

SCHEMATIC PITCH CODING: A NEW, MORE EFFICIENT METHOD FOR MEASUREMENT OF INFANT-DIRECTED SPEECH

Gary S. Katz

California State University–Northridge

Jeffrey F. Cohn

University of Pittsburgh

Christopher A. Moore

University of Washington

Infant-directed speech is an important but difficult-to-measure aspect of caregiver-infant interaction. Specialized training in speech science is required and processing is computationally intensive and time consuming. We developed a computer-assisted method of measuring infant-directed speech that requires little specialized training, has high concurrent validity with alternative approaches, and has improved efficiency in estimating frequency and temporal parameters. In Schematic Pitch Coding, an operator uses a computer interface to play digitized speech samples, view the corresponding narrow-band spectrograms, and mark points of inflection in fundamental frequency (f_0) using a computer mouse. Custom software generates graphic representations of f_0 contours and quantitative estimates of the mean and standard deviation of f_0 , utterance duration, and turn-taking pauses. Our current system, which is not optimized for speed, can process a 30-second sample of infant-directed speech in approximately a half-hour, which makes feasible its application in large-scale studies of infant-directed speech.

prosody infant-directed speech quantitative measurement spectrogram schematic pitch coding
fundamental frequency pitch vocal acoustics

• **Gary S. Katz**, California State University–Northridge, Department of Psychology, 18111 Nordhoff Street, Northridge, CA 91330; Tel: (818) 677-2827; e-mail: gary.katz@csun.edu.

INFANT BEHAVIOR & DEVELOPMENT 22 (3), 1999, 283–295
Copyright © 2000 Elsevier Science Inc.

ISSN 0163-6383
All rights of reproduction in any form reserved.

The sound of infant-directed speech is unique. Relative to adult-directed speech, the variability and complexity of infant-directed speech provide infants and their caregivers with a highly salient communications medium with which to convey pragmatic and affective information (Ferguson, 1964; Fernald & Simon, 1984; Stern, Spieker, Barnett, & MacKain, 1983; Fernald, 1989; Katz, Cohn, & Moore, 1996). Newborns prefer infant-directed speech to adult-directed speech (Cooper & Aslin, 1990; Moon, Cooper, & Fifer, 1993; although cf. Cooper, Abraham, Berman, & Staska, 1997), with this preference appearing to be consistent throughout the first year of life (Fernald, 1984; Fernald, 1985; Fernald & Kuhl, 1987). Acoustically, researchers have examined the pitch, intensity, rhythmicity, and formant structure of infant-directed speech for specific components that may mediate infant preferences for this mode of interaction (Fernald & Kuhl, 1987; Fernald, 1989; Katz et al., 1996; Kuhl et al., 1997). Of these four features, pitch, as measured by vocal fundamental frequency (f_0), appears to be especially salient (Fernald, 1984; Fernald & Kuhl, 1987; Katz et al., 1996).

Fernald (1984; Fernald & Kuhl, 1987) and others (Katz et al., 1996; Stern et al., 1983) have speculated that changes in f_0 over time (f_0 contours) are the primary medium for communication between infants and their caregivers. Empirical findings support their speculations. Infants will reliably orient to speech that varies in f_0 (Fernald & Kuhl, 1987). Rapidly rising f_0 contours orient infant attention (Fernald, 1989; Katz et al., 1996). Gently falling f_0 contours soothe distressed infants (Fernald, 1989; Katz, 1998; Papousek, Papousek, & Symmes, 1991). Given the importance of f_0 contours in moderating infant behavior, reliable, valid, and efficient methods of measuring f_0 are essential.

Two approaches have been used to describe features of f_0 contours. One approach is visual classification of f_0 shape. Fernald (1985), for

instance, identified four regularly occurring shapes, including "rising," "falling," "bell-shaped," and "complex." A second approach is the measurement of quantitative f_0 features. Typically, this measurement requires digitization of speech samples, preliminary extraction of f_0 values over the length of the speech sample, error-correction of the extracted f_0 values, and smoothing of the resulting error-corrected values (Moore, Cohn, & Katz, 1994). Summary statistics, such as mean f_0 or f_0 variability, are then calculated from the resultant f_0 contours (Fernald & Simon, 1984; Stern et al., 1983).

There are strengths and limitations to these two approaches to describing f_0 contours. Visual inspection provides rapid identification of f_0 contour shape but no quantitative measurements. Quantitative features require accurate pitch extractions, which are difficult to obtain, labor intensive, and prone to error (Moore et al., 1994; Reed, Buder, & Kent, 1992). We present a computer-assisted method of acquiring infant-directed pitch contours that allows for rapid identification of contour shape and quantitative measurement. Although Schematic Pitch Coding is a subjective method of measuring pitch changes over time, it has high concurrent validity with objective quantitative algorithms for pitch extraction. Because of its increased efficiency relative to quantitative pitch extraction, Schematic Pitch Coding is especially well suited for processing large databases of infant-directed utterances.

This manuscript describes the Schematic Pitch Coding System and evaluates its reliability, validity, and efficiency. The researcher interested in large-sample studies of infant-directed speech should consider Schematic Pitch Coding as a strong alternative to conventional pitch measurement techniques. Schematic Pitch Coding produces nearly identical results when compared with conventional pitch extraction techniques at a fraction of the cost in human labor and post-processing time.

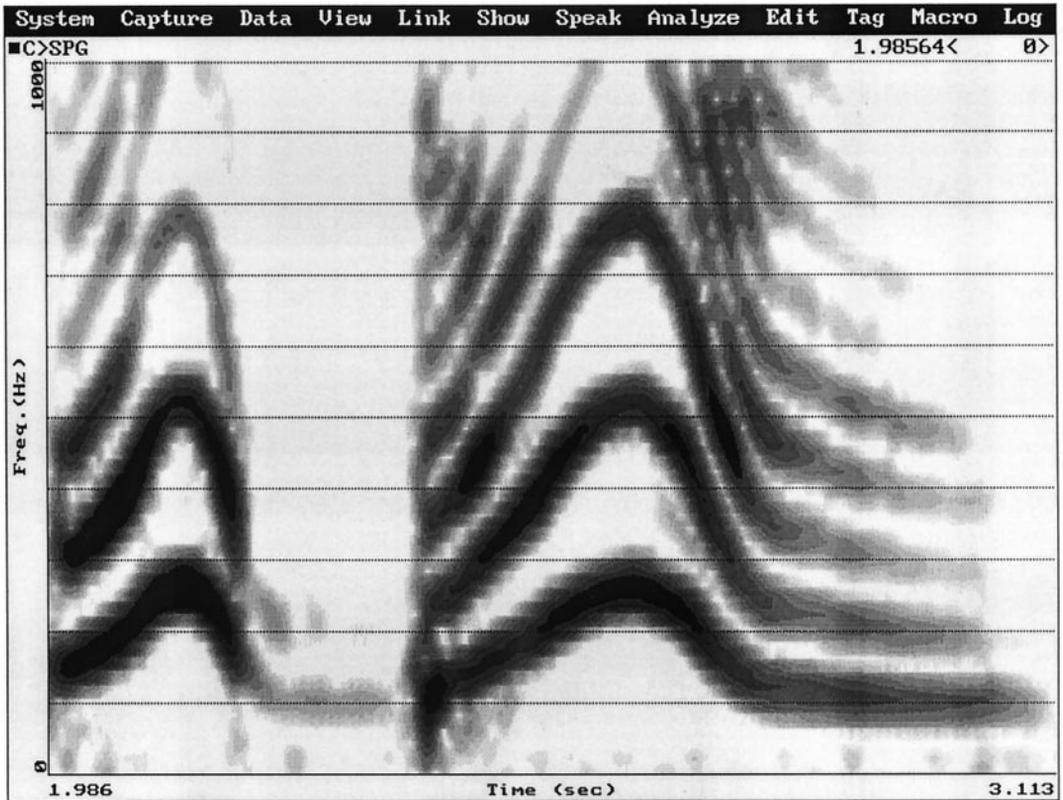


FIGURE 1

Sample Spectrogram

Method

Spectrographic Analysis

Schematic Pitch Coding captures the structure of f_0 contours based on the representation of f_0 changes depicted in a spectrogram. The spectrogram (shown in Figure 1) is one of the most common methods of illustrating the three most salient components of any speech act: frequency (corresponding to the percept of pitch), intensity (corresponding to the percept of loudness), and timing characteristics (corresponding to the percept of rhythmicity). Taken together, these three components are major elements of prosody—paralanguage describing the *sound* of language, as opposed to

the lexical components themselves. In Figure 1, time is represented by the values on the x-axis. On the y-axis, frequency increases from bottom to top. The relative darkness and lightness of the spectrogram at any pair of x (time) and y (frequency) values represents the intensity of the speech at that time and frequency. Darker shades indicate regions of greater intensity, and lighter shades indicate regions of lesser intensity. These graphic portrayals of the dynamic nature of speech sounds are relatively easy to compute and are available in a variety of commercially available speech and signal processing packages (e.g., MATLAB, CSRE, and the line of Kay Elemetrics CSL hardware and software).¹ To optimize the spectrogram for Schematic Pitch

Coding, we generated narrow-band spectrograms (typically 10–30 Hz) and displayed only the first 1000 Hz regions of energy for two-second samples of speech. Narrow-band spectrograms provide better frequency resolution than wide-band spectrograms, at the expense of some temporal resolution. At conventional sampling rates of 20 kHz or less, the loss of temporal fidelity is negligible for analyses of fundamental frequency. Displaying the range of 0 Hz to 1000 Hz ensures that we optimize the representation of the f_0 contour. With the physical size of the window and the display range constrained by the SPC code, we maintain a common reference frame for coding relative changes in pitch.

Coding

After generating the spectrogram, change points are manually delimited using a computer-marking device. Change points are used to identify onset points, offset points, and points of inflection in the vocalized speech sample. Each critical point is logged using a mouse and cursor. Coders are permitted to listen to the 2-second sample of speech to supplement their decision-making in marking critical points. For each point logged, a descriptive code, time of occurrence, and frequency value is recorded. This information is used to create a schematic representation of the pitch contour and to compute summary measures of pitch and time values.

As shown in Table 1, Schematic Pitch Coding identifies both intra-utterance and inter-utterance changes in spectral energy as well as negative and positive points of inflection. Coders are asked to identify as few points as necessary to capture the salient changes in pitch in the utterance. Complete utterances, as defined by Papousek, Papousek, and Haekel (1987), are those “embedded in one global continuous prosodic pattern and separated from adjoining utterances by pauses that appeared as natural breaks” (pp. 496–497). Thus, all utterances begin with a major onset

TABLE 1
Schematic Pitch Codes

Code	Explanation
N	Major Onset. Identifies the beginning of a complete utterance.
F	Major Offset. Identifies the ending of a complete utterance.
n	Minor Onset. Identifies the beginning of an intra-utterance speech act.
f	Minor Offset. Identifies the ending of an intra-utterance speech act.
p	Peak. Identifies local pitch maxima—point of inflection from positive slope to negative slope.
t	Trough. Identifies local pitch minima—point of inflection from negative slope to positive slope.

(N). Following the N-code, salient pitch maxima are marked with a peak code (p) and salient pitch minima are marked with a trough code (t). Any intra-utterance pauses and resumptions of vocalization are marked by minor offset and onset codes (f- and n-codes, respectively). Finally, all utterances terminate with a major offset code (F).

Training

For the results presented below, three undergraduates were trained in the use of Schematic Pitch Coding. Two were speech science majors and a third was majoring in psychology. They received one hour of didactic instruction and one hour of applied instruction in Schematic Pitch Coding. Following training, they achieved reliability with the first author after one to three hours of additional practice. In one other study and in a current study, we have trained diverse liberal arts majors with similar efficiency (Zlochower & Cohn, 1996).

Reliability and Validity

With any new speech analysis tool, it is critical to ensure that measurements made with the tool are reproducible within this cod-

ing system (reliable) and comparable to those made using conventional approaches to speech analysis (construct-valid). To assess the reliability of Schematic Pitch Coding, we randomly selected 16 2-minute mother-infant vocal interactions from a larger sample of 50 mother-infant dyadic interactions and had the coders described above independently code infant-directed speech samples. Two of these operators, as noted above, were undergraduate students in speech science, and one was an undergraduate psychology student. We examined both the qualitative (i.e., the sequences of descriptive codes) and the quantitative (i.e., the f_0 values selected by the coders) aspects of Schematic Pitch Coding.

To validate schematic coding, precise f_0 contours were extracted from 90 utterances randomly selected from an earlier study of mother-infant vocal interactions (Katz et al., 1996). The 90 schematically coded contours were compared with the conventionally extracted pitch contours of Katz et al. (1996).

Inter-rater Reliability

There are two areas of inter-rater reliability related to Schematic Pitch Coding. First, are coders identifying similar descriptive codes (e.g., onsets, offsets, and points of inflection) at the same points on the spectrogram? Second, are the descriptive codes reliably centered within the frequency bands? If f_0 contours are to be estimated from Schematic Pitch Coding descriptive codes, both the temporal location and the frequency estimates must be reliably recorded.

To determine whether coders identified similar features within the f_0 contours, we examined agreements and disagreements in the descriptive codes within a 100-ms window. Descriptive codes that differed in location by more than 100 ms were coded as a "disagreement" while codes differing in location by less than 100 ms were coded as an "agreement." Both X^2 analyses and κ -coefficients indicated strong concordance among

coder pairs (average $X^2 = 2882.1$; average $\kappa = .77$, all $p < .0001$), representing high reliability. No systematic errors were found within or across pairs of coders.

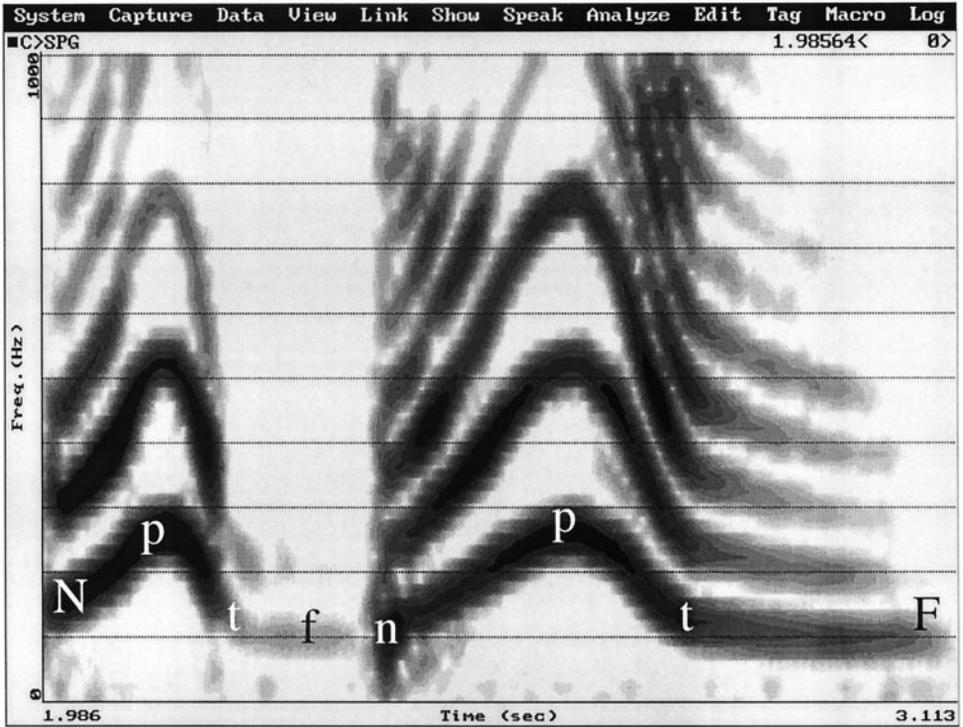
To determine whether coders were identifying similar points on the f_0 contour, we correlated f_0 estimates made by coder pairs within a 100-ms window. Coder pairs showed near-perfect correlations in identifying specific f_0 estimates from the spectrographic images (average $r = 0.93$, all p 's $< .0001$). When individual coder's means and standard deviations of estimated f_0 contours over the nearly 2000 points logged were compared, means and standard deviations of estimated f_0 contours across coders were found to differ by less than 5 Hz.

CONSTRUCT VALIDITY I

Schematic Pitch Contour Summary Measures

Summary measures of pitch mean traditionally require the complete extraction of f_0 values over the length of the entire utterance. Schematic Pitch Coding produces mean pitch estimates by weighted averaging of pitch values obtained from measurements of relatively few intra-utterance speech acts. Figure 2 illustrates how mean pitch estimates are calculated from Schematic Pitch Coding and through error-corrected methods of pitch extraction algorithms.

As shown in Figure 2, Schematic Pitch Coding estimates of mean pitch correspond closely with traditional spectrographic estimates of mean pitch used in traditional pitch extraction methods (Katz et al., 1996). We computed mean pitch from Schematic Pitch codes using weighted averages (Figure 2) and computed pitch contour duration from Schematic Pitch codes by computing difference scores from the times of a major onset (N-code) to a major offset (F-code). Mean pitch estimates between the two methods were nearly identical ($r = .95$; $p < .0001$), as were



<u>Time (s)</u>	<u>Frequency (Hz)</u>	<u>Code</u>	<u>Segment</u>	<u>Mean f_0 (Hz)</u>	<u>Weight (s)</u>	<u>Product</u>
1.99	144	N				
2.13	261	p	N→p	203	.14	28.42
2.24	107	t	p→t	184	.11	20.24
2.36	97	f	t→f	102	.12	12.24
2.41	102	n				
2.64	261	p	n→p	182	.23	41.86
2.80	121	t	p→t	191	.16	30.56
3.08	95	F	t→F	108	.28	30.24
				Sum	1.04	163.56
				Overall Mean f_0		157.27

Notes:

Weight = duration of vocalized segment in seconds

Product = Mean $f_0 \times$ Weight

Mean f_0 = overall weighted average for each vocalized segment

Overall Mean f_0 = sum of weighted f_0 measures / total of f_0 weights

FIGURE 2

Calculation of Mean f_0 from Schematic Pitch Codes

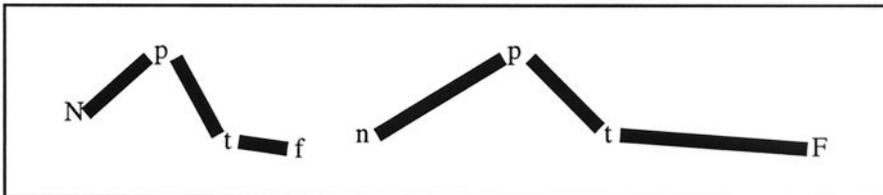
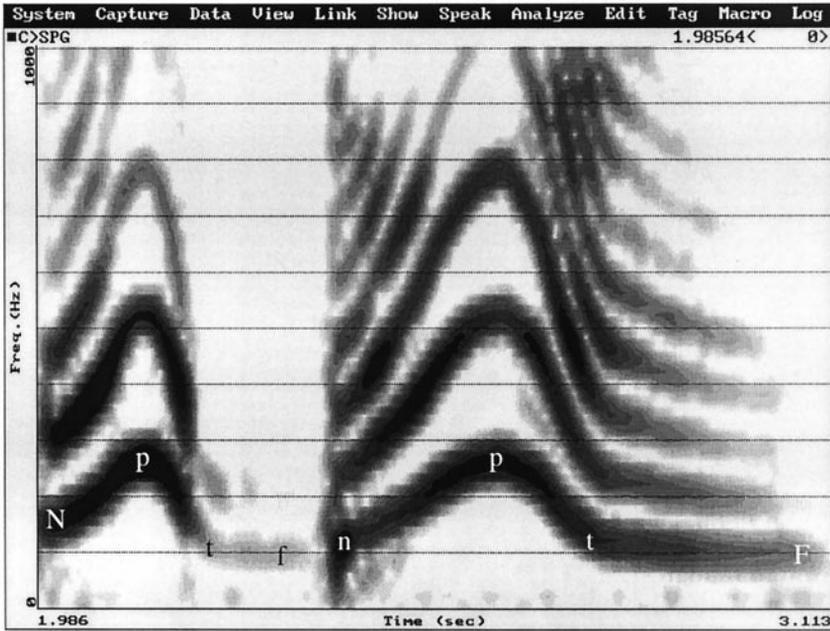


FIGURE 3

Extraction of Pitch Contour Shape from Schematic Pitch Codes

pitch contour duration estimates ($r = .98; p < .0001$).

CONSTRUCT VALIDITY II

Schematic Pitch Contour Shapes

By interpolating appropriate points from the time and frequency values logged with the

descriptive codes, schematic contours (Figure 3) can be generated from spectrograms. Schematic representations of pitch change provide the fine-structured shape of the pitch contour while sacrificing only high-frequency variation. Even though descriptive codes have been omitted from Figure 3 to illustrate the quantitative aspect of Schematic Pitch Coding, cursory inspection of Figure 3 shows a pitch

TABLE 2
Analytical Comparisons: Conventional Pitch Measures and Schematic Pitch Measures

	<i>Conventional Pitch Extraction Measures</i>	<i>Schematic Pitch Coding Measures</i>
Univariate Differences		
Duration	F = 18.65	F = 12.78
Pitch Variability	F = 13.33	F = 14.09
Multivariate Differences	Wilks' Λ = .451	Wilks' Λ = .512
Classification Rates	69% Correct	67% Correct

contour with two distinct peaks, two distinct troughs, and a definite intra-utterance pause of about 5 ms. The strength of Schematic Pitch Coding lies in its ability to faithfully represent the overall shape of the pitch contour (with the descriptive codes) without sacrificing quantitative measurement.

To ensure that pitch contours generated from Schematic Pitch Coding would be comparable with pitch contours extracted using conventional means, we re-examined the 90 utterances with Schematic Pitch Coding and conventional pitch extraction methods. Ninety conventional pitch contours were computed using precise pitch measurement and error-correction algorithms (Katz et al., 1996; Moore et al., 1994).

To obtain comparable time and frequency measures, we used linear interpolation between successive Schematic Pitch Codes on a 10-ms time base. Pitch estimates for each utterance from interpolated Schematic Pitch Coding measures and conventional extraction and schematic pitch coding were correlated. The overall correlation of pitch estimates made from the same speech samples using both methods averaged .93 ($p < .0001$).

Application of Quantitative Summary Measures From Schematic Contours

In an earlier study of the pragmatic use of infant-directed pitch contours (Katz et al., 1996), we found that a combination of shape and summary measures of infant-directed f_0 contours discriminated between three different

pragmatic categories (approval, attention, and comfort) in a large sample of infant-directed speech. These infant-directed f_0 contours were among those in the 90-utterance sample used to establish the validity of Schematic Pitch Coding generated contours.

To further evaluate the comparability of conventional and schematic pitch coding, we recoded a subset of utterances from the Katz et al. (1996) study and calculated analogous summary measures (estimated mean pitch, pitch variability, and pitch contour duration) from the schematic pitch contours. These summary measures were substituted for their respective pitch-contour measures (mean pitch, pitch variability, and pitch contour duration). Discriminative function analyses designed to classify infant-directed pitch contours into their respective pragmatic categories were then conducted using the analogous summary measures. As Table 2 shows, univariate measures of group differences based upon duration, mean pitch, and pitch variability (operationalized as the SD of the set of f_0 values) were nearly identical when compared with conventional pitch measures. Multivariate group differences were also strongly concordant. When these measures were used to classify tokens into their respective pragmatic group, the error rates of classification differed by only 2% overall.

Exploratory Findings and Replications

Two-minute speech segments from 50 mother-infant dyads during a free play interaction

were coded schematically to explore both the qualitative and quantitative aspects of Schematic Pitch Coding. The sequencing and frequencies of descriptor codes motivated qualitative analyses, whereas the time and frequency information logged as data points called for more quantitative analyses. Results in both of these areas were consistent with previous research using conventional measurements of pitch and vocal timing.

Qualitative Analyses of Token Descriptor Strings

Fernald and Simon (1984) and Stern, Spieker, and MacKain (1982) described the majority of infant-directed pitch contours as “complex” in shape. These complex shapes often took the form of “bell-shaped” or “sinusoidal” f_0 contours. In schematic coding, these contour shapes may be identified by examining token descriptor strings for presence or absence of peak or trough codes. Tokens with at least one peak or trough code may be either bell-shaped or sinusoidal because they have at least one geometric point of inflection (a point in the contour where the first derivative equals 0). Tokens with both peak and trough codes must be at least sinusoidal because they have at least two geometric points of inflection. We examined the proportion of tokens that had at least one peak and/or at least one trough code. Out of 3467 total utterances, 54.8% had at least one peak and 48.6% had at least one trough. Nearly 38% of the utterances had both a peak and a trough code. These proportions are comparable with those of Fernald and Simon (1984), who report proportions of all exaggerated contour shapes at 77%.

Analyses of Intra- and Inter-utterance Pause Duration

Jaffe and Feldstein (1970) and Feldstein, Jaffe, Beebe, and Crown (1993) noted that 300 ms of silence is the critical point that differentiates inter-utterance and intra-utterance

TABLE 3
Percentile Rankings of Intra- and Inter-
Utterance Pause Durations

Percentile	Intra-Utterance	Inter-Utterance
	Pause Duration (ms)	Pause Duration (ms)
5th	18	212
10th	27	300
90th	214	2489
95th	280	3651

pauses. Schematic Pitch Coding identifies intra-utterance pauses as the time spent between subsequent f and n codes. Inter-utterance pauses are identified by the time spent between subsequent F and N codes. To explore the duration of inter-utterance and intra-utterance pauses, we examined time differences between $f \rightarrow n$ and $F \rightarrow N$ transitions.

As is seen in Table 3, the vast majority of our inter-utterance pauses exceeded 300 ms, whereas nearly all of our intra-utterance pauses fell below 300 ms. These results provide empirical support for the choice of 300 ms to delimit the upper end of intra-utterance pauses and the lower end of inter-utterance pauses.

Coding Efficiency

In any study of infant-directed speech, extraction of accurate and reliable estimates of pitch is extremely difficult (Reed et al., 1992). Iterative techniques and substantial post-processing may assist the researcher in examining the structure of pitch contours (Katz, Moore, & Cohn 1992). For highly trained operators, we have found that optimal rates of pitch-contour extraction approach 2.7 minutes per 2 seconds of sampled speech using conventional pitch-extraction algorithms and techniques (e.g., Kay Elemetrics CSL4300 PITCH routines). This estimate does not include the additional time necessary for post-processing and error-correction of extracted contours. We have found that approximately 1 to 5 minutes

of post-processing and error-correction per contour is typical, depending upon the degree of computer automation. Consequently, the total amount of time required to process 2 seconds of sampled speech ranges from 3.7 to 8.7 minutes.

Overall, schematic coding is more efficient than conventional pitch-extraction techniques, with obtained rates of 1.5 to 2.0 minutes per 2 seconds of sampled speech. An additional advantage of schematic coding is that no extensive post-processing or error-correcting algorithms need be applied to the extracted data, which are typically required for conventionally extracted pitch contours (Katz et al., 1992). Without the need for post-processing, schematic coding has a clear advantage in generating large samples of infant-directed pitch contours. Additionally, at present, commercially available speech processing packages have upper limits on the size of the speech sample that can be analyzed with pitch extraction techniques. By circumventing the pitch extraction algorithms in favor of more stable spectrographic analyses, there is no theoretical upper limit as to the length of the speech segment being analyzed in schematic coding. We have coded segments of digital speech as short as 2 seconds and as long as 4 minutes. Much of the processing time required to implement Schematic Pitch Coding is related to the idiosyncrasies of the Kay Elemetrics computer interface, which is not optimized for on-the-fly data logging. The development of custom-programmed spectrographic interfaces may further reduce the time required to acquire Schematic Pitch Coding f_0 contours.

Another area in which Schematic Pitch Coding demonstrates increased efficiency over existing pitch-extraction techniques is in training demands. With traditional speech-processing measures, it is essential that staff have either a background in speech science or training in fundamentals of speech science and the computational issues surrounding pitch extraction techniques. Although formal measures have not yet been compiled, anecdotal evi-

dence supports the assertion that individuals can be trained in Schematic Pitch Coding techniques in two 90-minute training sessions that cover the basic theory of spectrography and provide hands-on experience with the computerized speech lab used for spectrographic analyses.

FUTURE DIRECTIONS

The superior efficiency of schematic coding makes it a promising approach for investigation of topics such as description of prosodic contours. Schematic coding provides a syntax for the complete specification of contours as well as quantitative measurement. Conventional methods are limited to a small set of geometric shapes (i.e., rise, fall, bell, wave, and flat) that arbitrarily restrict the study of infant-directed pitch contours across a number of domains. Enhanced description of infant-directed pitch contours with complementary quantitative measures will assist researchers studying mother-child vocal synchrony. In language acquisition studies, pitch marking of salient points in the infant-directed f_0 contour will be more efficient and more reliable using Schematic Pitch Coding.

The study of mother-infant-interactions across many modalities (e.g., facial, gestural, and vocal expression) is another area of research that would benefit from a more efficient and precise method of describing infant-directed speech. The analysis of facial expression in mother-infant dyads has benefited from advances in technology used to record and classify facial expressions (Cohn, Zlochower, Lien, & Kanade, 1999; Lien, Kanade, Cohn, & Li, 1998); however, difficulty in quantifying vocal expression has hampered the integration of face and voice in research of infant-caregiver interactions.

The vocal expression of emotion is another area of speech analysis research that would benefit from an accurate and reliable method of f_0 contour description and measurement. Shape and summary features communicate a

major portion of the emotional meaning of an utterance (Frick, 1985; Scherer, 1986). Sadness tends to be expressed with downward sloping f_0 contours, whereas anger and happiness tend to be expressed with highly variable and frequently rising f_0 contours (Katz, 1998). Still, the manner in which f_0 contours are analyzed in this literature demands a more rigorous method of measurement to capture both the consistency of contour shapes in specific emotional contexts and also the considerable variability within and across these emotional contexts. We expect that the flexibility of Schematic Pitch Coding in capturing the essential or salient modulation in the f_0 of emotional utterances will enhance empirical study in this area. Indeed, the adoption of Schematic Pitch Coding may enhance progress in a broad range of vocal behavior studies.

Acknowledgments: This research was supported by NSF Grant #BNS-8919711 to the University of Pittsburgh. The authors would like to thank three anonymous reviewers for their editorial comments and Adena Zlochower and Joan West for their assistance in data coding.

NOTE

1. Actual macro and keyboard commands for the Kay Elemetrics CSL4300 are available by contacting the first author (see appendix for pseudocode).

REFERENCES

- Cohn, J. F., Zlochower, A., Lien, J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology*, *36*, 35–43.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *66*, 1584–1595.
- Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*, 477–488.
- Feldstein, S., Jaffe, J., Beebe, B., & Crown, C. (1993). Coordinated interpersonal timing in adult-infant vocal interactions: A cross-site replication. *Infant Behavior and Development*, *16*, 455–470.
- Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*, 103–114.
- 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. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195.
- Fernald, A. (1989). Intonation and communicative intent in mother's speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*, 279–293.
- Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*, 104–113.
- Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*, 412–429.
- Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue*. NY: Academic.
- Katz, G. S. (1998). Emotional speech: A quantitative study of the psychoacoustics and psychophysiology of emotional expression (Doctoral Dissertation: University of Pittsburgh, 1998). *Dissertation Abstracts International*, *59*(06B), 3062.
- Katz, G. S., Cohn, J. F., & Moore, C. A. (1996). A combination of vocal f_0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, *67*, 205–217.
- Katz, G. S., Moore, C. M., & Cohn, J. F. (1992, November). *Semi-automated processing of very large natural f_0 samples*. Poster session presented at the conference of the American Speech and Hearing Association, San Antonio, TX.
- Kuhl, P. J., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lac-

- erda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684–686.
- Lien, J. J., Kanade, T. K., Cohn, J. F., & Li, C. C. (1998). A multi-method approach for discriminating between similar facial expressions, including expression intensity estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 98, 853–859.
- Moon, C., Cooper, R., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495–500.
- Moore, C. A., Cohn, J. F., & Katz, G. S. (1994). Quantitative description and differentiation of fundamental frequency contours. *Computer Speech and Language*, 8, 385–404.
- 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 tone and stress languages. *Infant Behavior and Development*, 14, 415–440.
- Reed, C., Buder, E. H., & Kent, R. D. (1992). Speech analysis systems: An evaluation. *Journal of Speech and Hearing Research*, 35, 314–332.
- Scherer, K. R. (1986). Vocal affect expression: A review and model for further research. *Psychological Bulletin*, 98, 143–165.
- Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, 10, 1–15.
- Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, 18, 727–735.
- Zlochower, A., & Cohn, J. F. (1996). Vocal timing in face-to-face interactions of clinically depressed and nondepressed mothers and their 4-month-old infants. *Infant Behavior and Development*, 19, 373–376.

01 December 1999; Accepted 09 September 1999 ■

Appendix: Computer Pseudo-code for Schematic Pitch Coding System

Tasks Executed Upon Speech System Start-up

- 1) Initialize Analyses Windows
 - (A) Window A full screen width by 20% screen height for oscillogram.
 - (B) Window B full screen width by 80% screen height for spectrogram.
 - (C) Link cursor displays between Window A & B.
- 2) Initialize Analysis Settings
 - (A) Spectrogram Display Range 0 - 1000Hz, linear scaling.
 - (B) Spectrogram Contrast set to middle range.
 - (C) Spectrogram Intensity set to middle range.
 - (D) Spectrogram bandwidth set to minimum setting (10Hz or less)
 - (E) Spectrogram pre-emphasis set to 0db.
 - (F) Initialize Data Logging Settings
 - (1) Record Time
 - (2) Record Frequency
 - (3) Record User-Provided Text String
- 3) Initialize Function Key Settings
 - (A) Shift-F1 - Clear Windows, Load Data File
 - (B) Shift-F2 - Open Log File
 - (C) F1 - Add Point to Log Book
 - (D) F2 - Speak Displayed Data
 - (E) F4 - Show next 1sec of data (move 50% overlapping window by 1sec)
 - (F) F5 - Increase Spectrogram Contrast & Redraw
 - (G) Shift-F5 - Decrease Spectrogram Contrast & Redraw
 - (H) F6 - Increase Spectrogram Brightness & Redraw
 - (I) Shift-F6 - Decrease Spectrogram Brightness & Redraw
 - (J) F12 - Close Log Book
- 4) Prompt user to load speech file.

Tasks Executed During Speech System Execution

- 1) Monitor mouse movements.
- 2) Monitor function key presses.
- 3) Execute commands associated with function key presses.
- 4) Write log book entries (ASCII text) as requested by F1 key.

Tasks Executed Following Speech System Execution

- 1) Open data log file.
- 2) Read header.
- 3) Open SPC output file.
- 4) Read Data Points
 - (A) Upon onset, start pitch weighting routine.
 - (B) Continue pitch weighting until offset recorded.
 - (C) Write string of feature codes, requested summary measures
- 5) Close data log file.