Acoustic Patterns of Infant Vocalizations Expressing Emotions and Communicative Functions

C. Papaeliou
University of Crete (Rethymno)
School of Philosophy
Department of Philosophical and Social Studies
Crete, Greece

G. Minadakis
D. Cavouras
Technological Educational Institute of Athens
Department of Medical Instrumentation Technology
Athens, Greece

The present study aimed at identifying the acoustic pattern of vocalizations, produced by 7- to 11-month-old infants, that were interpreted by their mothers as expressing emotions or communicative functions. Participants were 6 healthy, first-born English infants, 3 boys and 3 girls, and their mothers. The acoustic analysis of the vocalizations was performed using a pattern recognition (PR) software system. A PR system not only calculates signal features, it also automatically detects patterns in the arrangement of such features. The following results were obtained: (a) the PR system distinguished vocalizations interpreted as emotions from vocalizations interpreted as communicative functions with an overall accuracy of 87.34%; (b) the classification accuracy of the PR system for vocalizations that convey emotions was 85.4% and for vocalizations that convey communicative functions was 89.5%; and (c) compared to vocalizations that express emotions, vocalizations that express communicative functions were shorter, displayed lower fundamental frequency values, and had greater overall intensity. These findings suggest that in the second half of the first year, infants possess a vocal repertoire that contributes to regulating cooperative interaction with their mothers, which is considered one of the major prerequisites for language acquisition.

KEY WORDS: *infant vocalizations, pattern recognition, communication, emotions*

The ability of infants, before the emergence of language, to communicate different kinds of messages through nonlinguistic aspects of voice can be investigated by examining whether vocalizations that have been classified in different message categories exhibit distinct acoustic features. Relevant studies have focused mainly on vocal expressions of the first 6 months of life and demonstrated that vocalizations classified by mothers as cry, discomfort, calm, pleasure, or surprise exhibit distinct acoustic patterns (Hsu, Fogel, & Cooper, 2000; Papousek, 1992; Petrovich-Bartell, Cowan, & Morse, 1982; Scherer, 1984; Shimura & Imaizumi, 1995; Stark, Rose, & McLagen, 1975; Trainor, Austin, & Desjardins, 2000).

On the other hand, there are relatively few studies that have investigated the capacity for vocally conveying distinct messages in infants older than 6 months. These studies concentrated on the infants' use of vocalizations to denote different emotions or different communicative functions as identified by the researchers (Delack & Fowlow, 1978; D'Odorico & Franco, 1991; Scherer, 1984). However, vocal differentiation between emotions and communicative functions might also reflect the infants' communicative competence in the second half of the first year. As the ability for intentional communication develops, older infants can express vocally not only emotions, but also communicative functions (Bates, Camaioni, & Volterra, 1975; Camaioni, 1993; Flax, Lahey, Harris, & Boothroyd, 1991; Harding & Golinkoff, 1979; Hubley & Trevarthen, 1979; Marcos, 1987). Thus, infants may become more effective regulators of interpersonal communication by expressing either emotions, which identify the quality of communication, or communicative functions, which identify the direction and the purpose of communication (Trevarthen, 1990). Consequently, infants' ability to vocally distinguish between emotions and communicative functions may serve as an index of their capacity for communication. Nevertheless, there are very few published studies examining this hypothesis (Shimura, Imaizumi, & Shutu, 1992).

The method of acoustic analysis used in previous studies on infants' use of vocalization to convey information mainly extracted acoustic features from the spectrogram. The spectrogram shows the general frequency and amplitude movement of a vocalization through time, but it does not provide precise values of these specific features, mainly because it does not produce high-resolution analysis, especially in low frequencies (Malloch, 2000). Moreover, earlier studies examined the role of single acoustic features in discriminating the distinct classes of messages, and not the discriminating power of feature combinations, which ultimately form the acoustic pattern of the vocalization.

The present study aimed at examining whether 7- to 11-month-old infants are capable of producing acoustically distinct vocalizations that convey either emotions or communicative functions. A second issue focused on mother interpretation of the vocalizations produced by their infants. A pattern recognition (PR) system was designed for the acoustic analysis of the vocalizations. A PR system not only calculates signal features, it also detects the arrangement of such features into patterns (Gonzalez & Woods, 1992). To our knowledge, the PR method has not been used for the acoustic analysis of infant vocalizations.

Method

Participants

The participants were 6 healthy first-born infants, 3 boys and 3 girls, and their mothers, 27–38 years of age. All participants were from middle-class English families. First-born infants were selected because it was desirable to include mothers who were likely to be naive about any changes that might occur in their infant's behavior during the period of the study.

Data Collection

The material of this study is part of the material collected for more extensive research on infant vocal behavior in different situations in the second half of the first year. In that study, infants were recorded every 2 weeks from age 30 weeks to 50 weeks under two conditions: (a) mother-infant in free play with and without objects and (b) infant alone in the presence of various objects. Each condition lasted approximately 7 minutes and was introduced twice. A total of 840 minutes of recorded mother-infant interaction was available in the present study.

Recording sessions took place in the infants' homes. Home environment is considered more appropriate for obtaining representative samples of the infants' vocal repertoire than the unfamiliar laboratory environment (Lewdag, Oller, & Lynch, 1994). Mothers were asked to play spontaneously with their infant, as they would normally do. Each session was video recorded using a Panasonic HiFi camera, mounted on a tripod. In order to provide signals of good enough quality for the acoustic analysis, an additional audio recording was made simultaneously using a SONY Digital Audio Recorder with an external one-channel microphone. The microphone was also mounted on a tripod and placed next to the camera. During the time of observed mother-infant interaction, infants produced a total of 516 vocalizations (mean rate of vocal production: 3.1 vocalizations/5 minutes), which constituted the corpus judged by the mothers.

Vocalizations were attributed meaning by interviewing the mothers about what they felt their own infant was expressing in each instance of the videotaped interaction. In this way, interviews simulated the natural conditions of communication, where meaning attribution relies on features of the vocalization as well as on the accompanying nonvocal behavior. The interviews took place in the laboratory between successive recording sessions at a time that was convenient for the mothers. Each mother was asked to stop the video when she felt that her baby had expressed or said something and indicate as soon as possible what she felt her baby had expressed or said. Each mother was asked to express, in her own words, what she felt her infant had expressed or said. The mothers' answers were brief, precise, and described similar messages. Some typical examples are presented below: (1) "She is delighted," (2) "He is expressing irritation," (3) "She is pleased," (4) "She is frustrated," (5) "He is expressing enjoyment," (6) "She is not happy," (7) "He is trying to get my attention," (8) "She is asking me to give her the music box," (9) "She is showing me the toy," and (10) "He is responding to what I am saying." Based on their verbal content, the mothers' answers were assigned to one of the two categories *emotion* or *communicative function*. Terms such as *delighted*, *irritated*, *pleased*, *frustrated*, *joyful*, and *happy* were classified as either "positive" or "negative" emotion on the basis of the semantic differential analysis carried out by Plutchik (1980). Answers such as the ones presented in examples 7–10 were categorized as "communicative function" following the classification of other studies on infants' vocal expression of communicative intent (Flax et al., 1991; Marcos, 1987). Intrarater reliability was determined by asking the mothers to judge all of vocalizations on each tape twice: (a) at the laboratory interview that followed the filming session and (b) at the following interview session. The percentage of agreement between the mothers' judgments varied from 88.3–94.2%. Interrater reliability was not attempted because meaning attribution that occurs in spontaneous communication may also rely on factors that are not available to any person who does not participate in the dyad (e.g., any information that was acquired from the previous communicative experience of the dyad and is assumed to be shared by the infant and the mother).

Acoustic Analysis

Only vocalizations that were produced by aggressive voicing and formed fully resonant nuclei were considered for analysis. Cries, laughter, vegetative sounds, and word approximations, as well as vocalizations that overlapped with the mother's voice or external noise were excluded from the sample. From the original corpus of 516 vocalizations, 394 vocalizations met the above criteria. Because one of our purposes was to develop a new software PR system for the analysis and classification of infant vocalizations, it was felt more appropriate for the design and the pilot application of this system to use a small number of vocal signals. Thus, 79 vocalizations out of 394 (20%), were randomly selected so as to form a sample representative of all participants' vocal production. Of the 79 vocalizations analyzed, 41 vocalizations were categorized from the mothers' interviews as "emotions" and 38 were categorized as communicative functions. Fifty-seven percent of the vocalizations included in the category "emotion" conveyed positive emotion, and 43% conveyed negative emotion.

Acoustic analysis was performed on a Pentium III computer employing a customized software system. Initially, a spectrogram was obtained as follows: Each vocalization was sampled at 44.1 kHz. Using a 1024-point fast Fourier transform (FFT), an analysis window of about 25 ms that overlapped every 5 ms was chosen. The resulting amplitude spectrum of each FFT was smoothed by means of a 5-point averaging operation, and the first 512 points of the resulting amplitude spectrum were employed in forming the vocalization spectrogram (Forrest, Weismer, Milenkovic, & Dougall, 1988). The number of analysis intervals varied depending on the duration of the vocalization. Due to the nature of voice production, the signal at the edges is "noisy" and exhibits extraneous values that are not representative of the tonal movement of the vocalization. These areas were identified from the spectrogram and were cut off. An FFT of the new signal was then computed. The fundamental frequency was calculated using a threshold-based detection algorithm; accordingly, the algorithm examines local amplitude differences in each amplitude spectrum and decides on the first maximum (i.e., F0) on the basis of a predetermined threshold value that may be altered interactively. Additionally, there is provision for interactive intervention on the spectrogram for fine adjustment. A PR system for discriminating infant vocalizations that convey emotions and communicative functions on the basis of their acoustic pattern was developed. The design of the PR system was performed in two stages: (a) feature generation and (b) classifier design. In the first stage, 14 features were computed. Five features were from first order statistics and were calculated from each acoustic signal: mean value, variance, standard deviation, skewness, and kurtosis. Nine other features were selected as the most commonly used features in studies of acoustic analysis of infant vocalizations (Scherer, 1984). These nine features were calculated from the spectrogram: duration of the vocalization; beginning, final, peak, lowest, and mean fundamental frequency (F0) values; range of fundamental frequency; standard deviation of fundamental frequency; and ratio of standard deviation of fundamental frequency to duration of the vocalization. In the second stage of the PR-system design, a commonly used class-discriminating algorithm, the Least Squares Minimum Distance (LSMD) classifier, was employed to automatically characterize the vocalizations on the basis of the 14 generated features.

The reliability of the classifier was assessed by applying the leave-one-out method. The classifier was designed by all but one token, which was left out to be employed as test sample for the classifier. After classification, the left-out token was re-inserted into the data pool and the classifier was redesigned, this time leaving out the next token for classification. This procedure was repeated until all tokens had been tried (Gonzalez & Woods, 1992; Press, Flannery, Teukolsky, & Vetterling, 1990). The leave-one-out method was applied for all possible feature combinations (i.e., 2, 3, 4, etc. feature combinations) in order to determine the optimum feature combination that achieved the highest classification accuracy with the minimum number of features.

Results

According to the results presented in Table 1, the percentage of agreement between the mothers' interpretations and the results obtained by the PR system was 85.4% for vocalizations expressing emotions and 89.5% for vocalizations expressing communicative functions. The overall discriminating accuracy of the system was 87.34%.

The optimum feature combination that distinguished between the two message categories was the following: skewness, kurtosis, and duration of signal, peak and final fundamental frequency values, standard deviation of fundamental frequency of the vocalization, and ratio of standard deviation of fundamental frequency to duration of the vocalization. Table 2 shows the mean values of the best features employed by the PR system both for vocalizations that convey emotions and for vocalizations that convey communicative functions. In particular, compared to vocalizations expressing emotions, vocalizations expressing communicative functions exhibited lower kurtosis, shorter duration, and lower peak and final fundamental frequency values, but higher standard deviation of fundamental frequency relatively to duration. Moreover, vocal signals that convey communicative functions are negatively skewed, whereas vocal signals that convey emotions are positively skewed.

Figures 1 and 2 demonstrate vocalizations and their respective spectrograms interpreted by mothers as conveying emotion and communicative function, respectively. In both figures, the fundamental frequencies are depicted in the bottom line.

Table 1. Two-way truth table showing the results of the PR-system classification.

Message categories	Emotions	Communicative functions	Sums	Percent accuracy
Emotions	35	6	41	85.4
Communicative functions	4	34	38	89.5
Overall accuracy				87.34

Table 2. Mean values of the best features employed by the PR system for emotions and communicative functions.

	Message categories	
Acoustic features	Emotions	Communicative functions
Skewness	0.07	−0.04
Kurtosis	5.16	4.01
Duration (ms)	0.89	0.35
Fundamental frequency (peak) (Hz)	739.78	612.39
Fundamental frequency (final) (Hz)	531.02	366.54
Fundamental frequency (*SD*) (Hz)	63.58	65.3
Fundamental frequency (*SD*/duration) (Hz/ms)	94.88	264.1

Discussion

The present study investigated whether prelinguistic infant vocalizations that were interpreted by mothers as emotions or communicative functions exhibit distinct acoustic patterns. A pattern recognition system was designed for the classification of vocalizations on the basis of their acoustic features. Results showed that the PR system classified vocalizations interpreted as emotions and vocalizations interpreted as communicative functions with an overall accuracy of 87.34% (Table 1). This finding supports the hypothesis that at the threshold to language human infants are capable of distinctly communicating emotions and communicative functions through nonlinguistic aspects of voice. Similarly, Shimura and colleagues (1992) showed that prelinguistic infant vocalizations that were perceptually judged as "speaking," "affect request," and "negative emotion" differ acoustically. However, their result was obtained employing a single acoustic feature, namely, the fundamental frequency type (fall, rise, bell). In our study, there was disagreement between the mothers' meaning identification and the classification obtained from the PR system for about 13% of the vocalizations. It might be the case that for the interpretation of these few vocalizations the mothers relied not on the vocal pattern, but rather on accompanying nonvocal behaviors, such as facial expressions or gestures. That is what also occurs in spontaneous communication.

The optimum combination of features that demonstrated the highest discriminating accuracy were skewness, kurtosis, and duration of signal, peak and final fundamental frequency values, standard deviation of fundamental frequency, and ratio of standard deviation of fundamental frequency to duration of the vocalization. Duration of the vocalization and peak and final values of fundamental frequency are commonly used features in studies on infant vocal repertoire (Hsu et al., 2000; Papousek, 1992; Petrovich-Bartell et al., 1982; Scherer, 1984; Shimura & Imaizumi, 1995; Stark et al., 1975). Moreover, duration and peak fundamental frequency value have been shown to be reliable features in discriminating vocalizations produced by 2-month-old infants that have been characterized as *fuss* or *cry*, *pleasant* or *discomfort*, and *singing* or *speaking* (Papousek, 1992; Petrovich-Bartell et al., 1982; Shimura & Imaizumi, 1995). In contrast to duration, standard deviation of fundamental frequency has not been extensively used and, moreover, it has not been proven to be a powerful discriminating feature (Petrovich-Bartell et

Figure 1. Spectrogram of an infant vocalization interpreted as emotion, and F0 determination (bottom line).

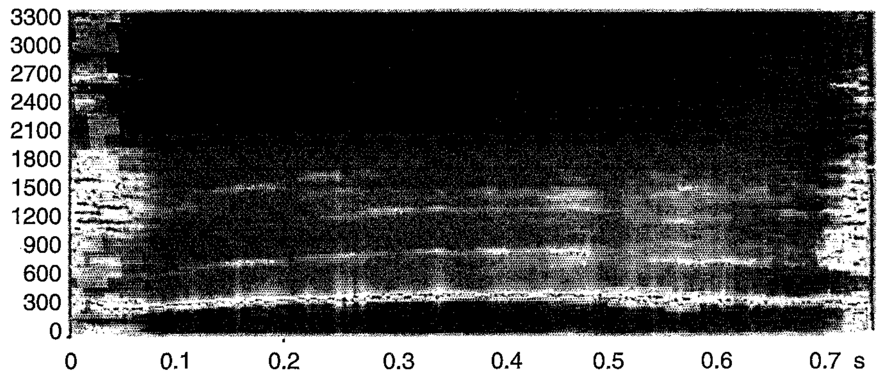
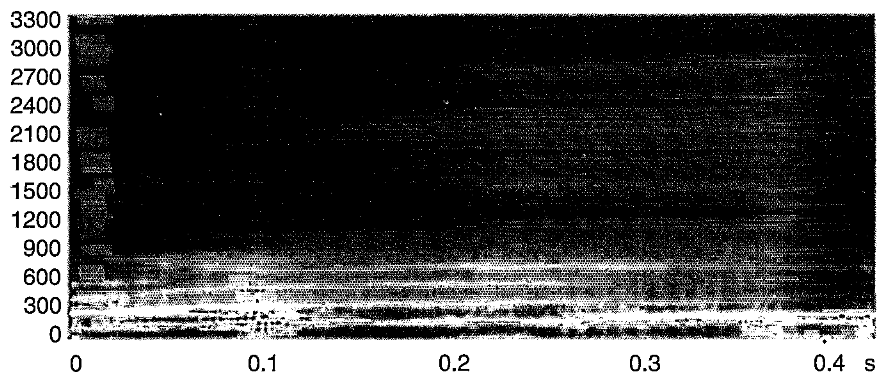


Figure 2. Spectrogram of an infant vocalization interpreted as communicative function, and F0 determination (bottom line).



al., 1982). However, standard deviation of fundamental frequency is supposed to reflect maturation in fundamental frequency control (Shimura, Imaizumi, Saito, & Ichijama, 1990). Additionally, the relative value of standard deviation of fundamental frequency depends on the duration of the vocalization. For this reason, the ratio of standard deviation of fundamental frequency to duration of the vocalization was used in the present study, and it was found to be a powerful discriminating feature. Specifically, vocalizations interpreted as communicative functions display shorter duration and lower fundamental frequency values, but more fundamental frequency fluctuations, than do vocalizations interpreted as emotions (Table 2, Figures 1 and 2). Also, it was demonstrated that vocal signals that convey communicative functions display greater overall intensity compared to vocal signals that convey emotions (Table 2).

These findings suggest that in the second half of the first year, infants' vocalizations that are interpreted by their mothers as emotions or communicative functions exhibit distinct acoustic patterns. This finding reflects a developing capacity for intentional use of voice in more complex forms of communication, which in turn is considered one of the prerequisites for language development (Bruner, 1975; Dore, 1983; Halliday, 1975; Locke, 1993, 1999). Infants are capable of vocally expressing discrete emotions from very early in life (Papousek, 1992; Petrovich-Bartell et al., 1982; Scherer, 1984; Shimura & Imaizumi, 1995; Stark et al., 1975). Mechanisms for the expression of emotions are ontogenetically primary and are regulated from motives for intersubjective communication (Trevarthen, 1982). This ability helps human infants to form and facilitate intimate relationships with their caregivers, which are necessary for their survival (Locke, 1999; Van Ijzendoorn, Dijkstra, & Bus, 1995). However, before 6 months, infants are not capable of systematically expressing communicative functions through any expressive modality, including vocalizations. The expression of communicative functions requires the highly evolved social ability to infer dispositions and intentions, which has not developed by 6 months of age (Bates et al., 1975; Hubley & Trevarthen, 1979; Locke, 1999). In the second 6 months of life, a change is observed in the infants' motivational organization for communication (Trevarthen, 1990, 1993), as well as an increment in their cognitive ability (Bates et al., 1975; Camaioni, 1993), which make them capable of conveying communicative functions. In particular, infants show growing awareness of how to coordinate their self-centered orientation to events with a partner's other-centered orientation. Progressively they gain the ability to direct the other's attention to a topic of their own interest as well as to follow the other's interest in a topic, and demonstrate an understanding of the other's feelings, interests, and intentions. (Hubley & Trevarthen, 1979). In the second half of the first year, infants are capable of using interchangeably vocalizations that convey emotions in order to identify the quality of communication and vocalizations that convey communicative functions in order to specify the direction and the purpose of their activities. This ability may serve as an index of their developing capacity for more elaborated forms of interpersonal communication.

The use of a PR system for detecting distinct acoustic patterns in vocalizations that are attributed different messages may provide information about the infants' meaningful vocal repertoire. Thus, the tool developed in the present study may be used to identify changes in this repertoire as referential ability develops and first words emerge. Moreover, the PR system may be used to study the vocal repertoire of atypically developing infants (e.g., William or Down syndrome infants), who suffer from communicative and emotional disorders (Locke, 1993; Thompson, Cicchetti, Lamb, & Malkin, 1985).

References

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21*, 205–226.

Bruner, J. (1975). From communication to language: A psychological perspective. *Cognition, 3*, 255–287.

Camaioni, L. (1993). The development of intentional communication: A re-analysis. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development*. London: Routledge.

Delack, J. B., & Fowlow, P. J. (1978). The ontogenesis of differential vocalization: Development of prosodic contrastivity during the first year of life. In N. Waterson & C. Snow (Eds.), *The development of communication*. New York: Wiley.

D'Odorico, L., & Franco, F. (1991). Selective production of vocalization types in different communicative contexts. *Journal of Child Language, 18*, 475–499.

Dore, J. (1983). Feeling, form and intention in the baby's transition to language. In R. Golinkoff (Ed.), *The transition from prelinguistic communication*. Hillsdale, NJ: Erlbaum.

Flax, J., Lahey, M., Harris, K., & Boothroyd, A. (1991). Relations between prosodic variables and communicative functions. *Journal of Child Language, 18*, 3–19.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84*, 115–123.

Gonzalez, R. C., & Woods, R. E. (1992). *Medical image processing*. New York: Wesley.

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. London: Edward Arnold.

Harding, C. G., & Golinkoff, R. M. (1979). The origins of intentional vocalizations in prelinguistic infants. *Child Development, 50*, 33–40.

Hsu, K., Fogel, A., & Cooper, R. P. (2000). Infant vocal development during the first 6 months: Speech quality and melodic complexity. *Infant & Child Development, 9*, 1–16.

Hubley, P., & Trevarthen C. (1979). Sharing a task in infancy. In I. C. Uzgiris (Ed.), *Social interaction during infancy*. San Francisco: Jossey-Bass.

Lewdag, V. L., Oller, K. D., & Lynch, M. P. (1994). Infants' vocalization patterns across home and laboratory environments. *First Language, 14*, 49–65.

Locke, J. L. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

Locke, J. L. (1999). Towards a biological science of language development. In M. Barrett (Ed.), *The development of language*. Sussex, U.K.: Psychology Press.

Malloch, S. (2000). Mothers and infants and communicative musicality. *Musicae Scientiae, Special Issue*, 29–53.

Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language, 14*, 255–268.

Papousek, M. (1992). Early ontogeny of vocal communication in parent-infant interactions. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches*. Cambridge, U.K.: Cambridge University Press.

Petrovich-Bartell, N., Cowan, N., & Morse, P. A. (1982). Mothers' perceptions of infant distress vocalizations. *Journal of Speech and Hearing Research, 25*, 371–376.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience. Vol. 1: Theories of emotion*. New York: Academic Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1990). *Numerical recipes in C: The art of scientific computing*. Cambridge, U.K.: Cambridge University Press.

Scherer, C. (1984). The assessment of vocal expression in infants and children. In C. E. Izard (Ed.), *Measuring emotions in infants and children*. Cambridge, MA: Cambridge University Press.

Shimura, Y., & Imaizumi, S. (1995). Emotional information in young infants' vocalizations. *Proceedings of the International Conference of the Phonetics Society, 3*, 412–415.

Shimura, Y., Imaizumi, S., Saito, K., & Ichijama, T. (1990). Infants' vocalization observed in verbal communication: Acoustic analysis. *Proceedings of the International Conference on Spoken Language Processing, 2*, 1329–1332.

Shimura, Y., Imaizumi, S., & Shutu, T. (1992). Perceptual evaluation of emotional aspects of infants' vocalizations. *Annual Bulletin of the University of Tokyo, 26*(3), 87–96.

Stark, R. E., Rose, S. N., & McLagen, M. (1975). Features of infant sounds: The first eight weeks of life. *Journal of Child Language*, *2*, 205–222.

Thompson, R., Cicchetti, D., Lamb, M., & Malkin, C. (1985). The emotional responses of Down syndrome and normal infants in the strange situation: The organization of affective behavior in infants. *Developmental Psychology*, *21*, 828–841.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*, 188–195.

Trevarthen, C. (1982). The primary motives for cooperative understanding. In G. Butterworth & P. Light (Eds.), *Social cognition: Studies of the development of understanding*. Brighton, U.K.: Harvester Press.

Trevarthen, C. (1990). Signs before speech. In T. A. Sebeok & J. U. Sebeok (Eds.), *The semiotic web*. Berlin, FRG: Mouton de Gruyter.

Trevarthen, C. (1993). Infant semiosis. In W. Noth (Ed.), *Origins of semiosis*. Berlin, FRG: Mouton de Gruyter.

Van Ijzendoorn, M. H., Dijkstra, J., & Bus, A. G. (1995). Attachment, intelligence, and language: A meta-analysis. *Social Development*, *4*, 115–128.

Received March 26, 2001

Accepted January 11, 2002

DOI: 10.1044/1092-4388(2002/024)

Contact author: C. Papaeliou, PhD, 54 Faradaton Street, Ampelokipi 115-27, Athens, Greece. E-mail: papailiou@phl.uoc.gr