A Combination of Vocal f_0 Dynamic and Summary Features Discriminates between Three Pragmatic Categories of Infant-Directed Speech

Gary S. Katz, Jeffrey F. Cohn, and Christopher A. Moore

University of Pittsburgh

KATZ, GARY S.; COHN, JEFFREY F.; and MOORE, CHRISTOPHER A. *A Combination of Vocal f_0 Dynamic and Summary Features Discriminates between Three Pragmatic Categories of Infant-directed Speech.* CHILD DEVELOPMENT, 1996, **67**, 205–217. To assess the relative contribution of dynamic and summary features of vocal fundamental frequency (f_0) to the statistical discrimination of pragmatic categories in infant-directed speech, 49 mothers were instructed to use their voice to get their 4-month-old baby's attention, show approval, and provide comfort. Vocal f_0 from 621 tokens was extracted using a Computerized Speech Laboratory and custom software. Dynamic features were measured with convergent methods (visual judgment and quantitative modeling of f_0 contour shape). Summary features were f_0 mean, standard deviation, and duration. Dynamic and summary features both individually and in combination statistically discriminated between each of the pragmatic categories. Classification rates were 69% and 62% in initial and cross-validation DFAs, respectively.

Parent-infant interaction is a reciprocal process in which each partner influences the attention, affect, and activities of the other (Maccoby & Martin, 1983; Tronick, 1989). Because infant linguistic competence is limited, this process depends on paralinguistic features, which include prosody, facial expression, and other expressive modalities. Contemporary research has emphasized parent and infant facial expression to the relative neglect of other modalities (e.g., Malatesta, Culver, Tesman, & Shepard, 1989; Matias & Cohn, 1993). This bias toward facial expression has been informed by the primacy of facial expression in theories of emotion (e.g., Ekman, 1984; Izard & Malatesta, 1987) and the availability of well-developed measurement systems for describing dynamic changes in facial expression (Ekman & Friesen, 1978; Izard, 1983; Izard, Dougherty, & Hembree, 1983; Oster & Rosenstein, in press). This relative neglect of prosody is unfortunate because acoustic stimuli are surely among those communicative markers most salient to infants (Caron, Caron, & MacLean, 1988; Lewkowicz, 1988).

Several observations support the representation of audition as a primary mode of parent-infant communication. Hearing develops and differentiates ontogenetically earlier than the other senses (Bronson, 1982; Gottlieb, 1971), and infants can detect subtle differences in highly similar speech signals (Kuhl, 1987). In contrast, infants in the first half year have relatively poor visual acuity and low contrast sensitivity, which limit their ability to make subtle visual discriminations (Banks & Salapatek, 1983; Lewkowicz, 1988). Smell, movement, and touch are variously developed in the infant, but it is difficult to encode highly specific messages in these modalities alone (Goldstein, 1989). For these reasons, it is likely that hearing is perceptually dominant and affords a critical channel for communicating with infants.

This research was supported by NSF grant no. BNS-8919711 to Jeffrey Cohn and Christopher Moore and NIMH grant no. MHR0140867 to Susan Campbell and Jeffrey Cohn. Portions of these data were presented at the biennial meeting of the Society for Research in Child Development in New Orleans, LA, in March 1993, and at the annual meeting of the American Speech, Language, and Hearing Association in San Antonio, TX, in November 1992. Thanks are due to Lourdes Caro-Martinez, Susan Ranier, Joan West, and Adena Zlochower for assistance with data collection and analysis and to Michael Pogue-Geile, Adena Zlochower, Anne Fernald, Robin Cooper, and anonymous reviewers for constructive suggestions at various points of the research. Correspondence concerning this article should be addressed to Gary Katz at the Department of Psychology, 604 OEH, 4015 O'Hara Street, Pittsburgh, PA 15260.

[*Child Development*, 1996, **67**, 205–217. 0009-3920/96/6701-0018$01.00]

Prosody is a composite of several suprasegmental (i.e., beyond the lexical, syntactic, and semantic content of the signal) acoustic variables of speech: pitch, the psychoacoustic percept corresponding roughly to vocal fundamental frequency; loudness, the psychoacoustic percept of intensity; and rhythm, which is related to speech rate and patterning. In speech directed to infants, these features are highly exaggerated relative to adult-directed speech (Fernald & Simon, 1984). The ubiquity of prosodic exaggeration in infant-directed speech (Fernald et al., 1989), together with what is known about auditory perception in infants, suggests that suprasegmental variation may be a primary or dominant feature of communication directed to infants. In the research reported here, we focus on one feature of prosody, vocal fundamental frequency, in relation to the pragmatic meaning of infant-directed utterances. Vocal fundamental frequency (f_0) may be the most salient of the prosodic components (Fernald & Kuhl, 1987), and it is the one that has received the most attention in this context.

Fernald and her colleagues (Fernald, Kermanschachi, & Lees, 1984; Fernald & Kuhl, 1987; Fernald et al., 1989) and Stern and his colleagues (Stern, Spieker, & MacKain, 1982) have proposed that parents vary dynamic features of vocal f_0 to communicate messages to their infant. According to this hypothesis, four predictions can be made: (1) rising contours are used to elicit infant attention or response; (2) bell-shaped, or rise-fall, contours are used to maintain infant attention or show approval; (3) slowly falling contours are used to soothe or comfort a distressed infant; and (4) rapidly falling contours are used to communicate prohibition (e.g., "stop").

Three lines of evidence are consistent with one or more of these predictions about dynamic features of f_0. First, all four contour types are common in infant-directed speech. Rising, bell-shaped, and falling contours were among five contour types Fernald and Simon (1984) found in mothers' speech to 3–5-day-old infants and were among seven Papousek and Papousek (Papousek, Papousek, & Bornstein, 1985; Papousek, Papousek, & Haekel, 1987) found in mothers' speech to 3-month-old infants. Stern et al. (1982) found rising, falling, and two types of bell contours in mothers' speech to 3–4-month-old infants; they also found sinusoidal contours, which they considered a variation of bell-shaped. The descriptive criteria in all these studies have been qualitative, and an exhaustive taxonomy of contour shapes has not been undertaken. Quantitative criteria, when used, have been arbitrary and not necessarily comparable between studies. Stern et al., for instance, rejected from their analyses contours that had less than a 128-Hz excursion overall. Fernald and Simon (1984) rejected contours that had less than a 13 semitone-per-second excursion. At basal frequency of about 260 Hz (the mean f_0 reported by Fernald & Simon, 1984), 13 semitones is more than twice the threshold used by Stern. Nevertheless, the contour types specified by the f_0 contour hypothesis above have been found repeatedly in infant-directed speech.

The second line of evidence comes from observational studies that find correspondences between f_0 contour type and infant attention or affect. Stern et al. (1982) found that mothers were more likely to use rising contours when their infant was looking away, and sinusoid or bell-shaped contours when their infant was looking toward them and positive in affect. Papousek, Papousek, and Symmes (1991) also found an increase in mothers' use of rising contours when their infants were looking away. Papousek et al. (1985), however, found that rising, bell-shaped, and sinusoidal contours were all more likely during infant "pleasant" affect than when the infant was fussy or in other states. They did not report whether infant gaze within the pleasant state was correlated with contour type. Because babies are more likely to have positive affect when their attention is toward the mother (Cohn & Tronick, 1987; Kaye & Fogel, 1980), this finding suggests that rising contours may not be unique to times when the baby is looking away and the mother wants to elicit her attention. Papousek et al. (1985) reported that mothers more often used falling contours when their infant was fussy, which is consistent with the f_0 contour hypothesis. The observational studies, then, are suggestive of specific correspondences between contour shape (i.e., dynamic features of vocal f_0) and the parent's communicative intent.

A third line of evidence comes from two mechanistic studies designed to determine whether f_0 contours have specific effects on infant attention and affect by manipulating f_0 contour and evaluating infant response. Using synthetic stimuli, Sullivan and Horowitz (1983) found that infants respond with increased attention to rising versus falling contours. Papousek, Bornstein, Nuzzo, Papousek, and Symmes (1990), using natural adult vocal sounds, also found that rising contours oriented infant attention, and sharply falling contours inhibited infant looking.

Although these findings suggest that dynamic features of f_0 are critical in discriminating between pragmatic categories, discrimination may depend on other features of f_0 as well, including mean level, variability or range, and duration. Thus, an alternative hypothesis is that low-temporal resolution, or summary, features of f_0 contribute to pragmatic meaning as well (Papousek et al., 1991). Some data are consistent with this suggestion. Papousek et al. (1991) found numerous relations between summary measures of f_0 and both the pragmatic meaning of an utterance and behavioral context. For instance, attention bids had higher mean f_0 and wider f_0 range than "turn-encouraging rises," which were used once infant attention was achieved. Comfort utterances differed in mean f_0 and f_0 range from "contingent rewarding" utterances (i.e., approvals), and Papousek and colleagues (1990) found that approvals were longer in duration than disapprovals. Fernald (1989) also reported relations between summary measures and pragmatic category. Although no statistical tests were reported, she found that mean f_0 was higher for approvals than for attention bids, which in turn were higher than prohibitions or comforts. The f_0 range showed a similar rank ordering. Approvals also differed from other pragmatic categories in their longer duration. These findings suggest that pragmatic category is encoded in both f_0 contour and summary characteristics.

To assess the relative contribution of dynamic and summary features to the pragmatic meaning of infant-directed utterances, we instructed mothers of 4-month-old infants to use their voice to elicit their baby's attention, to show approval, and to provide comfort. A fourth condition, prohibition, was used in pilot testing only because too few mothers of 4-month-old infants were able to comply with this request. Anecdotally, the infants often responded with negative affect to prohibitions. Fernald (1993), too, informally reported strong negative affect in response to prohibitory utterances. The instructions specified the pragmatic content of the mothers' utterances but did not prescribe what words the mothers were to use or how they were to speak them.

To describe dynamic features, we used convergent methods. One was the standard method of visual classification (e.g., Fernald & Simon, 1984; Masataka, 1992; Papousek et al., 1990; Stern et al., 1982). A second was quantitative modeling. Quantitative modeling has not been used before with f_0 contours, but previous research has used mathematical descriptors (e.g., sinusoidal) to describe contour shapes. These descriptors refer to simple mathematical functions that could be tested mathematically. Although an infinite number of curves might be fit, we chose quantitative descriptors based on the qualitative shapes described in previous research, namely, rising, falling, bell-shaped, and sinusoidal or wavelike. Quantitative description has several potential advantages that may increase its concurrent and predictive validity. These include measurement on a continuous scale, rather than the mutually exclusive nominal codes in visual judgment methods, and increased reliability, by reducing the subjectivity inherent in human-observer-based methods. Another motivator for using quantitative methods was that they have been used to model similar contours in a range of biologic systems (e.g., transfer functions to model neural receptor sensitivity and periodic functions to model EEG and ultradian and circadian rhythms).

In addition to dynamic features, we also measured standard summary features of f_0. Previous research suggests that they too may vary with pragmatic category. By including both dynamic and summary measures of f_0, we were able to evaluate the relative importance of each in statistically discriminating between pragmatic categories.

Method

Subjects

Subjects were 49 mother-infant pairs who had participated in a telephone survey conducted by Campbell and Cohn (1991) when their infants were 6 weeks of age. The mothers were contacted when their infants were 3½ months of age and invited to participate in the study. The mothers were primiparous, married and living with their spouse, caucasian, and had not experienced a postpartum depression, which may influence infant-directed speech (Bettes, 1988). Their average age was 28 years (SD = 3.3 years), and each had a minimum of a high school education. The infants were full-term and averaged 18 weeks of age (SD = 1 week) at the time of the study.

Setting and Recording Equipment

The observation room was equipped with an infant seat mounted on a table, a facing chair for the mother, and two videocameras connected to a split-screen generator, which provided frontal views of mother and infant. Mothers' vocalizations were recorded using a lavaliere microphone. Beginning with the twenty-fifth subject, mothers also wore accelerometers (about .25 inches in diameter and affixed to the midline of the mother's throat using double-sided tape), which transduced vibrations in the skin above the vocal folds into audible signals. Accelerometric signals are free of simultaneous speech and contain little acoustic information other than intensity and fundamental frequency (i.e., no formants). To afford unrestricted motion, microphone and accelerometric signals were transmitted to a recording station via wireless transmitters, where they were recorded along with split-screen video on a Panasonic AG-7400.

Procedure

The procedure consisted of 5 min of face-to-face play followed by three experimental conditions, or vignettes, presented in a pseudo-randomized order. Mothers were instructed to use their voice to (1) direct their baby's attention to a red ring, (2) indicate approval when their baby reached out and grasped the ring (all babies spontaneously grasped the ring), and (3) verbally comfort their baby as if the baby were upset. The instructions, which were adapted from Fernald (1989) and presented on cue cards in a predetermined sequence, specified the pragmatic function of utterances within each vignette (e.g., "Call your baby's attention to the red ring"), but did not prescribe what words the mother was to use or how she was to speak them. The vignettes were presented in random order, with the exception of approval, which followed attention bids.

Analytical Method

Digitizing and parsing.—The microphone and accelerometric signals were digitized at 5,000 samples/sec using a Computerized Speech Laboratory (CSL) (Kay Elemetrics; for a comprehensive review and additional technical information, see Read, Buder, & Kent, 1992). Tokens were defined as expressing the target message (e.g., comfort), embedded in a continuous utterance, and separated from adjoining utterances by pauses that appeared as natural breaks (Papousek et al., 1987). For example, the following vocalizations from the comfort condition were coded as three separate tokens: "Awww . . . That's Ok. You're Ok."

From digitized speech samples, coders who were blind to the study's hypotheses marked the onset and offset times of tokens within each condition. Onset and offset times of extraneous utterances were not marked. Extraneous utterances included reading the instructions out loud, talking to the experimenter, and tokens inappropriate to the experimental condition (e.g., comfort in the approval condition). Also omitted were nonvoiced sounds (e.g., whispers) and simultaneous mother and infant vocalizations, which could not be analyzed with the CSL.[1] Infant gaze or other behavior was not a criterion in selecting tokens.

To assess the reliability of token parsing, vignettes from 21% of the subjects were randomly selected for parsing by a second coder. Within each vignette, an agreement was counted when two coders agreed that the target token occurred (e.g., a comfort token in the comfort condition). Disagreements were counted when one coder identified an utterance that the other coder did not. Interobserver agreement for was 87%, 72%, and 80% for attention, approval, and comfort, respectively. Because extraneous and unanalyzable utterances were omitted, kappa could not be calculated. The mean difference in identifying onset times was 35, 66, and 55 msec for attention, approval, and comfort, respectively. The corresponding mean difference in identifying offset times was 24, 95, and 78 msec.

Vocal f_0 extraction.—Subjects produced a variable number of tokens of each type, for a total count of 711. Of these, 6% were excluded because they contained nonvoiced sounds or concomitant infant vocalizations, and less than 1% were excluded because of experimenter error. Fundamental frequency was extracted using the CSL. The parameters (e.g., spectral window, which varied be-

[1] To assess the relative frequency of extraneous and unanalyzable utterances, 15% of the utterances were randomly selected and then coded into three mutually exclusive categories: token, extraneous utterance, and nonvoiced or simultaneous mother and infant utterances. Sixty percent of the utterances met criteria for a token, 31% were extraneous utterances, and 8% were nonvoiced utterances or contaminated by simulations speech.

tween 5 and 20 msec) were interactively optimized for each token. Most of the f_0 contours, even though recording conditions and f_0 extraction parameters were optimized, contained predictable types of errors, including dropouts (i.e., periods during which the algorithm failed to detect voicing) and harmonic errors (i.e., periods during which the resultant frequency was in error by some factor, usually .5, which yielded the wrong source harmonic as the fundamental). These errors were corrected automatically using custom software (Katz, Moore, & Cohn, 1992; Moore, Cohn, & Katz, 1994).

The f_0 contours were smoothed using a Fast Fourier Transform.[2] After smoothing, 5% of the tokens contained 25 or fewer data points, which were too few for reliable curve fitting and were omitted from further analysis. The total set consisted of 621 tokens from which two samples were drawn for hypothesis testing and cross-validation.

Visual classification of f_0 contour.—Six hundred twenty-one contours were visually classified by three trained coders into one of five categories: rising, falling, bell, wave, and flat. The contours were displayed on a computer screen in random order and coders were blind to any identifying features (i.e., subject number and condition during which it was recorded). All contours were displayed on a common scale (2 sec in length and 1200 Hz in height). Coding was done independently, with disagreements resolved by discussion. Average agreement between pairs of coders was quantified with kappa coefficients, which correct for chance agreement. The effective reliability of the consensus classification was quantified using coefficient alpha (Nunnally & Bernstein, 1994). The consensus classification was used in the data analyses reported below. The average intercoder kappa was .72, and effective reliability for the consensus codes was .89. In the data analyses, token classification was dummy coded (e.g., if a token was coded as a rise, a score of 1 was assigned to the rise category and a score of 0 assigned to the fall, bell, and wave categories).

Modeling of f_0 contour.—Each contour was subjected to a commercially available line-fitting algorithm (Table Curve, Jandel Scientific). Equations were chosen that corresponded to the contour shapes described in previous research (i.e., rise, fall, bell-shaped, and sinusoidal or complex). Linear, power, transfer, decay, and exponential equations can represent both rising and falling shapes based on different values of parameters. Peak equations represent bell contours, and waveform functions represent sinusoidal or complex contours. For each of seven equation types (Fig. 1), we calculated a degrees-of-freedom adjusted R^2 statistic, which represented the proportion of variance accounted for by each respective equation type. Using degrees-of-freedom R^2 statistics controlled for differences in f_0 resolution. For the transfer, peak, and waveform equation types, the R^2 index was the average of several variants of the equation (e.g., sine- and sine-squared equations, in the case of the waveform index). Coefficient alphas for these composite indices were all .8 or higher.

R^2 statistics for the five equation types representative of rising or falling shapes (i.e., linear, power, transfer, decay, and exponential) were moderately correlated (mean r = .60). To eliminate redundancy of measures, the R^2 statistics for these five equations were averaged (alpha = .88) and split into two indices of trend. Trend was determined mathematically. Figure 1 details the exact combination of coefficients that represent slope for each equation. When trend was up, the rise index was coded as the average R^2, and the fall index was coded as 0. When trend was falling, the fall index was coded as the average R^2, and the rise index was coded as zero.

This data reduction resulted in four R^2 goodness-of-fit measurements for each contour. These represented rising, falling, bell, and waveform fits. The data analytic question was which fit, or shape, best discriminated among tokens.

Summary measures.—Mean f_0 was the average of the f_0 values from spectral frames containing voicing. Voicing was determined by the presence of periodic zero-crossings in the criterion range of the spectral frame.

[2] The fundamental frequency of an utterance varies continuously, but estimates of f_0 are analytically constrained to a range of discrete values based on the conditions under which the original speech sample was acquired (e.g., sampling rate) and analytic parameters used in f_0 extraction (e.g., type or size of window). The discontinuities between the discrete values are error. Fast Fourier Transform (FFT) smoothing is one of several digital signal processing techniques that adjust for these errors with a minimum of signal distortion.

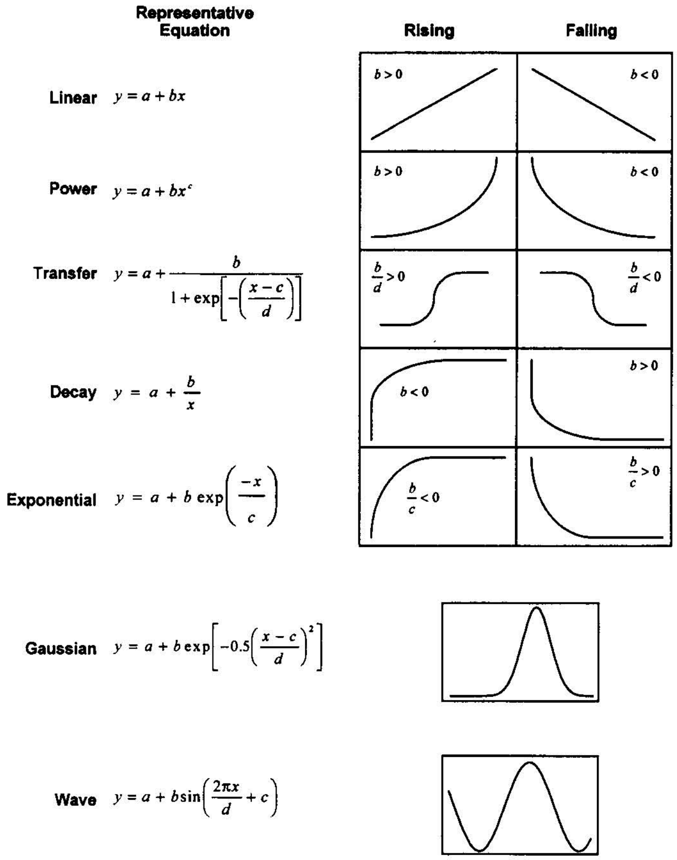

FIG. 1.—Modeling of f_0 contours by curve fitting

Unvoiced segments and pauses were omitted from the computation of average. The measure of variability in f_0 was the standard deviation of these f_0 values. Duration was measured from the f_0 contour, not the waveform.

Data analysis.—The test set consisted of 131 tokens randomly selected within subjects (one token of each type per subject). Two subjects were missing an approval, six subjects were missing an attention, and eight subjects were missing a comfort. The test data were screened for within-subject correlation. As this was found lacking (mean $r = .02$, SD = .14, $ps > .05$, 20 of 21 $ps > .20$), the tokens in the test ($n = 131$) and cross-validation samples ($n = 491$) were analyzed as independent observations. The data were next screened for any effect of f_0 source (microphone vs. accelerometer). No effect of f_0 source was found (in a series of univariate ANOVAs, $ps > .10$), so the data were collapsed across this variable in all subsequent analyses. Unless otherwise indicated, results refer to the test sample.

Results

Preliminary Results

Table 1 presents the correlations within and between the three sets of variables. Correlations between visual classifications and curve-fit or summary measures are point-biserial. Within each set, a number of variables were intercorrelated. These included positive correlations between f_0 mean and standard deviation and curve-fit indices of bell and wave. Several variables were negatively correlated, which reflected constraints in coding. For example, a bell contour could not also be coded as wave.

The visual classification and curve-fit indices both assess contour shape, and we anticipated moderate point-biserial correlations between these two methods. Moderate correlations between methods were found for both rise and fall but not for bell or wave.

TABLE 1

INTERCORRELATION MATRIX OF VOCAL f_0 CONTOUR SHAPE AND SUMMARY MEASURES

	Visual Classification				Curve-Fit Indices				Summary Measures	
	Rise	Fall	Bell	Wave	Rise	Fall	Bell	Wave	f_0 Mean	f_0 SD
Visual classification:										
Fall	−.14	...								
Bell	−.23*	−.29*	...							
Wave	−.25*	−.31*	−.53*	...						
Curve-fit indices:										
Rise	.57*	−.20	−.12	−.07	...					
Fall	−.16	.53*	−.14	−.14	−.43*	...				
Bell	.09	.10	.27*	−.33*	.21	.14	...			
Wave	.04	.20	.14	−.24*	.11	.23*	.36*	...		
Summary measures:										
f_0 Mean	−.12	−.10	.04	.18	−.07	−.10	−.03	−.16	...	
f_0 SD	−.13	−.07	.01	.21	−.03	.13	.16	.10	.45*	...
Duration	−.18	−.03	−.25*	.46*	−.25*	−.05	−.21	−.39*	.11	.17

NOTE.—Number of contours = 621.
* $p < .01$.

Correlations between visual classifications and summary measures were nonsignificant or low. Pearson correlations between the curve-fit indices and summary measures showed a similar pattern, which suggests that shape and summary measures of f_0 potentially contribute independent variance to the statistical discrimination of pragmatic category.

Do Shape and Summary Measures Vary between Pragmatic Categories?

Visual classification, curve-fit, and summary measures were analyzed in a MANOVA with condition as the between-subjects factor. Wilks's lambda was significant (Table 2). Subsequent univariate ANOVAs on each of the dependent measures were significant for two of the visual classi-

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR VOCAL f_0 CONTOUR SHAPE AND SUMMARY MEASURES

	Pragmatic Category			
	Attention	Approval	Comfort	*F*
Visual judgment:				
Rise	.23 (.43)[a]	.00 (.00)[b]	.07 (.26)[b]	7.68**
Fall	.09 (.29)	.15 (.36)	.20 (.40)	.88
Bell	.37 (.49)	.23 (.43)	.39 (.49)	1.49
Wave	.19 (.39)[a]	.62 (.49)[b]	.27 (.45)[a]	11.83**
Curve fit:				
Rise	.18 (.24)[a]	.03 (.08)[b]	.04 (.12)[b]	13.70**
Fall	.14 (.20)[a]	.28 (.25)[b]	.22 (.23)	4.53*
Bell	.70 (.33)	.59 (.30)	.63 (.29)	1.45
Wave	.68 (.29)	.56 (.30)	.63 (.30)	1.81
Summary measures:				
Mean f_0 (Hz)	295.35 (86.02)[a]	385.29 (115.81)[b]	307.47 (88.32)[a]	11.14**
f_0 variability (Hz)	52.31 (31.54)[a]	84.98 (31.05)[b]	56.05 (36.37)[a]	13.33**
Duration(s)	.65 (.32)[a]	1.12 (.44)[b]	1.09 (.43)[b]	18.65**

NOTE.—Wilks's lambda = .451. Univariate *F* ratios have df = 2, 128. Within rows, pairs marked with different superscripts significantly differ ($p < .05$, Tukey HSD).
* $p < .025$.
** $p < .01$.

fications (rise and wave), two of four curve-fit indices (i.e., rise and fall) and all three summary measures. Both shape and summary measures varied between pragmatic categories.

Visually classified rising contours were more common for attentions than for approvals or comforts. The curve-fit rise index showed a similar pattern: higher for attentions than for approvals or comforts. The curve-fit fall index was lower for attentions than for approvals or comforts. Visually classified wave but not bell contours were more common for approvals. The bell and wave indices, however, were high for all three token types, which suggests that bell and possibly wave shapes may lack specificity for pragmatic category.

Approvals had higher mean and SD f_0 than attentions and comforts. Attentions were of shorter duration than approvals and comforts.

Discrimination between Pragmatic Categories

To evaluate the relative contribution of shape and summary measures to discriminating between pragmatic categories, we used discriminant function analysis (DFA). Shape and summary variables were entered in two separate blocks. Because order of entry and the number of predictors may influence results, the analysis was conducted twice, once with the shape indices entered first and a second time with the summary measures entered first. We included only those variables that had significantly differed between pragmatic categories in the univariate ANOVAs. Because the rise and fall classifications and corresponding curve-fit indices were moderately correlated, we included only the latter in the DFA.

Shape and summary measures both individually and in combination statistically discriminated between each of the pragmatic categories (Table 3), and order of entry did not substantially influence the discrimination. Wilks's lambda at step 1 was similar for both shape and summary measures, and the combination of both sets of variables at step 2 resulted in significantly improved discrimination.

With three pragmatic categories (i.e., approval, attention, and comfort), there are two discriminant functions. Both functions were significant and together accounted for 61% of the total variance (canonical correlations = .67 and .40, $p < .001$, respectively). Both shape and summary variables contributed to both functions. Standardized discriminant coefficients, however, were marginally higher for the summary variables (Table 4). On function 1, f_0 mean and duration and the

TABLE 3

DISCRIMINANT FUNCTION ANALYSES: STEP-DOWN *F* RATIOS

Step and Block Entered	Group Contrasts			Wilks's Lambda
	AP vs. AT	AP vs. CO	AT vs. CO	
Analysis no. 1—original order:				
1:				
Rise index Fall index Wave consensus	17.46	6.14	6.32	.663
2:				
Mean f_0 f_0 variability Duration	16.65	6.84	7.08	.463
Analysis no. 2—reversed order:				
1:				
Mean f_0 f_0 variability Duration	19.38	7.47	8.48	.620
2:				
Rise index Fall index Wave consensus	16.65	6.84	7.08	.463

NOTE.—AT = Attention, AP = Approval, CO = Comfort; df = 3, 124 for change in Wilks's lambda, $ps < .001$; df = 3, 126 and 6, 126 for pairwise contrasts at blocks 1 and 2, respectively, $ps < .001$.

TABLE 4

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

Variable	Function No. 1	Function No. 2
Rise curve fit	−.344	.373
Fall curve fit	.426	.166
Wave consensus code	.368	.644
Mean f_0	.425	.255
f_0 variability	.284	.297
Duration	.437	−.759
Canonical correlations	.66*	.38*

* $p < .001$.

fall index were the largest coefficients. On function 2, f_0 duration and wave shape were the largest coefficients.

Classification results.—The two discriminant functions correctly classified 69% (kappa = .53, $p < .0001$) of the three token types (Table 5). All three token types were reliably discriminated from each other. Discrimination between approvals and attentions was the most accurate, with 88% correctly classified.

The classification results from the test and cross-validation samples were consistent. Applying the unstandardized discriminant function weights from the analysis of the test sample to the token measures in the cross-validation sample, all three token types were classified reliably again (correct classification rate of 62%, kappa = .43, $p < .0001$). Kappa coefficients between the two samples did not differ statistically ($z = 1.45$, $p > .10$).

To evaluate whether the pattern of correct and incorrect classifications was equivalent in the two samples, we conducted a log-linear analysis of the three-way table: sample (test vs. cross-validation) by actual category by predicted category. Two interaction terms were included in the model: one for the association between the actual and predicted classifications and the other for the association between the frequency of pragmatic categories and sample. The latter was included because the marginal distribution of categories varied little in the test sample but varied by a factor of over two in the cross-validation sample (Attention was underrepresented). The model omitted the three-way interaction among actual and predicted classifications and sample. Thus a nonsignificant likelihood ratio chi-square would signify that the pattern of classification did not vary between samples, which is what we found. The likelihood ratio chi-square was 11.60, $df = 6$, $p = .07$, with no standardized residual greater than 1.6.

Discussion

Theory and some data about the communicative function of infant-directed speech have emphasized the role of dynamic

TABLE 5

CROSS-CLASSIFICATION BETWEEN ACTUAL AND PREDICTED TOKENS FOR THE TEST AND CROSS-VALIDATION SAMPLES

	PREDICTED TOKEN		
	Attention	Approval	Comfort
Test sample: actual:			
Attention	31 (72%)	5 (12%)	7 (16%)
Approval	4 (9%)	35 (75%)	8 (17%)
Comfort	5 (12%)	12 (29%)	24 (59%)
Cross-validation sample: actual:			
Attention	61 (63%)	10 (10%)	26 (27%)
Approval	23 (11%)	126 (61%)	58 (28%)
Comfort	34 (18%)	31 (17%)	121 (65%)

NOTE.—Row percentages in parentheses.

changes in f_0 in discriminating between pragmatic categories. In its strongest version, this hypothesis suggests that the melody is the message (Fernald, 1989), and this perspective has influenced research over the past decade (Masataka, 1992; Stern et al., 1982; Sullivan & Horowitz, 1983). Our findings, together with those of others (Fernald, 1989; Papousek et al., 1991), strongly suggest that both dynamic and summary features are critical to the accurate discrimination of pragmatic categories.

Dynamic and summary features both individually and in combination statistically discriminated between each of the three pragmatic categories we studied. Discrimination was greatest when both sets of features were considered, and the two sets contributed relatively independent information. Wilks's lambda in the DFA improved by similar amounts regardless of whether dynamic or summary measures were entered in the initial step, and both dynamic and summary variables were among the highest loadings on the two discriminant functions. These findings suggest that both sets of features contribute to pragmatic meaning in infant-directed speech. What features the infant uses is not well understood. We know that adults use dynamic and summary features as well as other prosodic information in making perceptual discriminations between vocal expressions of emotion (Frick, 1985; Scherer, 1986), but comparable knowledge of infant perception is not available. Developmental studies are needed that systematically vary dynamic and summary features and possibly others and evaluate the responses of infants and children. Some of our findings are at variance with specific predictions about the relation between contour shape and pragmatic category. Fernald (1989) hypothesized that attention bids have rising contours, but we found that attention bids were characterized better as rise-fall. Visually classified rise contours were more common and rise indices greater for attention than for approval or comfort, but visually classified bell contours were common in all pragmatic categories. This finding suggests that many of the attention contours were actually rises followed by a slight fall. The R^2 values of attention on the bell index also implied that attention contours have a rapid rise followed by a terminal fall. Visual inspection of the fitted contours confirmed this impression. Rise-fall patterns predominated. The bell classification and index were able to capture this feature.

An implication is that it may be important to distinguish between attention bids whose goal is to orient a baby to turn her attention toward the mother (from avert) versus those whose goal is to elicit infant response to an object. Rise-only contours may characterize the former condition (Stern et al., 1982), and rise-fall contours the latter. In this regard, Stern et al. (1982) reported bell-right contours similar to ours associated with vocalizations such as, "Come on" and "Where's your rattle . . . ." Fernald (1989), who like us asked mothers to elicit their baby's attention to a toy, published two attention contours, one of which was also bell-shaped, or sinusoidal, depending on interpretation. The other contour, however, was rise-only. Further research is needed to determine whether subtypes of attention-bids may exist.

Papousek (Papousek et al., 1985, 1987) and Fernald (Fernald, 1984; Fernald et al., 1989) and their colleagues have described comfort tokens as falling, which we did not find. Visually classified fall contours were no more common and the fall index no higher for comforts than for attention or approval. Comfort contours contained peaks that were fit well by bell and wave shapes. This could occur several ways, and we saw instances of each. Often, comforts began with a slight terminal rise and then fell; less often, they began with a long fall and ended with a terminal rise. These patterns were not anticipated by previous research, but they were strongly suggested by the pattern of R^2 indices, visual classifications, and visual inspection of the contours. We also saw instances of a falling contour with a bell at the trough followed by tail. Fernald (1989) shows an example of this sort. These findings highlight the need for further descriptive studies of the f_0 contours mothers use in communicating with their infants.

For the pragmatic categories we studied, discrimination was best for attention versus approval. In the initial and cross-validated DFAs, accuracy for this classification was 85% or higher. The dynamic features of rise, fall, and wave and all three summary features contributed to this discrimination. In homes and other child-care settings, attention and approval occur in temporal proximity to one another, and it would be adaptive for them to be readily discriminable by varying on a wide range of characteristics such as we found.

The pairwise misclassification rate for comfort and attention or approval was higher than that for attention and approval with each other. A reason may be that there is little adaptive advantage for these tokens to be differentiated acoustically. In an infant's home, comfort and attention or approval are not likely to occur in temporal proximity to each other. Even when they do (e.g., parental attempts to distract an upset infant), attention and approval may have different correlates in other modalities that aid in discrimination of pragmatic meaning. For instance, it is likely that physical contact more often accompanies comfort vocalizations than it does vocalizations of approval or attention. In particular, facial and kinesthetic expression may afford important guidance in discriminating between pragmatic categories.

To describe contour shapes, we used both qualitative and quantitative methods. Quantitative modeling has not been used before with f_0 contours, but previous research has used mathematical descriptors (e.g., sinusoidal) to describe contour shapes. We evaluated the aptness of those descriptors. The mathematical functions we chose corresponded both to the shapes described in previous research and ones that are common in other biologic systems. We found that the curve-fit indices for the various functions had high internal consistency and provided robust fits of contour shapes. Bell and wave indices, in particular, accounted for 56% to 70% of the variance in pragmatic categories. The concurrent correlations between curve-fit and visually classified rise and fall contours were in the moderate range, which suggests convergent validity. The convergent correlations for wave and bell measures, however, were disappointing. One reason may be that the quantitative approach is overly sensitive to low-amplitude variation in contour shape. Slight sinusoidal variation might be disregarded as "noise" by a human observer but included as reliable "signal" in a quantitative analysis. If this is the case, additional statistical filtering might serve to increase the specificity of quantitative modeling without loss of sensitivity. In the meantime, it might be best to avoid the use of descriptors that imply mathematical functions. Contour shapes may be more complex than the simple functions referred to in previous research.

Prosodic analyses are highly labor intensive, and this has discouraged their use in developmental studies. This is unfortunate. Prosody is a dominant feature of the infants' socioemotional and cognitive experience and is known to be a powerful mechanism for communicating affect (Frick, 1985), yet we know relatively little about its role in infant affective development. Consider that theories of emotion development and parent-infant interaction either concentrate on facial expression alone or use measures that are highly influenced by facial expression (Izard & Malatesta, 1987; Malatesta et al., 1989; Matias, Cohn, & Ross, 1989). The information we have about prosody suggests that it may be involved in affective development and developmental psychopathology (Bettes, 1988; Breznitz & Sherman, 1987; Jaffe, Beebe, Feldstein, & Anderson, 1992; Jasnow et al., 1988) and in language development (Fernald & Mazzie, 1991). One of the ways in which depression may adversely impact cognitive development (Sameroff, Seifer, & Zax, 1982) is through impairments in mothers' use of prosody in the context of language learning. Semiautomated methods of f_0 measurement (Moore et al., 1994) and further improvements in the measurement of f_0 contours may make way for rapid advances in our ability to study communication processes in socioemotional and language development.

In summary, we found that dynamic and summary features of vocal f_0 contributed approximately equally to discrimination between pragmatic categories. A combination of three shapes (rise, fall, and wave) and three summary features (f_0 mean, standard deviation, and duration) correctly classified over 60% of pragmatic categories in a cross-validated DFA. These findings are consistent with theory and data from studies of emotion perception in adults, which suggest that both dynamic and summary prosodic features communicate emotion meaning. Several of our specific findings were at variance with previous predictions. For instance, attention contours were more often rise-fall than rise only, and the use of some mathematical descriptors for contour shape was questioned. Quantitative modeling showed acceptable convergent/discriminant validity for rise and fall contours but not for bell or wave. The latter had high sensitivity (accounting for over 50% of the variance across categories) but low specificity. Further work in quantitative modeling may make feasible semiautomated methods of prosodic analysis.

References

Banks, M. S., & Salapatek, P. (1983). Infant visual perception. In M. M. Haith & J. J. Campos

(Eds.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol.* **2**. *Infancy and developmental psychobiology* (pp. 435–571). New York: Wiley.

Bettes, B. (1988). Maternal depression and motherese: Temporal and intonational features. *Child Development*, **59**, 1089–1096.

Breznitz Z., & Sherman, T. (1987). Speech patterning of natural discourse of well and depressed mothers and their young children. *Child Development*, **58**, 395–400.

Bronson, G. W. (1982). Structure, status, and characteristics of the nervous system at birth. In P. Stratton (Ed.), *Psychobiology of the human newborn* (pp. 99–118). New York: Wiley.

Campbell, S. B., & Cohn, J. F. (1991). Prevalence and correlates of postpartum depression in first-time mothers. *Abnormal Psychology*, **100**, 594–599.

Caron, R. F., Caron, A. J., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development*, **59**, 604–616.

Cohn, J. F., & Tronick, E. Z. (1987). Mother-infant interaction: The sequence of dyadic states at three, six, and nine months. *Developmental Psychology*, **23**, 68–77.

Ekman, P. (1984). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion*. Hillsdale, NJ: Erlbaum.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, R. Golinkoff, M. T. Greenberg, C. Harding, & J. Bohannon (Eds.), *The origins and growth of communication* (pp. 5–29). Norwood, NJ: Ablex.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, **60**, 1497–1510.

Fernald, A. (1993, March). *Experimental studies of infants' responsiveness to affective vocalizations*. Paper presented at the meetings of the Society for Research in Child Development, New Orleans, LA.

Fernald, A., Kermanschachi, N., & Lees, D. (1984, April). *The rhythms and sounds of soothing: Maternal vestibular, tactile, and auditory stimulation and infant state*. Paper presented at the International Conference on Infant Studies, New York.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10**, 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, **27**, 209–221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, **20**, 104–113.

Fernald A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, **16**, 477–501.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, **97**, 412–429.

Goldstein, E. B. (1989). *Sensation and perception*. Belmont, CA: Wadsworth.

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In E. Tobach, L. R. Aronson, & E. Shaw (Eds.), *The biopsychology of development* (pp. 67–126). New York: Academic Press.

Izard, C. E. (1983). *The Maximally Discriminative Facial Movement Coding System*. Unpublished manuscript, University of Delaware.

Izard, C. E., Dougherty, L. M., & Hembree, E. A. (1983). *A system for identifying affect expressions by holistic judgments*. Unpublished manuscript, University of Delaware.

Izard, C. E., & Malatesta, C. Z. (1987). Perspectives on emotional development: 1. Differential emotions theory of early emotional development. In J. D. Osofsky (Ed.), *Handbook of infant development* (2d ed., pp. 494–554). New York: Wiley.

Jaffe, J., Beebe, B., Feldstein, S., & Anderson, S. (1992). Do vocal dialogues of infants of depressed mothers predict infant attachment? *Infant Behavior and Development*, **15** (Special ICIS Abstracts Issue), 50.

Jasnow, M., Crown, C., Feldstein, S., Taylor, L., Beebe, B., & Jaffe, J. (1988). Coordinated interpersonal timing of Down and non-delayed infants with their mothers: Evidence for a buffeted mechanism of social interaction. *Biological Bulletin*, **175**, 355–360.

Katz, G., Moore, C., & Cohn, J. F. (1992, November). *Semi-automated processing of very large natural f_0 samples*. Paper presented at the American Speech, Language, and Hearing Association, San Antonio, TX.

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology*, **16**, 454–464.

Kuhl, P. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception: From perception to cognition* (pp. 275–382). New York: Harcourt Brace Jovanovich.

Lewkowicz, D. J. (1988). Sensory dominance in infants: 1. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology,* **24,** 155–171.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent-child interaction. In E. M. Hetherington (Ed.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 1–102). New York: Wiley.

Malatesta, C. Z., Culver, C., Tesman, J., & Sheppard, B. (1989). The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development,* **54**(1–2, Serial No. 219).

Masataka, N. (1992). Early ontogeny of vocal behavior of Japanese infants in response to maternal speech. *Child Development,* **63,** 1177–1185.

Matias, R., & Cohn, J. F. (1993). Are MAX-specified infant facial expressions during face-to-face interaction consistent with differential emotions theory? *Developmental Psychology,* **29,** 524–531.

Matias, R., Cohn, J. F., & Ross, S. (1989). A comparison of two systems to code infants' affective expression. *Developmental Psychology,* **25,** 483–489.

Moore, C. A., Cohn, J. F., & Katz, G. (1994). Quantitative description and differentiation of fundamental frequency contours. *Computer Speech and Language,* **8,** 385–404.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* New York: McGraw-Hill.

Oster, H., & Rosenstein, D. (in press). *Baby FACS: Analyzing facial movement in infants.* Palo Alto, CA: Consulting Psychologists Press.

Papousek, M., Bornstein, M. H., Nuzzo, C., Papousek, H., & Symmes, D. (1990). Infant response to prototypical melodic contours in parental speech. *Infant Behavior and Development,* **13,** 539–545.

Papousek, M., Papousek, H., & Bornstein, M. (1985). The naturalistic vocal environment of young infants: On the significance of homogeneity and variability in parental speech. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 269–297). Norwood, NJ: Ablex.

Papousek, M., Papousek, H., & Haekel, M. (1987). Didactic adjustments in fathers' and mothers' speech to their 3-month-old infants. *Journal of Psycholinguistic Research,* **16,** 491–516.

Papousek, M., Papousek, H., & Symmes, D. (1991). The meaning of melodies in motherese in tone and stress languages. *Infant Behavior and Development,* **14,** 415–440.

Read, C., Buder, E. H., & Kent, R. D. (1992). Speech analysis systems: An evaluation. *Journal of Speech and Hearing Research,* **35,** 314–332.

Sameroff, A., Seifer, R., & Zax, M. (1982). Early development of children risk for emotional disorder. *Monographs of the Society for Research in Child Development,* **47**(7, Serial No. 199).

Scherer, K. R. (1986). Vocal affect expression: A review and model for further research. *Psychological Bulletin,* **98,** 143–165.

Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology,* **18,** 727–735.

Sullivan J. W., & Horowitz, F. D. (1983). The effects of intonation on infant attention: the role of the rising intonation contour. *Journal of Child Language,* **10,** 521–534.

Tronick, E. Z. (1989). Emotions and emotional communication in infants. *American Psychologist,* **44,** 112–119.