Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants

Patricia K. Kuhl; James D. Miller


Stable URL:
http://links.jstor.org/sici?sici=0036-8075%2819751003%293%3A190%3A4209%3C69%3ASPBTVC%3E2.0.CO%3B2-N

Science is currently published by American Association for the Advancement of Science.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aaas.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.
bladders were decoded and the frequency of aggregation sites was compared to corresponding osmotic water flow measurements (Table 1). There is clearly a linear relationship between these variables. Figure 2 illustrates the relationship ($r = 9; P < 0.05$) between relative osmotic water flow and the relative frequency of aggregation sites. A similar relationship ($r = 8; P < 0.05$) is also found between absolute osmotic water flow and absolute frequency of aggregation sites.

This finding suggests that the ordered aggregations of intramembranous particles observed in luminal membrane of granular cells after vasopressin stimulation might represent sites for vasopressin-induced osmotic water flow. Indeed, Pinto da Silva (11) has shown that intramembranous particles of red cell membranes may provide preferential structural pathways for water passage. However, since vasopressin is known to enhance the movement of Na$^+$, urea, and other solutes across the toad bladder (12), it should not be concluded that the rate of osmotic water flow is necessarily a specific function of particle aggregation frequency. Clearly, additional study is necessary to establish the functional significance of vasopressin-induced intramembranous particle aggregation. This report indicates that the distribution of intramembranous particles of granular cell luminal membranes is altered in vasopressin-stimulated toad bladder and suggests that this alteration may underlie vasopressin-induced changes in transport function.

WILLIAM A. KACHADORIAN  
JAMES B. WADE  
VINCENT A. DIASCALA  
Membrane Research Laboratory,  
Renal Service,  
U.S. Public Health Service Hospital,  
Staten Island, New York 10304

References and Notes
1. A. Leaf, Ergeb. Physiol. Biol. Chem. Exp. Pharma- 
kol. 56, 216 (1965); P. J. Bentley, Biol. Rev.  
(Camb.) 41, 273 (1966).
(1962); J. S. Handler, R. W. Batcher, W. Suther- 
lund, J. Orloff, J. Biol. Chem. 240, 4524 (1965); S.  
Hynie and G. W. G. Sharp, Biochim. Biophys. 
Acta 230, 40 (1971); J. Orloff and J. S. Handler,  
3. D. R. DiBona, M. M. Civan, A. Leaf, J. Membr, 
Biol. 1, 79 (1969); M. M. Civan and D. R. DiBona, 
ibid. 19, 195 (1974); W. L. Davis, D. R. P. Good- 
man, J. H. Martin, J. L. Matthews, H. Rasmussen,  
(1966).
7. J. B. Wade, V. A. DiScala, M. J. Karnovsky, J. 
9. V. Speeth and F. Wunderlich, Biochim. Biophys. 
1339 (1973).
12. A. Leaf, J. Gen. Physiol. 43, 175 (1960); R. H. 
Invest. 39, 630 (1960); H. S. Frazier, E. F. Demp- 
sley, A. Leaf, J. Gen. Physiol. 45, 529 (1962); R. M. 
Hays and A. Leaf, ibid., p. 905; M. M. Civan and 
H. S. Frazier and D. R. DiBona, J. Biol. Chem. 
239, 1277 (1964); D. C. Mack- 
night, A. Leaf, M. M. Civan, J. Membr. Biol. 6,  
127 (1971).
13. We thank Ms. Carole Uiterwyk and Mr. Raffael 
Saccone for their helpful technical assistance. Ms. 
Eloise Allen and Ms. Iris Quintanilla for typing 
this manuscript, and Dr. Stephen Rudich for his 
timely suggestions. This work was supported by 
grant SI-74-26 from the Central Clinical Investi- 
gation Committee, PHS, BMS, Division of Hos- 
pitals and Clinics.

3 October 1975; revised 13 June 1975

Speech Perception by the Chinchilla:  
Voice-Voiceless Distinction in Alveolar Plosive Consonants

Abstract. Four chinchillas were trained to respond differently to /t/ and /d/ consonant-vowel syllables produced by four talkers in three vowel contexts. This training generalized to novel instances, including synthetically produced /da/ and /ta/ (voice-on- 
time of 0 and +80 milliseconds, respectively). In a second experiment, synthetic stimu- 
lus with voice-onset times between 0 and +80 milliseconds were presented for identifica- 
tion. The form of the labeling functions and the "phonetic boundaries" for chinchillas 
and English-speaking adults were similar.

Neither speech analysis nor speech syn- 
thesis techniques have led to a successful 
account of our perception of speech sounds 
in terms of invariant acoustic properties 
(i). For this reason and others (i), current 
thorists have hypothesized that at least 
some classes of speech sounds are recog- 
nized by "special processing" (2, 3). Spec- 
ulation as to the nature of this special-to-

speech processing varies: some believe that 
it involves "phonetic feature detectors" 
that presumably respond to rather com- 
plicated and abstract characteristics of the 
auditone signal (4); others have espoused a 
"motor theory" of speech perception (2, 3), 
suggest that one's tacit knowledge of 
the acoustic results of articulatory 
maneuvers somehow mediates the per- 
ception of speech. While variously de- 
scribed, these current theories suggest that 
speech perception is a species-specific be- 
"behavior, and thus, in large part, a uniquely 
manimal ability. As Liberman stated, "Un- 
fortunately, nothing is known about the 
way non-human animals perceive speech 
... however, we should suppose that, lack-
ing the speech-sound decoder, animals 
not would perceive speech as we do, even at 
the phonetic level" (2).

We asked whether the chinchilla, a 
mammal with auditory capabilities fairly 
similar to man's (5), but certainly without 
a phylogenetic history of "phonetic knowl- 
dge," either acoustic or articulatory, could 
correctly classify a large number of 
aturally produced syllables on the basis of 
the voice contrast. In experiment 1, we 
trained four chinchillas, using an avoid-
ance conditioning procedure, to respond 
differently to a variety of spoken /t/ and 
/d/ consonant-vowel (CV) syllables. Once 
trained, these animals correctly classified 
known instances of /t/ and /d/ syllables, 
including syllables produced by new talkers, 
those produced in new vowel contexts, and 
computer-synthesized /a/ and /a/ sylla- 
bles. In experiment 2, we presented syn- 
thetic stimuli that varied systematically 
from /da/ to /da/ to the animals trained 
on natural speech, to animals not trained 
the animal On natural speech, and to English-speaking 
adults for identification. The labeling func- 
tions and the "phonetic boundaries" were 
similar for all animal and human subjects.

The voicing feature, which distinguishes 
voiced (/bdg/) from voiceless (/ptk/) plo-
sives in English, is appropriate for investiga-
tions of speech perception by animals since 
this distinction has been examined in 
cross-language studies of adults (6, 7) and 
in adults (6, 7) and 
in (8, 9). The acoustic properties that 
distinguish voiced and voiceless plosives in 
absolute-initial, prevoiced, stressed posi-
tion can be most readily described as a 
timing difference between the onset of 
the plosive burst and the onset of voicing (6), 
termed the voice-onset time (VOT). In syn-
thetic speech, VOT can be varied along a 
continuum to produce plosives in which 
voicing precedes the plosive burst (pre-
voiced), begins nearly simultaneously 
without the burst (unvoiced), or lags behind 
the plosive burst (aspirated). The VOT is 
specific in milliseconds; negative values 
indicate that voicing leads and positive val-
es indicate that voicing lags. In English, 
prevoiced and unaspirated plosives consti-
tute the voiced phonemic category and as-
pirated plosives constitute the voiceless 
phonemic category. When English speak- 
ers identify synthetic tokens from the VOT 
continuum, perception changes abruptly 
from voiced to voiceless sounds; the VOT 
at which responses divide equally between 
voiced and voiceless is termed the "phonet-
ic boundary." These boundaries differ with 
the place of articulation of the voiced-
voiceless pair; the boundaries for labials
(/ba-pa/), alveolars (/da-ta/), and velars (/ga-ka/) are approximately +22, +35, and +41 msec, respectively (7, 10). Many languages divide the VOT continuum as we do in English (6), but in some languages, the division between voiced and voiceless categories occurs elsewhere. For example, in languages such as Kikuyu (6, 9), prevoiced plosives constitute the voiced category and unaspirated plosives constitute the voiceless category (aspirated plosives do not occur). However, infants 1 to 4 months old discriminate synthetic stimuli that fall on different sides of the English /b-p/ phonetic boundary (VOT of +22 msec) whether they are reared in an English-speaking environment where this boundary is phonemically relevant (8) or in a Kikuyu-speaking environment where it is not (9). These facts lend themselves to at least two interpretations: either young infants demonstrate this perceptual boundary because their "speech processor" responds to its potential phonemic relevance or because the boundary is a natural psychophysical one that could be demonstrated by a nonhuman as well. While our results do not rule out the first interpretation, they are consistent with the second.

In experiment 1, the CV syllables used during discrimination training (/ti, ta, tu, di, da, du/), were recorded twice by each of four talkers (two male and two female), and dubbed onto a disk pack for use in a digital recorder (11). The VOT measurements ranged from +40 to +128 msec for /t/ syllables and -200 to +28 msec for /d/ syllables.

The animals were tested in a double-grille cage suspended below a loudspeaker in a sound-treated booth (5, 12). Four chinchillas were deprived of water and trained to lick a drinking tube mounted at one end of the cage for their daily ration of water. Every third lick on the tube produced a drop of water. The animals were maintained at approximately 90 percent of their original weights. It had been demonstrated (12) that an ongoing activity, such as drinking, reduced the number of false alarms and intratrial avoidance responses during discrimination training.

For two animals, /t/ was the positive

![Figure 1](image-url)
stimulus and /d/ was the negative stimulus; for the other two animals, these roles were reversed. On positive trials, the animal was trained to flee the drinking tube and cross the midline barrier to avoid shock. When the animal crossed the barrier, lights positioned at the barrier ends were lit briefly and the stimulus was terminated. Failure to cross the barrier within the 2.5-second trial interval resulted in simultaneous presentation of buzzer and shock until the crossing response was made. On negative trials, no consequences were delivered during the 2.5-second trial interval. If the animal correctly refrained from crossing, the water valve opened, making “free” water available for 1 second, and lights above the drinking tube were lit.

The animals were given 24 trials daily over a 7-month period. Each trial consisted of two presentations, separated by 500 msec, of the CV syllable. The time between trials was varied from 10 to 30 seconds, and the sound levels of the syllables were varied from trial to trial (52 to 66 db, sound pressure level (SPL)). A masking noise (speech-shaped, 12 db SPL) was continuously presented in the test booth.

The experiment began with a single /tu/ and /du/ syllable; when this task was mastered, variation in the form of tokens (second repetition by the same talker), talkers (identical phonetic token by different talkers), and vowels (different vowel contexts by a single talker) were singly introduced (13). The criteria for progressing from one condition to the next were that the group’s percentage correct remain ≥ 90 percent for two consecutive days and that during this time no single animal’s performance was less than 80 percent correct. Failure to make the crossing response on a positive trial and making the crossing response on a negative trial were both scored as errors. The group’s percentage correct is displayed in the form of the labeling gradients and in the location of the phonetic boundaries for human and chinchilla subjects, respectively. The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2. The stimuli were labeled “ta” and “da.” The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2.

On half of the trials during generalization tests, the stimuli presented were those already mastered by the animals. On the other half of the trials, generalization stimuli were randomly presented. In this way, performance on familiar stimuli served as a control for performance on novel stimuli. When control stimuli were presented, all feedback previously in effect was maintained. When generalization stimuli were presented, neither shock nor the buzzer were used, and all other feedback was arranged to indicate a correct response. Four English-speaking adults identified the same stimuli as /da/ or /ta/. The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2.

The experiment began with a single /tu/ and /du/ syllable; when this task was mastered, variation in the form of tokens (second repetition by the same talker), talkers (identical phonetic token by different talkers), and vowels (different vowel contexts by a single talker) were singly introduced (13). The criteria for progressing from one condition to the next were that the group’s percentage correct remain ≥ 90 percent for two consecutive days and that during this time no single animal’s performance was less than 80 percent correct. Failure to make the crossing response on a positive trial and making the crossing response on a negative trial were both scored as errors. The group’s percentage correct is displayed in the form of the labeling gradients and in the location of the phonetic boundaries for human and chinchilla subjects, respectively. The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2.

On half of the trials during generalization tests, the stimuli presented were those already mastered by the animals. On the other half of the trials, generalization stimuli were randomly presented. In this way, performance on familiar stimuli served as a control for performance on novel stimuli. When control stimuli were presented, all feedback previously in effect was maintained. When generalization stimuli were presented, neither shock nor the buzzer were used, and all other feedback was arranged to indicate a correct response. Four English-speaking adults identified the same stimuli as /da/ or /ta/. The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2.

The experiment began with a single /tu/ and /du/ syllable; when this task was mastered, variation in the form of tokens (second repetition by the same talker), talkers (identical phonetic token by different talkers), and vowels (different vowel contexts by a single talker) were singly introduced (13). The criteria for progressing from one condition to the next were that the group’s percentage correct remain ≥ 90 percent for two consecutive days and that during this time no single animal’s performance was less than 80 percent correct. Failure to make the crossing response on a positive trial and making the crossing response on a negative trial were both scored as errors. The group’s percentage correct is displayed in the form of the labeling gradients and in the location of the phonetic boundaries for human and chinchilla subjects, respectively. The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2.

The experiment began with a single /tu/ and /du/ syllable; when this task was mastered, variation in the form of tokens (second repetition by the same talker), talkers (identical phonetic token by different talkers), and vowels (different vowel contexts by a single talker) were singly introduced (13). The criteria for progressing from one condition to the next were that the group’s percentage correct remain ≥ 90 percent for two consecutive days and that during this time no single animal’s performance was less than 80 percent correct. Failure to make the crossing response on a positive trial and making the crossing response on a negative trial were both scored as errors. The group’s percentage correct is displayed in the form of the labeling gradients and in the location of the phonetic boundaries for human and chinchilla subjects, respectively. The mean percentage of “d” responses to each synthetic stimulus by human and chinchilla subjects is displayed in Fig. 2.
human processing is not essential for these particular tasks, and they suggest that alternative explanations merit further consideration. The fact that the chinchillas respond to the synthetic speech as though an abrupt qualitative change occurs in the short voicing-lag region of the VOT continuum at precisely the place where many languages separate two phonemic categories lends support to the idea that speech-sound oppositions were selected to be highly distinctive to the auditory system. By this reasoning, one might infer that there is at least one other natural psychophysical boundary, located in the voicing-lead region of the VOT continuum, which serves as a basis for the phonemic distinction between prevocal and voiceless unaspirated plosives of languages such as Spanish, Thai, and Kikuyu. In any case, further experiments with animals should help to pinpoint which speech-perception tasks require “special processing.”

PATRICIA K. KUHL
JAMES D. MILLER
Central Institute for the Deaf,
St. Louis, Missouri 63110

References and Notes
2. The plosive consonant /k/ is often used as an example. The direction of the second formant (F₂) transition is believed to be a sufficient cue for differentiating place of articulation (/b, d, g/; that is, in the /l/ vowel context, rising F₂ transitions produce /dar/; and those remaining fairly steady produce /dɔr/). However, the set of F₂ transitions that produce a single plosive, like /d/, across vowel contexts varies extensively, in /dɔr/ the F₂ transition rises, in /dɔr/ it falls, and in /dɔr/ it remains fairly steady.
10. A new variation was introduced by adding only the positive stimulus, then only the negative stimulus, and finally both positive and negative stimuli until performance criteria were met (C. K. Burdick and J. D. Miller, ibid. 58, 415 (1975)).
11. Beginning with training on individual vowels, all positive and negative stimuli for a new condition were added simultaneously.
12. Supported by NIH grant NS03856 to Central Institute for the Deaf and NIH grant RR00396 to the Biomedical Computer Laboratory of Washington University. We thank A. M. Engedreth, C. K. Burdick, and R. J. Dooling for their assistance, and we acknowledge the cooperation of the Haskins Laboratories (NIH contract NIH-71-2420) in providing the synthetic stimuli.

January 1975; revised 22 April 1975

Ventricular Obstruction: Effect on Drinking Induced by Intracranial Injection of Angiotensin

Abstract. Lesions of the subformical organ (SFO) severely attenuate drinking induced by injections of angiotensin II into the lateral ventricles, but a few days (4 to 14) later a recovery of the drinking response is observed. A possible explanation for this is that other dopsogenic sites are involved which are beyond the interventricular foramen and that SFO lesions produce an obstruction by edema or debris at the foramen which blocks access of cerebrospinal fluid-borne angiotensin to those sites. This hypothesis is supported by tracer studies and by direct injection into the third ventricle of SFO-lesioned animals. Other studies reported implicate the anteroventral third ventricle as a likely site for angiotensin receptors.

Angiotensin has been shown to induce drinking when administered systemically or intracranially (1), but its site, or sites, and the mechanism of action remain in question. Early studies of intracranial mapping implicated a diffuse number of tissue sites (2). However, a more recent study indicates such data need reevaluation, since mapping with low doses of angiotensin revealed that the critical aspect of the intracranial injection was that the peptide must gain access to the brain ventricular system (3). Other evidence supporting ventricular involvement includes the finding that the ventricles and peri-ventricular tissue, but not other brain regions, show a rapid accumulation of radioactivity after systemic injection of labeled angiotensin (4). These results suggest that peripheral angiotensin gains access to peri-ventricular sites which are important central nervous structures for the dopsogenic response to angiotensin.

Recently Simpson and Routtenberg (5) have suggested that, in the rat, the subformical organ (SFO) contains the exclusive receptors for angiotensin-induced drinking. The SFO is one of several highly vascularized structures which lies outside the blood-brain barrier, and it protrudes into the third ventricle adjacent to the interventricular foramen (6). Simpson and Routtenberg observed a cessation of drinking in response to angiotensin injections in the basal telencephalon after lesions of the SFO. In two different laboratories, one at the University of Iowa and one at the University of Pittsburgh, we have independently investigated the role of the SFO in the mediation of drinking behavior (7); our data do not support the conclusion (5) that SFO is the exclusive receptor site for the dopsogenic action of angiotensin.

Both of our groups found independently that SFO lesions blocked drinking to lateral ventricle (LV) injections of angiotensin II on the first test after the lesion, but with subsequent tests (4 to 14 days later), a recovery of elicited drinking behavior was regularly observed. Adult male rats were implanted with 23-gauge guide cannulae to ensure delivery of angiotensin to the LV (3) and were screened to establish the drinking response to the hormone after injection through a 30-gauge inner cannula. Then lesions aimed at SFO were made and animals were again tested for response to angiotensin in two tests after the lesion (8). In the Pittsburgh experiments both water and a 1.8 percent solution of NaCl were available (9), and tests were carried out during the dark phase of the 24-hour cycle. In the Iowa experiments, only a bottle of water was available and testing was done during the light phase.

Of 55 rats that passed the screening test and were lesioned, 22 showed severe deficits in drinking in response to angiotensin on the first test after the lesion (see Table 1). In the Pittsburgh study, 12 of 12; Iowa study, P < .01. Most of the animals with drinking deficits on the first test after the lesion sustained extensive SFO damage (19 of 22 had 70 to 100 percent destruction). It is not likely that recovery or enhancement of function of residual SFO tissue could explain recovery of the drinking response to angiotensin since recovery was noted in animals with complete as well as partial lesions. Furthermore, each laboratory identified one or more animals (a total of four) that failed to show any deficit in angiotensin-induced drinking after 70 to 100 percent destruction of SFO (10).

The Pittsburgh group initiated research for brain sites from which the dopsogenic response to angiotensin did not vary as a function of SFO lesions. Pilot work in...