DEVELOPMENTAL SPEECH PERCEPTION

by

JANET FELDMAN WERKER

B.A., Harvard University, 1974

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES
(Department of Psychology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
December, 1978

© Janet Feldman Werker, 1978
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date December 19, 1978
ABSTRACT

Previous research has indicated that infants have the ability to categorically discriminate many of the distinctive features of speech sounds regardless of their exposure to a language in which such distinctions are important, whereas adults of one language group may have difficulty discriminating linguistic features that are important in a foreign language. This suggests a decline during development in linguistic perceptual abilities, during which the ability to discriminate non-relevant features may be lost. This study was designed to be the first in a series of tests of such a decline and involved comparing English-speaking adults, Hindi speaking adults, and six-month old infants on their ability to discriminate foreign and native speech contrasts.

Two pairs of Hindi sounds, and one pair of English sounds were investigated in this study. Infants were tested in a "visually reinforced infant speech discrimination" (VRISD) paradigm. This is a discrimination paradigm in which the infant is conditioned to turn its head when there is a change in the auditory stimulus. A variate of this paradigm was employed for the adult subjects using a button-press, rather than a head turn, as the critical behavioral response.

This study yielded support for the notion that infants have the ability to categorically discriminate distinctive features of speech sounds regardless of exposure to a language in which such distinctions are important. Some support was also given for the idea that there may be a decline in speech perceptual abilities with either age/or linguistic experience, but this finding was only significant with one of the two Hindi sound pairs.
Two explanations are offered for these results. It is suggested that the English population may have had some experience with the non-significant Hindi contrast. Alternatively, it is suggested that differential perceptual distance may have accounted for the differences between the two Hindi sound pairs. Two levels of processing, acoustic and linguistic, are then invoked in this explanation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF APPENDIX A TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF APPENDIX A FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td>19</td>
</tr>
<tr>
<td>RESULTS</td>
<td>29</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>36</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>44</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>45</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>53</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>58</td>
</tr>
<tr>
<td>TABLE</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>TABLE 1</td>
<td>Hindi Stop Consonants</td>
</tr>
<tr>
<td>TABLE 2</td>
<td>Fricative Discrimination by Age</td>
</tr>
<tr>
<td>TABLE 3</td>
<td>Analysis of Proportions for Retroflex/Dental Contrast</td>
</tr>
<tr>
<td>TABLE 4</td>
<td>Analysis of Proportions for Voiceless Aspirated/Breathy Voiced Contrast</td>
</tr>
<tr>
<td>TABLE 5</td>
<td>Multiple Comparisons on Retroflex/Dental Contrast</td>
</tr>
<tr>
<td>TABLE 6</td>
<td>Multiple Comparisons on Voiceless Aspirated/Breathy Voiced Contrast</td>
</tr>
<tr>
<td>TABLE 7</td>
<td>Infant Mean Number of Trials on Speech Contrasts</td>
</tr>
<tr>
<td>TABLE 8</td>
<td>Cell Means and Standard Deviations For Number of Trials to Criterion</td>
</tr>
<tr>
<td>TABLE 9</td>
<td>Source Table for Number of Trials to Criterion; Analysis of Variance</td>
</tr>
<tr>
<td>TABLE 1</td>
<td>The Prime Features</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1</td>
<td>Mean Number of Sucks Per Minute</td>
<td>17</td>
</tr>
<tr>
<td>FIGURE 2</td>
<td>Arrangement of the Experimental Site</td>
<td>21</td>
</tr>
<tr>
<td>FIGURE 3</td>
<td>Infant During Control Trial</td>
<td>22</td>
</tr>
<tr>
<td>FIGURE 4</td>
<td>Infant During Experimental Trial</td>
<td>23</td>
</tr>
</tbody>
</table>
# LIST OF APPENDIX A FIGURES

<table>
<thead>
<tr>
<th>FIGURE 1</th>
<th>Overlap of Articulatory Features</th>
<th>.......</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 2</td>
<td>Abbreviated &quot;d&quot; Spectograms</td>
<td>.......</td>
<td>54</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to give special thanks to Dr. Richard Tees, Dr. John Gilbert, and Keith Humphrey for giving so much assistance and encouragement throughout the project. Thanks are also due to Dr. Robert Frender and Dr. Tannis MacBeth Williams for their willing assistance during critical stages of this project.
INTRODUCTION

This study was designed to be a first step in testing the following hypothesis:

Humans are born with the ability to discriminate the universal set of distinctive linguistic features. Through development, there is a decline in this ability as the perceptual space becomes organized to approximate the phonemic categories used in the native language.

This idea (that some perceptual capabilities decline with age) conflicts with most approaches to perceptual development. According to Gibson (1967) for example, perceptual capacity becomes increasingly differentiated and refined through ontogeny. Perceptual learning results in an increase in ability to acquire information from the environment as a result of experience with the array of stimuli provided by the environment. Thus only those stimuli present in the environment lead to refinements in perceptual capacity. Additionally, perceptual learning is seen as an increase, rather than a decrease, in discriminatory ability.

Other contemporary theories of perceptual development also support the notion of increasing differentiation as a result of experience. For example Hebb and Piaget would both view perceptual development as reflecting the construction of "schemata" as organizations of the elements of sensory input. The schema is then capable of being modified with further experience and is thought to influence the perception of objects by "guiding" the individual to certain details and by aiding in classifying and understanding the sensory input (e.g., cf. Tees, 1976). Somewhat similarly, perception has been regarded as a process of category building.
and inference drawing (Bruner, 1957). As more categories are built, perception becomes less dependent upon the details of the sensory input, and more on the categories. Through development the categories become more accessible so that fewer details are necessary for classifying an input.

On the other hand, behaviorists such as Miller (1948) view perceptual development as resulting from positive feedback from the environment. A chance response to an appropriate stimulus results in positive feedback, and this motivates the individual to discriminate that stimulus from others a second time, etc.

Although these theories deal adequately with many facets of perception, they do not deal directly with discrimination. According to S.S. Stevens (1939), discrimination is the fundamental cognitive act. The ability to discriminate the fundamental features of speech is the necessary first step in language learning.

On the basis of elementary discriminations, then, we make our first rudimentary classes and in doing so we have the first step toward generalization. (Sevens, 1939, p. 58).

One, perhaps unique aspect of speech perception (1) is the accumulation of evidence suggesting that "categorization" of speech sounds is an innate propensity of the human organism. That is, rather than having to learn to differentiate phonetic features and then organize them into categories, human infants seem to respond from the beginning, to speech sounds according to phonemic categories. This kind of ability would enable the infant to break the continuous auditory input of speech into discrete segments, thus starting the process of the eventual identification of meaningful segments. For example, when given synthesized speech sounds varying along

---

(1) Before proceeding with this section it is necessary to understand the basic research in experimental psychology concerning the decoding of speech sounds and the reality of distinctive features. Interested readers are referred to Appendix A.
In each natural language only a subset of the universal set of phonetic distinctions is utilized to differentiate meaningful sound segments (phonemes). In an individual language, phoneme categories may broaden to include more than a single phonetic specification (allophonic variation). For example, in English, aspiration does not connote differences in meaning. Thus if an English speaker pronounced "pill" with an aspirated or an unaspirated voiceless stop consonant, other English speakers would understand the word to be "pill" and would ignore the irrelevant phonetic variations. It can be seen then, that in terms of learning one's own language, perceptual development can be viewed as a process of grouping initially distinct speech sounds into broader categories, thereby establishing the required set or class of phonemes relevant to one's native language.

In addition to allophonic variations, speakers of a particular language may omit a subset of phonetic vfeatures from use. There is some evidence that adult listeners have difficulty discriminating those features when hearing them for the first time in a linguistic context (Lisker & Abramson, 1968).

The significance of the discovery that young infants can discriminate some linguistically relevant distinctive features of speech does not necessarily imply a conscious recognition of speech sounds as language, although Eimas (1975) has stated that this discriminatory ability "indicates
that infants have some knowledge of the phonetic structure of language, specifically, they must have knowledge of at least some of the phonetic features." (p. 214). It is unfortunate, however, that at present there is no way to differentiate strictly acoustic from phonetic knowledge. In any event demonstration of discriminatory abilities that reflect phonemic features suggests that infants have at least the necessary first response capabilities to begin organizing their linguistic input into a finite set of categories. Whether the infant has knowledge of the phonetic structure becomes irrelevant. Michael Laine, in his *Introduction to Structuralism* (1970) stated that when people use their own language

"...they consistently and constantly apply its phonological laws (its structure, in other words) in their speech. They will not, unless they are versed in linguistics, be consciously aware of them. Nor, if asked, would they be able to supply those laws...... What the observer sees is not the structure, but simply the evidence and product of the structure. ...... There is in man an innate, genetically transmitted and determined mechanism that acts as a structuring force. Moreover, this inherent quality or capacity is so designed as to limit the possible range of ways of structuring. (Laine, 1970, p. 15)

Thus the response capacities of the human infant would limit the range of categories used for linguistic input.

The hypothesis that speech perceptual development may proceed from the universal to a more limited set of categories does not originate with this study. Similar speculations have been made by other researchers (Eimas, 1976; Gilbert, 1975; Pisoni, 1977, Trehub, 1976). It is necessary, however, to actually test whether or not infants can discriminate non-native speech contrasts with greater ease than adults in a single, properly designed study.
In order to examine this hypothesis, infants and adults were tested on two pairs of Hindi speech contrasts, and one pair of English speech contrasts in two very similar discrimination paradigms.

The English pair used was the common /ba//da/. These are both voiced stop consonants differing only in place of articulation (bilabial vs. apical). This is a relatively common distinction across natural languages and has been previously studied using infants.

The first Hindi pair was the unvoiced, unaspirated retroflex vs. dental stop /Ta//ta/ where place of articulation is the critical dimension. Retroflex consonants are produced by curling the tongue back and placing it posterior to the alveolar ridge. In contrast, dentals are produced by placing the tip or blade of the tongue against the back of the upper front teeth. Dental is a common place of production in English (and in most natural languages) whereas retroflection does not carry phonemic information in English. Additionally, retroflex consonants are uncommon across natural languages and their contextual distribution is restricted even in those languages in which they have import.

The second Hindi pair was the unvoiced aspirated, dental stop vs. the murmured dental stop /tʰ//dʰ/ in which a difference in voice onset time is the critical distinction. Voice onset time is a combination of the state of the glottis during an articulation and the presence or absence of a period of voicelessness during and after the release of an articulation. In English only two categories of voice onset time are differentiated—voiced (the vocal words are nearly together so that they vibrate) and voiceless (vocal cords are so far apart that they cannot vibrate at all). There are two additional categories in Hindi; voiceless unaspirated and
breathy voiced (murmured) stops. (See Table 1).

As can be seen, both pairs of Hindi sounds involve distinctions between categories that do not exist at all in English and are relatively uncommon to natural languages. Evidence has been provided (Singh & Black, 1966) that these pairs would thus be among the hardest for English speakers to discriminate. Additionally, S.E. Blumstein suggested that in gathering pilot data for the study of retroflex stop consonants (Stevens & Blumstein, 1975), English-speaking subjects did seem to demonstrate difficulty in discriminating the unvoiced-unaspirated dental vs. retroflex stop consonants (personal communication, February, 1978).

The discrimination paradigm used for the infants is called "visually reinforced speech discrimination" (VRISD). A variant of this paradigm was used for the adults. VRISD was first developed as a variant of the classic "play audiometry" of Dix and Hallpike (1947), in which a child was conditioned through association to push a button in response to a sound source to elicit the presentation of a pleasant picture. Suzuki and Ogiba (1960) modified the response to a simple head-turn in order to test the hearing ability of children under 3, calling it visual reinforcement audiometry (VRA). Although Suzuki and Ogiba (1961) reported low success rates for infants one year and younger, more recent work (Wilson, Lee, Owen & Moore, unpublished manuscript) indicated that 90% of infants as young as age 5-1/2 months can be tested with this paradigm. More recently researchers have begun using this paradigm in speech discrimination studies (Eilers & Minifie, 1975; Eilers, Wilson & Moore, 1977; Hillenbrand, Minifie & Edwards, 1977; Kuhl, 1976, 1977; Minifie, 1976). In these studies the infant was conditioned to turn his/her head toward a sound source when he/she detected
<table>
<thead>
<tr>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
<th>Voiced</th>
<th>Breathy voiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>bilabial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pal</td>
<td>p\textsuperscript{h}al</td>
<td>bal</td>
<td>h\textsuperscript{h}al</td>
</tr>
<tr>
<td>(take care of)</td>
<td>(edge of knife)</td>
<td>(hair)</td>
<td>(forehead)</td>
</tr>
<tr>
<td>dental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t\textsubscript{\text WCHAR{5C}}\textsubscript{\text WCHAR{5E}}an\textsuperscript{*}</td>
<td>h\textsuperscript{an}**</td>
<td>d\textsubscript{\text WCHAR{5C}}an</td>
<td>d\textsubscript{\text WCHAR{5C}}\textsuperscript{h}an**</td>
</tr>
<tr>
<td>(mode of singing)</td>
<td>(roll of cloth)</td>
<td>(charity)</td>
<td>(paddy)</td>
</tr>
<tr>
<td>retroflex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t\textsubscript{\text WCHAR{5C}}al\textsuperscript{*}</td>
<td>t\textsuperscript{h}al</td>
<td>d\textsubscript{\text WCHAR{5C}}al</td>
<td>d\textsubscript{\text WCHAR{5C}}\textsuperscript{h}al</td>
</tr>
<tr>
<td>(postpone)</td>
<td>(place for buying wood)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-alveolar affricate</td>
<td>t\textsubscript{\text WCHAR{5C}}al</td>
<td>t\textsubscript{\text WCHAR{5C}}\textsuperscript{h}al</td>
<td>d\textsubscript{\text WCHAR{5C}}al</td>
</tr>
<tr>
<td>(g\textsubscript{\text WCHAR{5E}})</td>
<td>(deceit)</td>
<td>(water)</td>
<td>(glimmer)</td>
</tr>
<tr>
<td>velar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kan</td>
<td>k\textsubscript{an}</td>
<td>gan</td>
<td>g\textsubscript{\text WCHAR{5C}}\textsubscript{\text WCHAR{5E}}an</td>
</tr>
<tr>
<td>(ear)</td>
<td>(mine)</td>
<td>(song)</td>
<td>(kind of bundle)</td>
</tr>
</tbody>
</table>

\* = comparison Retroflex/Dental

\*\* = comparison Voiceless Asp/Breathy Voiced
a change in the stimulus, and was rewarded with the appearance of a flashing, noise-producing toy animal. Not only could infants as young as five months perform successfully in this paradigm, but they also show the ability to transfer this "learning set" to new stimulus pairs (Kuhl, personal communication).

Infant Research

Although speech researchers are united in their belief that by age 2 months infants demonstrate a sophisticated ability to make phonetic discriminations, (Eimas, 1974, 1975a,b; Eilers & Minifie, 1975; Kuhl, 1976; Moffit, 1971; Trehub, 1973, 1976; Williams, in preparation), the meaning of this discriminatory ability is debated. According to the acoustic cue hypothesis, these apparent "phonetic discriminations" are based on lower order auditory discriminations and are viewed as no different from non-speech auditory perception (Stevens & Klatt, 1974). In contrast, proponents of the linguistic cue hypothesis claim that speech sounds are heard specifically as speech (Eimas, 1974; Gilbert, 1971; Moffit, 1971; Morse, 1976; Trehub, 1973), and that speech discrimination is an innate biological ability important in the process of language acquisition, (Eimas, Siqueland, Jusczyk & Vigorito, 1971). Additional elements of the linguistic cue hypothesis include the following: (1) perception of consonant-vowel sounds is categorical and linguistically relevant, (2) speech perception is non-developmental, i.e., infants respond to phonetic stimuli in the same way as adults, (3) non-speech sounds are not perceived categorically and (4) humans are the only species that can perceive speech sounds categorically.

While there is some evidence to suggest that infants and adults can perceive some nonspeech sounds categorically (Jusczyk, Rosner & Cutting,
that other animals can show categorical perception for phonetic features (Kuhl & Miller, 1975), and that within category perception is possible under extreme conditions (Samuel, 1977), the critical issues for the present study are whether or not perception of consonant-vowel sounds is organized according to phonemic categories and whether infants can respond to phonetic stimuli in the same way as adults.

Several studies have already been undertaken to examine these two critical issues. Two measures have been used extensively in these studies; heart rate deceleration (HRD) and high amplitude sucking (HAS) (both have been discussed extensively by Morse (1974). HRD is based on the assumption that heart rate decelerates in response to a novel stimulus. In contrast HAS involves a behavioral response by the infant (sucking). Both HRD and HAS are used in a habituation-dishabituation design, contingent upon reinforcing properties of the speech stimuli. Both have the following weaknesses. (1) It is difficult to interpret negative results. It is never known whether the child was unable to make a discrimination or did not find the stimuli intrinsically reinforcing. (2) It is impossible to compare data across individuals. (3) There is about a 65% rate of incompletion, that is, data can only be reported on about 35% of the sample. In addition, the HAS paradigm can only be used with infants up to 12 months of age. The HRD can be further criticized for its sensitivity to both the state of the infant and the characteristics of the speech stimuli (Morse, 1974).

In spite of the problems Eimas et al. (1971), using a HAS paradigm, were able to provide answers relating to the fundamental questions. For example, in an experiment in which voice onset time (VOT) was varied by (2) These findings are all taken as evidence that there is both a linguistic and an acoustic stage in processing, but do not negate the psychological reality of categorical perception in a linguistic context.
Intervals of 20 msec, infants aged 1-4 months were shown to discriminate between the voiced and voiceless stop consonants /b/ and /p/ along the same phonemic boundary as adults; that is, they showed both categorical discrimination and linguistically relevant perception. In addition to categorical discrimination along the VOT continuum, carefully designed follow-up studies have demonstrated that infants can categorically discriminate the /r/-/l/ distinction (Eimas, 1975b), stop consonants in final position (Eimas, 1974), place of articulation (Jusczyk, 1977a), most of the fricatives (Eilers & Minifie, 1975), some vowels (Trehub, 1973), glides in initial and medial position (Jusczyk, Copan & Thompson, 1977), and nearly every other feature on which they have been tested.

Not all the available evidence is supportive of the central hypothesis of this study. There is some suggestion that fricative pairs may be difficult for young infants to discriminate (Eilers & Minifie, 1975; Eilers, Wilson & Moore, 1977). Using visually reinforced infant speech discrimination (VRISD), infants aged 6-8 and 12-14 months of age were compared on their ability to discriminate natural pairs of both Fi/θi and Fu/θa. Their results are summarized in Table 2. A '+' signifies the infants could make the discrimination, and a '-' signifies that they could not.

<table>
<thead>
<tr>
<th>Fricative Discrimination by Age</th>
<th>6-8 mos.</th>
<th>12-14 mos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fi/θi</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Fu/θa</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Eilers suggests that these findings show that "improvement" does occur with age in fricative discrimination because the 6-8 month infants failed both pairs, while the 12-14 month infants passed the F1/θ1 discrimination.

Eilers et al. (1977) concluded that their findings call into question the universality of phonemic discriminatory abilities in young infants and suggest instead a more complex picture of perceptual development including both auditory capacity and learning.

This "negative" evidence is not unchallenged, however. Using the same paradigm (VR1SD), Holmberg, Morgan and Kuhl (1977) have found evidence of Fa/θa discrimination in a study of 6-month old infants. Holmberg et al. suggest experimental conditions may account for the difference between their results and those obtained by Eilers et al. (1977) in that the latter required that infants show evidence of discrimination within only six trials. Failure to do so was interpreted as inability to make the discrimination. Holmberg et al. claim that state fluctuation and variations in attentiveness may make the six trial requirement too stringent for younger infants. In any event, it should also be pointed out that fricatives have been shown to be difficult to discriminate whether in natural or synthesized forms in the case of the adults (Miller & Nicely, 1955) and children (Abbs & Minifie, 1969).

Proponents of the acoustic cue hypothesis interpret findings such as those presented by Eilers et al. (1977) as evidence that discrimination is purely auditory, not linguistically relevant and thus subject to learning. They suggest that the categorical discrimination of consonants is an artifact of the particular stimuli used. For example, in VOT studies, it has been suggested that the crucial cue is the presence or absence of the first
formant transition rather than the relative amount of voicing (e.g., Stevens & Klatt, 1974). However, Moffit (unpublished) investigated this specific question in a study of categorical perception of bilabial stop consonants. Using four pairs of synthetically produced stimuli, two pairs differing multidimensionally and two pairs differing in voicing only, Moffit found that infants aged 30-60 days were able to make linguistically relevant discriminations without the acoustic cue of presence or absence of the first formant transition.

Kuhl (personal communication) is investigating the specific dimensions of speech discriminative abilities in infants using a paradigm which is relatively new to infant speech perception research, the perceptual constancy paradigm (similar to VRISD), visually reinforced infant speech discrimination. She has demonstrated that infants can more quickly learn to organize their linguistic categories according to phonetic features while disregarding variations in both speaker and intonation than they can learn to organize categories according to these latter dimensions disregarding phonetic features.

Further support for the two critical issues of the linguistic hypothesis (as outlined previously) is found in cross-cultural studies. Babies of Kikuyu, Spanish and English speaking parents have been shown to discriminate linguistic features that are not relevant in their native languages (Lasky, Lasky & Klein, 1975; Streeter, 1976; Trehub, 1973, 1976).

Adult Research

In contrast to the results of research with infants, studies of adult speech discrimination have resulted in observations that adults have difficulty discriminating phonemes that are not used in their native language.
Japanese adults have been shown to have difficulty discriminating the English /r/-/l/. In a study by Miyawaki, Strange, Verberge, Liberman & Fujimura (1975), 21 Japanese and 39 United States adults were tested on a discrimination task using synthesized versions of /ra/ and /la/ in which the critical perceptual cue was starting frequency and transition of the third formant (F3). The results for identification and discrimination tasks with mature English speaking subjects from the United States, showed their perception to be nearly categorical. On the other hand, native-Japanese speaking adults were able to discriminate only slightly above the level of chance. A comparison of these same two groups on discrimination of non-speech counterparts of /ra/ and /la/ (F3) same as in speech token, but F1 and F2 amplitudes set at zero) yielded a different result. Both groups showed equally successful discrimination of all comparison pairs. Miyawaki et al. concluded that experience with language affects linguistic, but not acoustic perceptual abilities.

Studies with English speaking adult subjects have shown that they experience difficulty in discriminating test tokens that span a non-English lead boundary in VOT (Lisker & Abramson, 1968). Discrimination data for adult listeners (Lisker & Abramson, 1970) shows that Spanish speakers can only distinguish between two categories of voicing. In contrast, data presented by Lasky, Lasky and Klein (1975), indicate that infants being reared in Spanish speaking environments can discriminate three voicing distinctions. Additionally, Trehub (1976) has suggested that adults have difficulty discriminating the distinctive feature, + stridency, as exemplified in, for example, the Czech /za/ /ra/.
It has been suggested that since English speakers have only two categories for prevocalic stop consonants they may lose the ability to discriminate sounds involving three and four categories. (Ladefoged, 1975). For example, the Eastern Armenian language has three categories for prevocalic stops, and Hindi has four categories. Ladefoged's suggestion has been supported in the literature. For example, Singh and Black (1966) tested adult Japanese, Hindi, English and Arabic speakers less than a month after they arrived in the United States. After being trained for one hour on a series of consonant-vowel sounds, the subjects were asked to identify those same sounds in writing. All listeners identified the sounds of their native language best. The Hindi stimuli used in the present study were particularly difficult for non-Hindi speakers to identify.

It should be noted, however, that memory requirements may confound these results. Identification paradigms (such as used by Singh and Black) involve the use of short-term memory for labels and are thus inadequate for testing pure discrimination abilities. Since many of the cross-language adult findings are based on identification studies they could be misleading. In the studies reported by Trehub (1976) and Miyawaki et al., (1975) discrimination paradigms were used, and the adults still demonstrated difficulty with the non-English discriminations.

It can be argued that results of tests of perceptual abilities can be misleading under conditions of low motivation because subjects will be inclined to use "everyday" categories. With greater motivation, however, subjects may be able to make finer discrimination (see Brown & Lenneberg, 1958). Although many studies in adult speech perception have ignored this problem, research in psychophysical acoustics suggests that motivation is
critical. Samuel (1977) has shown that with training (and thus feedback),
adults may be able to discriminate within category speech stimuli with
short onset time. Although he cites this as evidence for perception in both
a phonetic (categorical) and acoustic (continuous) mode, it may be interpre­
ted as evidence for the importance of motivation.

In summary, a review of the literature suggests that infants may have
greater discriminative speech abilities than adults and that experience
with language may narrow discrimination abilities. To date, however, this
question has not been examined properly in a single study with identical
tokens and with procedures that allow for comparable individual data.

Infant/Adult Research

Trehub (1976) described a series of four different experiments that
were grouped for analysis that most closely address the hypothesis of this
study in that she tried to compare results of infant and adult experiments.
In experiments I and II, English-Canadian infants aged 5-17 weeks were
tested in a HAS paradigm for their ability to discriminate the oral nasal
vowel distinction which occurs only in French and Polish /pa//[^ba]/, and the
distinctive feature of stridency exemplified by the Czech /za//[^ra]/. The
group results suggested that infants could discriminate these features
(see Figure I). Trehub then employed a change-no change paradigm to deter­
mine whether English-speaking adults could discriminate the Czech pair. A
signal detection analysis implied substantial confusion for the two Czech
sounds (d'= .83 compared with d'=1.00 for the common English pair /ba//da/).
Trehub then compared the adult data to the infant data of Experiment II,
and to some 1972 /ba//da/ infant data. The two sets of infant data were
analyzed by an analysis of variance (group x the sucking rate in the 5 post­
decrement minutes x language). There was no significant interaction between
experimental and control groups and language suggesting equal discrimination abilities for English and Czech contrasts. When compared to the adult data this suggests English infants have greater discrimination abilities than the adults. As Trehub herself points out; however, the use of different paradigms in different labs makes such comparisons only speculative. Additionally, the infant findings were based on group (HAS) data; responses had to be averaged over the postdecrement time period whereas adult results were based on discrete individual data (although grouped for analysis).

Potentially relevant findings have also been reported very briefly by Bower (1977). When listening to the speech of their own language, adults perform a "dance" of subtle body movement which reflects the discrete units of the continuous speech input (Condon & Ogsten, 1971). Without giving any experimental details, Bower claims that while adults perform this "dance" only in response to their native language, infants "dance" to the speech of any language.

Indirect evidence supporting the hypothesis that infants may be better able to discriminate non-native sounds than adults is provided by studies of second-language learning in children. The biological argument of the strict critical-period hypothesis as put forth by Penfield and Roberts (1959) and by Lenneberg (1967) suggests that the critical period for language acquisition lasts from about age two until puberty, and is due to the lack of complete hemispheric specialization. Although this hypothesis has been criticized from the point of view of age of lateralization (Kimura, 1967) and on the basis of Burstall's hypothesis of a later optimum age for language learning (in McLaughlin, 1977), the evidence for a critical period
Fig. 1 — Mean number of sucks per minute, as a percentage of the maximum predecrement sucking rate, for 5 minutes before and after the decrement criterion.
with respect to accent is strong. It has been found that the younger the child, the more perfect the pronunciation in learning a second language (Asher & Garcia, 1969). Additionally, case studies of bilingual children reveal that younger children seem to do better on the acquisition of phonological features (McLaughlin, 1977), demonstrating a greater perceptual flexibility in younger than in older children. Finally, recent speech perception data of adult bilinguals suggests their perception of acoustic continua reflects the phonetic categories of both languages (Carramazza, Yeni-Komshian, Zurit & Carbone, 1973; Williams, 1975).

The hypothesis examined in this study is actually intermediate to the linguistic and the acoustic cue hypothesis. According to the present hypothesis, the infant has the ability to respond to the universal set of phonemic distinctions at birth. Whether this ability is purely acoustic or linguistic is not important. What is important is that since the infants can discriminate the universal set of linguistic features, they are able to break the continuous flow of speech input into its discrete features, and thus begin segmenting the meaningful aspects of linguistic input. This predisposes infants to acquire their native language with greater ease than if they had to learn to discriminate the linguistic features. The purpose of this study was to test the hypothesis that through development the individuals begin to organize their speech perception categories to more and more closely approximate the phonetic categories used in their own language. This ability may be purely auditory at birth and become encoded linguistically through ontogeny, or it may be a specific language ability at birth, as some neuropsychological (Glanville, Levenson & Best, 1977; Kimura, 1967) and anatomical (Witelson & Paillie, 1973) evidence suggests.
METHOD

Subjects

Four groups of volunteer subjects were tested in a discrimination paradigm (described in the next section) on two pairs of Hindi speech contrasts.

Group I consisted of five Hindi speaking adults (three males, two females) aged 22-35. The subjects were recruited through advertising and word-of-mouth from the Vancouver community.

Group II consisted of fifteen infants (six males, nine females) ranging in age from 6 months, 7 days to 7 months, 23 days, with an average age of 6 months, 28 days. The infants were recruited from the community by advertising in newspapers, at well-baby clinics, and by telephoning people listed in birth announcements. Although English was the principle language spoken in all the infant homes, additional languages were spoken in four of the homes.

Groups III and IV each consisted of ten English-speaking adults (six males, four females in each group) aged 22-35 recruited from the University of British Columbia campus. As it was difficult to find adults with no foreign language training, notes were taken on formal training and informal exposure for each adult. Group III consisted of "naive" adults, whereas Group IV was given limited feedback in the discrimination paradigm to make their task more comparable to the infant task.

Procedure and Apparatus

The procedure and apparatus for the testing of the infants is outlined in some detail as it is not widely available in the literature. This procedure is called visually reinforced infant speech discrimination (VRISD).
As the adult procedure was designed to approximate this infant procedure as nearly as possible, it will only be described briefly.

In the present study, the experimental set-up consisted of a sound-attenuated room with a one-way observation window adjoining the control room in which the experimenter and logic system were situated. The sound attenuated room contained one small table, in the middle foreground of the room. A chair was directly behind the table on which the parent and infant were seated with their backs to the experimenter. A second chair was located across the table and slightly to the left from the parent/infant on which the assistant was seated. A speaker was located in the back right corner of the room, and a visual reinforcer was located at a 45 degree angle from the child's left side (See Figure 2). To avoid influencing the infant, the assistant and parent both wore sound attenuating earphones (see Figures 3 & 4).

The visual reinforcer was an electrically activated toy animal contained in a smoked plexiglass box. The smoked glass made it possible to see the animal only when the visual reinforcer was activated and lights inside the box came on. Activation of the reinforcer also made the animal start moving and producing noise (toy bear drums, chimpanzee claps cymbals).

In the VRISD system, the experimenter presented a sound which did or did not change. The experimenter and assistant both voted as to whether the infant responded to a change. If both voted that the infant had responded within the required interval, a reinforcer was presented to the infant.

More specifically, the entire VRISD system was controlled by a logic system to ensure maximum experimental control in the test paradigm throughout the test period. Sound I (e.g., ba) was first played through the speaker at 2 sec. intervals. When the experimenter activated the logic system, the
FIGURE 2
Arrangement of the Experimental Site

A = Assistant
S = Speaker
VR = Visual Reinforcer
I = Infant
P = Parent
E = Experimenter
FIGURE 3. Infant Orientation During Control Trial
FIGURE 4. Infant Orientation During Experimental Trial
vote button held by the assistant lit up to indicate that a response interval was beginning. The experimenter then selected either track A (control track, no change in stimuli) or track B (experimental track, in which the stimuli changes, e.g., to 4 tokens of da) according to a predetermined randomized schedule. An adjustable timer was set at 4-1/2 sec and activated. During this 4-1/2 sec interval, if the infant turned its head toward the sound source, and if both the experimenter and the assistant independently pressed their vote buttons the discrimination was assessed to have been made, and the visual reinforcer was activated for 4 sec. It was necessary that both the experimenter and the assistant vote within the specified time interval, and that track B had been chosen for the visual reinforcer to be activated. If any of those conditions were not met, the reinforcer did not come on. (A manual override was also included in the system which made it possible to activate the visual reinforcer at any time to allow for flexibility during the conditioning stages of the VRA paradigm).

The VRISD procedure began with a request to the mother to sit behind the table with the infant on her lap. The mother was given headphones delivering music to prevent her from hearing the speech stimuli. The assistant (who was also wearing headphones) sat across the table to the child's right side and showed the child a series of toys to keep him/her happy and occupied during the testing session. The assistant's role was extremely important to the paradigm, as it was necessary to keep the child passively interested in the toys, but not so interested that he/she would disregard changes in the speech stimuli. When the infant was attentive and calm, with his head facing the toys, the experimenter began a trial by selecting either "start" track A (control) or track B (change) according to a prede-
terminated random schedule.

The conditioning portion of the paradigm proceeded as follows. A series of one syllable sounds (i.e., Ba) was played over the speaker. When the assistant indicated the infant was in a state of readiness (by pressing a silent foot button to activate a light that the experimenter could see), the experimenter changed to Track 2, and Da was played over the speaker. Immediately following the first token of Da, the toy animal was activated. Upon activation, lights came on in the plexiglas box, and the toy animal started performing. The noise of the animal attracted the infant's attention, and a head turn response was made to see the toy animal (see Figures 3 and 4). This procedure of presenting the sound stimulus and activating the reinforcer was repeated for 2 to 3 more trials. The experimenter then waited until after the second token of "Da" before activating the reinforcer to give the infant the additional seconds to respond. If no head turn response was made, the entire procedure was repeated. As soon as one head turn response was made upon presentation of the sound change before activation of the reinforcer, activation of the toy animal became contingent upon the infant making a head turn within 4-1/2 seconds after a changed stimulus, and the paradigm was taken over by the logic system. Sixty-eight percent of the infants formed the association within an average of nine conditioning trials.

Two to four sessions were required to complete the testing of each infant. Parents were requested to bring their infants on days when they had no evidence of colds or ear infections. Care was taken to ensure that each infant was comfortable in the experimental room before testing began, and observations were made on the infants ability to respond to a sudden onset of sound (they all did). On the first day of resting the infant was condi-
tioned to one or the other of the Hindi sound pairs chosen on a random-
ized basis.\(^{(3)}\)

Testing on the other Hindi sound pair was then completed on day 1, or one of the next three testing sessions. Infants that did not condition on the first day were given a second day of training. If they did not begin to condition within the first five minutes of testing on the second day, they were switched to the common English pair /Ba//Da/. If they did not learn with that pair during the session, they were not continued in the study. (Mothers would not continue bringing a baby out to an experi-
ment their baby would not succeed at). Failure to respond to a new sound change was followed in the same session by retesting the infant on a pre-
viously tested contrast. This was done to try to determine whether the sound pair, or the state of the infant, was responsible for the negative results.

Procedures as similar as possible to these were used with the adult subjects. The experimental set-up was nearly identical, with the subject sitting at the table facing the loud speaker. When the subject indicated readiness, the experimenter would begin the testing session. The subject's task was to push the vote button on a "vote" box when a change in stimulus was detected. The visual reinforcer contained the toy animal, and was ac-
tivated upon a correct vote in an experimental trial.

The criterion for discrimination was set at 8 out of 10 correct re-
sponses for change trials. The mode of presenting control trials varied

\(^{(3)}\)In the initial design, we had planned to condition infants on /Ba//Da/ and then transfer them to the Hindi pairs. Several problems arose. Most importantly, our mothers, tired of making the long trip to UBC would make up reasons not to return after the 3rd or 4th session. As it was primarily important to have a within group comparison on the 2 pairs of Hindi speech sounds, data was therefore collected on /Ba//Da/ only when possible.
slightly between the infants and the adults. As indicated earlier, the assistant indicated the infant's readiness by silently activating a light for the experimenter to see. During the period the light was activated, all head turn responses (including false positives) were recorded. At least four times during each series of ten change trials, the experimenter would not change the signal. During the other times the signal would be changed between 1 and 8 sec. after activation of the "readiness" light in a random fashion. The assistant never knew if a change trial would occur and would never know when in the observation period that change would occur. The actual criterion was therefore at least 12 out of 14.

For adults, change trials occurred in an irregular fashion ranging from 8 to 30 secs. All false positives were counted, so again the 8 out of 10 criterion is not an accurate indicator of the level of performance since guessing could have occurred after every token (every two sec). The criterion for infants and adults is defined in terms of experimenter delimited observation periods, rather than in absolute performance.

A second group of English adults was tested with limited feedback to provide a procedure more comparable to that of the infants. In this group, subjects were run, as in a conditioning paradigm, with the sound change paired with activation of the reinforcer for the first X trials (where X was the average number of trials to criterion for the infants). All adults were tested in one session.

Stimuli

Three stimulus pairs, one English /Ba//Da/, and two Hindi /ta//Ta/ and/tʰ//dʰ/, were used. The vowel (a) was used for all stimuli as it is common in both Hindi and English and is one of the most frequent vowels to
appear with retroflex consonants (Stevens & Blumstein, 1974). All tones were made in the Phonetics laboratory at the University of British Columbia. Each tape contained 8 natural exemplars of each sound. This was to ensure that variations in duration, fundamental frequency and information would be randomized both within and between categories.
RESULTS

Each individual in each group was recorded as having either reached or not reached the 8 out of 10 criterion on the two Hindi pairs. An analysis of proportions was then performed on this data (see Tables 3 and 4). This analysis is based on the Scheffe theorem (Marascuillo, 1966). It is used to compare discrete data in which the proportions of individuals falling into given categories is compared.

The null hypothesis was that there would be "no significant difference among the proportion of individuals reaching criterion in the four groups." The rejection level for this hypothesis was set at $p = .05$.

For the first comparison (the Hindi contrast Retroflex/Dental) the overall Chi-square obtained had a probability of less than .05 ($p = .0000016$) and thus enabled rejection of the null hypothesis. A series of multiple comparisons between each pair of groups yielded the results outlined in Table 5.

There was no significant difference between the proportion of Hindi adults and infants reaching criterion, nor was there a significant difference between the naive and the trained English speaking adults. All other comparisons (Hindi vs. each English group and infants vs. each Adult English group) were significant.

For the second comparison (the Hindi contrast unvoiced, aspirated dental vs. breathy-voiced dental), the overall chi-square obtained was not significant ($p .05, p = .0580$), so rejection of the null hypothesis was not possible. However, since the $p$ value was so close to .05, a series of multiple comparisons was performed to comb the data for trends. The results are summarized in Table 6.
Table 3
Analysis of Proportions for Retroflex/Dental Contrast

<table>
<thead>
<tr>
<th></th>
<th>Group I (Hindi Adults)</th>
<th>Group II (Infants)</th>
<th>Group III (Naive Eng. Adults)</th>
<th>Group IV (Trained Eng. adults)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reached criterion</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Did not reach criterion</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Total N</td>
<td>5</td>
<td>*12</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 4
Analysis of Proportions for Voiceless aspirated/ Breathy voiced Contrast

<table>
<thead>
<tr>
<th></th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reached criterion</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Did not reach criterion</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Total N</td>
<td>5</td>
<td>*12</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

* Altogether, 15 infants were tested in all, but 3 would not condition to the paradigm, even with the English Ba/ba sound pair. The remaining 12 conditioned and 9 passed both speech sound pairs.
Table 5

Multiple Comparisons on Retroflex/Dental Contrast

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Confidence Interval</th>
<th>Prob. of No Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>-0.306 to 0.140</td>
<td>.779</td>
</tr>
<tr>
<td>3-1</td>
<td>-1.165 to -0.635</td>
<td>* .000</td>
</tr>
<tr>
<td>3-2</td>
<td>-1.163 to -0.470</td>
<td>* .000</td>
</tr>
<tr>
<td>4-1</td>
<td>-1.163 to -0.470</td>
<td>* .000</td>
</tr>
<tr>
<td>4-2</td>
<td>-1.140 to -0.694</td>
<td>* .000</td>
</tr>
<tr>
<td>4-3</td>
<td>-0.365 to 0.165</td>
<td>.774</td>
</tr>
</tbody>
</table>

Table 6

Multiple Comparisons on Voiceless Aspirated/Breathy Voiced Contrast

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Confidence Interval</th>
<th>Prob. of No Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>-0.436 to 0.102</td>
<td>.494</td>
</tr>
<tr>
<td>3-1</td>
<td>-0.987 to 0.213</td>
<td>* .002</td>
</tr>
<tr>
<td>3-2</td>
<td>-0.905 to 0.038</td>
<td>.153</td>
</tr>
<tr>
<td>4-1</td>
<td>-0.662 to 0.062</td>
<td>.232</td>
</tr>
<tr>
<td>4-2</td>
<td>-0.585 to 0.318</td>
<td>.909</td>
</tr>
<tr>
<td>4-3</td>
<td>-0.230 to 0.830</td>
<td>.572</td>
</tr>
</tbody>
</table>
The only significant difference occurred between the Hindi adults and the naive English adults. The comparison between the infants and the naive English adults was the next closest to being a significant difference but still had a 15% probability of being due entirely to chance.

Since the VRISD paradigm is a conditioning paradigm, it can be criticized for allowing the infants a chance to "learn" the discrimination during the testing session. As outlined in the method section, Group IV (adults with training) was included to address this type of criticism. Additionally, two other sources of data were collected to try to understand the relative contribution of "learning" within the testing session. The first source was the inclusion of the common English /Ba/ vs. /Da/ sound pair whenever possible in infant testing. The second source was a comparison of the number of trials to either criterion or stopping for each of the four groups.

It was only possible to collect /Ba/-/Da/ data on four infants, two of whom were tested on the Hindi contrasts, and two of whom were only tested on the English contrasts. Since the two who were only tested on the English contrasts were never tested on the Hindi sounds, and since other infants who were tested on the Hindi sounds failed to reach criterion or were not tested on the Ba/Da, no chi-square comparisons were made. These data were analyzed in terms of mean number of trials to shaping and to criterion as summarized in Table 7. Only individuals who reached criterion were included in computing these means. T-tests for correlated data were then performed comparing the means listed in Table 7. No significant differences were found among the three different sound pairs in either number of trials to shaping, or number of trials to reaching criterion. Since the N was so small, a 2-group, one-tailed t-test was performed to compare
Table 7

Mean Number of Infant Trials on Speech Contrasts

<table>
<thead>
<tr>
<th>Infants</th>
<th>Trial to shaping</th>
<th>Trials to passing</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 4</td>
<td>Ba/Da</td>
<td>5.5</td>
</tr>
<tr>
<td>N = 11</td>
<td>Ret/Dent</td>
<td>10.9</td>
</tr>
<tr>
<td>N = 10</td>
<td>Un/ASD</td>
<td>8.1</td>
</tr>
</tbody>
</table>
/Ta/ /tə/ and /Ba/ /Da/. Although the difference between the means on number of trials to criterion was clearly not significant, the difference on number of trials to shaping almost reached significance at p .05 level (p = .06). A similar analysis on /tʰ/ /dʰ/ vs. /Ba//Da/ did not approach significance.

An analysis of variance was run on the number of trials to criterion (or stopping)\(^{(4)}\) for the four groups on the two pairs of Hindi sounds. Cell means are shown in Table 8, and the results of the analysis of variance are shown in Table 9. As can be seen, the main effects for both Group (I-IV) and Sound Pair were statistically significant. Using Tukey's method of planned comparisons, it was found that the group main effect could be accounted for by a significant difference between the Group I (Hindi adults) and all other groups on number of trials to criterion. No significant differences emerged between Groups II, III and IV. The other main effect was simply that overall, the unvoiced, unaspirated/unvoiced aspirated sound pair was easier than the retroflex/dental. (Ba//Da/ could not be included in the ANOVA as it would not fit into a repeated measures design).

\(^{(4)}\)Infants and adults were run as long as possible during a testing session in the hope that they would reach criterion. Whenever an infant became overly fussy, or sleepy during a testing session, we would stop. Conversely when an adult started signalling he/she was fatigued we would stop. Additionally, testing of adults who failed to show any recognition of the sound change was terminated after 25 trials.
### Table 8

Cell Means (M) and Standard Deviations (SD) for Number of Trials to Criterion

<table>
<thead>
<tr>
<th></th>
<th>Group I</th>
<th></th>
<th>Group II</th>
<th></th>
<th>Group III</th>
<th></th>
<th>Group IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Ta/ta</td>
<td>10.4</td>
<td>.548</td>
<td>27.58</td>
<td>9.65</td>
<td>25.6</td>
<td>10.38</td>
<td>27.7</td>
<td>3.498</td>
</tr>
<tr>
<td>Th/Dh</td>
<td>10.2</td>
<td>.447</td>
<td>22.5</td>
<td>11.31</td>
<td>18.5</td>
<td>5.60</td>
<td>19.3</td>
<td>5.056</td>
</tr>
</tbody>
</table>

### Table 9

Source Table for Number of Trials to Criterion; Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>MS</th>
<th>F ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Group</td>
<td>2247.099</td>
<td>3</td>
<td>749.03</td>
<td>13.307*</td>
</tr>
<tr>
<td>Error</td>
<td>1857.516</td>
<td>33</td>
<td>56.29</td>
<td></td>
</tr>
<tr>
<td>Sound Pair</td>
<td>446.842</td>
<td>1</td>
<td>446.84</td>
<td>6.797*</td>
</tr>
<tr>
<td>Grp x Snd pair</td>
<td>160,814</td>
<td>3</td>
<td>53.61</td>
<td>0.815</td>
</tr>
<tr>
<td>Within Groups</td>
<td>2169.508</td>
<td>33</td>
<td>65.74</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

The results of this study yield support for the hypothesis:

"Humans are born with the ability to discriminate the universal set of distinctive linguistic features. Through development there is a decline in this ability as the perceptual space becomes organized to approximate the phonemic categories used in the native language."

There was no difference between the infants and the Hindi adults, or between the two groups of English speaking adults, but there were significant differences in discriminatory abilities between all other pairs of groups for the retroflex/dental distinction. A higher proportion of the infants and the Hindi adults could discriminate this Hindi contrast than could either group of English speaking adults. English adults with limited training performed no different from English adults without training, showing that minimal training with adults did not facilitate the discrimination. Interestingly, the one adult English speaker who was able to discriminate the contrast reported that seven years prior to his being tested on the Hindi contrast he had spoken Twi for a period of three months. Retroflex/Dental is used in Twi to differentiate phonemic categories (Fromkin, 1974). This subject, like the Hindi speakers, reported the contrast as being very obvious, and was surprised others could not discriminate the difference.

Results from the comparison of the four groups on the other Hindi sound pair, \( /d^h/ \) and \( /t^h/ \) were less clear. Although the analysis of proportions did not quite reach significance planned comparisons were performed to try to determine whether there was any pattern to the possible difference between the four groups. The only significant difference
was between the Hindi adults and the naive English adults. It should be pointed out that two infants out of twelve failed to perform this discrimination. One of the infants, however, was very clearly able to make this discrimination, but simply failed to perform the required task. During the testing session, she would turn her eyes toward the speaker whenever the sound changed, but she would not perform a full head turn. Eye movements were not reinforced, nor could they be counted as correct responses. Had we been able to count them, however, this infant would have reached criterion, and the overall Chi-square for the comparison would have reached significance (p = .02). In addition, the infant vs. naive adults comparison would have reached significance (p = .03). The comparison between the English adults with training, and both the Hindi adults and English infants would still not have been significant, however. Thus, at best, the inclusion of this infant in the analysis would have lent only limited support to the main hypothesis.

In an attempt to determine whether the Hindi contrasts required a greater number of trials to criterion than did the English contrast /Ba/ /Da/, t-tests were performed on the infant data. Although the means were greater for the Hindi contrasts (see Table 7), this difference did not reach significance in a correlated one-tailed t-test. This finding was surprising since it was predicted that the English contrast would be somewhat easier, at least in shaping trials, since many of the infants were already repeating /ba/'s and /da/'s in both their imitative and spontaneous babbling. This lack of significance suggests that even though some language appropriate sounds have entered the infants' productive
repertoire, flexibility may still be maintained at the perceptual level.

An analysis of variance conducted on number of trials to criterion yielded two significant effects. "Groups" and "Sound Pairs". Planned comparisons showed only the difference between Hindi adults and all other groups to be significant. This is not surprising given that Hindi adults performed similarly to English speakers in a pilot experiment conducted using English contrasts as stimuli. That is, they reached perfect scores almost immediately on native contrasts. Of more interest, however, is the lack of significance between pairs of the other three groups. The infants may have faced a more difficult task than the adults for three reasons:

(1) The infants had to continually divide their attention between the toys being manipulated by the assistant and the speech sounds. The adults on the other hand, always had their attention directed at the speech sounds.

(2) Although the visual reinforcer may have been more rewarding to the infants than to the adults, presumably the fear of failure was not as strong. (Green and Swets, 1966) have suggested that fear of failure may be equal to monetary reinforcements as a motivating force for adults. Certainly, the apologies, rationales, and complaints offered by each adult subject who did not reach criterion on these contrasts yielded subjective support for this notion!

(3) Infants are described as being more affected by state fluctuations (Eisenberg, 1976) than adults and should thus be expected to make more errors on an equally easy task due to these state fluctuations. That is, being tired, hungry, etc..., is described as interfering much more
with an infant's attention to a task than would similar fluctuations in an adult. Additionally, day to day differences in qualities such as sociability, excitement, fear of strange places, etc. affect an infant's performance in a controlled experimental setting. The effects of such factors on the performance of adults is assumed to be minimal.

Since there was no significant difference between the number of trials to criterion (or stopping) for adults and in infants, it cannot be argued that the infants were simply given a greater number of trials to "learn" the contrasts. Thus the hypothesis that infants can discriminate non-native sounds with little or no learning although adults may not be able to do so was substantiated.

There is some question as to whether VRISD should continue to be considered a conditioning paradigm or whether it should be reconceived as a discrimination paradigm with feedback. Since the monkey in a box can hardly be seen as a potent adult reinforcer, one could argue that fear of failure was the motivating force behind adult performance. Additionally, infant behavior suggested some kind of "competence" motivation rather than a direct stimulus reinforcer as being primary. Once an infant had "learned" to perform a head turn response upon a change in the background stimulus the reinforcing value of the toy seemed to be diminished. After several trials, infants would typically concentrate on the assistant manipulating toys, would then swing their head around toward the speaker and back again when the speech sound changed and continue watching the assistant rather than the reinforcer. It seemed as if infants only wanted to know they could activate the toy animal,
but did not particularly care to watch it. In addition, it would be hard to argue that subjects could "learn" which acoustic signals they should heed given the multiple natural tokens and the limited number of trials.

The second main effect "Sound Pairs" was attributable to /tʰ//dʰ/ taking fewer trials to discriminate for all groups. Such a result is consistent with the results from both the Analysis of Porportions and from the t-tests in that both these statistical analyses gave some support for the notion that /tʰ//dʰ/ was intermediate in difficulty between the common English /Ba//Da/ and the more rare Hindi /Ta//ta/ contrasts. Two explanations could be offered for this difference.

(1) A "language experience" confound could have been possible for two reasons. First, voiceless, aspirated vs. breathy voiced stops are used phonemically across a wider range of natural languages than is the retroflex vs. dental place of articulation distinction. Thus adults, and even infants, would be more likely to have had exposure to this sound contrast.

Second, there is controversy as to the correct description of the four categories of voice-onset time used in Hindi. Although many linguists (e.g., Ladefoged, 1975) assert that breathy voicing must be described as a unique category of voice-onset time, some linguists have described breathy voicing as being a voiced, aspirated stop. If this latter description is correct, one could argue that the voiceless vs. voiced distinction could provide English listeners with a partial cue to the discrimination. (English differentiates voice vs. voiceless unaspirated stops).
(2) An explanation based on the notion of "perceptual distance" is also possible. It may be that although the human ear can necessarily discriminate all phonemic differences, the perceptual dissimilarity of some of these differences is greater for some feature distinctions than it is for others. In this regard a strictly phonological description (based on meaningful articulatory differences) may not always be complete. The reliance on such a description in traditional linguistics rather than on a phonetic description (based more on acoustic differences) may have been misleading. That is, a single feature difference is a useful tool for explaining phenomena that yield a perceptual (and meaningful) invariant but not an acoustic one. This does not, however, imply perceptual (or acoustic) equality among all feature differences.

A perceptual distance explanation initially appears to involve an acoustic rather than a linguistic, speech perception mechanism. On closer examination however, a more complex picture emerges, suggesting possible parallel levels of processing with differential access to an acoustic or a linguistic level depending upon the most effective strategy for the task.

In summary, the infants were found to discriminate all sound pairs according to phonemic category with comparable ease (with a slight suggestion that the Hindi /Ta//ta/ was more difficult to shape). It was also shown that Hindi adults could discriminate all sound pairs, whereas English adults could not discriminate /Ta//ta/ at all, and only some English adults could discriminate /tʰ//dʰ/. Overall, the /tʰ//dʰ/ discrimination appeared to be easier than the /Ta//ta/ discrimination. An important additional piece of information concerns the categorizing
phenomenon observed. That is, individuals who could perform to criterion for a sound pair would continue to perform at that level throughout a testing session without an increase in false positive rate. (Infants would tire after an additional 10 trials or so). Once a criterion was established, the individuals did not return to pre-criterion levels of performance. Because natural tokens were used, there were obvious differences between each token within a speech sound category. Those differences were only responded to by individuals who could not reach criterion on a sound pair. Individuals who categorized the sounds (reached criterion at the phoneme level) did not respond to within category variation, although they could hear differences when asked.

If it is accepted that neither Hindi sound pair involved tokens that are used phonemically in English (i.e., if the language experience confound explanation is discovered) recourse to a categorizing explanation based on experience with a language is not possible. The explanation I would like to offer is thus based on the notion of two-level of processing, an acoustic and a linguistic level, much as suggested by Wood (1974). When possible, people seem to employ a linguistic level of processing, and respond to the speech sounds in a categorical manner disregarding within category differences. When this strategy fails, they employ an acoustic level of processing.

An acoustic phonetic analysis of the speech sounds chosen for this study supports such an explanation. The phonological and phonetic descriptions of retroflex vs. dental sounds are identical. The phonetic description of voiceless aspirated vs. breathy voiced is different, however, from the phonological description. That is, although these sounds
differ in only a single phonemic feature, the voice onset time differ-
ence may be at least two steps apart on a continuum. This is because
the Hindi murmured stops are neither voiced nor aspirated. It could be
assumed that they actually differ in at least two ways from voiceless
aspirated stops (Ladefoged, 1975). It could then be argued that the per-
ceptual distance between /tʰ/ and /dʰ/ is greater than that for /Ta//ta/.
Thus employing an acoustic level of processing would be more successful
for this sound pair, and could lead to more accurate discrimination by
non-native speakers. Since the data suggested that phonemic discrimina-
tion of /tʰ//dʰ/ was easier (required fewer trials to reach criterion)
for all groups than /Ta//ta/, a linguistic level of processing can also
be invoked enabling listeners to ignore acoustic differences that are
irrelevant to natural languages.
CONCLUSIONS

This study yielded definite support for the idea that infants possess the propensity to discriminate the universal set of linguistic features. The checks built into the study make it difficult to argue that infants were simply learning these contrasts in a very few trials. The results therefore support the view that infants actively impose a structure (but a structure constrained by biological perceptual capabilities) on continuous auditory input. Such an innate constraint would allow the infant to segment the continuous input of speech into the units used to convey meaningful differences within the language environment to which the infant is exposed.

Some support was given for the idea that there may be a decline in speech perceptual abilities with either age or linguistic experience. The finding that this decline was more evident with one than the other Hindi sound pair makes interpretation more difficult. Two explanations were offered for these results. It was suggested that the Non-Hindi population may not have been naive with respect to the /t̪ʰ/ /d̪ʰ/ contrast. Alternatively, it was suggested that perceptual distance may have accounted for the differences, and two levels of processing were involved in this explanation.
REFERENCES


Moffit, A.R. Categorical discrimination among the acoustic cues for bilabial stop consonants by young infants. (Unpublished manuscript).


Stevens, S.S. Psychology and the science of science, *Psychological Bulletin,* 1939, 36, 221-263.


APPENDIX A

Background in Speech Decoding and in Distinctive Features

It is now well established that the cochlea is completely formed by 26 weeks in utero, and that by this time both the middle and inner ear structures have reached full adult size. (Eisenberg, 1976). Auditory nerve fibers begin to myelinate during the sixth month in utero, so that by birth even the auditory cortex is myelinated in the normal full-term infant (Eisenberg, 1976). Although it was once believed that the infant could not make any differential responses associated with the complex characteristics of auditory stimuli (Pratt, 1960), it is now well documented that most newborns, even premature babies and those with abnormalities of the central nervous system can discriminate sounds according to various acoustic variables (Eisenberg, 1976).

Several acoustic parameters related to the decoding of speech have been explored. Each of the parameters which may be particularly important in speech has been shown to be within the auditory capabilities of the newborn: (1) duration, or the total amount of time consumed by a stimulus; (Eisenberg, 1976); (2) frequency, in Hertz (Hz), or the number of sine wave repetitions per second in a pure tone signal (Trehub, 1973); (3) sound pressure level, or the physical intensity of the signal in decibels with reference to acoustic zero (Eisenberg, 1976), (4) dimensionality, or the kind and amount of variance within a complex auditory stimulus (Eisenberg, 1976).

Unfortunately, speech sounds cannot be easily analyzed in terms of their constituent parts for three main reasons.
(1) Speech is continuous, thus acoustic cues are highly overlapping, as shown by the diagram below.

Figure 1: Schematic diagram showing how the overlap of articulatory features produced encoding in the conversion to sound. (From Liberman, 1967).

(2) Phonetic segments do not have invariant properties. For example, the /d/'s in the continuum from /di/ to /du/ would each be perceived as a /d/ sound, but the acoustic dimensions of these /d/ portions would be entirely different as illustrated in the abbreviated spectograms below. (From Liberman, 1967).

Figure 2. Abbreviated "d" spectograms.

In addition, accents, voice quality, whispering, etc. all create variation in the acoustic signal, but do not alter a listener's invariant perception.

(3) Phonetic segments do not stand in a one-to-one relation to the acoustic signal. In most non-speech and "unencoded" speech sounds, it is possible to judge the manner in which the energy of the sound is distributed, whereas in encoded speech sounds, it is impossible to judge the
acoustic cues underlying the signal.

Although so far it has proved impossible to demonstrate invariant acoustic cues underlying segments of speech, there are (as indicated above) perceptual invariants. That is, many speech sounds (except vowels and fricatives) are perceived in a categorical fashion when presented in a linguistic context. Thus although a speech signal may vary along an acoustic continuum, it will not be perceived as a new category until it crosses a phoneme boundary. The meaning of this perceptual invariance is a matter of considerable debate. Some investigators posit a specific linguistic processes (Eimas, 1976), whereas others claim it to be a simple acoustic processor more sensitive to the complexities of multidimensional signals (Stevens & Klatt, 1974). Additionally, the finding of within-category discrimination under special circumstances has lead some investigators to posit two parallel modes of processing, acoustic and phonetic (Wood, 1974).

In an attempt to understand this perceptual invariance, traditional linguistics describes a structural analysis of the regularity between phonemes and articulation, and posits a universal set of phonemes in terms of articulatory features. Several classification schemes have been developed, some in terms of only phonemic description (Jakobson & Halle, 1956), and others including rules to account for recombination into morphemes (Chomsky & Halle, 1968). An example of feature description is given in Table 1.

As can be seen in Table 1, each phoneme can be described in terms of articulatory features and phonemic contrasts can be described in terms

*A stimulus dimension is considered to be perceived categorically if the spacing of signals along that dimension is found to be the same in discrimination experiments as in identification experiments.
<table>
<thead>
<tr>
<th>Feature name</th>
<th>Abbreviated definition of physical scale</th>
<th>Phonological terms</th>
<th>Exemplification</th>
<th>% value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Glottalic</td>
<td>Rate of upward movement of the glottis</td>
<td>s, t, r, l</td>
<td>t' Uduk t'ē</td>
<td>100</td>
</tr>
<tr>
<td>2 Velaric</td>
<td>Degree of suction of air in the mouth</td>
<td>s, t, r, l</td>
<td>t Uduk tēr</td>
<td>50</td>
</tr>
<tr>
<td>3 Voice</td>
<td>Degree of approximation of the arytenoid cartilages</td>
<td>s, t, r, l</td>
<td>t Uduk ċek</td>
<td>0</td>
</tr>
<tr>
<td>4 Aspiration</td>
<td>Time of onset of voicing with respect to release of the articulation</td>
<td>s, t, r, l</td>
<td>t Zulu ńbud</td>
<td>100</td>
</tr>
<tr>
<td>5 Place</td>
<td>Distance from the glottis to the first constriction of the vocal tract</td>
<td>s, t, r, l</td>
<td>t Zulu ńtā</td>
<td>0</td>
</tr>
<tr>
<td>6 Labial</td>
<td>Degree of approximation of the centers of the lips</td>
<td>s, t, r, l</td>
<td>t Javanese ńbuk</td>
<td>100</td>
</tr>
<tr>
<td>7 Stop</td>
<td>Degree of approximation of the articulators</td>
<td>s, t, r, l</td>
<td>b Hausa ńbērā</td>
<td>80</td>
</tr>
<tr>
<td>8 Nasal</td>
<td>Degree of lowering of the soft palate</td>
<td>s, t, r, l</td>
<td>b Hausa ńbērā</td>
<td>60</td>
</tr>
<tr>
<td>9 Lateral</td>
<td>Amount of airstream flowing over the side of the tongue</td>
<td>s, t, r, l</td>
<td>b Hindi ńbāl</td>
<td>20</td>
</tr>
<tr>
<td>10 Trill</td>
<td>Degree of vibration of an articulator</td>
<td>s, t, r, l</td>
<td>p Hindi ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>11 Tap</td>
<td>Rate of articulatory movement?</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>100</td>
</tr>
<tr>
<td>12 Sonorant</td>
<td>Amount of acoustic energy</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>50</td>
</tr>
<tr>
<td>13 Sibilant</td>
<td>Amount of high frequency energy (over 3000 Hz.)</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>100</td>
</tr>
<tr>
<td>14 Grave</td>
<td>Ratio of low to high frequency energy</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>15 Height</td>
<td>Inverse of frequency of the first formant</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>16 Back</td>
<td>Difference between frequency of formant two and formant one</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>17 Round</td>
<td>Inverse of distance between corners of the lips</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>18 Wide</td>
<td>Degree of advancement of tongue root</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>19 Rhotacized</td>
<td>Lowering of the frequency of the third formant</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
<tr>
<td>20 Syllabic</td>
<td>(No agreed physical scale)</td>
<td>s, t, r, l</td>
<td>p Thai ńpāl</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 1**
The prime features.
of feature differences. Support for such a system was presented by Miller and Nicely (1955) in a now classic study which demonstrated that under masking conditions, errors in perceptual responses increased as the number of feature differences decreased. Similar pairs (phonemes differing in only one articulatory feature) were judged to be the hardest to discriminate. An example of a highly similar pair of sounds would be /b/ and /p/; the only articulatory feature difference in voicing (voicing is then called the "distinctive feature"). Most current speech perception research involves phoneme pairs which differ in only one distinctive feature.
APPENDIX B: Consent Forms

Adult Consent Form

THE UNIVERSITY OF BRITISH COLUMBIA
2075 Wesbrook Mall
Vancouver, B.C., Canada
V6T 1W5

Department of Psychology

CONSENT FORM

This experimental procedure has been requested by

____________________________________________________________________

I have been informed of the procedures and understand them. I also understand that the procedures may be terminated at any time at my request.

PROCEDURE:

This is a study in speech perception. You will be sitting on a chair facing a loud speaker in the testing chamber. A series of one syllable speech sounds will be played over the speaker. Your task will be to press a button whenever you detect a change in the speech sounds. Every correct discrimination response will be signalled by flashing lights. A record of all your responses will be kept.

If at any time you desire, there will be a break in, or termination of, the testing session.

My signature below certifies that I consent to the experimental procedure which has been described and which is to be conducted on the following date: ________________________________

in the following place: ________________________________

and designated in the following manner: ________________________________

____________________________________________________________________

Date: ________________________________ Name: ________________________________

Signature: ________________________________
Infant Consent Form

THE UNIVERSITY OF BRITISH COLUMBIA
2075 Wesbrook Mall
Vancouver, B.C., Canada
V6T 1W5

Department of Psychology

CONSENT FORM

This experimental procedure has been requested by

________________________________________________________________________

I have been informed of the procedures and understand them. I also understand that the procedures may be terminated at any time at my request.

PROCEDURE:

The infant will be held on his/her guardian's lap in the testing chamber. He or she will see a series of toys, and will hear a series of one syllable speech sounds. The infant will be watched by the experimenter in the adjoining room, and by the assistant in the testing chamber. Whenever the infant turns toward the loud-speaker during a test trial, the experimenter and the assistant will press a button. If this head turn has occurred when there is a change in the speech stimuli, a toy animal inside the dark plexiglass box will begin performing. A record of all the baby's responses to the speech sounds will be kept. The baby will be held by, and will be under the control of, the guardian at all times. Whenever the guardian desires, there will be a break in or termination of the testing session.

My signature below certifies that I consent to the experimental procedure which has been described and which is to be conducted on the following date: ____________________________

in the following place: ____________________________

and designated in the following manner: ____________________________

________________________________________________________________________

Date: ____________________________ Name: ____________________________

Signature: ____________________________