DOCUMENT RESUME

ED 115 124 FL 007 325

AUTHOR Streeter, Lynn A.; Landauer, Thomas K.
TITLE The Effects of Learning English as a Second Language on the Acquisition of a New Phonemic Contrast.
PUB DATE 75
NOTE 19p.; Revised version of a paper delivered at meeting of the Acoustical Society of America (Austin, Texas, April 1975)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Articulation (Speech); Auditory Discrimination; Bantu Languages; Consonants; Contrastive Linguistics; Descriptive Linguistics; *Distinctive Features; *English (Second Language); Language Research; Phonemes; Phonemics; *Phonetic Analysis; Phonetics; Phonological Units; *Phonology; *Second Language Learning; Vowels
IDENTIFIERS *Categorical Perception; Kikuyu

ABSTRACT

Very sharp discrimination functions for the timing of voice onset relative to stop release characterize perceptual boundaries between certain pairs of stop consonants for adult speakers of many languages. To explore how these discriminations depend on experience, their development was studied among Kikuyu children, whose native language contains no stops in which voicing is substantially delayed relative to stop release (e.g., /p/). Kikuyu distinguishes stops in which voice onset substantially precedes release (prevoiced) from those in which voice onset is nearly simultaneous with release (voiced) for apical and velar places of articulation. However, the language has only a single prevoiced labial stop. Prior to exposure to English, children discriminated prevoiced from voiced labials and voiced from voiceless labials, although these distinctions are not phonemic in Kikuyu. Moreover, the voiced/voiceless discrimination for labials ([ba]) versus [pa]) improved markedly with schooling in English, rapidly surpassing the prevoiced/voiced distinction. Apparently, certain voice onset time differences are naturally discriminable, but it is also apparent that the very fine voiced/voiceless discrimination among adults for whom it is phonemic is largely attributable to experience. (Author)

ED115124

The Effects of Learning English as a Second Language on the Acquisition of a New Phonemic Contrast

by

L. A. Streeter
T. K. Landauer
Bell Laboratories
Murray Hill, New Jersey 07974

A. INTRODUCTION

For pure tones and other nonspeech stimuli, listeners discriminate among vastly more stimulus values that vary along a single dimension such as frequency than they can identify or name.[1,2,3,4] However, when subjects discriminate among certain speech sounds, in particular stop consonants, performance is excellent for sounds that are identified as members of different phoneme categories, but discrimination is very poor for sounds that are identified as members of the same phoneme category. In fact, it has been claimed that listeners do not discriminate between speech stimuli that belong to the same phoneme category any better than they can absolutely identify them. This hypothesized phenomenon has been called categorical perception.[5]

The striking superiority of discrimination across phoneme boundaries could be explained in at least two ways: 1) some differences between sounds that humans can make are simply easier to hear, and linguistic evolution has taken advantage of this in setting phoneme boundaries;

FL007325

and/or 2) speech sounds acquire distinctiveness or equivalence through experience. Because the evidence for "categorical perception" of speech sounds has all come from individuals whose language makes contrastive use of the phonemes involved, it is not possible to disentangle these factors. However, by studying individuals whose first language does not contain a particular contrast, but who learn a language that requires it, information as to both natural ease of discrimination and the effects of experience can be obtained. We report such a study here.

One phonetic feature that has been used to distinguish between stop consonants in initial, prevocalic position is voice onset time (VOT), which is defined as the time between the onset of vocal cord vibration and articulatory release of the stop. Linguists typically characterize the pairs /p/ and/b/, /t/ and /d/, and /k/ and /g/ as differing in voicing; /b/, /d/, and /g/ are voiced, whereas /p/, /t/ and /k/ are voiceless. One way in which this difference in voicing onset can be realized acoustically is by varying the onset of the first formant (F1) relative to the onset of the second (F2) and third (F3) formants.[6] When the onset of all three formants is simultaneous, one hears voicing at the time of the stop release, whereas when F1 onset is delayed substantially and F2 and F3 are excited by a "hiss" noise source when F1 is absent, the quality of aspiration characteristic of voiceless stops is achieved.

Lisker and Abramson[7] have claimed that this single feature, VOT, is universally sufficient for differentiating voicing of stop consonants. While VOT cannot completely characterize all differences within and between languages among stop categories, it appears to be successful in describing important differences among stop consonants in production and perception data from numerous languages. Spectrographic analyses of word-initial, prevocalic stops for speakers from diverse languages have shown that there are three consistent production ranges for stops: the prevoiced range (-125 to -75 msec VOT) in which voicing precedes articulatory release; the voiced range (0 to +25 msec VOT) in which voicing onset is nearly simultaneous with stop release; and the voiceless range (50 to 100 msec VOT) in which stop release substantially precedes voicing. In addition, VOT has been shown to be a perceptually sufficient cue for stop differentiation in several quite different languages. In these studies[8,9,10] subjects <u>identified</u> synthetically produced stimuli that varied in VOT with the phoneme labels appropriate for the language. On the whole, the identification and production functions obtained were consistent within each language.

Generally for the languages studied by Lisker and Abramson, discrimination data using VOT were also consistent with production data. In assessing discrimination, subjects heard VOT triads such as [ba], [ba], [pa], and were asked to indicate which one was different. The data for the languages studied for the most part showed that discrimination across phoneme boundaries (i.e., the VOT value at which half of the stimulus presentations were identified as one phoneme and half as another) was much sharper than discrimination within phoneme categories. Moreover, it appeared that "categorical perception" was influenced by linguistic experience, since the pattern of discrimination differed depending on which VOT categories the languages used for their stop categories.

To further explore how the discrimination of voicing categories is influenced by linguistic experience, we studied the development of children's discrimination of the three voicing categories (prevoiced, voiced, and voiceless) in labial stops among people whose first language made no such phonemic contrast. It was of interest to determine whether the discrimination pattern was altered by exposure to a second language, English, which does possess a phonemic voicing contrast for the labial place of articulation.

The children studied had as their native language Kikuyu, a Bantu language spoken in Kenya. Kikuyu has only one labial stop, /b/, in which voicing precedes articulatory release by an average value of 64 msec (as determined from spectrograms of production data). The language distinguishes between /d/ and /t/ and /g/ and /k/; however, for /d/ and /g/ voicing precedes release (prevoiced), and for /t/ and /k/ voicing is approximately coincident with release (voiced). Thus, Kikuyu has no stops in which voicing onset is substantially delayed relative to stop release as in the English voiceless, /p/, /t/, and /k/.

Kikuyu school children begin learning English as a second language in the second grade. By studying them one might determine whether or not the three voicing categories (prevoiced, voiced, and voiceless) are naturally discriminable, i.e., do not depend on specific linguistic experience and whether the discrimination pattern is altered in a systematic way by exposure to English.

B. METHOD

1. Subjects

The subjects were 128 children attending a periurban school in Kenya. There were 32 subjects from each of grade levels first, fourth, seventh, and high school. The average chronological ages were 7.5, 10, 13, and 15, respectively.

2. Stimuli

Stimuli were generated on the Haskins Laboratories' parallel resonance synthesizer by the same algorithm used by Lisker and Abramson. The stimuli were labial stops followed by the vowel [a]. The VOTs used were: -30 (pre-voiced), 0 (voiced), +10 (voiced), +40 (voiceless), +50 (voiceless), +80 (voiceless), with each stimulus 500 msec in duration.

For VOT -30 the first formant circuit in the synthesizer was used with a low-frequency value and a suppressed level, simulating the voice bar during the voiced period preceding stop release. For VOT 0 the onset of all three formants was simultaneous with the first formant transition and 50 msec in duration. For the positive VOTs the second and third formants were excited by a hiss source until voice onset. In addition, for positive VOTs the first formant transition was cut back until voice onset. Thus, for VOT greater than or equal to 50, the F1 transition was completely eliminated. All stimuli consisted of three steady state formant frequencies appropriate for the vowel [a].

VOT triads were recorded onto analog tape with 1/2 sec interstimulus intervals between triad members while successive triads were separated by 5 sec.

The triads were such that two of the members were identical and one was different, with the "different" member separated by 30 msec VOT from the "same" members. There were three distinct triad types: (1) prevoiced/ voiced with VOTs -30 and 0, (2) voiced/voiceless with VOTs 10 and 40, and (3) voiceless/voiceless with VOTs 50 and 80. Each of the 16 possible permutations within each triad type was used equally often.

In addition to the experimental set described above, "easy" triads in which the same and different members were separated by 80 msec VOT were interspersed in a random fashion among the experimental triads. The triad order was randomized separately for each block. There were two blocks, each consisting of 18 test and six easy discrimination items.

To familiarize subjects with the task procedure, a practice tape was presented. The practice tape consisted of consonant-vowel syllable triads (e.g., /na/ /na/ /wa/). The consonant ensemble included only admissable Kikuyu consonants. These CV triads were produced by a native Kikuyu male speaker. There were a total of 20 Kikuyu CV triads followed by four easy VOT discriminations in which the same and different members were separated by 100 msec VOT.

3. Procedure

The triads were presented free-field at a comfortable listening level. Two groups of approximately 16 subjects were tested from each grade level. Instructions were given by a Kikuyu assistant, who told the subjects that they would hear three sounds; two the same, one different. Their task was to select the odd member in the set and to cross out either "1", "2", or "3" on their response sheet depending on the different sound's position. If they were unsure, they were to make their best guess. Initially, subjects listened to the practice tape and were given feedback after every trial. After the training task, subjects listened to the two experimental blocks. Block order was counterbalanced across the two groups for each educational level.

C. RESULTS

"Oddity" judgments are not common tasks in non-literate societies, and are generally difficult for young children. Among the younger subjects there appeared to be some difficulty in understanding the requirements of the task. Consequently, we discarded data from subjects who showed no evidence of above-chance performance on the easiest of the triads. A total of 21 (first grade), 29 (fourth grade), 28 (seventh grade), and 32 (high school) subjects remained. It should be noted that the pattern of results was not altered by eliminating the poorer subjects.

The results based on the remaining data are shown in Figure 1. The proportion correct and the stan-

---

Insert Figure 1 about here

---

dard error of the mean for each of the three phonetic discriminations for each of the four grade levels are shown. The first grade children showed reliably above-chance performance ($p < .05$ in both cases) on the voiced/voiceless and on the voiced/prevoiced distinctions but not on the two voiceless labial stops. Discrimination between the voiced and voiceless stops (VOTs 10 and 40) increased steadily with age and school training in English. Discrimination accuracy for prevoiced versus voiced stops (VOTs -30 and 0) also increased despite the fact that there is no distinction between these two sounds in the Kikuyu language, nor is any present in the language being learned in school. In contrast, the accuracy of discrimination of the voiceless/voiceless triads (VOTs 50 and 80) actually declined somewhat, but not significantly, with age and training in English. The increase in discriminability was significantly greater for both the voiced/voiceless distinction and the prevoiced/voiced distinction than for the two voiceless labials ($\underline{F}$ = 42.48, df = 1,4, $\underline{p} < .005$ and $\underline{F}$ = 33.48, df = 1,4, $\underline{p} < .005$, respectively).

Moreover, discrimination of the voiced/voiceless condition increased rapidly relative to the prevoiced/voiced condition with discrimination of the voiced/voiceless triads significantly surpassing the prevoiced/voiced for the high school subjects ($\underline{t}_{31}$ = 2.810, $\underline{p} < .01$).

D. DISCUSSION

Discrimination of voiced from voiceless, and prevoiced from voiced labials is better than chance at the earliest ages measured. Without training, the discrimination is not better for one of these than the other. Since training in English had not begun for the first grade children, the above-chance level of performance on the voiced/voiceless discrimination can be taken as evidence that this phonetic difference is naturally easy to hear. Performance on it is as good as on a discrimination involving a phoneme found in the language, and is better than another contrast involving an equal VOT difference. Thus, it seems likely that there are certain regions along the VOT continuum that are naturally easier to discriminate than others. Discrimination of voiceless/voiceless triads (VOTs 50 and 80) was not reliably different from chance. One could conjecture that the wide use of the voiced/voiceless distinction in languages around the world is a reflection of nature's attempt to capitalize on the ease of this discrimination. Similarly, certain results seem to indicate that very young infants can discriminate [pa] from [ba].[11,12,13] These results, which have been taken as evidence of linguistic categorical perception, may owe much to natural discrimination ability.

On the other hand, the fact that the voiced/ voiceless discrimination becomes better with training in English indicates that the natural easiness of the discimination is not all that is involved. One should note that the discrimination between the two voiceless stops did not improve with age, so what we are seeing here is not merely an improvement in ability to perform the oddity task,or a general improvement in all discriminations, but a specific improvement in discriminability of the voiced/ voiceless contrast. The discriminability of this contrast increases markedly and to a greater extent than that of the prevoiced/voiced contrast. It is conceivable that the specific increase in performance on the voiced/voiceless distinction observed here was due to age alone, rather than to experience with English, since in this study age and exposure to English were confounded. However, there is other evidence that age is not enough. Streeter[13] found that monolingual Kikuyu adults, like the first grade children, but unlike the high school students and other

English speakers, performed no better on the voiced/voiceless than on the prevoiced/voiced discrimination. Thus, the excellent voiced/voiceless discrimination found in adult speakers of English--and other languages for which this contrast is phonemic--can reasonably be concluded to be a combination of natural ease and large amount of direct discrimination training.

What exactly is the source of the significant improvement in the prevoiced/voiced discrimination is more difficult to specify. However, the voiced phoneme is one of a pair being learned as a part of English training and this could be a sufficient experience for learning to discriminate it from the prevoiced /b/. There is another possibility, however. The Kikuyu language does contain a prevoiced/voiced contrast for the apical and velar places of articulation. Increasing experience with these distinctions could, conceivably, generalize to the labials.

In summary, we have found that the discrimination of prevoiced from voiced and voiced from voiceless labial stops is above chance before any training in English with children reared in a linguistic environment in which only one of these three phoneme types exists. With increasing age and training in English, the discriminations between the voiced/voiceless and prevoiced/voiced labials become better. The improvement is greater for the voiced/

voiceless distinction, corresponding to the phonemic distinction required in English and being learned by the children, than it is for the prevoiced/voiced distinction. We conclude that certain VOT distinctions are naturally easy to make, but that the very precise performance on this discrimination characteristic of English speaking adults is also to a large measure due to specific training.

ACKNOWLEDGMENTS

We thank Haskins Laboratories for the use of their facilities, and Pamela Cockey for her assistance. We thank John Whiting and Beatrice Whiting and the staff of the Child Development Research Unit, University of Nairobi for support, assistance, and advice. We thank Wilfred Mbogua and Lydia Wangome for help. A version of this paper was delivered at the Acoustical Society of America meeting, Austin, Texas, April 1975.

REFERENCES

[1]I. Pollack, "The information in elementary auditory displays," J. Acoust. Soc. Am. 24, 745-749 (1952).

[2]I. Pollack, "The information in elementary auditory displays II," J. Acoust. Soc. Am. 25, 765-769 (1953).

[3]G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for information processing," Psych. Rev. 63, 81-97 (1956).

[4]P. D. Eimas, "The relation between identification and discrimination along speech and nonspeech continua," Lang. Speech 6, 206-217 (1963).

[5]A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the speech code," Psych. Rev. 74, 431-461 (1967).

[6]A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some cues for the distinction between voiced and voiceless stops in initial position," Lang. Speech 1, 153-167 (1958).

[7]L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," Word 20, 384-422 (1964).

[8]A. S. Abramson and L. Lisker, "Discriminability along the voicing continuum: Cross-language tests," Proceedings of the Sixth International Congress of Phonetic Sciences (Academia, Prague, 1970), pp. 569-573.

[9]A. S. Abramson and L. Lisker, "Voice-timing perception in Spanish word-initial stops," Haskins Labs. Status Rep. Speech Res. SR 29/30, 15-25 (1972).

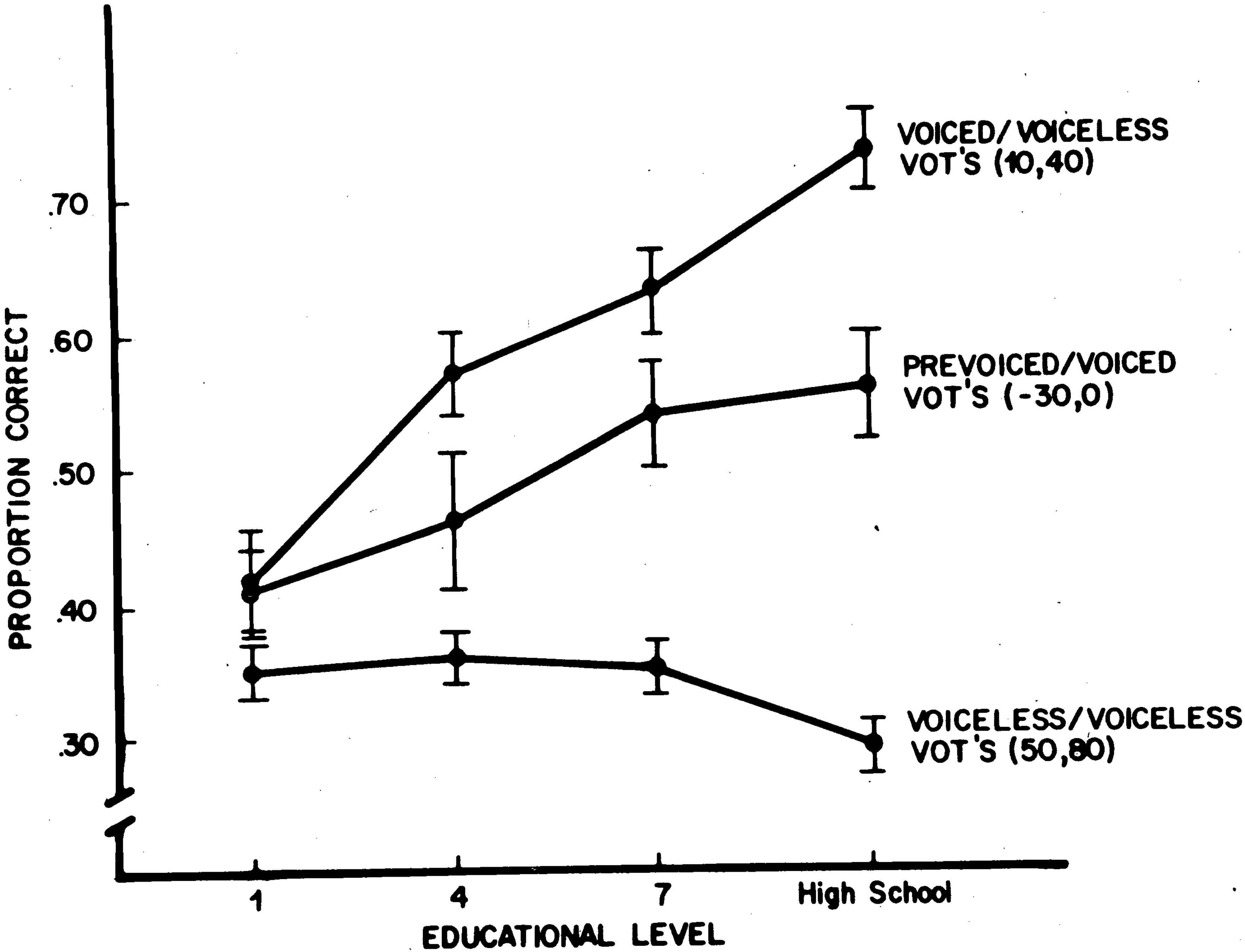
[10]L. Lisker and A. S. Abramson, "The voicing dimension: Some experiments in comparative phonetics," Proceedings of the Sixth International Congress of Phonetic Sciences (Academia, Prague, 1970), pp. 563-567.

[11]S. E. Trehub and M. S. Rabinovitch, "Auditory-linguistic sensitivity in early infancy," Develop. Psych. 6, 74-77 (1972).

[12]L. A. Streeter, "The effects of linguistic experience on phonetic perception," Unpublished Doctoral dissertation Columbia University (1974).

FIGURE CAPTION

Figure 1. Mean proportion correct for the three labial voicing distinctions plotted as a function of educational level.


VOICED/VOICELESS
VOT'S (10,40)
PREVOICED/VOICED
VOT'S (-30,0)
VOICELESS/VOICELESS
VOT'S (50,80)
PROPORTION CORRECT
.70
.60
.50
.40
.30
1
4
7
High School
EDUCATIONAL LEVEL