Journal of Experimental Psychology:
Human Perception and Performance
1983, Vol. 9, No. 6, 864–881

Invariances in the Acoustic Expression of Emotion During Speech

Leda Cosmides

Journal of Experimental Psychology:
Human Perception and Performance
1983, Vol. 9, No. 6, 864–881

Invariances in the Acoustic Expression of Emotion During Speech

Leda Cosmides
Harvard University

An experiment was designed to test whether different individuals produce similar voice patterns when they read the same emotional passage. Quantitative scoring criteria were developed that reflect the extent to which different individuals consistently produce similar constellations of acoustic attributes in response to the same emotional context. The scoring procedure was applied to the voice tracks of standard utterances produced by 11 subjects reading 10 different emotionally evocative scripts. The results supported the hypothesis that different individuals produce standard acoustic configurations to express emotions. Because acoustic properties reflecting contrastive stress consistently varied with emotional context over syntactically and semantically *identical* utterances, some factor related to emotional context other than syntax or semantics must account for the variations. An evolutionary argument that emotion communication can be seen as intention communication is presented to account for these variations. Implications for theories of emotions and of intentional generative semantics are discussed.

Researchers interested in the acoustic expression of emotion usually assume that different individuals express the same emotions in similar ways. Yet, this has never been empirically demonstrated. Furthermore, there is no a priori theoretical reason why the acoustic expression of emotion must manifest cross-culturally universal or even culturally shared but nonuniversal acoustic patterns. Emotionally charged vocal patterns could be idiosyncratic, requiring a period of acquaintance with the speaker to decode. This experiment was designed to address this issue by detecting the extent to which different individuals consistently produce similar constellations of acoustic attributes in expressing a particular emotion.

Just as human and nonhuman primates share facial expressions of emotion (Jolly, 1972, pp. 158–159), it is empirically likely that they share certain acoustic expressions of emotion. Shrieks of fear in chimps and in humans are likely to share high frequency due to muscles tensed for fight or flight (Scherer, 1981b) and high amplitude, as befits a call for help or a warning. Averaged acoustic measures, like mean fundamental frequency (F_0) and mean amplitude, are likely to uncover any such homologies and therefore be a source of interindividual similarity among humans.

As a call for help or warning of danger, the two primate shrieks also share some semantic meaning. The human shriek of fear, however, is a platform for further semantic content because it can contain words: One can shriek "Don't kill me" or "I didn't do it" or "He did it," depending on the particulars of the fear-inducing situation. This presents the possibility that some aspects of the acoustic expression of emotion in humans may be particularly adapted to our species-specific language capacity.

More specifically, linguists have noted that acoustic stress is often used for pragmatic implication. Although the sentence "Jane spoke

The acoustic analysis equipment used was paid for, in part, by National Institutes of Health Grant NS 15059; the research was supported by private sources.

I gratefully thank William Cooper, Sheldon White, John Tooby, and Stephen Kosslyn for the intellectual stimulation and material support they gave this project; William Epstein, Carol Fowler, Kenneth Stevens, and Laura Petitto for their thoughtful comments on the manuscript; and Douwe Yntema, Robert Rosenthal, Penelope Greene, Victor Solo, and Stan Klein for their statistical advice. This project could not have been completed without the technical expertise and assistance of Scott Bradner. I also thank S. T. F. Randolph, Kenny Vaughan, and Jordan Smoller.

Requests for reprints should be sent to Leda Cosmides, Department of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts 02138.

to Alex" logically and grammatically entails both (a) "someone spoke to Alex" and (b) "Jane spoke to someone," a speaker who believes the listener already knows that someone spoke to Alex will say "*Jane* spoke to Alex," whereas one who believes the listener already knows that Jane spoke to someone will say "Jane spoke to *Alex*" (Smith & Wilson, 1979, p. 154). The acoustic stress highlights which of the utterance's logicogrammatical entailments the speaker considers most important: It distinguishes "new" from "given" information (Bolinger, 1972; Clark & Clark, 1977, p. 32; Gunter, 1982; Hornby, 1972; Jones 1962, p. 108). Essentially, acoustic stress can be a clue that allows the listener to select which interpretation the speaker intends.

Entailments are always derived by the application of rules or procedures to a background of knowledge that the conversants are presumed to share. In linguistic and cognitive theories these rules are usually grammatical and/or logical, and they are applied to constituents of the utterance in relative isolation from contextual elements. For example, in the above case, (a) and (b) are grammatically specified by "Jane spoke to Alex" through the substitution of appropriate indefinite phrases at nodes of the sentence's phrase structure (Smith & Wilson, 1979, p. 159). In the case of lexical meaning, the propositional calculus is applied to the word's descriptors: Because "all uncles are men" is true by virtue of the lexical meaning of "uncle," the sentence "my uncle spoke" deductively entails the proposition "A man spoke."

However, the highlighting of logically and grammatically derived entailments is often not sufficient for the interpretation of utterances. Since Bartlett (1932), psychologists have acknowledged that context often plays a central role in linguistic interpretation. Although the study of contextual factors in language production and comprehension has been granted a subfield—"pragmatics"—little theoretical attention has been given to the types of procedural knowledge that mediate these factors. While researchers from Bartlett to Schank and Abelson (1977) have posited that these procedures are represented in the form of schemas or scripts—domain specific inference structures—they have provided little insight into their specific content. Indeed, if such scripts are the product of idiosyncratic personal experiences, elucidating their content would be a pointless academic exercise.

Interestingly, recent developments in evolutionary biology suggest that many *emotion* scripts are not idiosyncratic, that some of them lie at the core of what we think of as human nature. These developments may provide some insight into the specific content of these inference procedures. Game theory, with its emphasis on the incentives and intentions of actors, lies at the heart of the current Darwinian revolution in the understanding of social behavior that has already hit anthropology and behavioral biology (cf. Hamilton, 1964; Williams, 1966; Maynard Smith, 1979; Dawkins, 1982; Trivers, 1974; Popp & DeVore, 1979; Chagnon & Irons 1979; Alcock, 1979). These game theories provide reasonably specific hypotheses about the content of the inference procedures organisms use to reason about situations involving large fitness costs and benefits. Furthermore, they emphasize the importance of *signaling* costs, benefits, and behavioral intentions to conspecifics in negotiative interactions. Ethologists have traditionally considered such signaling the primary function of emotional expression, studying intention movements, courtship dances, agonistic displays, and aggressive interactions in mammals, birds, reptiles, fish, and insects.

Thus evolutionarily important contexts—ones involving sex, pair bonding, death, aggression, relatives, friendship, parenting, resource accrual—are likely to be emotional contexts, and are precisely the sort of domains for which one would expect humans to possess a variety of specialized, highly structured inference procedures. Such inference procedures would allow two or more conversants to derive relatively uniform context-dependent "entailments" of utterances in emotional situations. Acoustic stress may play a role in the interpretation of emotional speech similar to that proposed for grammatically derived entailments in nonemotional speech. Namely, acoustic stress might be used by a speaker to highlight which socioemotional "entailment" he or she intends. If this is the case, (a) emotional contexts are particularly likely to produce great conformity in stress patterns, and (b) even when the syntactic and semantic structure of an utterance is held constant, stress patterns should differ with emotional context. To see if this is true, one wants to look not only at averaged acoustic measures but also at ones that can vary with the words and relations expressed by the sentence's semantic structure.

Accordingly, the experiment reported in this article was designed to explore three questions: (a) Do different individuals consistently produce similar constellations of acoustic attributes in reading the same emotional passage? (b) Which are the consistent acoustic properties? (c) Are any of the consistent properties ones that vary with the words and relations expressed by the sentence's semantic structure?

Although theoretically oriented linguists have long hypothesized a relation between intonation and emotion (Bolinger, 1972, 1982; Gunter, 1982), acoustic studies of emotion communication are rare. Averaged acoustic measures like mean F_0, amplitude, and tempo are thought to be associated with anger (Davitz, 1964; Huttar, 1968; Markel, Bein, & Phillips, 1973; Williams & Stevens, 1972), benevolence and competence (Brown, Strong, & Rencher, 1973a, 1973b), depression (Markel et al., 1973), confidence (Scherer, London, & Wolf, 1973), deception (Ekman, Friesen, & Scherer, 1976), anxiety and stress (Hauser, 1976; Scherer, 1981a; Utsuki & Okamura, 1976), fear (Fairbanks & Pronovost, 1939), and grief (Davitz, 1964; Eldred & Price, 1958; Huttar, 1968; Williams & Stevens, 1972).

The agreement among many of these studies argues that different individuals do produce standard configurations of acoustic attributes in expressing particular emotions. To demonstrate interindividual similarities decisively, however, one needs to compare detailed acoustic information on standardized utterances across a *number* of subjects. Unfortunately, the technical difficulty of such analyses—especially prior to the use of digital computer systems—has tended to severely limit the number of subjects per study. Scherer (1981a) reviewed the most relevant research; since 1970 no published acoustical analysis with standard text has included more than three subjects (although Scherer & Wallbott, Note 1, are preparing a study using six). The only published study in this period that attempts to compare responses to standard text across subjects is Williams and Stevens's (1972) work with three male actors. Although Williams and Stevens looked at a number of sophisticated acoustic variables, the only ones they quantitatively compared across subjects were mean and median F_0, F_0 span, and mean rate of articulation. Their comparison of spectrograms between subjects was qualitative, and they did not attempt to compare the F_0 contours of different individuals.

In the experiment reported here, on two different occasions 11 subjects read a standard utterance, "I'll do it," which had been embedded in 10 different emotional contexts ("scenes"). I looked at six acoustic parameters of the "I'll do it"s, five of which could vary with the words and relations of the sentence's semantic structure. An *acoustic emotion configuration* (AEC) was defined as a constellation of acoustic attributes that is consistently produced by many individuals in expressing a particular emotion. I considered a parameter to contribute to an AEC if it varied with emotional context in a consistent, replicable manner across subjects. A quantitative scoring procedure that captures these criteria is presented in the Method section. My hypothesis was that the operation of the evolutionarily predicted inference procedures on the emotional scenes would structure subjects' acoustic responses. If this is true, a number of acoustic properties, including ones that vary with elements of semantic structure, should fulfill the scoring criteria, establishing the existence of AECs.

Method

Subjects

Subjects were 11 Harvard undergraduates, 6 male and 5 female, who had answered an ad posted in the psychology building offering payment for participation in an experiment on acting techniques. They had no formal acting training. Subjects were divided into two groups, one that had an imagery ("I") session first and another that had a no-imagery ("N") session first (see the Procedure section). The first group contained 2 females and 3 males, the second, 3 females and 3 males.

Stimulus Materials

Asking subjects to simulate emotional responses, to "be angry" or "be sad," invites stereotyped, theatrical responses. Therefore I followed Williams and Stevens's (1972) procedure of having the subject play a role in a script. The hope in such a procedure is that the emotional response will arise naturally and subtly from the situation described. Each subject read 10 scripts, 500–700-word scenes excerpted from novels by Ursula K. LeGuin (cf. LeGuin, 1980). LeGuin is a prize-winning author of science fiction and fantasy who writes concrete, easily imageable, emotionally evocative descriptions (Pavio, Yuille, & Madigan, 1968). The emotive content of the scripts is not easily labeled, again, to minimize the possibility that subjects' tacit semantic knowledge of emotive terms would produce stereotyped responses. The only line in the script for the subject to read aloud was "I'll do it," which constituted the last three words of every script. "I'll do it" is easily adapted to different emotional contexts and takes less than half a second to say. Looking for acoustic configurations that might vary with the semantic structure of a sentence requires detailed acoustic information. Therefore I recorded and used every glottal pulse in the analysis (approximately 200 pulses per sec for females, 120 per sec for males). This would have been impractical with a longer utterance. Furthermore, the case for AECs is even stronger if they can even be found in very short, grammatically simple sentences.

Procedure

Each subject underwent two experimental sessions consisting of 10 scripts per session. Subjects were tested individually in a soundproof booth with computer-controlled recording equipment. The sessions lasted about 2.5 hours each and were conducted at least 1 week apart. They were identical except that the subject was asked to use imagery in one and refrain from using imagery in the other. This manipulation was related to aspects of the experiment reported elsewhere (Cosmides, Note 2) but irrelevant to the analysis reported in this article. Here, the "N" condition was treated as a replication of the "I" condition (and vice versa) for two reasons: (a) As the Results section shows, there are no main effects or interactions associated with the imagery manipulation in two-way analyses of variance (ANOVAS) on acoustic parameters (where the other factor is "scenes"), and (b) because the scoring criteria (see the Quantitative Analysis section) demand similarity between the two conditions before a parameter is considered to contribute to an AEC, any variation due to the imagery manipulation would *lessen* the probability of finding AECs. Thus, whatever the consequences of the imagery manipulation, they cannot, in principle, have affected the conclusions arrived at in this article.

The portions of the instructions to subjects relevant to this analysis are as follows:

> The purpose of this study is to learn about various acting techniques. In front of you there is a pile of scripts and a pile of questionnaires. You will be playing one of the characters in each script and answering some questions about your experience. At any point you can leave the experiment if you choose.
>
> The scripts are numbered 1 through 10, and all but number 3 are written in the first person. You will play the character who refers to her or himself as "I" in all the scripts except script number 3 where you will play the woman/man. Don't be alarmed if some of the scripts mention wizards, magery, or strange beasts or places—some were taken from works of fantasy. In each script you will have one line to say aloud—the same line in each script—"I'll do it." It will appear near the end of the script in boldface, enclosed in brackets. When you are comfortable and ready to begin, read the first script through once to yourself, then push the "0" button on the box. This button will set a timer for 4 minutes. During the 4 minutes, read the script through as many times as you like, but don't read out loud, gesticulate, or get up from your chair.

(Imagery or no-imagery preparation instructions were here.)

> Don't spend your time worrying over or planning the way you will say your line.
>
> At the end of the 4 minutes you will hear a short beep. The beep indicates that your 4 minutes are up and the tape recorder is on. When you hear it, stop what you are doing and read through the script as you did during your 4 minute preparation, up to the point that your line appears, then say your line—"I'll do it"—into the microphone. There's no need to feel inhibited or embarrassed, no one will be listening to you or criticizing your acting. Just let your line come out, naturally and spontaneously; don't try to be "theatrical." Read the script through and say your line two more times, for a total of 3 repetitions of the line. Each time read it through as you did during the 4 minute preparation (imagery and all). When you are through, say the number of the *next* script so we have a record on the tape and hit the "0" button again—this will turn the tape recorder off.

(Instructions for answering questionnaires were here.)

> When you have finished with the questions for the first script, repeat the procedure for the other scripts. It is important that you do the scripts in order from 1 to 10. [Summary of procedure here.]

A copy of the summary was posted in the room as a reminder. Subjects read the line three times per scene so I would have some record of how much the line varied in repetition, but only the first rendition was analyzed. The time required to acoustically analyze all three renditions for each scene would have been prohibitive.

The tape recorded "I'll do it"s were analyzed by the Fundamental Period (FPRD) program, which was developed at the Massachusetts Institute of Technology by W. L. Henke (Cooper & Sorensen, 1981, p. 23), in the Computer Based Laboratory, William James Hall, Harvard University. This program finds the amplitude "strokes" associated with closure of the vocal cords and uses them to compute the duration of each glottal pulse (to the nearest microsecond). It estimates the fundamental frequency (rate in Hz) and relative amplitude associated with each glottal pulse. The amplitude measure uses a linear scale with units of magnitude arbitrarily defined by the program. The fundamental frequency and relative amplitude measures provided the raw data for the quantitative analysis.

Quantitative Analysis

I considered six different acoustic parameters in the analysis of each utterance: mean F_0, frequency span (highest minus lowest F_0), amplitude ratio (highest divided by lowest amplitude), duration of utterance (total duration, duration of "I'll," duration of "doit" [acoustically, "do it" is continuous, so I treated the two words as one unit], and the duration of the space between "I'll" and "doit"), a measure of the frequency fall–rise pattern, and a measure of the amplitude fall–rise pattern (fall–rise patterns display the relative variation of F_0 or amplitude over the length of the utterance). All parameters except mean F_0 (an averaged value) and total duration of utterance can vary with surface elements of the sentence's semantic structure (F_0 and amplitude fall–rise patterns are the most detailed measures of this variation). For analysis of the single-valued nontime parameters, the "I'll" and "doit" were each normalized for duration: they were each divided into 10 equal time segments and a mean F_0 or amplitude taken for each segment. For fall–rise analyses the utterances were also normalized for span on a 10-point linear scale. An interval scale from 1 (corresponding to the lowest F_0 or amplitude for the utterance) to 10 (corresponding to the utterance's highest F_0 or amplitude) was constructed for each utterance, and the F_0 or amplitude value for each of the utterance's 20 time segments was assigned a scale number corresponding to the interval it fell into. (For amplitude fall–rise one gets the same results whether the scale is constructed out of a ratio or span measurement.) See Figure 1 for an illustrative calculation. This normalization procedure allows fall–rise patterns to be considered separately from F_0 span or amplitude ratio.

Intuitively, one would not want to say that AECs exist unless vocal responses to particular emotional contexts (in this case, different scenes) vary consistently across subjects. Vocal parameters that vary idiosyncratically across different subjects within a scene fail the consistency criterion, whereas ones that adhere to a single value or form regardless of emotional context are obviously not being used to express the different scenes' varying emotional contents. Both of these possibilities are eliminated if the analysis of an acoustic parameter for a session yields a significant F ratio (MS[scenes]/MS[Scenes × Subjects]) from a one-way ANOVA with repeated measures on subjects (since each subject underwent all 10 scenes). The error term accounts for any idiosyncratic variation; a significant F ratio thus represents variation accounted for only by emotional context, and it indicates that subjects are not adhering to a single acoustic form regardless of emotional context.

To determine (a) which scenes are contributing to the effect, (b) whether the acoustic parameter is high, middling, or low for those scenes, and (c) whether many scenes or just a few scenes are varying, the scene effect was decomposed into single df comparisons between scene means through the use of contrasts (Winer, 1971, p. 170). Because consistency of emotional expression is at issue in this article, a set of contrasts was considered valid only if it yielded a significant F ratio in both conditions ("I" and "N") and in a combined two-way ANOVA with repeated measures on both factors, where the imagery manipulation is the second factor. The requirement that the contrasts also produce a significant effect in the two-way ANOVA allows a choice in cases where inspection of the maximal contrasts in the "I" and "N" conditions suggests competing sets of contrasts for the two replications.

Because this was an exploratory study to see whether there is any consistency at all in the acoustic expression of emotion, I did not start out with hypotheses regarding which scene means would be high, middling, or low. Although there was the constraint that contrasts in one replication mirror those in the other, technically, they were unplanned contrasts—contrasts derived by looking at the data (in fact, "hypotheses" for the contrasts were derived by computing maximal contrasts for the two-way ANOVA according to Winer, 1971, p. 176). To avoid Type I errors, an F ratio derived from unplanned contrasts must pass the more conservative Scheffé test (Winer, 1971, p. 198). Although an F derived from contrasts has only one degree

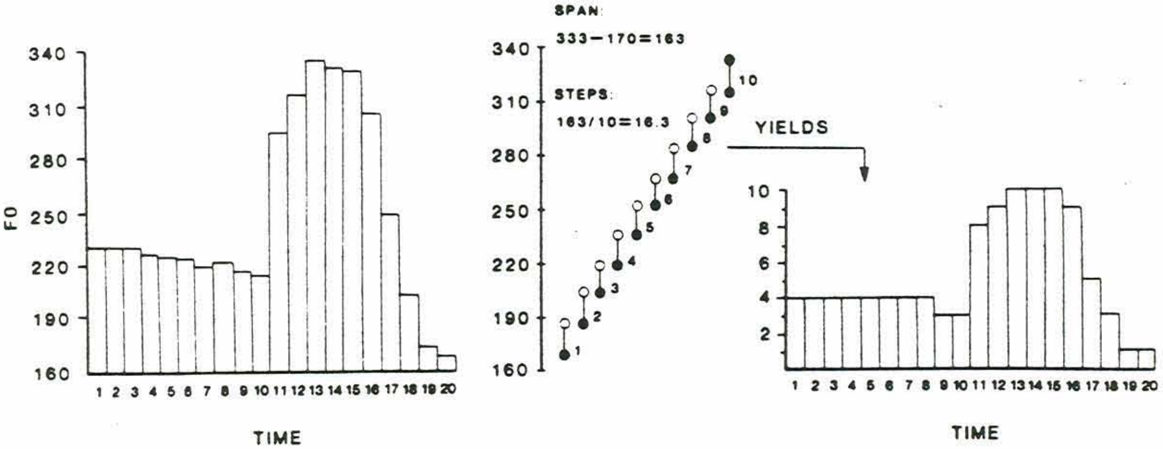


Figure 1. How raw fundamental frequency (F_0; or amplitude) fall–rise patterns were normalized for span. (The first graph pictures the initial F_0 fall–rise pattern [normalized for time]. The second graph shows the mapping algorithm: this utterance's span [highest minus the lowest F_0] is 333 − 170 = 163. Division by 10 yields the 10 equal steps from 170 to 333 shown in the second graph. The third graph shows the span-normalized fall–rise pattern that results when F_0 values for each of the 20 time segments are mapped onto the 10 equal steps of the second graph; e.g., F_0s of 223 and 222 on the first graph fall in the fourth interval, 330 and 333 fall in the tenth, and so on.)

of freedom, the critical value for the Scheffé is $F = (k - 1)(F(k - 1, nk - k))$ where n is the number of subjects and k the number of treatment groups.

Operationally, therefore, an acoustic parameter was considered to contribute to an AEC only if there was a set of contrasts describing the variation of its scene means that passed the Scheffé test for both replications and the combined two-way ANOVA. As a further check on the validity of the scene mean differences, and to see if there is any regularity to the pattern of scene means for parameters that did not pass the stringent Scheffé test, I also calculated Spearman correlations for the scene mean ranks for the two replications (Siegel, 1956).

The 20 values that constitute a fall–rise pattern must be condensed to one meaningful measure in order to apply these criteria. First, a "mean shape" was computed for each scene by averaging all 11 subjects' interval F_0 or amplitude values for the first time segment, then the second, and so on for each of the 20 time segments (see Figure 2 for a sample mean shape calculated from the fall–rise patterns of two subjects). The relevant question for a configuration analysis is as follows: Do subjects' fall–rise patterns deviate randomly from this mean shape, or do they tend to adhere to it? To quantify this I computed a "deviation score" for each subject's fall–rise pattern. For each time segment the fall–rise pattern's interval value was subtracted from the corresponding value for the mean shape for that scene, and the magnitudes (absolute values) of these deviations were summed (e.g., the deviation score for fall–rise pattern A in Figure 2 would be 19). The smaller the deviation score, the closer the fall–rise pattern adheres to the scene mean shape. The ANOVAS described were run on these deviation scores.

For a fall–rise pattern to be said to vary consistently across subjects, its scene mean deviation score should be lower than a standard value representing deviations that are random with respect to scene. For comparison purposes, a "standard" score can be computed from a "grand mean shape"—a mean shape computed from the fall–rise patterns of all subjects in all scenes in a session. The advantage of this shape as a standard is that it takes into account any global similarities common to all utterances. A deviation score from this grand mean shape can be computed for each fall–rise pattern. For each subject, the "grand mean" deviation scores for each scene are averaged. If it looks like all the scene mean deviation scores are low (indicating that within every scene subjects are producing very regular fall–rise patterns), these grand mean deviation scores can be included for comparison purposes as an 11th level of the scene factor in the ANOVA. Otherwise, an average of these values can be used in constructing contrasts to decide which scene mean deviation scores should be considered high. (This is how the grand mean deviation score was used in this experiment.)

If fall–rise patterns for certain scenes are found to be regular, one wants to make sure they are not all the same shape; if they are, subjects are not using different fall–rise patterns to express different emotional contents. Scene mean shapes were compared pairwise through the computation of deviation scores. Two shapes were considered "very similar" if their deviation score fell within confidence intervals set around the average deviation score for scenes matched across the "I" and "N" conditions. The assumption that this amount of deviation is attributable to error rather than to different expressed emotions is supported if the variance of deviation scores for these matched scenes is lower than that for the "I" or "N" different scene deviation scores. This assumption was tested using a homogeneity of variance test (Winer, 1971, p. 38).

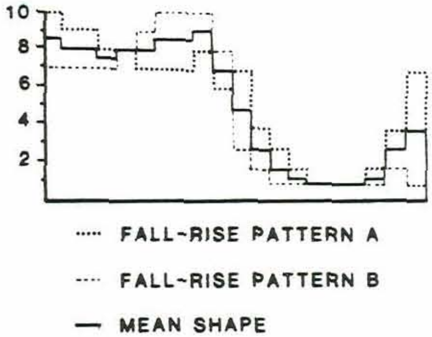
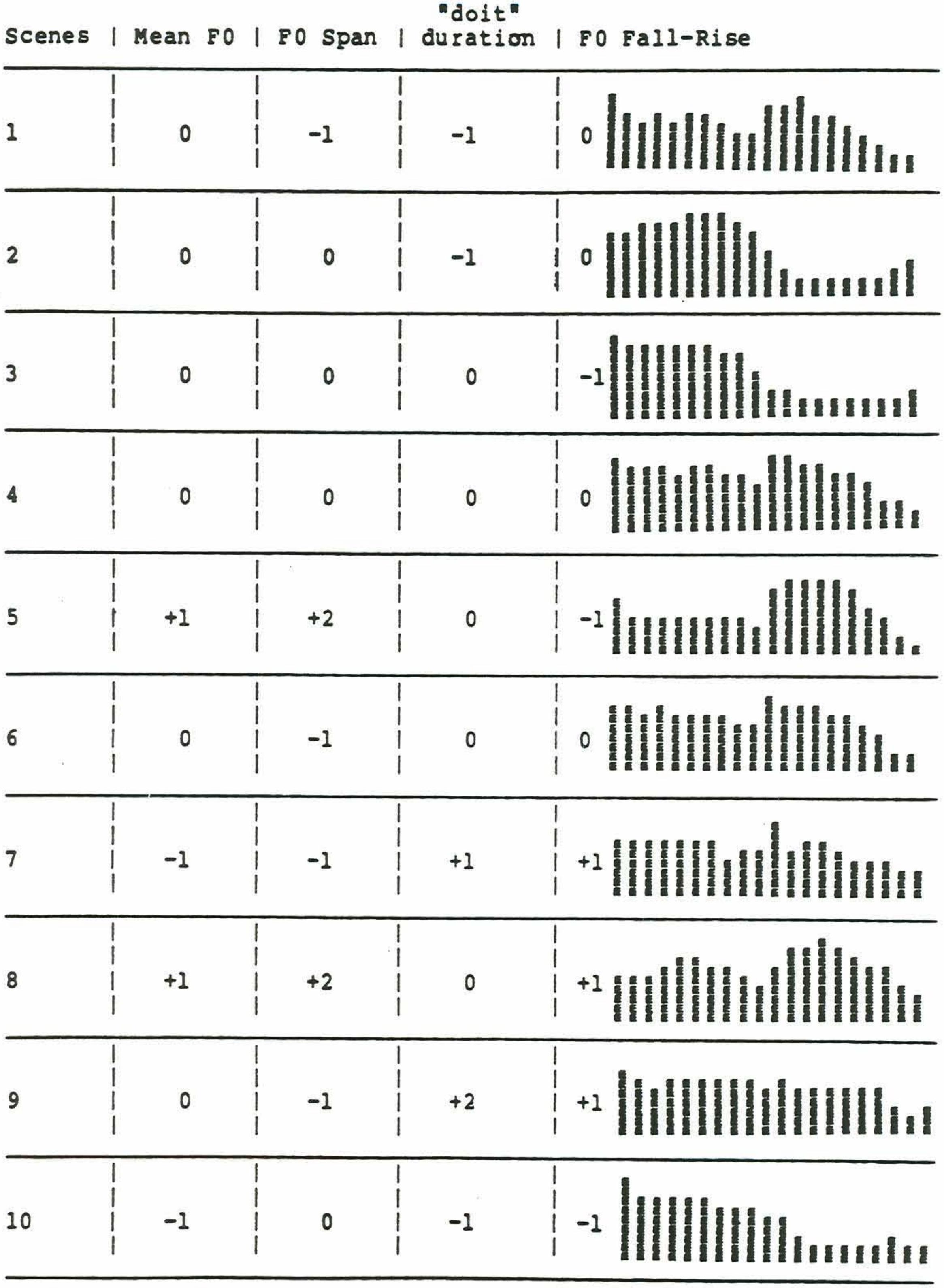


Figure 2. The mean shape that would result from averaging two fall–rise patterns, A and B.

Finally, "reliability 1" (r_1), using *MS*(scenes) and *MS*(Scenes × Subjects) was computed for each acoustic parameter that passed the AEC test (Winer, 1971, p. 287, formula 8; Rosenthal, Note 3). r_1 can vary from zero to one, and there is no significance value associated with it. The higher it is for a particular acoustic parameter, the better that parameter is at discriminating among the scenes (i.e., the more subjects are using that parameter to differentiate emotional contexts). Thus r_1 is useful for making *relative* comparisons between different acoustic parameters; the higher the r_1 for an acoustic parameter, the more important it is in creating AECs.

Results

Figure 3 is a scene by scene summary of the acoustic properties that fulfilled the AEC criteria and their associated contrasts. For mean F_0, F_0 span, and "doit" duration, a positive contrast indicates that the parameter was higher (or longer, for duration) than average for that scene, a zero contrast indicates that it was average, and a negative contrast indicates that it was lower (or shorter) than average. Contrasts have a different meaning for the F_0 fall–rise pattern. Because these contrasts refer to scores indicating an individual's *deviation* from the scene's mean shape, a negative contrast indicates that individual fall–rise patterns for a scene adhered closely to the scene's mean shape. A zero fall–rise contrast indicates middling adherence, whereas a positive contrast indicates little or no adherence. The mean shapes for each scene are also pictured, though one should bear in mind that mean shapes for scenes with positive contrasts are constructed from a mishmash of quite distinct fall–rise patterns (i.e., there is a sense in which positive



Scenes	Mean F0	F0 Span	"doit" duration	F0 Fall-Rise
1	0	-1	-1	0
2	0	0	-1	0
3	0	0	0	-1
4	0	0	0	0
5	+1	+2	0	-1
6	0	-1	0	0
7	-1	-1	+1	+1
8	+1	+2	0	+1
9	0	-1	+2	+1
10	-1	0	-1	-1

Figure 3. Contrasts for acoustic emotion configuration (AEC) parameters. (For all parameters except fundamental frequency [F_0] fall–rise, a negative contrast indicates that the scene mean is below average, a zero contrast indicates that the scene mean is average, and a positive contrast indicates that the scene mean is above average. For F_0 fall–rise, a negative contrast indicates that individual fall–rise patterns show good adherence to the scene mean shape, a zero contrast indicates middling adherence, and a positive contrast indicates poor adherence.)

fall–rise contrast scenes do not have "true" mean shapes). (The shapes pictured are from whichever condition produced the most reliable [as measured by average deviation scores] mean shape for that scene.)

Mean F_0

The mean F_0 of an utterance was computed by taking the mean of the F_0 values associated with the 20 time segments of the "I'll doit." Because the average mean F_0 for males and females differ, I ran the analyses on scores that had been standardized within subjects. The criterial $F(9, 90)$ value for the Scheffé test is $9 \times F(9, 90)$, or 23.49 at the $p < .01$ level, and 17.82 at the $p < .05$ level. The set of scene contrasts pictured in Figure 3 are significant at the $p < .01$ level for the "I" condition, $F(1, 90) = 57.89$, the "N" condition, $F(1, 90) = 52.24$, and in the two-way ANOVA, $F(1, 90) = 42.72$. The imagery manipulation yielded no main effects or interactions in the two-way ANOVA: main effect, $F(1, 10) = .34$; interaction, $F(9, 90) = .55$). The nonparametric Spearman test on ranks corroborates the conclusion that subjects are consistent in their application of different mean F_0s to different emotional contexts; the correlation between ranks for the "I" and "N" conditions is .82 ($p < .005$, one-tailed). Mean F_0 therefore fulfills the AEC criteria. For mean F_0, r_1 is .49.

Frequency Span

The F_0 span of an utterance was computed by subtracting the lowest F_0 value of the 20 segments from their highest value. Because scene variances were quite unequal for this parameter, I first did the nonparametric analog of the one-way ANOVA with repeated measures, the Friedman two-way ANOVA by ranks (Siegel, 1956, p. 166; "subjects" are the second factor). The Friedman chi-square ($df = 9$) was 28.13 for the "I" condition ($p < .001$) and 18.14 for the "N" condition ($p < .05$), indicating significant scene effects. Because these less powerful tests showed an effect, I transformed F_0 spans (which had been standardized within subjects) to equalize their variances using a power derived from the slope of a log(mean) versus log(square root of variance) plot. An F_0 span score was transformed by adding 200 to its standard score and raising this value to the .3 power. The ANOVAS were run on the transformed scores. The scene contrasts pictured in Figure 3 for F_0 span are significant at the $p < .01$ level for the "I" condition and the two-way ANOVA: $F(1, 90) = 46.34$, $F(1, 90) = 25.00$, respectively. Scene contrasts are significant at the $p < .05$ level for the "N" condition, $F(1, 90) = 22.56$. The Spearman rank test on the transformed scores yielded a correlation of .71 ($p < .025$, one-tailed). Thus F_0 span fulfills the AEC criteria. The r_1 was .32. There were no effects of the imagery manipulation: main effect, $F(1, 10) = .14$; interaction, $F(9, 90) = .77$).

Amplitude Ratio

The amplitude ratio was computed by dividing the highest amplitude reading for the 20 segments by the lowest. Division is more appropriate than subtraction for amplitude because the amplitude reading depends on the volume at which the tape is read into the computer, and low amplitude utterances have to be read in at higher volume to do the FPRD analysis. As in the F_0 span parameter, the scene variances were quite unequal. The nonparametric Friedman test gave a chi-square ($df = 9$) of 13.29 for the "I" condition (not significant) and of 24.01 in the "N" condition ($p < .01$). The "I" results for this parameter thus fail one of the AEC criteria—that scenes vary from one another. A two-way ANOVA run on the untransformed scores shows a significant scene effect, $F(9, 90) = 2.09$, $p < .05$, but not even an $F(1, 90) = 9.39$ calculated from the maximal contrasts (those that will give the largest F ratio; see Winer, 1971, p. 176) passes the Scheffé criterion. The Spearman rank correlation is .08 (*ns*), indicating that ranks in the "I" and "N" conditions are uncorrelated. Amplitude ratio therefore fails to qualify as contributing to AECs. There were no main effects or interactions due to the imagery manipulation: main effect, $F(1, 10) = .16$; interaction, $F(9, 90) = 1.48$.

"I'll" Duration

All durations are measured in 10,000ths of a second, which is the accuracy of the FPRD program. The two-way and "N" condition ANOVA on "I'll" durations yielded significant scene effects, $F(9, 90) = 3.00$, 3.38, respectively, $p < .01$, but the "I" condition ANOVA showed no scene effect, $F(9, 90) = 1.01$. This means the "I'll" duration parameter fails the AEC criteria. The fact that not even the maximal contrasts for the two-way ANOVA yield an F ratio that passes the Scheffé test, $F(1, 90) = 14.00$, corroborates this conclusion. The two-way ANOVA showed no effects of the imagery manipulation: main effect, $F(1, 10) = 1.02$; interaction, $F(9, 90) = .76$.

The Spearman test did yield a significant correlation between ranks for the "I" and "N" conditions ($r_S = .77, p < .025$ one-tailed). This suggests that "I'll" duration is capable of varying with emotional context but that the effect is simply not strong enough for this particular choice of emotional scenes.

"Space" Duration

All three ANOVAS yielded a significant scene effect for the "space" duration parameter: $F(9, 90) = 4.20, p < .01$ (two-way); $F(9, 90) = 5.43$, $p < .01$ ("I" condition); $F(9, 90) = 2.35$, $p < .05$ ("N" condition). The maximal contrasts from the two-way ANOVA, however, yield an F that is just barely significant with the Scheffé test at the .05 level, $F(1, 90) = 19.22$, and the integer contrasts they suggest (0, −1, −1, 0, +1, 0, +1, 0, 0, 0) do not pass the Scheffé test in the two-way analysis, $F(1, 90) = 16.4$. These contrasts do pass the Scheffé test in the "I" condition, $F(1, 90) = 34.36$, $p < .01$, and just miss the .05 level in the "N" condition. In addition, the Spearman rank correlation is .77 ($p < .025$, one-tailed), indicating a correlation between ranks in the two conditions. There were no effects of the imagery manipulation: main effect, $F(1, 10) = .54$; interaction, $F(9, 90) = 1.43$).

Strictly speaking, this parameter does not pass the AEC criteria. Yet the many regularities that it does show suggest that scene effects are there, but they are a bit too weak to pass the stringent criteria. The "space" duration varies around 100 msec, suggesting that it reflects the stop closure of the /d/ in "doit." The variation in results may be due to emphasis of the /d/ in "doit" (Stevens, Note 4); the fall–rise pattern of Scene 7, for example, shows a spike in the first interval of the "doit," even though the mean F_0 and F_0 span are lower than average for this scene.

"Doit" Duration

The set of scene contrasts for "doit" pictured in Figure 3 yield an $F(1, 90)$ of 35.34 in the "I" condition, 25.86 in the "N" condition, and 32.37 in the two-way ANOVA. All three are significant at $p < .01$. Furthermore, the Spearman rank correlation is .93 ($p < .001$, one-tailed), corroborating the ANOVA results that indicate that the two conditions are highly correlated. The duration of "doit" therefore fulfills the AEC criteria. There were no effects due to the imagery manipulation in the two-way analysis: main effect, $F(1, 10) = .33$; interaction, $F(9, 90) = 1.71$). The reliability was .29.

Total Duration of Utterance

All three ANOVAS yielded significant Fs for the scene factor: $F(9, 90) = 3.24, p < .01$ (two-way); $F(9, 90) = 3.23, p < .01$ ("I" condition); $F(9, 90) = 2.30, p < .05$ ("N" condition). The Spearman rank correlation was .83 ($p < .005$, one-tailed). However, not even the maximal contrasts for the two-way analysis passed the Scheffe test, $F(1, 90) = 14.79$. This means that the total duration of utterance parameter does not pass the AEC criteria. The regularities may be due to the "doit" duration's contribution to the totals. For the "I" condition the correlation between ranks for "doit" and ranks for total duration is .64 ($p < .05$, one-tailed), for the "N" condition, .73 ($p < .025$, one-tailed).

Frequency Fall–Rise Pattern

The scene contrasts for F_0 fall–rise pictured in Figure 3 are significant at the $p < .05$ level for the two-way ANOVA, and at the $p < .01$ level for the "I" and "N" conditions: $F(1, 90) = 23.46$, 25.11, and 31.46, respectively. These contrasts are supported by the Spearman results: Although the rank correlation is .55 ($p < .1$), the lack of correlation is due to Scene 2 adhering to the mean shape in the "I" condition but not the "N" condition. When Scene 2 is eliminated from the calculation, the rank correlation is .867 ($p < .005$, one-tailed). For F_0 fall–rise, r_1 was .30.

The Figure 3 contrasts indicate that subjects' individual fall–rise patterns were quite similar to the mean shapes for Scenes 3, 5, and 10, which are shown in Figure 3. For Scenes 1, 2, 4, and 6, individual fall–rise patterns adhered somewhat to the scene mean shape, as the zero contrasts indicate. But the positive contrasts for Scenes 7, 8, and 9 suggest that these mean shapes are averages of a number of quite distinct fall–rise patterns. This conclusion is corroborated by the fact that the average deviation scores for these scenes are quite similar to those from a grand mean shape constructed from all the fall–rise patterns in a condition (see the Method section). Deviations from such a shape represent the maximum amount of deviation one can expect, given any properties common to all utterances. The average deviation from the grand mean shape is 48 in the "I" condition—quite similar to the scores of 50, 48, and 49 for Scenes 7, 8, and 9—and 47 in the "N" condition—similar to the scores of 45, 45, and 46 for these 3 scenes in the "N" condition. One can see the operation of the averaging of diverse patterns in Scene 9. The situation involved a clinging and irritable mother who is unreasonably demanding things of her guilt-ridden adolescent child. It suggests two distinct emotional responses: irritated and defiant or tired and resigned. Accordingly, individual fall–rise patterns split roughly in half, one set adhering to a shape like Scene 5's (defiant), the other to one like Scene 10's (resigned). Although the "I" and "N" deviation scores for Scene 9 taken as a whole are quite high, the average deviations for the two sets considered separately are within the range considered low to medium by the ANOVA ("I", 37 and 37 with low–medium averages of 32 and 39; "N", 32 and 30 with low–medium averages of 29 to 42). This type of analysis allows some insight into how different people interpret and react to situations. Other scenes with high deviation scores did not seem to split naturally into two categories, suggesting that the situations portrayed in them were not as easily categorized by the human emotional system.

Not only did subjects tend to adhere to particular mean shapes, shapes for different scenes differed from one another. Table 1 shows the 45 pairwise deviation scores for mean shapes for both the "I" and "N" conditions. The diagonal shows the matched scene scores. The Pearson product-moment correlation between the "I" and "N" condition matrices is .86 ($p < .0005$, one-tailed).

Provided the variance for the matched scene scores is lower than that for the "I" and "N" matrices in Table 1 (see the Method section), a deviation score near the matched scene mean is the criterion for judging two mean shapes "very similar." The variance requirement was met. The $p < .05$ cutoff for a homogeneity of variance test (Winer, 1971, p. 37) is $F(44, 9) = 2.89$; the ("I" or "N" variance)/(matched scene variance) ratio exceded this value for both conditions: "I", $F(44, 9) = 10.54$; "N", $F(44, 9) = 5.37$. The mean matched scene score was 15, with a 95% upper confidence limit of 18.5 ($df = 9$). The average similarity of mean shapes was 33.73 for the "I" condition and 32.53 for the "N" condition. Because both are higher than the "very similar" cutoff of

Table 1
Deviation Scores for Pairwise Comparisons of Scene Mean Shapes

Scene	1	2	3	4	5	6	7	8	9	10
1	**14**	56*	50*	9	33*	14	12	16	12	40*
2	37*	**19**	14	59*	83*	52*	46*	56*	46*	16
3	50*	21*	**10**	55*	81*	48*	44*	54*	44*	12
4	18	37*	42*	**19**	28*	13	13	17	15	43*
5	28*	63*	72*	34*	**15**	37*	37*	29*	39*	69*
6	15	24*	39*	27*	41*	**25**	14	18	14	36*
7	15	32*	45*	21*	33*	14	**7**	12	10	32*
8	28*	47*	62*	26*	26*	33*	29*	**22**	16	42*
9	18	29*	38*	20*	36*	15	13	28*	**8**	32*
10	39*	22*	13	33*	61*	28*	34*	51*	27*	**11**

Note. The scores on the diagonal (in boldface type) are for matched scenes. The upper matrix is for the imagery ("I") condition, the lower for the no-imagery ("N") condition.
* $p < .05$.

18.5, pairs of scenes did, on average, differ in mean shape. More specifically, 27 pairs of scene mean shapes differed in the "I" condition and 37 pairs in the "N" condition. These scenes are underlined in Table 1.

There were no effects of the imagery manipulation in the two-way ANOVA: main effect, $F(1, 10) = .18$; interaction, $F(9, 90) = 1.40$).

Amplitude Fall–Rise

Although the two-way and "N" condition ANOVAS yielded significant scene effects, $F(9, 90) = 3.40, p < .01, F(9, 90) = 2.66, p < .05$, respectively, the "N" condition did not, $F(9, 90) = 1.74$ (*ns*). Not even the maximum contrasts for the two-way analysis passed the Scheffé test, $F(1, 90) = 16.74$. The Spearman rank correlation of .37 (*ns*) also argues for a lack of consistency. The average deviations from the grand mean shape were 40 ("I") and 39 ("N"), lower than the corresponding frequency fall–rise scores. This suggests that the lack of a scene effect is due to a great similarity between different (scene) mean shapes. This view is supported by the fact that pairwise deviation scores for mean shapes were, on average, lower than for the F_0 mean shapes ("I", 23.04; "N", 23.62). These results suggest that, although amplitude fall–rise does show some interindividual regularity, it does not vary much with *emotional context*. Thus it does not contribute to an AEC in this experiment. The two-way ANOVA showed no effects of the imagery manipulation: main effect, $F(1, 10) = .84$; interaction, $F(9, 90) = 1.19$.

Summary

Four acoustic parameters—mean F_0, F_0 span, "doit" duration, and F_0 fall–rise pattern—met the AEC criteria. The reliability score for mean F_0 was higher than those of the other three parameters, which were quite similar, suggesting that mean F_0 is one of the most consistently used parameters in the acoustic expression of emotion.

Discussion

Subjects tended to produce particular acoustic configurations in expressing particular emotions. Evidence for this claim is summarized in Figure 3. Mean F_0, F_0 span, F_0 fall–rise pattern, and "doit" duration all contributed to AECs. Other acoustic parameters either failed to vary with emotional context or their pattern of variation was inconsistent from one trial to the next.

The existence of AECs is made all the more interesting by the fact that the scenes were chosen to have complex and subtle emotional material that would be difficult to label. Subjects produced similar acoustic patterns in spite of the fact that the material did not lend itself to stereotypical conceptions.

Mean F_0 was the only "averaged" acoustic property to contribute to the AECs. The other three reflect varying emphasis on the words and relations of the sentence's semantic structure. For example, a large F_0 span represents an increase in the magnitude of F_0 variation over the duration of the sentence, the highest F_0 values usually being associated with stressed words (Cooper & Sorensen, 1981, p. 17). F_0 fall–rise, a property independent of the absolute magnitude of the F_0 span (because fall–rise was normalized for span), specifies the changing directions of emphasis. Varying word duration, as reflected in the "doit" duration parameter, can be used to de-emphasize or call attention to a word. Thus, speakers can use all three of these parameters to stress or underplay the words or relations in the sentence's semantic structure.

Three explanations have been advanced in the linguistic and acoustic literature to account for the deployment of acoustic parameters that reflect changes in stress or emphasis over the course of the sentence: syntax (Bresnan, 1971; Chomsky & Halle, 1968; Trager & Smith, 1951), semantics (Lehiste, 1970, p. 151; Smith & Wilson, 1979, p. 162), and intentions (Bolinger, 1972, 1982; Pike, 1945, p. 21). Although the role of syntax and semantics has been experimentally verified (cf. Cooper & Sorensen, 1981; Jones, 1962, p. 108), that of intentions, which has occasionally been proposed in the linguistic literature (though not necessarily in the context of emotion communication), has not been tested acoustically. The experiment reported here was designed such that neither syntax nor semantics varied in the test sentences, so these factors cannot, even in principle, account for the variations produced. Consequently, the results demonstrate that there is at least one additional factor, aside from syntax and semantics, that is regulating speakers' use of acoustic parameters reflecting stress.

Figure 3 shows that variations between AECs are a function of emotional context. The question then is, what is it about emotional context that can account for these between scene variations? Evolutionary biology provides meta-theoretical support for the notion that people share procedural knowledge for reasoning about emotional domains (see Introduction). Furthermore, it suggests that the encoding and decoding of intended courses of action is the primary function of emotion communication. Following this view, I will argue that emotions express the speaker's *intentions* and that in emotional speech, these intentions factor into the rules for mapping stress-related acoustic parameters onto the words and relations expressed by the sentence.

By intentions I mean the *speaker's evaluative relationship to aspects of the semantic structure of the sentence actually produced.* If conversants share "emotion scripts"—procedural knowledge for making inferences about evolutionarily crucial social domains—they will be able to translate information about the speaker's valuations into predictions about the speaker's intentions and their consequences. Thus the speaker's evaluative attitude toward the actions, state of being, or persons represented by the agent–action–object relations of the sentence's semantic structure embodies his or her behavioral intentions. "I'LL do it" said in a resigned tone of voice does not simply mean that the actor will do the action, it means "I'll do it because I am too tired of fighting you about it, but if I could easily avoid doing it I would." "I'll DO it" said in an irritated, edgy tone of voice really means "I'll do it this time to get you off my back, but I'm getting fed up and might not do it next time." "I'LL do IT" said in a lilting tone, emphasizing the "I'll," means that the person is happy to do it, especially because it is you, whom he or she likes, who wants it done. Although semantic theories are usually adequate for representing what the speaker intends *to say,* they are not adequate for representing factors controlling the paralinguistic communication of what the speaker intends *to do* about the state of affairs represented by the utterance.

Consideration of Scenes 5 and 10 (see Appendix) illustrates how an intentional explanation would explain some of the data of Figure 3. Although the emotions aroused in Scene 5 are a complex jumble of anger, jealousy, insecurity, spite, and pride, the situation itself is, from an evolutionary point of view, a classic agonistic encounter. The speaker's ability to *do* certain actions has been called into question by an older rival trying to assert his dominance in front of his peers. He tries to publicly humiliate the speaker. To save face, the speaker must state emphatically that he intends to *demonstrate* that his rival is wrong (that he can in fact "do it" and do it better than his rival). Accordingly, the mean F_0 is high (making the utterance easy to hear), and the "doit" is emphasized over the "I'll" in the fall–rise pattern. Furthermore, it is emphasized strongly, as the very large F_0 span indicates. Although the high mean F_0 and F_0 span are consistent both with Williams and Stevens's (1972) and Fairbanks and Pronovost's (1939) findings for anger, and with Scherer, London, and Wolf's (1973) findings for dominance, knowledge of the predominant emotion aroused does not allow one to predict *which* of the words is going to be emphasized relative to others (i.e., the fall–rise pattern). This requires an understanding of the particulars of the situation and the specific behavioral intentions engendered by the speaker's valuation of those particulars. The game-theoretic logic of an agonistic encounter *could* allow you to predict that a speaker asserting dominance is going to use a large F_0 span (for special emphasis), a high mean F_0 if he or she wants to be heard by all, and even which of the words or relations the speaker is likely to stress.

The logic of Scene 10 is almost opposite that of Scene 5. The speaker and his or her sister vacillate between guilt, shame, disgust, and revulsion. They are under the obligation of doing a revolting and guilt-provoking task, and one of the two must volunteer to "do it." Neither wants to do it. Unlike Scene 5, there is no reason to emphasize the action itself, no audience to announce it to, no reason to shout. The salient thing to be communicated is the intention of the speaker to perform the act *for* his or her sister, however reluctantly. As the low mean F_0 indicates, the sentence is barely, reluctantly, said. Although the fall–rise pattern shows that the "I'll" is emphasized over the "doit," as one would expect in an offer, it is not emphasized much, as the middling F_0 span indicates—after all, this is not an enthusiastic offer, but one of willingness in spite of unpleasantness. Most tellingly, the "doit"—the statement of the action that must be performed—trails off into a long, barely audible whisper. Again, although the work of Williams and Stevens and Fairbanks and Pronovost predicts low mean F_0 and span for sorrow, they have no prediction regarding fall–rise pattern. Furthermore, although Scene 10 is clearly not a happy situation, it is not clear that it should be considered "sorrowful" either; revulsion, guilt, disgust, and self-hatred are all prominent in this scene.

I am not suggesting that no aspect of emotional expression can be predicted by knowing a simple semantic descriptor of the speaker's emotional state. After all, certain emotional states are accompanied by general physiological reactions. Physiological correlates of emotional states can alter the larynx and articulatory apparatus, and this, in turn, can change vocal properties. For example, the copiousness and consistency of lubricating mucus in the larynx and of the mucal lining of the vocal folds during sexual arousal affects the efficiency of vibration in both men and women, making the voice more whispery and fine pitch control more difficult (Laver & Trudgill, 1979). Sympathetic activation in stressful circumstances deepens respiration, dilates the bronchi, and increases muscle tension, leading to increased amplitude and fundamental frequency (Scherer, 1981a). Such temporary physiological changes in the state of the vocal apparatus that affect properties of entire utterances can be expected to produce cross-culturally invariant (or even primate-wide) characteristics of emotional communication.

But even here, a knowledge of the emotion descriptors will tell one nothing about the fall–rise pattern, whereas an understanding of the game-theoretic nature of emotion-laden social interactions *can* predict which physiological changes in state will occur in which situations, in addition to predicting fall–rise pattern. For example, emotion scripts can suggest which situational parameters are likely to trigger the sympathetic activation of a "fight–flight" response and its corresponding acoustic correlates. Moreover, the same theoretical framework, through the elucidation of the specific inference structures characteristic of emotion scripts, can also explain the types of stress patterns found in this experiment. Knowledge of an emotion descriptor cannot tell one both.

Although this information-processing view of emotions in terms of procedural knowledge is biologically inspired, it is not a physiological theory of emotion. In fact, a physiological theory would have difficulty accounting for the acoustic data. If emotions are primarily transformations "of chemical or physical energy at the sensory output level into autonomic or motor output" (Zajonc, 1980, p. 154) or experienced somatic reactions in the James–Lange tradition (James, 1890)—if emotional expression in humans is the product of general inchoate experiences or preferences like "feeling angry" or "feeling sad"—unstructured by the cognitive appraisals emphasized by some psychologists (e.g., Lazarus, 1982)—one would expect acoustic emotion communication to consist *only* of acoustic parameters that reflect temporary physiological changes in the vocal apparatus. One would not predict that acoustic parameters reflecting differential stress patterns would be major contributors to AECs, as found in this experiment. And the mapping of acoustic properties onto a list of simple emotion descriptors would be a relatively trivial matter.

Yet linguists have not been able to construct systematic rules for assigning stress in emotion speech (Bolinger, 1972, 1982; Lieberman, 1967, pp. 121–122). Moreover, the somatic-type view cannot even account for the highly situation-specific, though stereotyped, acoustic patterns used by nonhuman primates. Vervet monkeys, for example, have three different alarm calls for their three most dangerous predators: snakes, birds of prey, and predatory cats (Seyfarth, Cheney, & Marler, 1980). Clearly the effects of one autonomic state (due to a "flight–fight" reaction) cannot account for the existence of three distinct acoustic patterns. However, the evolutionary, game-theoretic, emotion-scripts view of emotion expression as intention expression can not only *account* for (a) the importance of stress patterns in emotion speech, (b) the existence of situation-specific acoustic patterns in other primates, and (c) the failure of linguists to find simple mapping rules for emotion speech, it *predicts* them.

If emotions are intimately tied to specialized evaluative inference procedures, and emotional expression involves the expression of intentions, or of valuations from which intentions can be deduced by a hearer who is using similar inference procedures, then signaling that your intentions differ from those the hearer presupposes (or may presume in the absence of new information) is an instance of highlighting new or important information. Linguistically, this function calls for the use of contrastive stress. So, the emotion-scripts view of emotion communication as intention communication predicts that the acoustic expression of emotion will include acoustic parameters that reflect the use of contrastive stress. Furthermore, the use of similar inference procedures predicts great uniformity in subjects' stress patterns for a particular emotional situation. On this view, a one-to-one mapping of acoustic properties onto a list of simple emotions would be misguided because stress patterns depend on how the particulars of the situation feed into the emotion script. For example, I could create a situation where you were just as "angry" as the speaker in Scene 5, by arbitrarily and cruelly deciding to let someone else do something you really want to do. Presumably, you would shout "I'LL do it," producing a fall–rise pattern with emphasis on the "I'll" rather than on the "doit," as in Scene 5. Thus the intentional view would explain the difficulty linguists have had in finding systematic rules for assigning stress in emotion speech. Furthermore, in cases where it is important for highly social, though nonlinguistic species, like nonhuman primates, to differentiate situations that may engender the same autonomic response (such that acoustic correlates of autonomic response are not sufficiently informative), the intentional view predicts the existence of regular, differentiated acoustic patterns for expressing the different situation-specific information.

The intentional view stresses the importance of expressing not only the relations between words, as generative semantics, for example, already does, but also the speaker's valuing of those relations and the actions they represent. As currently construed, semantic theories cannot represent this intentional information such that it can be readily incorporated into a performance model of AEC production. Schank's (1972) conceptual dependency model and Fillmore's (1968) case grammar, for example, use semantic representations in which relations are internally defined, that is, defined between classes of words or "ideas" that function similarly. However, their approaches do not capture the *evaluative* relationship the speaker has to these agent–action–object relationships. A conceptual dependency or case grammar representation of "I'll do it" does not tell how the speaker *feels* about doing it—does he or she want to or not, will he or she do it in the future or not, does he or she consider doing it to be a positive benefit or a way of avoiding unwanted cośts, and so on. Not even Schlesinger's (1971) "intentional grammar" expresses this relationship. His "I-markers," like the base structures in generative semantics, specify the relations between elements (thereby expressing what the speaker intended *to say*), but contain no evaluative position explicating the speaker's *behavioral* intentions regarding the actions, people, or states being discussed.

At least three types of representations (plus emotion-script processing systems) could embody such evaluative information:

1. Semantic representations like Schank's or Fillmore's could be fed through an "emotion processor" that evaluates the intensions expressed by the representation so intentional acoustic markers can be appropriately assigned. This seems like putting the cart before the horse, however. After all, the speaker's emotional attitude toward the referent of the utterance is often part of the reason he or she wants to make the statement in the first place. Therefore, from the perspective of a performance model, it would be odd to assign emotion values after the decision regarding what to say has been made.

2. Evaluations could be represented as propositions separate from the agent–action–object relations of the sentence actually uttered. A "grammatical" transformation would then be applied, superimposing acoustic patterns corresponding to the evaluative propo-

sitions on the utterance expressing the central proposition. The transformational rule of attachment, for example, says that the two deep structures "I'll do it" and "I don't want to do it" can be combined into one surface structure, "I'll do it but I don't want to." Logically, however, the same transformation is happening when you say "I'll do it" in a tired, resigned tone of voice. You are attaching the same two propositions, but substituting a particular tone of voice for the phrase "but I don't want to."

3. The relations between agent, action, and object in the underlying semantic representation could be labeled with the speaker's evaluations of them. Intonational rules for producing the acoustic configurations would be applied accordingly. In this system, motivational factors are an integral part of the semantic representation. Assume the representation of "I'll do it" is as in a case grammar. If I said it in response to the hundredth request you made today, the agent relation (corresponding to "I") might be labeled with a negative evaluation and emphasized, but not the verb or object because what I object to is the fact that I am doing anything at all for you, regardless of what it is I am doing. On the other hand, if said in reply to a request to clean up vomit, the verb might be labeled negatively and emphasized because it is the process of doing the action rather than the fact of who I'm doing it for or the resulting state (having the floor clean—the referent of "it") that is unpleasant. This system captures the fact that it is often specific words that are emphasized in particular ways, a convenient property for a performance model.

The discovery that different individuals adhere to reasonably specific acoustic patterns in expressing different emotions promises to add a new dimension to the study of language. These patterns cannot be explained by current syntactic or semantic theories. Interpreting them in terms of the speaker's valuing of the relations, persons, actions, or entities represented in the sentence, that is, in terms of the speaker's intentions, opens the way for the reconciliation of two views of language: language as a formal system of rules, and language as an evolutionary adaptation. If the relation between emotion and language is pursued, this reconciliation may take the form of a theory of intentional generative semantics.

Reference Notes

1. Scherer, K. R., & Wallbott, H. G. *Cues and channels in emotion recognition.* Manuscript in preparation, University of Giessen, Giessen, West Germany.
2. Cosmides, L. *Feelings, emotions, and the voice: An experimental approach.* Unpublished manuscript, Harvard University, 1982.
3. Rosenthal, R. Personal communication, April 1983.
4. Stevens, K. Personal communication, January 1983.

References

Alcock, J. *Animal behavior: An evolutionary approach* (2nd ed.). Sunderland, Mass.: Sinauer, 1979.

Bartlett, F. C. *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press, 1932.

Bolinger, D. Accent is predictable (If you're a mind reader). *Language,* 1972, *48,* 633–644.

Bolinger, D. Intonation and its parts. *Language,* 1982, *58,* 505–533.

Bresnan, J. Sentence stress and syntactic transformations. *Language,* 1971, *47,* 257–280.

Brown, B. L., Strong, W. J., & Rencher, A. C. Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America,* 1973, *55,* 313–318. (a)

Brown, B. L., Strong, W. J., & Rencher, A. C. Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *Journal of the Acoustical Society of America,* 1973, *54,* 29–35. (b)

Chagnon, N. A., & Irons, W. *Evolutionary biology and human social behavior: An anthropological perspective.* North Scituate, Mass.: Duxbury Press, 1979.

Chomsky, N., & Halle, M. *The sound pattern of English.* New York: Harper & Row, 1968.

Clark, H. H., & Clark, E. V. *Psychology and language.* New York: Harcourt Brace Jovanovitch, 1977.

Cooper, W. E., & Sorensen, J. M. *Fundamental frequency in sentence production.* New York: Springer, 1981.

Davitz, J. R. *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Dawkins, R. *The extended phenotype.* San Francisco: Freeman, 1982.

Ekman, P., Friesen, W. V., & Scherer, K. R. Body movement and voice pitch in deceptive interaction. *Semiotica,* 1976, *16,* 23–27.

Eldred, S. H., & Price, D. B. A linguistic evaluation of feeling states in psychotherapy. *Psychiatry,* 1958, *21,* 115–121.

Fairbanks, G., & Pronovost, W. An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monographs,* 1939, *6,* 87–104.

Fillmore, C. J. The case for case. In R. Jacobs & T. Harms (Eds.), *Universals in linguistic theory.* New York: Holt, 1968.

Gunter, R. Review of D. R. Ladd's "The structure of intonational meaning." *Language in Society,* 1982, *11,* 297–307.

Hamilton, W. D. The genetical evolution of social behaviour. *Journal of Theoretical Biology,* 1964, *7,* 1–52.

Hauser, K. O. The use of acoustical analysis for identification of client stress within the counseling session (Doctoral dissertation, North Texas State University, 1975). *Dissertation Abstracts International,* 1976, *36,* 5149–5150. (University Microfilms No. 76-43,97)

Hornby, P. A. The psychological subject and predicate. *Cognitive Psychology,* 1972, *3,* 632–642.

Huttar, G. L. Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research,* 1968, *11,* 481–487.

James, W. *Principles of psychology.* New York: Holt, 1890.

Jolly, A. *The evolution of primate behavior.* New York: Macmillan, 1972.

Jones, D. *The phoneme: It's nature and use.* Cambridge, England: W. Heffer & Sons, 1962.

Laver, J., & Trudgill, P. Phonetic and linguistic markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech,* Cambridge, England: Cambridge University Press, 1979.

Lazarus, R. S. Thoughts on the relations between emotion and cognition. *American Psychologist,* 1982, *37,* 1019–1024.

LeGuin, U. K. *The beginning place.* New York: Harper & Row, 1980.

LeGuin, U. K. *The wind's twelve quarters.* New York: Harper & Row, 1975.

LeGuin, U. K. *A wizard of Earthsea.* Boston: Houghton Mifflin, 1968.

Lehiste, I. *Suprasegmentals.* Cambridge, Mass.: MIT Press, 1970.

Lieberman, P. *Intonation, perception, and language.* Camridge, Mass.: MIT Press, 1967.

Markel, N. N., Bein, M. F., & Phillips, J. A. The relationship between words and tone-of-voice. *Language and Speech,* 1973, *16,* 15–21.

Maynard Smith, J. Game theory and the evolution of behavior. *Proceedings of the Royal Society of London.* 1979, *205,* 475–488.

Pavio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph,* 1968, *76*(1, Pt. 1).

Pike, K. *The intonation of American English.* Ann Arbor: University of Michigan Press, 1945.

Popp, J. L., & DeVore, I. Aggressive competition and social dominance theory: Synopsis. In D. A. Hamburg & E. R. McCown (Eds.), *The great apes.* Menlo Park, Calif. Benjamin/Cummings, 1979.

Schank, R. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology,* 1972, *3,* 552–631.

Schank, R., & Abelson, R. P. *Scripts, plans, goals, and understanding.* Hillsdale, N.J.: Erlbaum, 1977.

Schlesinger, I. Production of utterances and language acquisition. In D. Slobin (Ed.), *The ontogenesis of grammar.* New York: Academic Press, 1971.

Scherer, K. R. Speech and emotional states. In J. K. Darby (Ed.), *Speech evaluation in psychiatry,* New York: Grune & Stratton, 1981. (a)

Scherer, K. R. Vocal indicators of stress. In J. K. Darby (Ed.), *Speech evaluation in psychiatry,* New York: Grune & Stratton, 1981. (b)

Scherer, K. R., London, H., & Wolf, J. J. The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality,* 1973, *7,* 31–44.

Seyfarth, R. M., Cheney, D. L., & Marler, P. Monkey responses to three different alarm calls: Evidence for predator classification and semantic communication. *Science,* 1980, *210,* 801–803.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Smith, N., & Wilson, D. *Modern linguistics: The results of Chomsky's revolution.* Bloomington: Indiana University Press, 1979.

Trager, G. L., & Smith, H. L. *Outline of English structure: Studies in linguistics, No. 3.* Norman, Okla.: Battenburg, 1951.

Trivers, R. L. Parent–offspring conflict. *American Zoologist,* 1974, *14,* 249–264.

Utsuki, N., & Okamura, N. Relationship between emotional state and fundamental frequency of speech. *Reports of Aeromedical Laboratory,* 1976, *16*(4), 179–188.

Williams, C. E., & Stevens, K. N. Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America,* 1972, *52,* 1238–1250.

Williams, G. C. *Adaptation and natural selection: A critique of some current evolutionary thought.* Princeton, N.J.: Princeton University Press, 1966.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Zajonc, R. B. Feeling and thinking: Preferences need no inferences. *American Psychologist,* 1980, *35,* 151–175.

(Appendix follows on next page)

Appendix

Script 5[1]

I kept up my foolishness for the laughter's sake, laughing with them, for after those two long nights of dance and moonlight and music and magery I was in a fey and wild mood, ready for whatever might come.

Jasper, who never laughed aloud, looked at me. "I am sick of boys and noise and foolishness," he said.

"You're getting middle-aged, lad," Vetch remarked from above.

"If silence and gloom is what you want," put in one of the younger boys, "you could always try the Tower."

I said to him, "What is it you want, then, Jasper?"

"I want the company of my equals," Jasper said. "Come on, Vetch. Leave the prentices to their toys."

I turned to face Jasper. "What do sorcerers have that prentices lack?" I inquired. My voice was quiet, but all the other boys suddenly fell still, for in my voice, as in Jasper's the spite between us now sounded plain and clear as steel coming out of a sheath.

"Power," Jasper said.

"I'll match your power act for act."

"You challenge me?"

"I challenge you."

Vetch had dropped down to the ground, and now he came between us, grim of face. "Duels in sorcery are forbidden to us, and well you know it. Let this cease!"

Both Jasper and I stool silent, for it's true that we knew the law of Roke, and we also knew that Vetch was moved by love, and ourselves by hate. Yet our anger was balked, not cooled. Presently, moving a little aside as if to be heard by Vetch alone, Jasper spoke, with his cool smile: "I think you'd better remind your goatherd friend again of the law that protects him. He looks sulky. I wonder, did he really think I'd accept a challenge from him? a fellow who smells of goats, a prentice who doesn't know the First Change?"

"Jasper," said I, "What do you know of what I know?"

For an instant, with no word spoken that any heard, I vanished from their sight, and where I had stood a great falcon hovered, opening its hooked beak to scream: for one instant, and then I stood again in the flickering torchlight, my dark gaze on Jasper.

Jasper had taken a step backward, in astonishment: but now he shrugged and said one word: "Illusion."

The others muttered. Vetch said, "That was not illusion. It was true change. And enough. Jasper, listen—"

"Enough to prove that he sneaked a look in the Book of Shaping behind the Master's back: what then? Go on, Goatherd. I like this trap you're building for yourself. The more you try to prove yourself my equal, the more you show yourself for what you are."

At that, Vetch turned from Jasper, and said very softly to me, "Sparrowhawk, will you be a man and drop this now—come with me—"

I looked at my friend and smiled, but said nothing. "Now," I said to Jasper, quietly as before, "what are you going to do to prove yourself my superior, Jasper?"

"I don't have to do anything, Goatherd. Yet I will. I will give you a chance—an opportunity. Envy eats you like a worm in an apple. Let's out the worm. Once by Roke Knoll you boasted that Gontish wizards don't play games. Come to Roke Knoll now and show us what it is they do instead. And afterward, maybe I will show you a little sorcery."

"Yes, I should like to see that," I answered, cooly. "What would you like me to do, Jasper?"

The older lad shrugged, "Summon up a spirit from the dead, for all I care!"

"I will."

"You will not." Jasper looked straight at me, rage suddenly flaming out over his disdain. "You will not. You cannot. You brag and brag—"

"By my name, [I'll do it!]"

Script 10[2]

There is a city called Omelas. How can I tell you about the people of Omelas? We have almost lost hold; We can no longer describe a happy man, nor make any celebration of joy. But in Omelas . . .

The festival of summer! A marvelous smell of cooking goes forth from the red and blue tents of the provisioners. The faces of small children are amiably sticky: in the benign grey beard of a man a couple of crumbs of rich pastry are entangled. The youths and girls have mounted their horses and are beginning to group around the starting line of the race course. An old woman, small, fat, and laughing, is passing out flowers from a basket, and

[1] Text adapted from the book A WIZARD OF EARTHSEA by Ursula K. LeGuin, published by Houghton Mifflin Company, Boston. Copyright © 1968 by Ursula K. LeGuin. Reprinted by permission of the publisher.

[2] Text adapted from the book THE WIND'S TWELVE QUARTERS by Ursula K. LeGuin. Copyright © 1975 by Ursula K. LeGuin. Reprinted with permission of Harper & Row, Publishers, Inc.

tall young men wear her flowers in their shining hair. A child of nine or ten sits at the edge of the crowd, alone, playing on a wooden flute. People pause to listen, and they smile, but they do not speak to him, for he never ceases playing and never sees them, his dark eyes wholly rapt in the sweet, thin magic of the tune.

The people of Omelas have compassion, too. They have compassion because of the existence of another child, a child locked away, out of sight. It is because of the child that they are so gentle with children. They know that if the wretched one were not there, sniveling in the dark, the other one, the flute player, could make no joyful music as the young riders line up in their beauty for the race in the sunlight of the first morning of summer.

They all know it is there, all the people of Omelas. Some of them have come to see it. I came to see it.

In a basement under one of the beautiful public buildings of Omelas, there is a room. It has one locked door, and no window. A little light seeps in dustily between cracks in the boards, secondhand from a cobwebbed window somewhere across the cellar. In one corner of the little room a couple of mops, with stiff, clotted, foul-smelling heads, stand near a rusty bucket. The floor is dirt, a little damp to the touch, as cellar dirt usually is. The room is about three paces long and two wide: a mere broom closet or disused tool room. In the room a child is sitting. It could have been a boy or a girl. It looked about six, but actually was nearly ten. It picked its nose and occasionally fumbled vaguely with its toes or genitals, as it sat hunched in the corner farthest from the bucket and the two mops. It was afraid of the mops. It found them horrible. It shuts its eyes, but it knows the mops are still standing there; and the door is always locked; and nobody ever comes, except that sometimes the door rattles terribly and opens, and a person, or several people, are there. One kicks the child to make it stand up. The others never come close, but peer in at it with frightened, disgusted eyes. The food bowl and the water jug are hastily filled, the door is locked, the eyes disappear. The child used to scream for help at night, and cry a good deal, but when I came it only made a kind of whining, "eh-haa, eh-haa." It is so thin there are no calves to its legs; its belly protrudes; it lives on a half-bowl of corn meal and grease a day. It is naked. Its buttocks and thighs are a mass of festered sores; as it sits in its own excrement continually.

My sister held the child's bowl. "I can't," she said. "I can't even look at it."

"Give me the bowl," I said. ["I'll do it."]

Received November 15, 1982
Revision received June 8, 1983 ■