ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Experimental Social Psychology 38 (2002) 618–625

Journal of Experimental Social Psychology

www.academicpress.com

Inferring speakers' physical attributes from their voices

Robert M. Krauss,* Robin Freyberg, and Ezequiel Morsella

Department of Psychology, Columbia University, 1190 Amsterdam Avenue, New York, NY 10027, USA

Received 10 September 2001; received in revised form 15 January 2002

Abstract

Two experiments examined listeners' ability to make accurate inferences about speakers from the nonlinguistic content of their speech. In Experiment I, naïve listeners heard male and female speakers articulating two test sentences, and tried to select which of a pair of photographs depicted the speaker. On average they selected the correct photo 76.5% of the time. All performed at a level that was reliably better than chance. In Experiment II, judges heard the test sentences and estimated the speakers' age, height, and weight. A comparison group made the same estimates from photographs of the speakers. Although estimates made from photos are more accurate than those made from voice, for age and height the differences are quite small in magnitude—a little more than a year in age and less than a half inch in height. When judgments are pooled, estimates made from photos are not uniformly superior to those made from voices.

Most people have had the experience of seeing for the first time a speaker whose voice is familiar (from telephone conversations, the radio, etc.), and being surprised by that person's appearance. The fact that people are surprised in such situations suggests they expect their mental images of speakers to have some degree of verisimilitude. To what extent are such expectations justified? More generally, what do we know about the inferences listeners make from speakers' voices?

It has long been known that, quite apart from what is said, a speaker's voice conveys considerable information about the speaker, and that listeners utilize this information in evaluations and attributions. Giles and Powsland (1975) provide a useful (albeit now somewhat outdated) review of research on this topic. Perhaps the most familiar example of how listeners spontaneously use variations in speakers' voices is the biasing effect of dialects associated with social class. Status variation in language use occurs in most societies (Guy, 1988), and it is remarkable how accurately naïve listeners can utilize these variations to identify a speaker's socioeconomic status (SES). Judgments of SES based on hearing speakers read a brief standard passage are highly correlated with measured SES, and even so minimal a speech sample as counting from 1 to 10 yields reasonably accurate judgments (Ellis, 1967). Lower (and working) class speakers tend to be judged less favorably than middle-class speakers (Smedley & Bayton, 1978; Triandis & Triandis, 1960), and middle-class judges perceive themselves to be more similar to middle-class speakers than to lower class speakers (Dienstbier, 1972).

One might expect that research on the inferences listeners make from speech would be part of the study of speech perception, but for interesting reasons that is not the case. For speech perception researchers, the fundamental issue has been one that is common to all psychological studies of perception: *constancy*. Spoken language shows variability in its realization, but stability in its perception, and the primary goal of speech perception research is to explain how this is accomplished—how a perceiver arrives at a stable percept from a highly variable stimulus. Goldinger makes the point with regard to word recognition:

> Most theories of spoken word identification assume that variable speech signals are matched to canonical representations in memory. To achieve this, idiosyncratic voice details are first normalized, allowing direct comparison of the input to the lexicon (Goldinger, 1995, p. 1166).

* Corresponding author. Fax: +1-212-854-3609.
E-mail address: rmk@psych.columbia.edu (R.M. Krauss).

0022-1031/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S0022-1031(02)00510-3

Comprehending speech requires the hearer to distinguish between variability in the acoustic signal that is linguistically significant (i.e., that contributes to comprehension of the utterance's intended meaning) and variability that is not. A great deal of the variability found in speech does not contribute to comprehension, while at the same time tokens of the same linguistic type (that must be perceived as equivalent for purposes of comprehension) can differ markedly in their realization.

Some of this variability is the result of language-specific coarticulation rules and typically goes unnoticed by the listener, but some of it reflects important attributes of the speaker that can serve as a basis for inferences about his or her identity, attitude, emotional state, definition of the situation, etc. For example, systematic variation in the articulation of certain phonemes distinguishes dialects and accents. Dialects are associated with speech communities, and reflect regional origin and SES. Stereotypes associated with the speech communities (Southerners are stupid, New Yorkers are venal and rude, poor people are lazy) affect the way the speaker's behavior is perceived (Giles & Powsland, 1975). Variation in fundamental frequency (F_0), amplitude, rate and fluency may be related to momentary changes in the speaker's internal state. The most intensively investigated of these internal states is affective arousal. F_0, amplitude and syllabic rate increase, and fluency decreases, when arousal is high (Hecker, Stevens, von Bismarck, & Williams, 1968; Streeter, Krauss, Geller, Olson, & Apple, 1977; Streeter, Macdonald, Apple, Krauss, & Galotti, 1983; Williams & Stevens, 1972)—but it is likely that finer distinctions could be made.

Anatomical differences constitute another source of variability. Speakers' vocal tracts differ, and each produces a signal that is acoustically distinctive, although the audible differences between any pair of voices may be small and not readily discernible. Gross differences in the vocal tract are related to inter-individual differences on a number of personal attributes. Perhaps the most familiar is age. The physiological changes that mark the progression from infant to toddler to adolescent to adult are paralleled by striking changes in voice quality; only slightly less familiar are the vocal changes that accompany the transition from adulthood to old age (Caruso, Mueller, & Shadden, 1995; Ramig, 1986; Ramig & Ringel, 1983). Anatomy also accounts for some of the difference among the voices of speakers of the same age. Just as children's voices deepen as their size increases, adult speakers who are large tend to have lower, more resonant voices than speakers who are small, although the correlation is far from perfect. In all likelihood there are other acoustic correlates of size and physique, although they are not uncomplicated.

Several investigators have reported relationships between naïve listeners' estimates from voice samples of such attributes as age, height, and weight and the actual values (Allport & Cantril, 1934; Lass & Colt, 1980; Lass & Davis, 1976; van Dommelen, 1993). Unfortunately, differences in method, sample characteristics and measures make it difficult to reach general conclusions about how accurate naïve listeners' estimates are. In the typical study, a relatively large number of listeners hears samples of speakers' voices, and estimates each speaker's age (or some other attribute). The mean estimate for each speaker is calculated, and the average difference between mean of the estimated ages and the actual ages is used as a measure of accuracy. Although such statistics are often presented as an index of people's accuracy in estimating age from voice, what they really reflect is the accuracy of judges' pooled estimates. For example, Lass and Colt (1980) reported a mean difference between a speaker's actual height and height estimated from voice to be -1.4 in for female speakers and -0.49 in for male speakers. These values represent the difference between the mean of judges' estimates of speakers' heights and the mean actual height in the sample of speakers, and tell us little about how accurately the height of an individual speaker is likely to be estimated by the average judge.

Nearly all of the previous studies have used speech samples drawn from college populations, which restricts the range of such variables as age. In the experiments reported here, we took pains to obtain a more heterogeneous sample of speakers. Using this sample, we examined the ability of listeners to match speakers' pictures to their voices and to estimate speakers' physical attributes from their voices. In Experiment I, naïve listeners heard speakers reading standard test sentences, and then saw a pair of pictures. Their task was to identify the pictures of the speaker. In Experiment II, judges heard the test sentences and estimated the speaker's age, height, and weight. For comparison purposes, another set of judges made the same estimates from photographs of the speakers.

Experiment 1. Speaker identification

Method

Collection and processing of stimulus materials

Weekend strollers in New York City's Central Park were asked to participate in a research project described as a study of voices. People who were under 20, were involved in athletic activities, or who were not native speakers of English were excluded. About 90% of those approached agreed to participate. Although an attempt was made to draw a representative sample, the exigencies of working in this natural setting did not permit implementation of a formal sampling plan and the experimenter was allowed to exercise some discretion in deciding whom to approach. Means, standard devia-

Table 1
Descriptive statistics for speaker sample

	Age (years)			Height (in.)			Weight (lbs)		
	Mean	Range		Mean	Range		Mean	Range	
		Min	Max		Min	Max		Min	Max
Male speaker ($n = 20$)	32.3 (6.50)	25	52	70.6 (3.25)	66	78	176.3 (40.92)	110	260
Female speaker ($n = 19$)	30.6 (9.71)	20	60	65.3 (3.25)	61	72	127.9 (14.07)	107	160

Values in parentheses are standard deviations.

tions, and ranges for speakers' age, height, and weight are shown in Table 1.

Participants first completed a short questionnaire that asked their height, weight, and age, their region of origin, and their own and their parents' years of educaion. Then, they recorded two test sentences: "Joe took father's shoe bench out" and "She is waiting at my lawn"[1] using a Sony WM-D3 cassette recorder and a handheld Sony ECM-MS907 microphone. They did this twice. Finally, a full length, frontal view photograph was taken of the participant in front of a neutral background. We took care that no objects that could serve as cues to size were visible in the foreground. A total of 40 participants (20 males and 20 females) constituted the sample of speakers for this research.

The better of each speaker's two speech samples was digitized and edited on a Macintosh 7100/80AV computer, and converted to 44.1 kHz, 16 bit, System seven sound files.[2] The photographs were digitized, and edited to standardize image size and brightness.

Participants

Fifteen Columbia undergraduates (7 males and 8 females) performed the identification task. Their participation fulfilled an undergraduate course requirement.

Procedure

Stimuli were presented and responses recorded on a Macintosh 7100 AV computer using the PsyScope software package (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants first entered their name and sex, and then read instructions. The experiment consisted of a series of 120 trials. On each trial a voice was heard reading the two test sentences, followed 1000 ms later by the display of two photographs. One of the photos (the target) was of the person whose voice had just been heard, the other (the distracter) was randomly selected from the remaining 19 speakers of the same sex as the target. The side of the screen on which the target and distracter appeared varied randomly. Participants went through three blocks of 40 trials. In each block, each speaker's voice was heard only once and each speaker's photograph appeared only once as a target.

Results

Because one of the digitized sound files turned out to be defective, the analysis is based on the remaining 39 voice samples. On average, the speaker's photograph was selected on 76.5% of the trials.[3] Although participants differed considerably in how accurate they were (67–81% correct), all 15 were reliably more accurate than the chance expected value of 50%, and males and females were equally accurate ($F(13) < 1$).

Female speakers were identified marginally better than male speakers (79% vs. 74.1%), but the difference was not statistically reliable ($t(37) = 1.15$, $p = .26$). A speaker's age was positively correlated with how accurately he or she was identified ($r(37) = .32$, $p < .05$). Neither height nor weight were correlated with identification accuracy.

Discussion

It is clear that naïve listeners can match speakers to photographs with considerable (although less-than-perfect) accuracy. The correlation we found between age and accuracy probably is an artifact of the positively skewed age distribution of our sample of speakers. Since most speakers were in the 20–35 year age bracket, older targets were likely to be paired with younger distracters,

[1] These sentences were chosen because they provide a good sampling of the American-English vowel space. Physical differences among speakers are most likely to be seen in vowels, which reflect the resonant properties of the vocal tract.

[2] Despite their having been made in a public setting, only a minimal amount of background noise is audible in the speech samples. Having speakers record the speech samples twice permitted us to select the one that minimized background noise and speaker dysfluency.

[3] Judging from their photographs, 5 of the 40 speakers were African-Americans. In urban areas in the northeastern US, many African-Americans speak an identifiable dialect (Labov, 1996) and, because that dialect is associated with a visible feature, we considered removing African-American speakers from the data analysis on the grounds that their presence would artificially inflate accuracy. In fact, accuracy with African-American speakers removed was marginally *higher* (77.7 vs. 76.5%), so we decided to include all speakers in this and subsequent analyses.

making discrimination relatively easy. Because height and weight were more symmetrically distributed, they were less useful cues. The finding suggests the possibility that listeners performed the identification task by estimating speakers' characteristics from their voices, and then selecting the photograph that most closely matched these estimates. Experiment 2, in which listeners estimate speakers' physical attributes from their voice samples, allows us to examine this possibility more directly.

Experiment 2. Judging speaker attributes from voice

Method

Participants

Forty Columbia University undergraduates (14 males and 26 females) served as judges. Their participation fulfilled a course requirement.

Procedure

Twenty judges (8 males and 12 females), seated in front of a computer monitor, heard the voice samples used in the Speaker Identification task presented in random order. After each sample was played, in response to on-screen prompts, judges entered their estimates of the speaker's age, height, and weight using the computer keyboard.[4] The order in which the attributes were presented was varied randomly.

An additional 20 judges (6 males and 14 females) made the same estimates from the speaker's photograph. Except that the attributes were judged from photographs rather than voice samples, the two conditions were identical.

Results

Selecting a measure to index accuracy is not a completely straightforward matter, because exactly what constitutes accuracy in social perception is not self defining. As Cronbach pointed out in a series of classic papers (Cronbach, 1955; Gage & Cronbach, 1955), correlations between actual and estimated scores (a common way of indexing accuracy in social perception research) can be decomposed into several independent components of variance, each of which taps an aspect of what might meaningfully be regarded as accuracy. For example, in some circumstances it might be more important for judges to be able to rank order individuals correctly than to assign absolute numerical values to them. In other circumstances, the ability to estimate the group mean for a category of individuals may be more important than the ability to distinguish among members of a category.

The most direct index of accuracy for our purposes is the average of the absolute difference between estimated and actual values (AD)—the mean of the absolute differences between judges' estimates of speakers' values on an attribute and the speakers' actual value. The AD measure indexes judges' average error in estimating a particular attribute. It answers the question "How close is the average estimate of attribute X to the actual value of X?" Another measure of interest is the mean of the pooled absolute differences between estimated and actual values (PAD)—the mean of the absolute differences between the average of estimates and the actual value. This index reflects how close, on average, the means of judges' pooled judgments are to the actual values. An index used in much previous research in this area is what we will call the mean algebraic difference (MD)—the mean of the differences between judges' estimates and actual values. This index reflects how close the mean of the distribution of estimates is to the mean of the distribution of actual values.

The AD index seems to capture the intuitive sense of accuracy, while the PAD measure provides an index that might be useful for some practical purposes. The MD measure seems to be of least theoretical or practical value, since the accuracy of a group of people in estimating the mean of a distribution is not often of great interest. The way these indexes are calculated constrains their magnitudes. In terms of their relative magnitudes, AD $\geqslant$ PAD $\geqslant$ MD. Although correlation essentially reflects a judge's ability to rank order the samples on an attribute, which is a different from the kind of accuracy we are interested in, we also performed a correlational analysis to allow comparison of our findings with those of prior studies.

We calculated a 2 (speaker sex: males vs. females) $\times$ 2 (medium: voice vs. photo) ANOVAs with AD and PAD for age, height, and weight as dependent variables. The means and standard deviations are shown in Table 2. Looking first at AD, speakers' age and height are judged slightly more accurately from photos than from voice. Although the differences are statistically reliable ($F(1,37) = 6.65$, $p < .01$ and $F(1,37) = 8.50$, $p < .01$, for age and height, respectively) they are quite small in magnitude—a little more than a year in age and less than a half inch in height. For neither attribute do the effects of speaker's sex or the interaction of sex and medium (voice vs. photo) approach statistical significance (Fs < 1). Weight estimates are more complicated. A male speakers' weight is much more accurately estimated from his photo than from his voice, although both estimates have a substantial margin of error. Female speakers' weights are more accurately estimated than males', but only slightly better from voice than from a photo. For weight,

[4] Listeners also were asked to indicate whether the speaker was male or female. Since all judgments were correct, we will not present these data.

Table 2
Average absolute difference (AD) and average pooled absolute difference (PAD) between estimated and actual age, height, and weight judged from voice and photograph

	Average absolute difference (AD)		Average pooled absolute difference (PAD)	
	Voice	Photo	Voice	Photo
Age				
All speakers	7.11 (3.49)	5.89 (3.77)	4.39 (4.38)	4.59 (4.43)
Male speakers	6.68 (2.30)	5.59 (2.87)	3.50 (3.07)	4.21 (3.69)
Female speakers	7.57 (4.44)	6.20 (4.60)	5.33 (5.36)	4.98 (5.177)
Height				
All speakers	2.94 (1.45)	2.46 (1.40)	2.41 (1.78)	1.96 (1.74)
Male speakers	2.81 (1.39)	2.50 (1.56)	2.16 (1.80)	1.88 (1.97)
Female speakers	3.07 (1.53)	2.42 (1.25)	2.68 (1.77)	2.04 (1.51)
Weight				
All speakers	25.59 (18.10)	19.95 (12.53)	22.13 (20.30)	14.95 (15.01)
Male speakers	34.76 (20.43)	23.27 (13.82)	31.37 (23.53)	16.07 (17.68)
Female speakers	15.93 (7.67)	16.45 (10.23)	12.40 (9.50)	13.76 (11.96)

Values in parentheses are standard deviations. For males speakers, $n = 20$; for female speakers, $n = 19$.

ANOVA reveals statistically significant effects due to sex ($F(1, 37) = 9.40$, $p < .01$), medium ($F(1, 37) = 12.17$, $p < .01$), and their interaction ($F(1, 37) = 13.79$, $p < .01$).

Examination of the results for PAD presents a slightly different picture. Pooling judges' estimates yields a closer approximation to the actual value of the attribute. Also, with PAD as index, in most cases the differences between estimates made from photos and voice are smaller than was the case for AD, and estimates made from photos are not uniformly the more accurate. For age, neither sex, nor medium, nor their interaction differ reliably (all Fs < 1). For height, estimates made from photos are more accurate than those made from voice ($F(1, 37) = 4.565$, $p = .0393$), although the average difference is less than a half inch. Females' weights are more accurately estimated than males' weights both from voice and from photos ($F(1, 37) = 4.546$, $p = .04$). Estimates of males' weights made from photos are considerably more accurate than those made from voice, but for females' weights the differences are negligible (Interaction $F(1, 37) = 18.51$, $p = .0001$). As would be expected, height and weight are correlated in our sample, but the relationship is stronger for males ($r = .83$) than for females ($r = .345$).

Individual (by judge) correlations between estimated and actual age, height, and weight parallel the results found for the difference measures. The values are shown in Table 3. Estimates of age made from voice computed on all speakers are highly correlated with speakers' actual age; the mean value was 0.61, and all 20 individual correlations were significant beyond the

Table 3
Mean of individual correlations between estimated and actual age, height, and weight judged from voice and from photograph

	Voice		Photo	
	Mean r	$p < .10$	Mean r	$P < .10$
Age				
All speakers	0.61 (0.09)	20	0.62 (0.10)	20
Male speakers	0.70 (0.11)	20	0.63 (0.13)	18
Female speakers	0.59 (0.14)	19	0.63 (0.10)	19
Height				
All speakers	0.54 (0.14)	19	0.67 (0.10)	20
Male speakers	0.29 (0.19)	6	0.52 (0.17)	17
Female speakers	0.04 (0.32)	5	0.44 (0.18)	14
Weight				
All speakers	0.55 (0.09)	20	0.77 (0.07)	20
Male speakers	0.16 (0.29)	6	0.78 (0.05)	20
Female speakers	0.09 (0.39)	4	0.52 (0.12)	15

Values in parentheses are standard deviations. Also shown are the number of judges (out of 20) whose estimates produced rs in the predicted direction associated with $p \leqslant .10$.

.05 level. The magnitude of these correlations is roughly the same as those for estimates made from photos. For height and weight, correlations made from voice, while substantial (0.54 and 0.55, respectively), are somewhat smaller than estimates made from photos (0.67 and 0.77). Because the distributions of height and weight differ for men and women, computing correlations on the two categories separately truncates the range of the variable, with a predictable effect on the correlation coefficient. The magnitude of correlations for age (which is distributed comparably in the two samples) is not affected in this way, although halving the *df* reduces slightly the number of correlations that are significant.

How do listeners perform the picture identification task in Experiment 1? One possibility previously mentioned is that they estimate a speaker's age, height, and weight from his or her voice, and then select the photograph that seems closest on those attributes. If that were the case, one would expect that how accurately a speaker's attributes were estimated in Experiment 2 would predict how reliably that speaker was identified in Experiment 1. Such a relationship does seem to exist. A multiple regression model with AD for age, height, and weight as the independent variables accounted for 28% and 12% of the variance in identification accuracy for female and male speakers, respectively. Apparently estimates of age, height, and weight do contribute to our listeners' ability to identify a speaker's photograph, but they account for only a small part of it.

General discussion

After hearing a brief voice sample, naïve listeners can select the speaker's photograph from a pair of photographs with better-than-chance accuracy. Naïve listeners also can estimate a speaker's age, height, and weight from a voice sample nearly as well as they can from a photograph. When judges' judgments are pooled, estimates made from voice are about as accurate as estimates made from photographs.

Since all speakers said the same test sentences, judgments of speakers' age, height, and weight had to have been based on acoustic variation that is not linguistically significant. Such variation can derive from at least two sources. One source is anatomical—differences in speakers' size, shape and physical condition can produce differences in the way they sound. The point is easiest to illustrate with the variations that make it possible to identify a speaker's sex. Men and women differ anatomically, and some of these differences affect the sounds they produce. Men tend to be larger and more muscular than women, and this has consequences for the thickness of their vocal chords and the architecture of their vocal tracts that affect the pitch and timbre of their voices. However, identifying the acoustic features that enable listeners to distinguish male from female voices is not a simple task (Klatt & Klatt, 1990). Most likely a configuration of attributes, each of which is less-than-perfectly related to the criterion, is involved. The acoustic features that serve as cues to age, height, and weight are considerably more diffuse, and correspondingly more difficult to specify.

A second source of acoustic cues is cultural. People learn to use their voices in ways that are culturally determined. Although the architecture of the vocal tract constrains the sounds a speaker can produce, the range of possibilities that remain is quite considerable. As is the case with other behaviors performed in social situations, some of this variability is under normative control—that is to say, cultures designate "ways of talking" that are considered appropriate or desirable for particular categories of speakers. Some of the difference in the way men and women speak is accounted by differences in the way they use their voices. For example, a speaker's range is constrained by larynx mass, but cultural norms may dictate where within that range the speaker "places" his or her voice. Japanese women traditionally have been expected to speak more politely than men, and one way of expressing politeness is by using the upper range of the register. One might expect the speech of Japanese males and females to become less differentiated as differences in gender roles diminish, and there is some evidence that this is occurring (Horvat, 2000). English-speaking males and females also may differ in how they place their voices. The correlation between basal F_0 (the lowest tone a speaker's can produce) and F_0 while speaking is considerably larger for men than for women, probably a result of women trying to place their voices in their midranges and men favoring the lower part of their range (Gradol & Swann, 1983). Our speakers may have been identifiable as males or females because they articulated the test sentences in a stereotypically masculine or feminine manner. However, while it is possible that culturally defined speech norms helped listeners judge speakers' gender and, conceivably, age, the idea that there are speech norms related to height or weight is considerably less plausible. In any event, we cannot specify with any confidence the acoustic properties of voices that made it possible for listeners to estimate speakers' attributes as well as they did.

Any generalization about accuracy must take into account the way the estimated attribute is distributed in the sample. For example, the fact that AD for speakers' ages was 7.1 years would be unimpressive if the estimates were based on a sample of undergraduate speakers, where so large an interval might include 95% of the population. Given our more heterogeneous sample, and the fact that estimates made from photos

are only marginally better, our naïve listeners' accuracy is more interesting. The fact that estimates of height from voice are within three inches of the speaker's actual height (and only a half inch less accurate than estimates made from photos) is particularly remarkable.

It should be noted that virtually all of the studies reported in the literature have drawn their participants from undergraduate populations, a limitation that constrains not only the distribution of age, but of such attributes as weight, social class, regional origin, and, of course, education. All of these can be reflected in speech. Although our sample is considerably more heterogeneous than those used in any other studies of which we are aware, it certainly is not a representative sample of the US population. Not surprisingly, New York City and its environs is the region of origin for most of our speakers. The speakers in our sample averaged about 2 in taller and 12 lbs lighter than the means for their age categories in the US population according to norms published by the Center for Disease Control. And the fact that the speakers in our sample chose to spend their Sundays in the park rather than engaged in other pursuits may produce a bias whose effect we can't assess.

The finding that pooled group estimates of speaker attributes made from voice samples were about as accurate as those made from photographs of the speakers suggests a possible practical application. In an effort to identify anonymous callers who have phoned in bomb threats, harassing messages, etc., law enforcement authorities often turn to speech experts for clues to the speaker's identity. Our findings suggest that quite accurate estimates of the speaker s age, height, and weight could be obtained simply by having a dozen or so naïve listeners judge these attributes, and averaging their estimates. Although dialect specialists probably can identify subtle clues to a speaker's regional origin that a naïve listener could not detect, it is difficult to imagine them improving on the accuracy of our naïve judges' pooled estimates of age, height, or weight.

Acknowledgments

The data reported here were gathered as part of an undergraduate Honors Research project at Columbia University by Robin Freyberg, who is now at Rutgers University. A pilot study conducted by Rachel Wohlgelernter contributed to the planning of this research. We gratefully acknowledge the comments and suggestions of Julian Hochberg, Jennifer Pardo, Lois Putnam, and Robert Remez, the technical advice of Niall Bolger and Elke Weber, and the assistance of Anne Ribbers, Ariel Dolid, and Anna Marie Nelson.

References

Allport, G. W., & Cantril, H. (1934). Judging personality from voice. *Journal of Social Psychology, 5*, 37–55.

Caruso, A., Mueller, P., & Shadden, B. B. (1995). Effects of aging on speech and voice. *Physical and Occupational Therapy in Geriatrics, 13*, 63–80.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, and Computers, 25*, 257–271.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin, 52*, 177–193.

Dienstbier, R. A. (1972). A nodified theory of prejudice emphasizing the mutual causality of racial prejudice and anticipated belief differences. *Psychological Review, 79*, 146–160.

Ellis, D. S. (1967). Speech and social status in America. *Social Forces, 45*, 431–437.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62*, 411–422.

Giles, H., & Powsland, N. F. (1975). *Speech style and social evaluation*. New York: Academic Press.

Goldinger, S. D. (1995). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 22*, 1166–1183.

Gradol, D., & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech, 26*, 351–366.

Guy, G. R. (1988). Language and social class. Linguistics: The Cambridge survey. In F. J. Newmeyer (Ed.), *Language: The sociocultural context* (pp. 37–63). Cambridge, UK: Cambridge University Press.

Hecker, M. H. L., Stevens, K. N., von Bismarck, G., & Williams, C. E. (1968). Manifestations of task-induced stress in the acoustical signal. *Journal of the Acoustical Society of America, 44*, 93–101.

Horvat, A. (2000). *Japanese beyond words: How to walk and talk like a native speaker*. Berkeley, CA: Stone Bridge Press.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*, 820–857.

Labov, W. (1996). The organization of dialect diversity in North America. Paper given at the Fourth International Conference on Spoken Language Processing. (Available at: http://www.ling.upenn.edu/phono_atlas/ICSLP4.html).

Lass, N. J., & Colt, E. G. (1980). A comparative study of the effect of visual and auditory cues on speaker height and weight identification. *Journal of Phonetics, 8*, 277–285.

Lass, N. J., & Davis, M. (1976). An investigation of speaker height and weight identification. *Journal of the Acoustical Society of America, 60*, 700–704.

Ramig, L. A. (1986). Aging speech: Physiological and sociological aspects. *Language and Communication, 6*, 25–34.

Ramig, L., & Ringel, R. (1983). Effects of physiological aging in selected acoustic characteristics of voice. *Journal of Speech and Hearing Research, 26*, 22–30.

Smedley, J. W., & Bayton, J. A. (1978). Evaluative race-class stereotypes by race and perceived class of subject. *Journal of Personality and Social Psychology, 36*, 530–535.

Streeter, L. A., Krauss, R. M., Geller, V. J., Olson, C. T., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology, 35*, 345–350.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America, 73*, 1354–1360.

Triandis, H. C., & Triandis, I. M. (1960). Race, social class, religion and nationality as determinants of social distance. *Journal of Abnormal and Social Psychology, 61*, 110–118.

van Dommelen, W. A. (1993). Speaker height and weight identification: A re-evaluation of some old data. *Journal of Phonetics, 21*, 337–341.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 233–248.