Bulletin of the Psychonomic Society
1981, Vol. 17(5), 217-220

Recognition memory for accented and unaccented voices

ALVIN G. GOLDSTEIN, PAUL KNIGHT, KAREN BAILIS, and JERRY CONOVER
University of Missouri, Columbia, Missouri 65211

Laboratory research has established that face recognition memory performance for own-race faces is better than for other-race faces. Three studies are reported exploring the possibility that the other-race effect will generalize to voice recognition memory. Recognition memory performance for non-native American speakers speaking both English and their native languages was compared with memory for native American speakers. With relatively long speech samples, accented voices were no more difficult to recognize than were unaccented voices; reducing the speech sample duration decreased recognition memory for accented and unaccented voices, but the reduction was greater for accented voices.

Face recognition memory and speaker recognition memory are both terms referring to an important social skill, namely, the ability to remember whether a person is or is not someone who has been seen before or whose voice has been heard before. In short, the capacity to remember faces and voices allows us to decide whether or not a person is familiar to us. The conditions that allow us to develop familiarity with a face have received much attention recently; in contrast, speaker recognition performance has attracted relatively little research interest (Clifford & Bull, 1978; Hecker, 1971; Mann, Diamond, & Carey, 1979).

One intriguing finding that emerges from the face recognition research is that unfamiliar exemplars of faces of people from one's own race (or ethnic group) become familiar more efficiently than faces of people who are clearly members of "foreign" groups (Chance, Goldstein, & McBride, 1975; Malpass & Kravitz, 1969). This finding, which has been labeled the other-race effect, is most likely the laboratory equivalent of the commonly held belief that, for example, all Chinese look alike. In spite of the fact that we have not heard anyone say "they all sound alike to me," the similarities of the face recognition problem and the voice recognition problem motivated us to ask whether the other-race effect could be demonstrated to influence speaker recognition performance. In particular, research reported here represents the results of a series of exploratory studies in which we compared speaker recognition memory performance when listeners heard speakers with and without "accented" voices and when speakers were speaking a foreign language.

Portions of the article were read at the annual meeting of the Psychonomic Society, November 13, 1980, in St. Louis, Missouri. The authors are grateful to June Chance for her critical reading of earlier drafts of this paper. Reprint requests should be sent to Alvin G. Goldstein, Psychology Department, McAlester Hall, University of Missouri, Columbia, Missouri 65211.

The general question of voice recognition by listeners has attracted unusually little research interest, which is surprising because the problem has figured in many court cases, including the world-famous Lindbergh kidnapping case. In that case, Bruno Hauptmann was found guilty at least partially on the basis of Lindbergh's testimony that he could recognize Hauptmann's voice as the voice of the kidnapper (after more than a 2-year interval).[1] The voice recognition research that has been reported has most often been concerned with specifying the factors that influence listeners as they try to aurally identify the voice of a familiar person, such as one of the listener's friends or business associates (e.g., Bartholomeus, 1973; Bricker & Pruzansky, 1966; Pollack, Pickett, & Sumby, 1954; Stevens, Williams, Carbonell, & Woods, 1968). Some research on recognition of unfamiliar voices had been reported (e.g., Carterette & Barnebey, 1975; Hecker, 1971; Mann et al., 1979), with conflicting results.

Direct comparison of the results of voice recognition studies with those of face recognition studies should be attempted with caution because of the vast differences in procedures employed in the two areas of investigation. However, if that warning is temporarily held in abeyance, the meager amount of relevant voice recognition data that is available suggests that recognizing unfamiliar voices is more difficult than recognizing unfamiliar faces. For example, in contrast to face recognition, voice recognition performance is more likely to be adversely affected by the number of voices attended to in the study trial and by the number of distractor voices presented in the test trial. Performance also seems to be sensitive to the length of the speech utterance in a way not easily compared with face recognition performance; speech samples less than about 1 sec in duration (and thus with few phonemes) are more difficult to recognize than are voice samples of about 1.5 sec. Increasing the length of the sample beyond about 1.5 sec seems to add little to the level of performance (Bricker & Pruzansky, 1966; Pollack et al., 1954).[2]

0090-5054/81/050217-04$00.65/0

We have been able to find only one report of research that appears to be closely related to the question explored in the present research. In one condition of McGehee's (1937) experiment, subjects listened (without benefit of an electronic recorder) to a native German speaker reading a 56-word paragraph in accented English and were asked 48 h later to identify the speaker's voice from among five accented speakers. Eighty-one percent of the subjects (N = 68) discriminated the correct German speaker from a Chinese, Greek, Russian, and a second German distractor speaker. This result, which was almost identical to the recognition rate McGehee had obtained for unaccented American voices (83%) in an identical task, seems to imply that recognition of accented voices is at least as accurate as recognition of native voices. However, McGehee's test is faulty because the group of foreign voices might have been more heterogeneous than the group of native American voices used for comparison.

Before initiating the three studies reported here, recognition memory for tape-recorded voices of foreign speakers (accented group) and American-born speakers (unaccented group) were tested in a series of pilot studies. Both groups of voices were aurally heterogeneous. Mild regional accents could be detected in some of the voices in the American-born group; the foreign-born speakers' voices were strongly accented. Portuguese was the native language of most of the foreign-born speakers; the remaining speakers were born in Jordan, Pakistan, India, Mexico, and Nicaragua. The study and test sentences read by each speaker and heard by the listeners were identical 15-word sentences, but the two sentences were tape recordings of two separate utterances. Depending on the study, target-to-distractor ratios ranged between 1 in 6 and 1 in 13.

Immediate recognition memory tests, following exposure to the accented and unaccented voices, resulted in overall hit rates averaging about 55%; moreover, recognition of accented voices was not appreciably different from recognition of unaccented voices. False alarms were quite frequent, varying between 2.0 and 3.7 per subject, and they were also distributed equally between the two sets of voices.

These data, which were consistent with McGehee's (1937) results, clearly implied that immediate recognition memory for accented voices and that for unaccented voices were about equal. Considering the relatively undemanding conditions of the experiment, subjects' performances were surprisingly poor. As in McGehee's (1937) experiment, heterogeneity of voices within the group of foreign-born speakers was troublesome; performance in response to the accented voices could have been spuriously enhanced by differences in the kinds of accents.

For this reason, a new set of accented voices was collected and tape-recorded. Native Taiwanese speakers were recruited from among students on the campus at the University of Missouri. All speakers were Chinese, all were born in Taiwan, and all had learned English in their school system before coming to America.

EXPERIMENT 1

Method

Voices and listeners. Twelve of the 36 speakers were Taiwanese, 12 were white Americans, and 12 were black Americans. Black and white speakers were native Americans living in either St. Louis or Kansas City, Missouri. All speakers were college-aged men. All voice samples within an ethnic group were recorded in one session to minimize within-group nonvocal cues. In this and all subsequent studies reported in this article, an Akai Model 1722W magnetic reel-to-reel tape recorder was used to record and play back the speaker's voices. Listeners always heard the voices, amplified through the recorder's built-in speaker system, in a small quiet room. Sound intensity differences among the voices were reduced to a minimum by appropriate manipulation of the volume control during retaping. A total of 67 listeners were recruited from general psychology courses. This group was composed of 20 white men, 25 white women, 4 black men, and 16 black women.

Procedure. Testing was done in small groups. Listeners recorded their responses (heard before or not heard before) on special answer sheets. During the study session, subjects heard this sentence: "The University of Missouri is located in Columbia, halfway between St. Louis and Kansas City." The test sentence, which followed the study sentence by a few seconds, was: "Joe took father's shoe bench out." Four different speakers uttered the test sentence; one of the four speakers was always the target speaker. A sequence of study voices followed by test voices was repeated for a total of 36 trials.

Results and Discussion

Overall correct identifications of target speakers by white listeners averaged 83%. Differences in response accuracy as a function of voice ethnicity were statistically nonsignificant and negligible (Taiwanese, 81%; blacks, 82%; whites, 85%). The 20 black listeners averaged 77% correct overall, and their performance, like the whites' performance, hardly varied across the three voice groups (blacks, 78%; whites, 78%; Taiwanese, 75%). The 6% black-white average subject difference in performance was not significant. These results and the results of the pilot studies imply that, for short retention periods, accented voices, voices that sound "foreign," are no more difficult to recognize than are unaccented voices.

EXPERIMENT 2

Memory for a speaker's voice is to an important degree dependent on the duration and complexity of the speech sample (e.g., Pollack et al., 1954). Thus, increasing phonemic content of a study stimulus should improve memory for the voice. It is reasonable to suppose that, even if both accented and unaccented voices are equally well recognized when speech samples are long, reducing sample length might affect accented voices more than unaccented voices.

Method

The 36 voices used in the previous investigation were presented to 27 native American white listeners (19 women, 8 men). The study stimulus was the single word "impossible," and the test sentence, which followed immediately after the study word, was "Joe took father's shoe bench out." All testing procedures duplicated those used in the previous study.

Results and Discussion

As expected, by reducing the sample of the talker's speech repertoire contained in the study stimulus, overall accuracy of performance was reduced from 83% to 50% correct. Recognition memory was almost identical for white and black voices (56% and 55%, respectively), but Taiwanese target voices were correctly identified significantly less often (37% correct; $p < .01$).

These data, when considered in the context of the earlier findings, suggest that reduced acoustic information has a greater effect on memory processing for accented than for unaccented voices. The results also suggest that memory for only heavily accented voices is affected by reduced speech duration. Voices of black and white speakers were equally well remembered, yet voices of black speakers are (for white listeners) slightly accented and can be reliably distinguished from white speakers' voices by white listeners (Cordell, 1973).

EXPERIMENT 3

It is reasonable to argue, as we have, that accented voices are the counterparts of other-race faces. It is just as reasonable to argue that the true auditory analog of a foreign-looking face is a foreign-speaking voice. In this experiment, we compared listeners' recognition memory performance for the voices of native Spanish speakers speaking English (with a heavy accent) and also speaking Spanish. Because the results of Experiment 1 offered no evidence for believing that accented voices were less well remembered than were unaccented voices, a comparison with American speakers was not included in this experiment.[3]

Method

Speakers and listeners. Ten male speakers' voices were tape-recorded during one 2-h session. All speakers were born in either Central or South American countries and had lived in those countries most of their lives. All spoke Spanish as their first language. Familiarity with English varied between 3 months and 30 years. All spoke English with a noticeably heavy accent. Volume of the master recording was adjusted during initial recording by increasing or decreasing the gain (and monitoring the VU meters) so as to equalize loudness across the 10 voices. Further adjustment to equalize intensity across voices was done when the voices were transferred to the test tape. Sixty-eight listeners were recruited from introductory psychology courses. Listeners were all native Americans who spoke English as their first language. Formal knowledge of Spanish, or even slight familiarity with it as a spoken language (e.g., friends who are Spanish speaking), was reason enough to reject a prospective listener.

Procedure. Of the 10 voices, 6 were randomly selected to serve as targets in both the Spanish and English conditions. During the study trial, each listener, however, heard only two speakers in each language condition, and for any one listener, the two voices speaking Spanish were never the same two voices speaking English. Every listener heard the two speakers saying a brief sentence in accented English ("The pretty sky is blue with large white clouds") and two speakers reading a short sentence in fluent Spanish ("El tiró la pelota grande a la niña de vestido rojo"). During the two 10-min retention intervals between study and test trials in both language conditions, listeners were given a visual task to occupy their time. In the appropriate test trials, listeners heard 10 voices speaking accented English and 10 voices speaking fluent Spanish. The test sentences were identical to the two sentences presented to the listeners in the study trials, but the recordings of the target voices were not identical copies of the study session recordings; each speaker in the original taping session had repeated the stimulus sentence twice, so that two recordings were made of every voice. Listeners were instructed to identify those voices heard earlier in the study trial. Thirty-five listeners were given instructions that told them simply to decide for each of the 10 voices whether it was or was not one of the voices heard in the study session (ambiguous instruction). Thirty-three listeners were given these same instructions, except they were also told that exactly 2 of the 10 voices were target voices and that 8 were voices they had not heard earlier (unambiguous instruction). The order of presenting Spanish and English sentences was counterbalanced across subjects. Subjects were tested in small groups varying from three to seven persons. Subjects recorded their responses on special answer sheets.

Results and Discussion

Listeners identified 58% of the Spanish-speaking voices and 57% of these voices when speaking accented English; false alarms associated with both these hit rates were identical, 18%. Instructions had an important influence on both correct responses and false alarms, but the instruction condition did not interact with the language conditions. As expected, ambiguous instructions produced increased hit rates in the Spanish- and English-speaking conditions (63% and 61%, respectively) compared with unambiguous condition (Spanish, 52%; English, 52%). Correspondingly, in the ambiguous condition, false alarm rate was almost double that of the unambiguous condition for Spanish (22% vs. 13%) and for English, (24% vs. 12%) voices. Clearly, these data demonstrate that accented voices speaking a familiar language are as well remembered as are voices speaking incomprehensible words in a foreign language. Statistical analysis of the results supports this interpretation (all p values except those associated with the instruction variable were nonsignificant).

CONCLUSION

A confused picture emerges from these preliminary explorations directed at testing the hypothesis that the other-race effect influences voice recognition memory. On the one hand, a small amount of evidence was uncovered that suggests that the other-race effect generalizes from faces to voices. By reducing the amount of speech information available to listeners, accented voices were less well remembered than were unaccented voices. On the other hand, recognizing a voice of a native Spanish speaker uttering unintelligible sounds is no better or worse than recognizing the voice of a native American speaker talking English. Interestingly, the latter finding is not without prece-

dent. Bricker and Pruzansky (1966) and others (see Hecker, 1971, p. 42) have reported that voices of familiar speakers can still be identified when the tape-recorded speech sample is electronically reversed by playing the magnetic tape backward. Although voice recognition of backward speech is much poorer than recognition of forward speech, it is above chance, and for certain kinds of samples, it is appreciably above chance. These results, and the results of Experiment 3, suggest that the speech signal need not be understood for recognition to take place. In conclusion, although some evidence for an auditory other-race effect was discovered, the effect appears to be much less readily obtained with voices than with faces. Practically speaking, voice recognition is just as good (or as poor) for foreign voices as it is for native voices.

Is face recognition memory more or less efficient than voice recognition memory? Although these studies never directly addressed that question, constructive speculation about the answer is possible because of the evidence derived from the research reported here and from other experiments conducted in our laboratory. Data from these sources strongly imply that correctly recognizing a voice heard only once is a very difficult task for a listener. For example, with a memory load of only one or two voices and a few-second retention interval, listeners' performance levels are much below levels achieved by observers in face recognition tasks, in which memory loads are several times larger and retention intervals are much longer. Another example exists in support of this view: In a study reported elsewhere (Goldstein, Knight, Bailis, & Conover, Note 1), subjects were not aware that they had heard in the span of only a few minutes one voice (the target) repeatedly embedded among several different sets of four test voices. If this procedure were to be used in an analogous face recognition experiment, every subject would have been aware of the repeated appearance of the target face on every test trial. Finally, even testing procedures (described by Hecker, 1971) commonly employed in voice recognition research suggest that voices are harder to recognize than faces. With those procedures, voice recognition seldom approached 100% accuracy; if faces were substituted for voices, accuracy levels would almost certainly approach perfection.

If the foregoing speculations about the relative difficulty of recognizing voices and faces are accurate, then future investigators will be confronted with an interesting paradox. Recognizing voices of familiar speakers (e.g., friends, business associates) is a remarkably precise memory function, a fact supported by both laboratory demonstrations (e.g., Compton, 1963) and common experience. As we have seen, memory for unfamiliar voices is quite poor. The paradox, then, is this: Why do we have so much difficulty remembering an unfamiliar voice and so little difficulty remembering a familiar voice? Stated in other terms, how do we finally learn to remember voices so effectively when, during the early stages of the process, we appear to be so very inefficient?

REFERENCE NOTE

1. Goldstein, A. G., Knight, P., Bailis, K., & Conover, J. *The auditory analog of the other-race effect: They all look alike, but do they all sound alike?* Paper presented at the annual meeting of the Psychonomic Society, St. Louis, Missouri, November 13, 1980.

REFERENCES

BARTHOLOMEUS, B. Voice identification by nursery school children. *Canadian Journal of Psychology,* 1973, **27**, 464-472.

BRICKER, P. D., & PRUZANSKY, S. Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America,* 1966, **38**, 1441-1449.

CARTERETTE, E. C., & BARNEBEY, A. Recognition memory for voices. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception.* New York: Springer Verlag, 1975.

CHANCE, J., GOLDSTEIN, A. G., & MCBRIDE, L. Differential experience and recognition memory for faces. *Journal of Social Psychology,* 1975, **102**, 243-253.

CLIFFORD, B. R., & BULL, R. *The psychology of person identification.* London: Routledge & Kegan Paul, 1978.

COMPTON, A. J. Effects of filtering and vocal duration upon the identification of speakers, aurally. *Journal of the Acoustical Society of America,* 1963, **35**, 1748-1752.

CORDELL, J. *The identification of voice of speakers belonging to two ethnic groups.* Unpublished doctoral dissertation, Ohio State University, 1973.

HECKER, M. H. L. Speaker recognition: An interpretive survey of the literature. *American Speech and Hearing Association Monographs,* 1971, **16**, 1-122.

MALPASS, R. S., & KRAVITZ, J. Recognition for faces of own and other "race." *Journal of Personality and Social Psychology,* 1969, **27**, 330-334.

MANN, V. A., DIAMOND, R., & CAREY, S. Development of voice recognition parallels with face recognition. *Journal of Experimental Child Psychology,* 1979, **28**, 153-165.

MCGEHEE, F. The reliability of the identification of the human voice. *Journal of General Psychology,* 1937, **16**, 249-271.

POLLACK, I., PICKETT, J. M., & SUMBY, W. H. On the identification of speakers by voice. *Journal of the Acoustical Society of America,* 1954, **26**, 403-406.

STEVENS, K. N., WILLIAMS, C. E, CARBONELL, J. R., & WOODS, B. Speaker authentication and identification: A comparison of spectrographic and auditory presentation of speech material. *Journal of the Acoustical Society of America,* 1968, **40**, 1596-1607.

NOTES

1. In March 1932, a few days after his son was kidnapped, Lindbergh, in the company of a go-between, and in accordance with the kidnapper's orders, went to a cemetery to deliver the ransom money. While sitting in the car he had come in, Lindbergh heard someone in the graveyard calling to the go-between, "Hey, doctor," or "Hey, doctor, over here," depending on the account one reads. In October 1934, Lindbergh was called by police to listen to Bruno Hauptmann say the words heard in the cemetery. Lindbergh later testified that Hauptmann's voice was the voice of the person who had spoken in the cemetery. Lindbergh made this identification knowing that the police had other information about Hauptmann that connected him to the crime. All laboratory voice recognition evidence collected to date would suggest that Lindbergh could not have made a reliable identification of Hauptmann's voice after an interval of more than 2 years.

2. This conclusion is based on data collected only from subjects listening to speakers who were familiar to them through daily contacts. Further research is needed to determine the optimal speech duration for recognizing barely familiar voices.

3. The authors are most grateful to Carolyn Backer for making the master tape and for collecting the data in this experiment.

(Received for publication January 30, 1981.)