The Sounds and Sights of Intelligence: A Lens Model Channel Analysis

D'Arcy J. Reynolds, Jr.
Robert Gifford
University of Victoria

The links between 13 auditory and visual behavioral cues, measured intelligence, and observer judgments of intelligence in a zero-acquaintance context were examined in a lens model study. Auditory-plus-visual, auditory-only, and visual-only information conditions, in addition to a transcript-only control condition, were employed to determine whether auditory or visual cues encode measured intelligence more strongly and which are used more in judgments of intelligence. Five cues (of both types) accounted for nearly half the variance in measured intelligence, but it was much more strongly associated with auditory than visual cues. Observers' judgments of intelligence were also much more strongly related to auditory than visual cues. Visual cues may even depress accuracy; accuracy was higher in an auditory-only condition than in an auditory-plus-visual condition.

The judgment of intelligence in zero-acquaintance contexts is a very old topic, dating back to the early years of the 20th century. Then, as now, intelligence and closely related concepts such as competence often were judged during and after brief interactions in person, as in job interviews, on the telephone, or from written materials such as job applications. The practical importance of the issue is enormous; the consequences of being thought intelligent or not can strongly affect one's present and future success (cf. Rosenthal & Jacobson, 1968). On the perceiver side, the consequences of misjudging the intelligence of a student, job applicant, or colleague can be similarly crucial to the future of the interviewer's organization.

The present study extends this old research tradition by gathering numerous auditory and visual indicators of both measured and judged intelligence from previous studies and examining their role in the zero-acquaintance assessment of intelligence using the lens model (Brunswik, 1956; Cooksey, 1996; Gifford, 1994) and multiple information channels. The lens model (see Figure 1) includes the ecological-validity (or left) half (i.e., which objective behavioral cues are correlated with measured intelligence), the cue-utilization (or right) half of the lens model (i.e., which of the same cues are related to observers' inferences about others' intelligence), and achievement (i.e., the accuracy of observers' estimates of the targets' intelligence).

This research area began with studies that measured no specific cues as indicators of intelligence and has progressed to increasingly sophisticated investigations of visual and auditory cues to judged and measured intelligence.

Early Research: Photographs With No Measurement of Cues

The earliest studies that examined strangers' perceptions of intelligence used photographs as stimuli. The belief that individuals can estimate intelligence from photographs was not well supported in several early studies (Cook, 1939; Gurnee, 1934; Laird & Remmers, 1924; Pintner, 1918). However, two early studies that selected targets with a wide range of intelligence showed that judges could at least discriminate intelligence above chance levels (Anderson, 1921; Gaskill, Fenton, & Porter, 1927). Thus, observers' accuracy depends in part on

Authors' Note: We wish to thank David Pitre (principal) and the students of Spectrum Community School for their participation, and Sharon Ali, Joel Tourigny, and Ryanne Conley for their assistance in various phases of the study. Thanks to Del Paulhus for his early discussions on this topic. Eugene Deen, Michael Hunter, and Lorne Rosenblood's statistical expertise was invaluable during the revision of the article, as were the comments of the anonymous reviewers. We are grateful to the Social Sciences and Humanities Council of Canada, which supported this project, in part. Correspondence concerning this article should be addressed to Robert Gifford, Psychology Department, University of Victoria, Victoria, British Columbia, Canada V8W 3P5; e-mail: rgifford@uvic.ca.

PSPB, Vol. 27 No. 2, February 2001 187-200

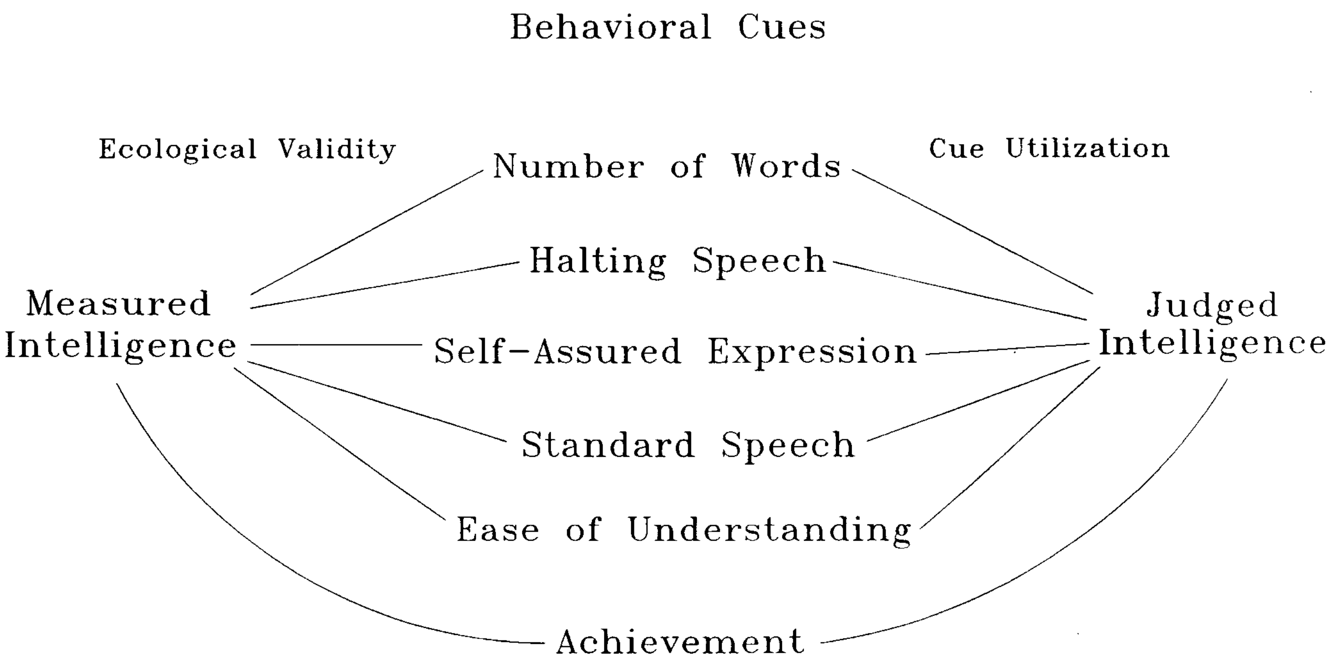

Figure 1 **A hypothetical lens model for the transmission of measured intelligence through observable cues to create impressions of intelligence.**

having a relatively wide range of target intelligence to assess, a goal toward which we strove in this study.

Judged Intelligence Based on Specific Visual Cues

Midcentury researchers began to examine whether specific cues, such as wearing glasses, smiling, and eye contact, lead observers to judge targets as more or less intelligent. Thornton (1943, 1944) found that persons who wore glasses in a photograph or in person were rated as more intelligent. However, length of exposure may affect the impact of glasses: Argyle and McHenry (1971) found that the rated intelligence of targets who wore glasses was higher when they were seen for 15 seconds than when they were seen for 5 minutes. A Japanese follow-up study reported that targets who wore glasses were judged more intelligent than targets without glasses in both 15-second and 3-minute conditions (Saito, 1978). Thus, wearing glasses seems to enhance judged intelligence, but the effect may diminish after 3 minutes and before 5 minutes.

Targets who smiled less while listening to a confederate speak were perceived as significantly more intelligent (Campbell & Rushton, 1978). In another study, a confederate who looked more while listening led judges to infer that he was low in power (Dovidio & Ellyson, 1982). The act of looking while listening may signal low power to observers but also is a marker of a person's tendency to pay attention to others while they speak, which may be related to intelligence; therefore, we investigated it as a possible indicator of intelligence.

Judged Intelligence Based on Specific Auditory Cues

In the 1960s and 1970s, the role of auditory cues such as speech rate and speech fluency was studied. For example, the synthetic voices of targets were judged as more competent when speech rate was faster (Smith, Brown, Strong, & Rencher, 1975). Judges who listened to speeches devoid of nonfluencies (i.e., "uh" and syllable repetition) rated speakers as significantly more competent than did judges who heard speeches that included nonfluencies (Miller & Hewgill, 1964). Sereno and Hawkins (1967) extended this idea by incorporating additional types of nonfluencies (i.e., word correction, sound correction, sound repetition, and word repetition) and concluded that Miller and Hewgill's findings probably could be extended to most major categories of nonfluency.

These studies, however, were concerned with the inference of intelligence, not the encoding of measured intelligence in specific, objective cues. A full understanding of the inference process requires a three-phase study: measure intelligence, record objective cues that may indicate or encode measured intelligence, and obtain observer judgments of intelligence that may be based on the objective cues.

Measured and Judged Intelligence: Lens Model Studies

The lens model approach can help identify the objective basis of valid and invalid social judgments (e.g., Brunswik, 1956; Gifford, 1994) (see Funder's, 1995, RAM model for a modern theoretical integration). Until recently, only Brunswik's (1956, pp. 26-29) study, originally reported in 1945, examined the social perception of intelligence in this three-phase manner. Using photographs, Brunswik investigated certain visual features of targets (stature, forehead height, and nose height) as cues, albeit without much success; the largest cue-criterion relation found was $r = .27$, and achievement was low.

Brunswik's (1956) study remained the only full, three-phase study of intelligence until the mid-1990s. In Germany, Borkenau and Liebler (1993, 1995) examined many auditory and visual cues of both measured and judged intelligence. Three groups of judges were shown videotapes of targets who walked on-screen and read a standard script. The targets' intelligence had been measured with a standard German intelligence test. One group viewed a sound videotape and a second group viewed a silent videotape. Both groups rated the targets' intelligence, among other attributes. The third group of judges rated 48 auditory and visual cues displayed by the targets.

Judgments of target intelligence based on visual information only (silent videotapes) were significantly related to target attractiveness, although attractiveness was not significantly related to measured intelligence. In the visual-information-only condition, in which the seven auditory cues that were related to measured intelligence (effortful reading, lower pitched voice, more pleasant voice, less halting speech, greater ease of understanding, less hectic speaking, and more standard language) were not available, the links between visual cues (that were unrelated to measured intelligence) and judgments of intelligence were stronger. For example, the correlation between friendly behavior by targets (which was not an ecologically valid cue of intelligence) and judged intelligence increased when the ecologically valid auditory information was absent. Judges in the visual-information-only condition may have relied on available but presumably invalid heuristics such as "beauty encodes intelligence" when they lacked valid auditory information.

Presumably, such errors are facilitated when few ecologically valid visual cues exist. Indeed, in Borkenau and Liebler's study, only a few visual cues were significantly related to measured intelligence: plainer (less showy) clothing, more self-assured expression, and certain aspects of the targets' walking behavior and reading abilities. The judges in the audio-plus-visual (sound videotape) condition also appeared susceptible to the misattribution of intelligence to beauty, but much less so than judges in the visual-information-only condition. This probably occurred because the seven ecologically valid auditory cues were available; in fact, five of these auditory cues were significantly related to judgments of intelligence.

As would be expected from this pattern of results, achievement was lower in the visual-information-only condition than in the audio-plus-visual condition. Zero-acquaintance judges apparently infer intelligence from visual as well as from auditory cues, but auditory cues seem to be stronger ecologically valid indicators of measured intelligence.

The Present Study

We sought to identify numerous specific objective cues that are responsible for accuracy (or the lack of it) in zero-acquaintance judgments of intelligence. We used the lens model approach, with some auditory and visual cues drawn from Borkenau and Liebler's work, some from the earlier studies, and some new ones. In selecting the cues, our overall goal was to account for measured and perceived intelligence as fully as possible. Thus, we used every cue that had been shown to predict intelligence in the previous studies reviewed above and added one (looking while listening) that seemed likely to be related to intelligence. Our specific goals were to (a) determine the ecological validity of these cues, (b) understand how these same cues are related to the judgments of target intelligence by unacquainted observers, (c) investigate the level of achievement between measured and judged intelligence, and (d) do so in a context designed to elicit responses from targets that reflects their degree of intelligence.

In addition, this study examines the assessment process separately for three information channel conditions: visual-only, auditory-only, and auditory-plus-visual condition (in addition to a transcript-only control condition). Borkenau and Liebler (1995) concluded that the auditory channel was most important for judge accuracy but did not include an auditory-only condition; rather, their conclusion was indirectly drawn from a comparison of auditory-plus-visual and visual-only conditions.

A contrasting view of the communication of personal qualities (other than intelligence) might be called the cross-modality hypothesis. Berry (1991), for example, found that auditory and visual channels independently predicted targets' self-rated power, which suggests that personal qualities are carried or encoded in both auditory and visual channels. Borkenau and Liebler (1992) also presented evidence that personality is manifested consistently across channels.

A key question, then, is whether measured intelligence is mainly manifested in auditory cues or in both auditory and visual cues. Other questions concern which specific cues reflect measured intelligence, which cues appear to signal intelligence to observers, the extent to which these two sets of cues match, and how judges' cue use appears to vary with the availability of visual and auditory information.

In addition, one might ask whether inferences from auditory information are related to the content of a target's statement, to noncontent paralinguistic cues, or to both. Borkenau and Liebler (1993) did not address this question; instead, they controlled content by asking every target to read the same script. Although this was one way to control content, content may be an important signaler of intelligence. In this study, targets were allowed to express themselves freely in response to standard questions but a written transcript condition was introduced as a control condition.

In sum, this study is similar to Borkenau and Liebler's studies in that they examined the relations between (a) some of the same cues with judges' impressions of intelligence in two independent conditions, (b) these cues and measured intelligence, and (c) the intelligence judgments in the two conditions with measured intelligence. The present study extends the work of Borkenau and Liebler in four main ways. First, a context specifically designed to elicit responses indicative of the targets' intelligence is established; second, judges are provided with an auditory-only condition to convincingly determine whether the auditory channel is the most important for accuracy; third, the role of targets' utterances is examined by allowing targets to express themselves freely in response to thought-provoking questions; and fourth, a different culture (North American rather than German) is studied.

METHOD

Participants

The targets were obtained from five classes of a local high school. We deliberately chose a high school population so that a relatively wide and representative spectrum of intelligence might be found; we wished to strengthen the generalizability of our results. The school principal assisted in this goal by giving us access to classes known to have both strong and weak students. Thirty students from these classes in the last 2 years of high school (13 women and 17 men) volunteered to participate as targets. The judges (total $N = 103$) were undergraduate university students, most of whom were enrolled in psychology classes. Most of the students received a small amount of course credit for their participation in the study. Subgroups of the 103 judges independently assessed various cues and rated the targets' intelligence in the three channel conditions and the transcript control condition (see details below).

Procedure

The targets were administered Form A of the Wonderlic Personnel Test (WPT) (Wonderlic, 1992), a brief intelligence test, as a group, in their classrooms. For their interviews (see below), each student was individually led to an unoccupied adjacent room and seated, and the video camera was then turned on. The experimenter sat down across from the target and conversed informally with the student for approximately 5 minutes to reduce his or her sensitivity to the camera.

The interview was carefully designed in three parts to suit the measurement of the cues yet not to appear contrived to the targets. First, after the informal conversation, the experimenter announced that he would describe the study and delivered a carefully scripted 1-minute speech.

Second, the targets answered five general-knowledge trivia questions. This part of the interview was taped but not presented later to judges; its purpose merely was to prime the targets for the third phase of the interview, which was designed to elicit intelligence. We assumed that judge accuracy would be optimized if the targets were presented in a context that would allow them to demonstrate their level of intelligence; to the best of our knowledge, no previous study has given targets a clear opportunity to do so.

In the third phase of the taped interview, the targets were asked three thought-provoking questions and were given 20 seconds to respond to each one. If a target spoke for less than 20 seconds, the experimenter prompted her or him once (only) for more information; if the target spoke for 20 seconds, the experimenter gently interrupted to move on to the next question. The three questions were as follows: (a) "What should be the role of the media, like TV, radio, and newspaper, in today's world?" (b) "What do you see as the future of the world environment?" and (c) "What things are essential for a good life? These 'things' can be anything you think are important for a good life."

Next, the raw videotape footage of each interview was edited down to two 1-minute segments for each target. Segment 1, during which the target only listened to the experimenter describe the study, was presented to the three independent judge groups (see Table 1) who assessed the while-listening cues (e.g., smiling and looking) and the visual impression cues (e.g., attractiveness). The second segment, during which the target answered the three thought-provoking questions for 20 seconds each, was presented to nine further groups of judges (see Table 1) who counted the number of nonfluencies,

TABLE 1: Psychometric Properties of the Main Variables

Variable	*Mean*	*Standard Deviation*	*Reliability*	*Judge Group*[a]	*Judge (N)*	*Based on Segment*
Number of nonfluencies[b]	3.71	2.43	.92[c]	1	3	2
Number of words spoken[b]	77.67	15.13	.93	2	4	2
Speech rate[d]	95.89	11.13	.96	3	2	2
Facial regard while listening[e]	55.33	6.23	.92	4	4	1
Smiling while listening[e]	10.80	11.33	.96	5	3	1
Easy-to-understand speech[f]	4.54	1.42	.85	6	7	2
Halting speech[f]	4.19	1.24	.83	6	12	2
High-pitched voice[f]	2.36	0.85	.81	6	12	2
Standard language[f]	5.33	0.98	.78	6	12	2
Loudness[f]	4.83	1.59	.93	7	2	2
Attractiveness[f]	4.32	1.04	.96	8	57[g]	1
Large body[f]	3.50	1.20	.93	8	19	1
Self-assured expression[f]	5.00	0.93	.89	8	19	1
Judged intelligence:[f]						
Audio + visual condition	4.89	1.14	.80	9	7	2
Audio-only condition	5.26	1.18	.89	10	10	2
Visual-only condition	5.20	0.58	.78	11	28	2
Transcript condition	5.41	1.05	.91	12	9	—
Measured intelligence[h]	21.23	6.45	.88 to .94[i]	—	—	—

a. Twelve independent groups of judges rated the various variables.
b. Frequency in 1 minute.
c. Based on Shrout and Fleiss's (1979) formula ICC(3, k).
d. Words per minute.
e. Number of seconds out of 1 minute.
f. Rating (0-9) with larger values labeled as the variable name.
g. This composite cue is composed of 57 ratings from 3 independent former cue groups of 19 judges.
h. Wonderlic raw score.
i. Internal consistencies from the Wonderlic manual.

the number of words spoken, and the length of speech in time (to compute speech rate), and rated the auditory cues (such as vocal pitch) and judged intelligence in the four conditions.

Three independent groups of judges who judged the targets' intelligence watched and listened (in the video-plus-audio condition), listened only (in the audio-only condition), or watched only (in the visual-only condition) the videotapes on a 15-in. television monitor. The fourth group of independent judges, in the transcript control condition, read verbatim transcripts (which included the targets' nonfluencies) of the tapes presented to auditory-only judges. Judges were randomly assigned to rating groups. See Table 1 for more information on the groups of judges and what they rated.

Measures

Measured intelligence. The WPT is a relatively brief measure of intelligence that correlates very strongly (*r*s = .75 to .96) with the Wechsler Adult Intelligence Scale–Revised Full Scale IQ (Wonderlic, 1992). It contains 50 questions that parallel the types of items in the Wechsler Adult Intelligence Scale–Revised (WAIS-R).

Judged intelligence. To assess judged intelligence, 54 judges (of the 103 total) were presented with one of the four information conditions and asked to rate how accurately the word *intelligent* described each of the targets on a scale ranging from 0 to 9. The scale was labeled *not at all like him or her* for the 0 to 2 range of the scale, *moderately like him or her* for the 3 to 6 portion, and *very much like him or her* for the 7 to 9 portion. Seven of the 54 judges rated the targets in the auditory-plus-visual condition, 10 in the auditory-only condition, 28 in the visual-only condition, and 9 in the written transcript condition. Different numbers of judges were used to attain adequate standards of interrater reliability (see Table 1).

Cues. Initially, a total of 16 cues were considered, although later 1 was discarded for lack of variability and 3 others that were highly intercorrelated were combined, leaving 13 cues. Eight independent groups of judges (49 of the 103 total; see Table 1 for details) assessed the cues. Five of these groups counted or measured cues. Three judges counted the number of nonfluencies (i.e., the number of times each target uttered "ah" or "uh"); four judges counted the number of words (i.e., the number of distinguishable words, excluding the before-mentioned nonfluencies) uttered by each target; two judges measured the time spent speaking so that speech rate could be computed; four judges measured the duration of facial regard while listening (i.e., the total number of seconds the target looked at the experimenter's face, excluding brief glances, while the experimenter spoke);

TABLE 2: Intercorrelations Between Behavioral Cues

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Attractiveness	—												
2. Easy to understand	–5	—											
3. Facial regard while listening	–10	35	—										
4. Halting speech	–9	–43	–17	—									
5. High-pitched voice	15	4	7	–3	—								
6. Large body	–39	–11	21	18	–45	—							
7. Nonfluencies	–6	–7	30	36	–24	34	—						
8. Self-assured expression	44	42	49	–20	4	–3	36	—					
9. Smiling while listening	–7	–3	30	7	24	11	13	36	—				
10. Standard language	11	35	–1	–65	18	–36	–38	–9	–14	—			
11. Number of words spoken	–45	37	37	–40	–18	19	–2	–1	–20	12	—		
12. Speech rate	–27	29	6	–37	13	4	–37	–18	5	33	44	—	
13. Loudness	–37	59	13	–16	–25	5	6	28	3	–9	35	5	—

NOTE: Decimal points are omitted. See Table 1 for variable codings. $r > .36 = p \leq .05$. $r > .47 = p < .01$.

and three judges measured the duration of smiling while listening (i.e., the overall number of seconds the target smiled while the experimenter spoke). Whether the targets wore glasses was coded by the experimenter alone.[1]

Three judge groups rated, rather than counted or measured, the following 10 auditory and visual cues on 10-point scales that were identical to the before-mentioned judged-intelligence scale, with the same instructions to rate how accurately the word or words described the target.

One group rated auditory attributes; this included seven judges who rated ease of understanding (i.e., "each word is clear") and the same seven judges in addition to five more judges (to improve reliability) who assessed each target's use of standard language (i.e., "lack of slang"), higher vocal pitch, and halting speech (i.e., "pauses, repetitions, and uneven pace"). Two judges in a second rating group rated vocal loudness. The third group ($N = 19$) rated five visual attributes of the targets: attractiveness (by the judge's own standards), showy clothing (i.e., "expensive, flashy clothes or accessories"), large body, well-proportioned body (i.e., "athletic, muscular, graceful"), and self-assured expression.

RESULTS

Reliability

Cues. Interrater agreement (Shrout & Fleiss, 1979, formula ICC[3, k]), averaging over clips and using targets as the unit of analysis, was .92, .93, .96, .92, and .96, respectively, for nonfluencies, number of words spoken, speech rate, facial regard while listening, and smiling while listening. Those for halting speech, ease of understanding, use of standard language, voice pitch, and vocal loudness were .83, .85, .78, .81, and .93, respectively. For attractiveness, showy clothing, large bodied, well-proportioned body, and self-assured expression, reliabilities were .91, .93, .93, .93, and .89, respectively.

The cues were examined for redundancy. Attractiveness, showy clothing, and well-proportioned were intercorrelated, $r = .53$ to .79, so a composite cue was created (scale alpha = .96); it was named attractiveness. A few other pairs of cues were correlated at about $r = .50$ but were not conceptually similar so were not combined. Correlations between the 13 cues are presented in Table 2.

Judged intelligence. Interrater agreement for judged intelligence (Shrout & Fleiss, 1979, formula ICC[3, k]) was .80, .89, .78, and .91, respectively, for the auditory-plus-visual, auditory-only, visual-only, and written transcript conditions.

Psychometric Properties of the Key Variables

The basic psychometric characteristics of all the variables are presented in Table 1. The mean intelligence raw score on the WPT was 21.2 (out of 50), very close to the mean score of 21.1 reported in the WPT manual, which is based on more than 118,000 administrations of the test. The targets' raw scores ranged from 9 correct (5th percentile, based on the manual) to 39 correct (99th percentile). In more familiar terms, these scores correspond to full-scale WAIS IQ scores of 78 to 136, with a mean of 102. Thus, our goal of obtaining a target sample with wide-ranging intelligence and a representative mean was fulfilled remarkably well.

The mean judged intelligence scores were 4.9, 5.3, 5.2, and 5.4 on the 0-to-9 rating scale in the auditory-plus-visual, auditory-only, visual-only, and written transcript conditions, respectively.

Lens Models for Measured and Judged Intelligence

Pearson correlation coefficients were computed for each of the three parts of the lens model; those representing ecological validity, cue utilization, and achieve-

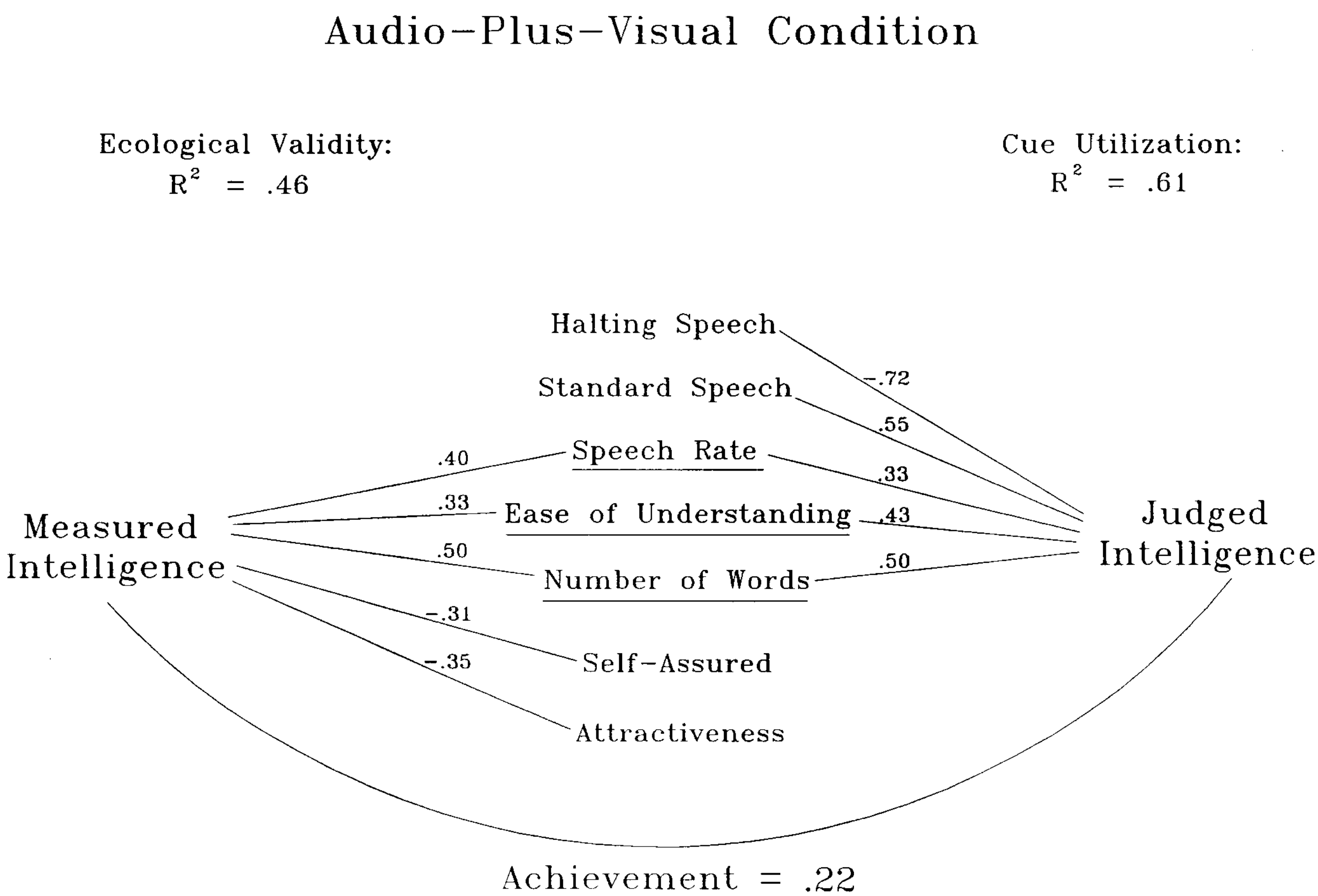

Figure 2 **The lens model for measured and perceived intelligence when judges are presented with auditory and visual information.**

ment are presented in Figures 2, 3, and 4 for each of the three conditions.[2] Matched links (where cues are underlined) represent cues that are significantly linked to both measured and rated intelligence.

Ecological validity. The left side of the lens model describes the links between measured intelligence and the visual and auditory cues. Higher measured intelligence was significantly correlated with two cues: speaking more words (r = .50, p = .005) and speaking faster (r = .40, p = .03).[3] Three other cues were marginally related to measured intelligence: greater ease of understanding (r = .33, p = .07), lower self-assurance (r = –.31, p = .09), and less attractiveness (r = –.35, p = .06).

The magnitude of ecological validity, computed as the multiple correlation between measured intelligence and these 5 cues, was R = .68 (R^2 = .46, adjusted R^2 = .35). Thus, a considerable portion of the variation in measured intelligence was accounted for by just 5 objective cues, which, it should be noted, include both visual and auditory cues. When all 13 cues were used to predict measured intelligence, R^2 was .67 (although the adjusted R^2 dropped to .39, which can be expected when 8 predictors with little predictive value are included). However, this analysis does allow for an estimate of each cue's relative relation (as beta weights) to measured intelligence after adjusting for all the other cues investigated in this study (see Table 3). Of course, these relative cue strengths are not absolute; they could change if cues beyond those examined in this study were to be investigated.

Auditory versus visual cues and measured intelligence. Are auditory cues or visual cues, as categories of information, more closely linked to measured intelligence? Hierarchical multiple regression was employed to examine this question, using the cues that were at least marginally significant, as listed above. First, the two visual cues (self-assured and attractive) were entered, followed by the three auditory cues. The addition of the auditory cues resulted in a large, significant increase (from R^2 = .16 to R^2 = .46) in variance accounted for, $F(3, 24)$ = 4.48, p = .012.

Second, the three auditory cues were entered first, followed by the two visual cues. The visual cues produced a much smaller increment in variance explained, although the increment was significant; R^2 increased

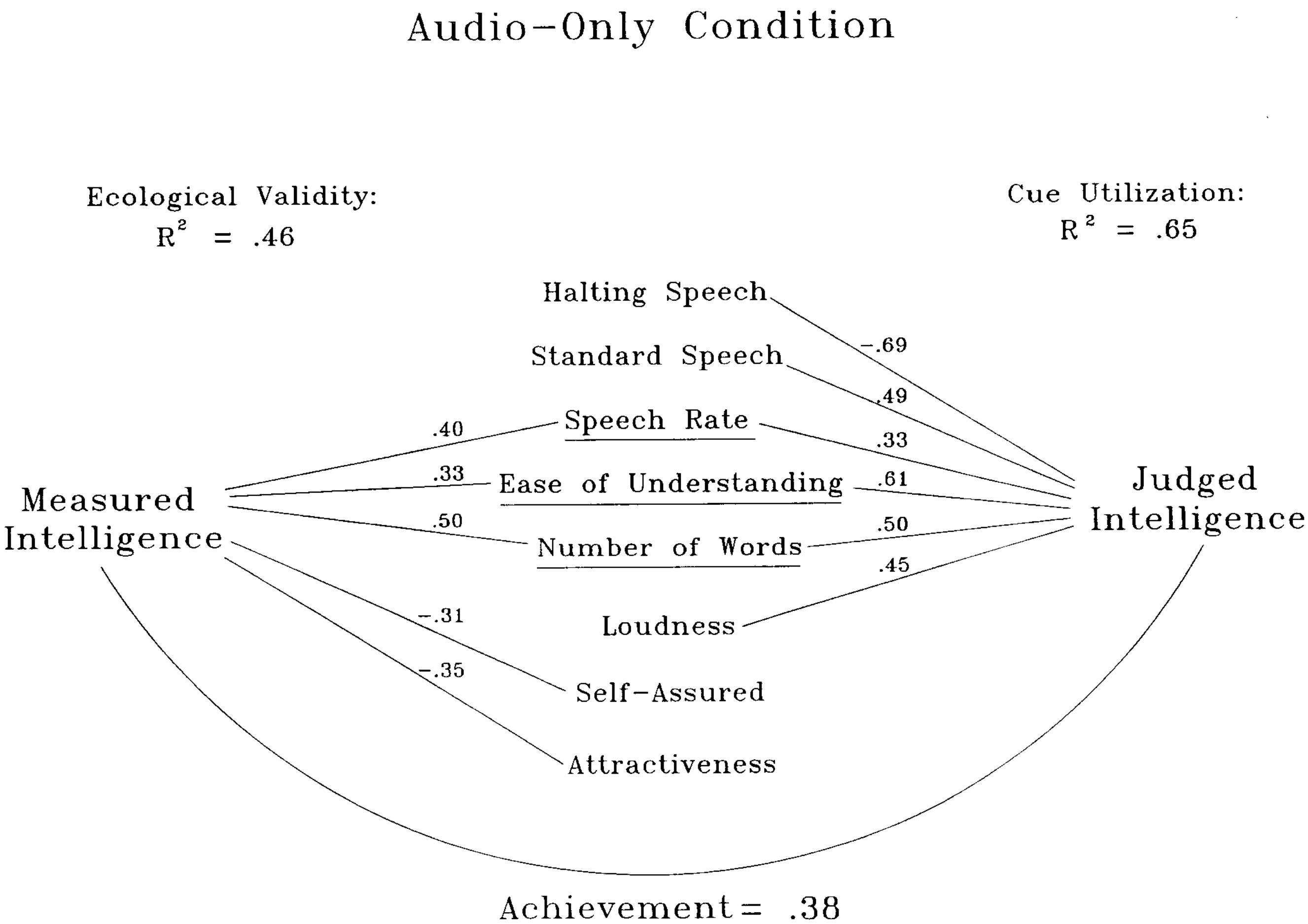

Figure 3 **The lens model for measured and perceived intelligence when judges are presented with auditory information only.**

from .31 to .46, $F(2, 24) = 3.42$, $p = .049$. Thus, both sets of cues produce a significant increase beyond that predicted by the other set. However, auditory cues accounted for twice the variance increment (R^2 increase = .30) in predicting measured intelligence that the visual cues did (R^2 increase = .15), at least among the cues examined in this study.

Cue utilization in the auditory-plus-visual condition. Judges' ratings of higher intelligence in the auditory-plus-visual condition were significantly correlated with four cues and marginally significant with one cue, all of which are auditory: more words spoken ($r = .50$, $p = .005$), less halting speech ($r = –.72$, $p < .001$), more easily understood speech ($r = .43$, $p < .05$), more standard use of language ($r = .55$, $p < .01$), and faster speech rate ($r = .33$, $p = .07$). The multiple correlation between judged intelligence and these six cues was $R = .78$ ($R^2 = .61$, adjusted $R^2 = .53$).

Auditory versus visual cues in judged intelligence. It is readily apparent that auditory cues are much more closely related to judgments of intelligence than are visual cues in an auditory-plus-visual context. No visual cues were even marginally significant predictors of judged intelligence. Thus, in a broad (channel-level) context, the judges were ecologically correct in apparently relying mainly on auditory cues and much less on visual cues.

Auditory-plus-visual information probably is the most common channel context for everyday assessments because that is what we experience in face-to-face meetings and interviews. The cue-utilization results for the auditory-only and visual-only conditions are reported next because in some contexts (e.g., telephone interviews or watching people in public) observers infer intelligence based solely on auditory or visual information.

Cue utilization in the auditory-only channel. Higher judged intelligence was significantly related in the auditory-only channel to five auditory cues and marginally related to one additional auditory cue—greater vocal loudness ($r = .45$, $p < .05$)—and the same five auditory cues that were significant in the auditory-plus-visual condition—more

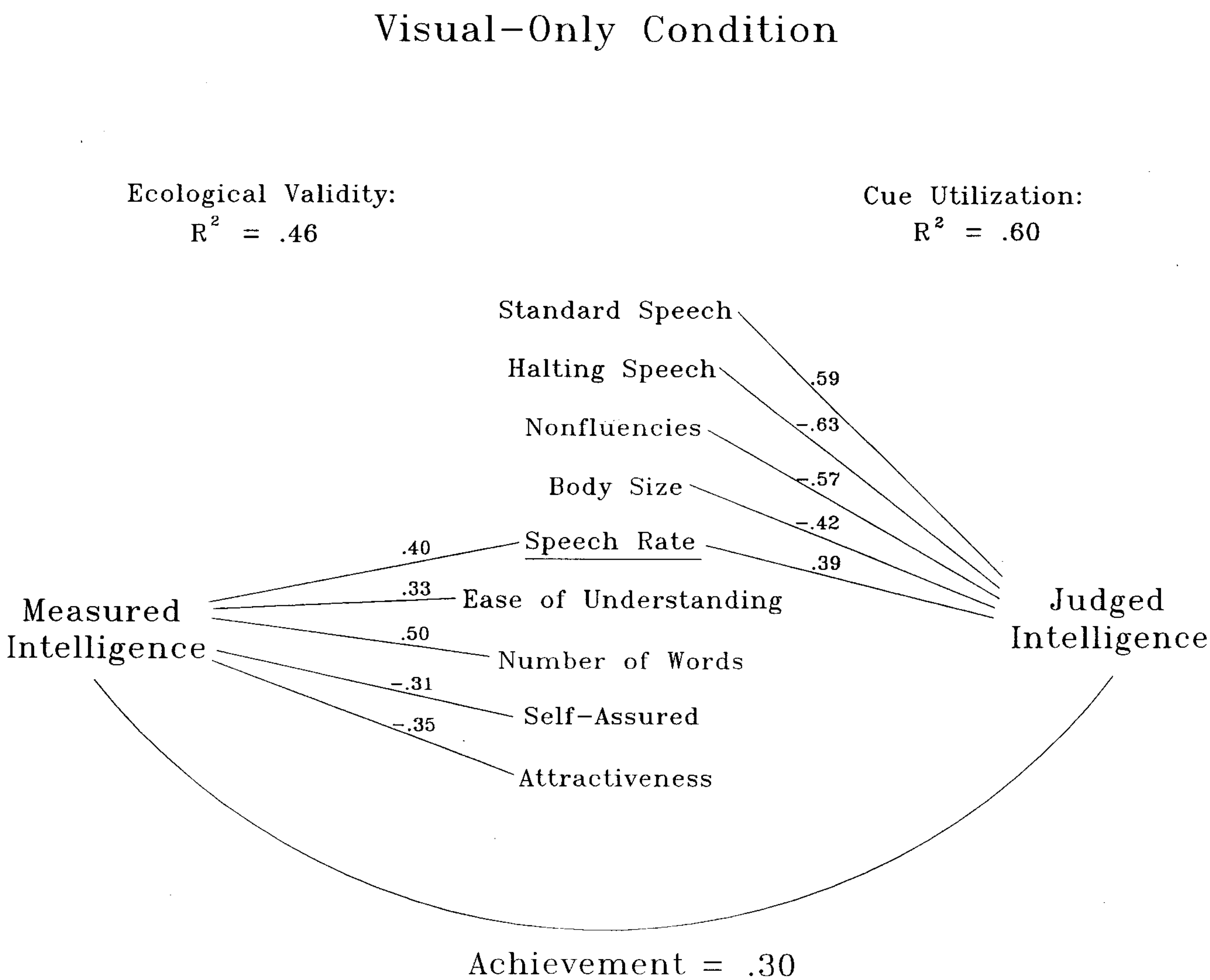

Figure 4 **The lens model for measured and perceived intelligence when judges are presented with visual information only.**

words spoken ($r = .50$, $p = .005$), less halting speech ($r = -.69$, $p < .001$), more easily understood speech ($r = .61$, $p < .001$), more standard use of language ($r = .49$, $p < .01$), and faster speech rate ($r = .33$, $p < .08$). The multiple correlation between assessed intelligence and these cues in the auditory-only condition was $R = .81$ ($R^2 = .65$, adjusted $R^2 = .56$).

Content versus style. One might ask whether, in the auditory channel, the influence of the paralinguistic cues examined (such as pitch or loudness) were affected by the content of the targets' statements. In Borkenau and Liebler's (1995) study, targets read a standardized text with content controlled so it could have no effect. This study adopted a more naturalistic approach, in which content varied because targets naturally responded to the questions in different ways. Thus, content might have influenced the relation between paralinguistic cues and judgments of intelligence.

We examined this question from two different perspectives. First, we calculated the partial correlations between judgments of intelligence in the auditory-only condition and each of the eight paralinguistic (auditory) cues investigated in this study, one at a time, while controlling for the judgments of intelligence made in the transcript-only (content) condition.[4] The links between six of the eight cues and judged intelligence changed very little (i.e., either remained significant if they had been so previously or remained nonsignificant if they had been so previously), which suggests that content had little effect on the way these cues were related to judged intelligence. However, standard speech and speech rate declined to nonsignificance as predictors of judged intelligence when content was partialed, which suggests

TABLE 3: All Cues Regressed on Judged and Measured Intelligence: Beta Weights

Cue	*Audio + Visual*	*Audio-Only*	*Visual-Only*	*Wonderlic Score*
Number of nonfluencies[a]	.28	.20	–.16	.08
Number of words spoken[a]	.09	–.05	–.01	.56[b]
Speech rate[c]	.09	–.03	.14	.04
Facial regard while listening[d]	.16	.17	–.23	.01
Smiling while listening[d]	–.20	–.03	–.05	.40
Easy-to-understand speech[e]	.06	.34	–.04	.59[f]
Halting speech[e]	–.65[g]	–.66[h]	–.42	.03
High-pitched voice[e]	–.04	–.05	.24	.04
Standard language[e]	.02	–.04	.15	–.35
Loudness[e]	.01	.10	.18	.11
Attractiveness[e]	–.12	–.27	.09	.37
Large body[e]	–.36	–.24	–.06	–.23
Self-assured expression[e]	–.15	–.26	.00	–.95[i]
All 13 behavioral cues:				
Multiple R	.86	.88	.82	.82
Multiple R^2	.75	.77	.67	.67
Adjusted multiple R^2	.54	.58	.40	.39

a. Frequency in 1 minute.
b. $t = 2.36$, $p < .05$.
c. Words per minute.
d. Number of seconds out of 1 minute.
e. Rating (0-9) with larger values labeled as the variable name.
f. $t = 2.23$, $p < .05$.
g. $t = –2.93$, $p = .01$.
h. $t = –3.10$, $p < .01$.

that content was an important part of their relation to judged intelligence.

Second, we conducted a multiple regression in which the eight paralinguistic (auditory) cues were entered first, as a group, and the intelligence judgments from the transcript condition were entered second. This time, the beta weights of five cues were unchanged but those of three cues were altered by the addition of the transcript condition. Content apparently acted as a suppressor variable in two cases: nonfluencies and loudness became marginally significant and significant predictors, respectively, of judged intelligence when they had not been so before; in contrast, halting speech declined from significance to nonsignificance.

Thus, the importance of most paralinguistic cues to judged intelligence in an auditory-only condition was unaffected by content in a naturalistic method that allowed targets to speak freely. However, content appears to influence the connection between judged intelligence and some auditory cues. The role of verbal content in judged intelligence requires further study.

Cue utilization in the visual-only condition. Judges' assessments of higher intelligence in the visual-only condition were significantly linked to less halting speech ($r = –.63$, $p < .001$), more standard use of language ($r = .59$, $p < .001$), smaller body size ($r = –.42$, $p < .05$), fewer nonfluencies ($r = –.57$, $p = .001$), and faster speech rate ($r = .39$, $p < .05$). The multiple correlation between assessed intelligence and these cues was $R = .77$ ($R^2 = .60$, adjusted $R^2 = .51$), slightly lower than in the other conditions.

Four of these five cues are auditory. How can auditory cues possibly affect ratings in a visual-only condition? Two of the most important of these five cues ($R^2 = .73$), based on a stepwise regression analysis, were halting speech and nonfluencies. Although these nominally are auditory cues, they may have important visual components, that is, observers may be able to see hesitations and breaks in speech and use them to form impressions of (lower) intelligence (cf. Bernieri & Rosenthal, 1991, p. 405 et seq.).

These results are consistent with a form of the cross-modality hypothesis: Observers may use visually detectable aspects of nominally auditory cues to reach the same conclusion about a target's personality that would be reached by judges who hear only the auditory cues. This is noteworthy in that it suggests that judgments of intelligence in a zero-acquaintance context may be heavily related to nominally auditory cues even in a visual-only condition. However, the cross-modality is heavily weighted toward the auditory channel, and in fact, the visual channel's effects appear almost epiphenomenal.

Might any of these "visually detectable" cues have been found among the visual cues measured in this study? If they did mediate between the auditory cues and judged intelligence, this should be revealed by the methods outlined by Baron and Kenny (1986). Using their method, we checked all the nominally visual cues measured in this study, and none significantly mediated the relations between measured auditory cues and judgments of intelligence. Therefore, if there are important visual aspects of the important auditory cues in this condition, they must be other than the visual cues measured in this study.

Achievement. Accuracy, or in Brunswikian terms, achievement, computed as the correlation between measured and judged intelligence, was $r = .22$ ($p = .23$), $r = .38$ ($p < .05$), and $r = .30$ ($p = .11$) for the auditory-plus-visual, auditory-only, and visual-only conditions, respectively. Although the differences between these accuracy correlations do not reach statistical significance, it is worth noting that the highest level of accuracy occurred in the auditory-only condition rather than in the auditory-plus-visual condition, which obviously contains the greatest amount of information.[5]

DISCUSSION

This study examined 13 objective cues as potential encoders of measured intelligence and the assessment of intelligence by three separate groups of zero-acquaintance judges, based on the same 13 cues, in auditory-plus-visual, auditory-only, and visual-only conditions, in addition to a content-control condition with a fourth independent set of judges. Targets who answered intelligence-eliciting questions were high school students whose collective mean, range, and distribution of measured intelligence was very similar to that of the 118,000 people on whom the WPT intelligence test was standardized.

One purpose of the study was to determine whether measured and perceived intelligence appear to be signaled primarily through auditory or visual cues. It appears that both auditory and visual cues are important indicators of measured intelligence; sets of both types of cues significantly added to the variance accounted for by sets of the other type. Nevertheless, auditory cues added twice as much variance to that explained by the visual cues as vice versa. We conclude that both auditory and visual cues significantly indicate measured intelligence but auditory cues are much more important.

Another purpose was to determine the apparent roles of specific auditory and visual cues in the communication of intelligence. The next three sections elaborate.

Ecological Validity

Higher measured intelligence was reasonably related to three auditory and two visual cues, but the auditory cues clearly accounted for more variation in measured intelligence than did the visual cues. Almost half the variation in measured intelligence could be predicted by these 5 cues; about two thirds could be predicted from the full set of 13 cues, although the adjusted variance in both cases was about the same: approximately one third of the variance. To account for a robust one third of the variance in measured intelligence from just a handful of independent, objective behavior cues seems a valuable insight into the ways that measured intelligence is manifested in observable behavior.

The specific cue findings both agree and disagree with those reported by Borkenau and Liebler (1995), who, in the most relevant and recent study on this topic, studied adult Germans reading a standard text. Ten similar cues were examined in the two studies. Four of these 10 cues (halting speech, standard speech, nonfluencies, and high-pitched voice) predicted measured intelligence in Borkenau and Liebler's study but not in the present study, although the result was vice versa for another cue (attractiveness). One reversal occurred: More self-assured expression was associated with higher measured intelligence in the other study but was marginally associated with lower measured intelligence in this study. These six discrepancies may be related to cultural differences in the judges or targets, to differences in the intelligence tests used, or to differences in the targets' tasks: All targets read the same weather report in Borkenau and Liebler's study; in this study, targets extemporaneously responded to thought-provoking, intelligence-eliciting questions.

On the positive side, two cues were associated with measured intelligence in both studies: faster speech and speech that was easier to understand. Presumably, these cues possess a robustness that transcends cultural, test, and task differences and should be regarded as having a more stable relation to measured intelligence. Finally, two cues were associated with measured intelligence in neither study: body size and vocal loudness. Thus, four cues had the same outcome in both studies, and the two indicator findings and two nonindicator findings therefore appear to be robust across cultures and paradigms.

We examined three additional cues that were not included in the Borkenau and Liebler study. Two of these, facial regard while listening and smiling while listening, were unrelated to measured intelligence. However, speaking more words, which is only moderately correlated with speech rate, was strongly correlated with measured intelligence.

Overall, the cues in this study were, collectively, quite successful in predicting the measured intelligence of a target sample that is very representative of measured intelligence levels in the general population. Of course, these were high school students in a particular culture and the judges were young people from the same culture; therefore, some relations may change with other age groups or cultures. Also, only 13 cues were examined in this study. Thus, there is room to discover other objective cues that encode other objective measures of intelligence in other groups; those are challenges for future studies.

Such research also will have to consider the very nature of intelligence, a difficult question. The measure of intelligence used in this study (the WPT) correlates very strongly with overall scores on the WAIS, a standard general measure of intelligence. Nevertheless, the relations found in this study may not hold for other objective measures of intelligence. Also, the WPT was not designed to measure specific components of intelligence; it is too brief. If we assume that the components of intelligence measured by the WAIS-R reflect most key components of intelligence (while accepting that some theorists suggest yet other possible components of intelligence), subsequent studies might examine the links between behavioral cues and particular WAIS-R components or other conceptions of intelligence. For example, if intelligence also includes broad nonverbal components, an interesting challenge would be to discover behavioral cues that might reflect that component. How do people with differing spatial intelligence or form perception look and act?

Cue Utilization

Associations between judges' impressions of intelligence and particular behavior cues were very similar in the auditory-plus-visual and auditory-only conditions. Five of the same auditory cues were related to the judges' inferences in the same way in the two conditions, and a very similar amount of variance was accounted for, even though independent groups of judges were used in the two conditions.

Oddly, in the visual-only condition, the main correlates of intelligence assessments were auditory cues, which probably covary with visual cues. We checked all the visual cues examined in this study and found that none of them acted as mediators to the useful auditory cues; therefore, the identity of the visual cues that do mediate the auditory cues remains to be discovered. In the visual-only condition, a noteworthy discrepancy occurred between the beauty-equals-intelligence stereotype, which found support in the Borkenau and Liebler (1993, 1995) studies but not in this study.

The answer to the question of auditory cue dominance versus cross-modality in cue utilization seems clear. Observers' inferences of intelligence were associated much more with auditory than visual cues in all three conditions. Because the three conditions involved three separate sets of judges, this might be considered three replications of this conclusion. Also, visual cues, considered on their own, simply bore very little relation to any intelligence assessments. Furthermore, if there is cross-modality of any sort, it comes in the form of the auditory cues' dominance even in a purely visual condition.

Finally, a less-direct indication that assessments primarily are made though the auditory channel is that judges with access to auditory information made stronger distinctions in their assessments: As Table 1 shows, the standard deviation of judgments made with auditory information were about twice as large as the standard deviation of the assessments made by judges who were presented with visual information only.

Achievement

Predictability, unfortunately, is not the same as accuracy. Judges' inference policies were quite predictable from the objective cues (R^2 was large), yet their accuracy was modest. However, in the auditory-only condition, accuracy ($r = .38$, $p < .05$) was moderate (Cohen, 1988). Given that the judges had only 1 minute of experience with the targets, this is impressive, but it is not particularly surprising (cf. Ambady & Rosenthal, 1992). Our deliberate attempt to elicit intelligence by asking thought-provoking questions and the relatively wide range of measured intelligence among the targets may account for this moderate level of accuracy.

The 13 cues examined here accounted for about two thirds of the variance in measured intelligence; presumably there exist cues we did not examine that could increase that. This suggests that measured intelligence is, indeed, quite detectable from observable behavior cues. However, judges even in the best (auditory) condition did not detect intelligence near the potential limit set by the ecology. Only 9 of 45 judges (20%) detected intelligence above chance levels, albeit from a thin slice of behavior. Also, these judges were untrained undergraduates, not professionals. Whether the average teacher, personnel officer, or professional psychologist is markedly better at detecting intelligence is an empirical question of great practical importance, one that should be pursued. Regardless of whether professionals can do this, it appears that judges could be trained to recognize intelligence at higher levels than these judges did by training them to look for ecologically valid cues and to ignore ecologically invalid cues. This is another potentially fruitful avenue of investigation.

Based on the general principles of lens models, assessments will be more accurate when judges use appropriate (ecologically valid) cues and less accurate when (a) few ecologically valid observable cues are available, (b) judges fail to use ecologically valid cues, (c) judges use invalid cues, and (d) judges use a valid cue in its opposite (incorrect) direction. Beyond lens model principles, judges' assessments will agree with the criterion less if their fundamental conceptualization of intelligence differs from that measured by the test used (in this case, the WPT).

To illustrate the lens model principles in these data, in the auditory-only and auditory-plus conditions, judges' assessments of intelligence were related to three auditory cues that were, in turn, related to measured intelligence; these matched links (cues are underlined in the figures) presumably enhanced judge achievement or accuracy. However, in both conditions, judges' assessments of intelligence also were related to two cues (halting speech and standard speech) that were not related to measured intelligence (and auditory-only judges' inferences were correlated with loudness, which also was not correlated with measured intelligence). Thus, these judges' inferences were linked to ecologically invalid cues, which tend to depress accuracy.

Achievement also was presumably hampered because the judges' inferences of intelligence were not correlated with other visual cues that were ecologically valid (lack of self-assurance and lower attractiveness). Apparently, the inference of intelligence largely is based on auditory cues, but observers probably also must attend to certain visual cues, such as those just listed, if they are to improve their accuracy.

Given that auditory cues appear to be so important in the encoding and decoding of intelligence, more detailed attention should be paid to them in the future. The content of target utterances, although complex and difficult to measure in scientifically useful ways, should be examined more closely given that it seemed to affect the relation of some important paralinguistic (noncontent) auditory cues to judged intelligence. Personality studies (e.g., Gifford & Hine, 1994) have already demonstrated the utility of investigating content in the judgments of dispositions.

NOTES

1. This cue was excluded from further examination because only 2 of 30 targets wore glasses to the interview.

2. We checked the zero-order correlations for gender differences. In general, when target sex was controlled, the correlations were very similar. Out of 55 correlations (13 cues in the ecological validity condition and 13 in each of the three cue utilization conditions, in addition to the three achievements between measured and judged intelligence), only two showed significant ($p \leq .05$) sex differences. Speaking more words was an ecologically valid indicator of measured intelligence for men ($r = .70, p < .01$) but not for women ($r = -.06$, *ns*). This is a very interesting and potentially important finding. However, this study was not about sex differences, and these correlations are based on only 17 men, so this finding needs to be replicated. The other difference was observed in the visual-only condition: for men, but not women, more fluent speech ($r = -.66, p < .01$, vs. $r = .00$, *ns*) was associated with judgments of greater intelligence. The importance of the latter difference is questionable because it involves a nominally auditory cue in a visual-only condition.

3. The number of words spoken and speech rate were only moderately correlated ($r = .44$), so the two cues were considered separately.

4. All eight auditory cues were examined, even those that were not significantly related to judged intelligence, in case the transcript (content) condition acted as a suppressor.

5. A reviewer correctly noted that this study focuses on the mean achievement of consensus judgments and suggested that we also report the mean achievement for the sample of judges. These achievements across the judges ranged from $r = -.11$ to .40 ($M = .18$, r-to-Z transformed) in the auditory-plus-visual condition, $r = .09$ to .50 ($M = .28$) in the auditory-only, and $r = -.24$ to .49 ($M = .14$) in the visual-only condition. The accuracy of a single observer is significantly greater than 0 ($p < .05$) when the correlation between the intelligence judgments and the Wonderlic scores exceeds $r = .36$. Thus, at the zero-order acquaintance level, among the 45 untrained judges in the above three conditions, 9 (1 in the auditory-plus-visual condition, 3 in the auditory condition, and 5 in the visual-only condition) assessed the intelligence of others with statistically significant accuracy.

REFERENCES

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256-274.

Anderson, L. D. (1921). Estimating intelligence by means of printed photographs. *Journal of Applied Psychology, 5*, 152-155.

Argyle, M., & McHenry, R. (1971). Do spectacles really affect judgments of intelligence? *British Journal of Social and Clinical Psychology, 10*, 27-29.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Bernieri, F., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior.* New York: Cambridge.

Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information. *Journal of Personality and Social Psychology, 61*, 298-307.

Borkenau, P., & Liebler, A. (1992). The cross-modal consistency of personality: Inferring strangers' traits from visual or acoustic information. *Journal of Research in Personality, 26*, 183-204.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*, 546-553.

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*, 1-25.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Campbell, A., & Rushton, J. P. (1978). Bodily communication and personality. *British Journal of Social and Clinical Psychology, 17*, 31-36.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cook, W. C. (1939). The judgment of intelligence from photographs. *Journal of Abnormal and Social Psychology, 34*, 384-389.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications.* San Diego, CA: Academic Press.

Dovidio, J. F., & Ellyson, S. L. (1982). Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening. *Social Psychology Quarterly, 45*, 106-113.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670.
Gaskill, P. C., Fenton, N., & Porter, J. P. (1927). Judging the intelligence of boys from their photographs. *Journal of Applied Psychology, 11*, 394-403.
Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology, 66*, 398-412.
Gifford, R., & Hine, D. W. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispositions. *Journal of Research in Personality, 28*, 115-132.
Gurnee, H. (1934). An analysis of the perception of intelligence in the face. *Journal of Social Psychology, 5*, 82-89.
Laird, D. A., & Remmers, H. (1924). A study of estimates of intelligence from photographs. *Journal of Experimental Psychology, 7*, 429-446.
Miller, G. R., & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech, 50*, 36-44.
Pintner, R. (1918). Intelligence as estimated from photographs. *Psychological Review, 25*, 286-296.
Rosenthal, R., & Jacobson, L. (1968). *The Pygmalion effect.* New York: Holt, Rinehart & Winston.
Saito, K. (1978). An experimental study of personality judgment: Effect of wearing glasses. *Japanese Journal of Experimental Social Psychology, 17*, 121-127.
Sereno, K. K., & Hawkins, G. J. (1967). The effects of variations in speakers' nonfluency upon audience ratings of attitude toward the speech topic and speakers' credibility. *Speech Monographs, 34*, 58-64.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech, 18*, 145-152.
Thornton, G. R. (1943). The effect upon judgments of personality traits of varying a single factor in a photograph. *Journal of Social Psychology, 18*, 127-148.
Thornton, G. R. (1944). The effect of wearing glasses upon judgments of personality traits of persons seen briefly. *Journal of Applied Psychology, 28*, 203-207.
Wonderlic, E. F. (1992). *Wonderlic Personnel Test and Scholastic Level Exam user's manual.* Libertyville, IL: Wonderlic Personnel Test, Inc.

Received May 26, 1999
Revision accepted December 12, 1999