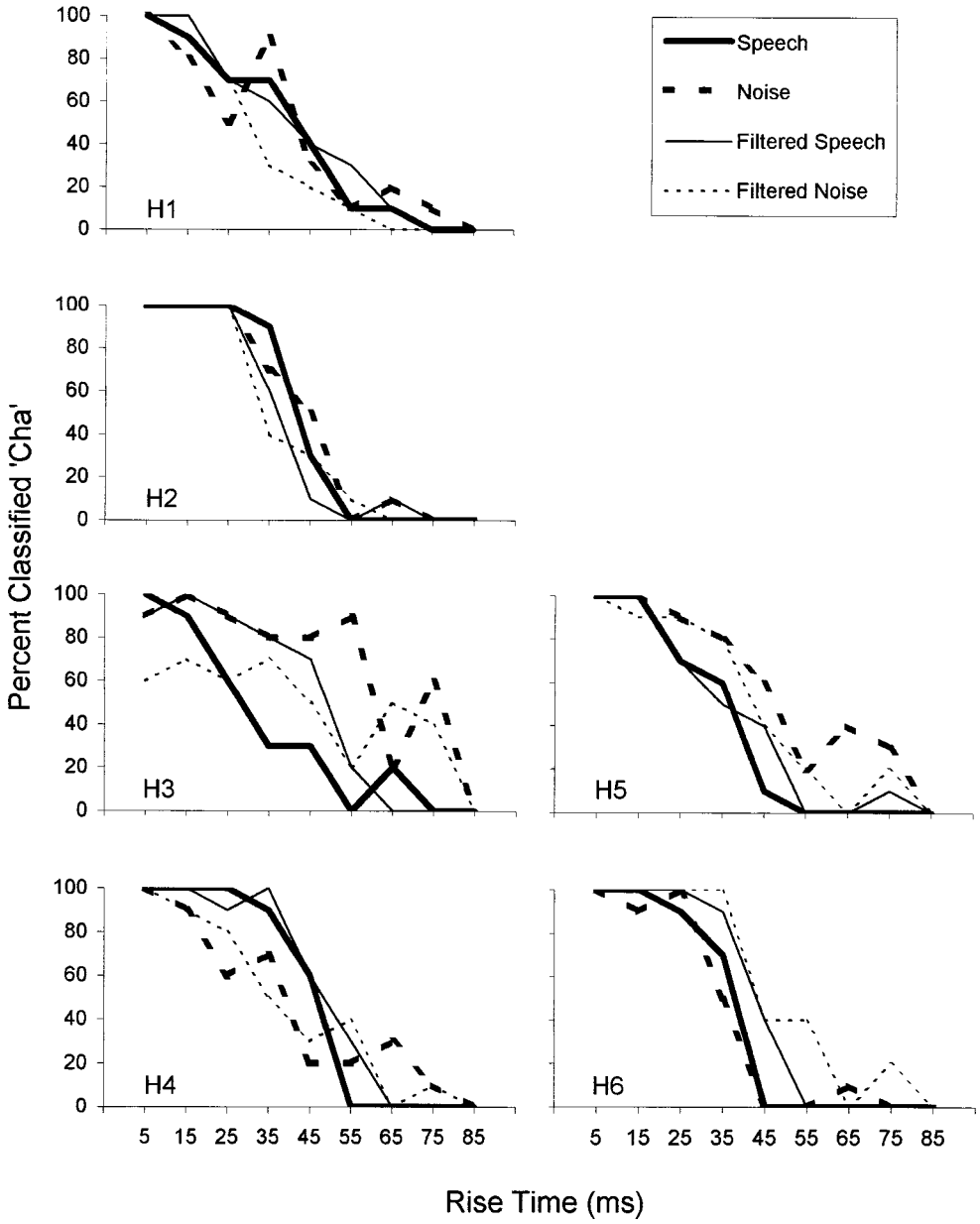

Fig. 3. Results of 6 human listeners labeling speech–speech and noise–noise continua before and after filtering.

DISCUSSION

The results demonstrate that all of the rats employed learned the speech distinctions, and that 8 out of the 9 rats transferred this ability to nonspeech continua that varied in rise time. The rat subjects performed these discriminations of speech sounds as well as humans. Also, rat performance was functionally similar to the performance of humans in that both rat and human performance varied with rise time (albeit different absolute rise times being responsible for controlling the 50% categorization point). That performance was similar across species suggests that there is no specialized mechanism for the perception of speech sounds that is uniquely human. This corroborates the findings obtained from other species using other pho-

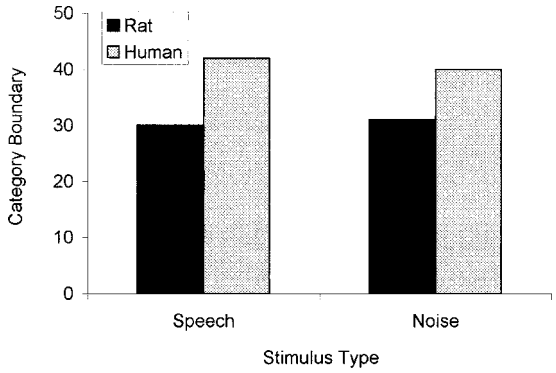

Fig. 4. Group-mean (and standard errors) for the category boundaries for the speech and noise continua for the rat and human subjects.

nemic distinctions (e.g., Kluender et al., 1987; Kuhl & Miller, 1975; Lotto et al., 1997). In doing so, using rats for the first time, these data also add to the growing body of evidence that nonhuman animals discriminate between speech sounds in a similar manner to humans (see Kuhl, 1987, for a review). To this extent, these data give support to auditory theories of speech perception and are problematic for motor theories of speech perception.

The present results implicate rise time as the salient dimension that allows classification of the affricate/fricative continuum because labeling ability was still demonstrated when duration was neutralized as a possible discriminative cue. This conclusion is also indicated by Cutting and Rosner's (1974) results with nonspeech sounds. The nonspeech sounds in their experiment consisted of a sawtooth waveform that varied in rise time. The continuum did not attempt to mimic duration differences signaled in speech sounds by change from aperiodic frication to quasi-periodic vocalic portions. Labeling performance was similar between speech and nonspeech continua. This result, unlike the discrimination peaks reported to occur on the nonspeech analogue, has been replicated (Rosen & Howell, 1981). The similarity in performance between speech and nonspeech across these studies indicates that discrimination of the speech and nonspeech sounds on these continua are based on the rise time attribute of the auditory stimulus, and that frication duration only plays a secondary role (Dorman, Raphael, & Isenberg, 1980).

Before discussing the implications of these data for theories of speech perception, two procedural issues should be addressed. The speech continua were presented to the subjects in the same order as one another each time. This may have introduced potential confounds into the data. It should be noted, however, that this order of continua was presented twice to the rats, which should serve to minimize such problems. Secondly, the procedure adopted used relatively rich reinforcement schedules. The rich schedules may have afforded the opportunity for discrimination between the training and test sessions (there being a lower frequency of reinforcement in the tests). In turn, this may have resulted in the weaker stimulus control seen in some of the rats. Because the vast majority of generalization curves were very similar to one another, however, this concern is somewhat offset.

The fact that rats and humans generalized between the speech and nonspeech continuum in highly similar ways to one another suggests corroborative evidence for a perceptual theory of speech perception, and further presents problems for a simple motor theory of speech perception. Of course, this interpretation holds as long as it is assumed that all sounds that come from speech are special, and that if one can find a single speech sound continuum that rats can discriminate in a similar manner to humans, then the motor theory is falsified. This strong version of motor theory claims that in order to discriminate speech stimuli, the subject must be able to both perceive and produce the sound. Although motor theories do not in themselves discriminate speech sounds in any other way than this, it may be that such a view is too simplistic. For example, it could be assumed that rats and humans do have similar auditory systems, but that humans lose sensitivity to phonemes that are not present in their language as they are not used. Rats, of course, would not lose sensitivity since they have no language. In this instance, rats would be able to discriminate all the sounds that humans can, plus those that the humans lose. For example, some Asian languages do not have separate "*r*" and "*l*" sounds, and native speakers may not be able to discriminate between the two. It would be worth extending

the analysis in future work to investigate this possibility.

The present data also suggest that duration may be learned as an attribute that differentiates affricates from fricatives. In fact, 1 rat provided data consistent with the interpretation that it used this duration cue to make the discrimination. Thus performance of this rat was random in the two continua that had the duration cue neutralized. Future work should seek to establish why the majority of animals respond on the basis of rise time in performing this task. It may be that rise time is the more salient dimension of the stimulus for this species, and that other species may differ depending upon their auditory sensitivities. Any species differences obtained in labeling of speech stimuli could, thus, reflect auditory differences rather than species-specific processing mechanisms. A practical point about this finding, however, is that it provides some indication that different subjects may learn different cues when learning the same speech sound contrast. If hearing loss in humans prevents use of certain sets of cues, then this suggests that it may be possible to base decisions on other cues.

REFERENCES

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice.* San Francisco: Holden-Day.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception and Psychophysics, 16,* 564–570.

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In R. Carlson & B. Granstrom (Eds.), *The representation of speech in the peripheral auditory system* (pp. 131–149). Amsterdam: Elsevier.

Dorman, M. F., Raphael, L. J., & Isenberg, D. (1980). Acoustic cues for a fricative-affricate contrast in word-final position. *Journal of Phonetics, 8,* 397–405.

Gerstman, L. J. (1957). *Perceptual dimensions for the friction portions of synthetic speech sounds.* Unpublished doctoral dissertation, New York University, New York.

Howell, P., & Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of Acoustical Society of America, 73,* 976–984.

Kelly, D. B. (1986). A motor theory of song perception. *Trends in Neurosciences, 9,* 149–150.

Kelly, J. B., & Masterton, B. (1977). Auditory sensitivity of the albino rat. *Journal of Comparative and Physiological Psychology, 91,* 930–936.

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 634–647.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237,* 1195–1197.

Kuhl, P. K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), *Categorical perception* (pp. 355–386). Cambridge, MA: Cambridge University Press.

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

Liberman, A. M. (1996). *Speech: A special code.* Cambridge, MA: The MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H-L. Teuber (Eds.), *Handbook of sensory psychology, Vol. VIII: Perception* (pp. 143–178). New York: Springer.

Lindblom, B. (1991). The status of phonetic gestures. In I. G. Mattingly, M. Studdert-Kennedy, & A. Liberman (Eds.), *Modularity and the motor theory of speech perception* (pp. 7–24). Hillsdale, NJ: Erlbaum.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America, 102,* 1134–1140.

Paliwal, K. K., Lindsay, D., & Ainsworth, W. A. (1983). Correlation between production and perception of English words. *Journal of Phonetics, 11,* 77–83.

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89–112). Cambridge, MA: Cambridge University Press.

Rosen, S., & Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception and Psychophysics, 30,* 156–158.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735–1736.

Sinnott, J. M., Brown, C. H., & Borneman, M. A. (1998). Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys. *Perception and Psychophysics, 60,* 1032–1043.

Sinnott, J. M., Brown, C. H., Malik, W. T., & Kressley, R. A. (1997). A multidimensional scaling analysis of vowel discrimination in humans and monkeys. *Perception and Psychophysics, 59,* 1214–1224.

Received October 31, 2000
Final acceptance June 30, 2003