THE EFFECTS OF TEMPO AND PITCH ON THE JUDGEMENT OF INFANT CRIES

Taeko Tsukamoto
Kobe University School of Medicine

Abstract

Two experiments were performed to examine the effects of tempo and pitch on the identification of infant cries by manipulating the duration and F_0 of the cries using a signal processing method. The subjects employed for these experiments were inexperienced in infant care, so they were given training on how to categorize cries at the beginning of each experiment. In experiment 1, 68 subjects, all of them university students, were asked to identify stimuli subjected to duration manipulation by making a forced choice among three categories: hunger, anger, and call. In experiment 2, 89 subjects, all of them university students, were asked to identify stimuli subjected to pitch manipulation or manipulations involving both duration and pitch. An analysis was made on the data of the 40 subjects for experiment 1 and 55 subjects for experiment 2 respectively, who successfully learned how to categorize the cries in their prior training. The results from the two experiments indicate that tempo is the predominant perceptual cue and pitch secondary, for discrimination between hunger and anger but not for call.

Key Words: infant cries, tempo, pitch, perceptual judgement

Infant cries in the preverbal stage are the primary mode to communicate the need for secure nurturance from the environment. They are, however, graded signals and do not convey discrete message with a single meaning (Murray, 1979; Zeskind, Sale, Maio, Huntington, & Weiseman, 1985). In this sense, the communicative value of infant cries has been studied through the feature analyses of cries and/or their perception by adult listeners. A number of analytical studies of infant cries have revealed that they are characterized by differences in the fundamental frequency (F_0) and duration: the mean and dynamic range of F_0 (Murry, Hoit-Dalgaard, & Gracco, 1983) and the temporal

This research was conducted when the author was with ATR Auditory & Visual Perception Research Laboratories in collaboration with Dr. Yoh'ichi Tohkura. It is noted that Experiment 1 in this study was partially presented at 119th meeting of the Acoustical Society of America, Pennsylvania, 1990, and Experiment 2 at the meeting of the 4th International Workshop on Infant Cry Research, Munich, 1992.

patterns of F_0 (D'Odorico, 1984 ; Rosenhouse, 1977 ; Wasz-Höckert, Lind, Vuorenkoski, Partanen, & Valanne, 1968), and the temporal structure of phonation and non-phonation (Kobayashi, Oda, & Murooka, 1986 ; Tsukamoto & Katagiri, 1988 ; Wolff, 1967, 1969). Many studies on the relationship between physical acoustic features and perception of the infant cries have demonstrated that these acoustic features are also related to the perception of cries with different impressions by the listener (Bates, Freeland, & Lounsbury, 1979 ; Bisping, Steingrueber, Oltmann, & Wenk, 1990 ; Gustafson & Green, 1989 ; Lounsbury & Bates, 1982 ; Wiesenfeld, Malatesta, & DeLoach, 1981 ; Zeskind, 1985 ; Zeskind & Lester, 1978 ; Zeskind & Marshall, 1988). These findings suggest that the infant cries provide certain information to listeners trying to judge why the infant is crying. On the other hand, several studies on judgment of infant cries have focused on whether adult listeners are able to identify the cry types correctly and, if so, to what extent (Müller, Hollien & Murry, 1974 ; Scherman, 1927 ; Wasz-Höckert et al., 1968). However, there is no agreement among their experimental results due to the methodological differences, i.e., the cry types and subject groups used in their experiments, between experiments. Another important requirement is to clarify the process of cry judgment : which acoustic features help to differentiate the infant cries and they interpreted by the adult listeners? In this view, our current question is : which acoustic features are relevant perceptual cues to the judgment of infant cries.

The cry of younger infants is fundamentally an involuntary vocal expression evoked by the inner stress related to the balance of homeostasis (i.e., biological and self-organizing functions to help adapt to the environment) and/or the state transitions of the sleep/wake cycle. It has been considered that, physiologically, cry production involves the larynx, vocal tract and respiratory system and, neurologically, probably the central nervous system (CNS) and anatomic nervous system (ANS) (Boukydis, 1985 ; Golub & Corwin, 1985 ; Lester, 1978, 1984 ; Lester & Zeskind, 1982 ; Michelson & Wasz-Höckert, 1980 ; Wolff, 1967, 1969). Physiological and anatomical maturation of the infant brings developmental differentiation and an acoustical change in the cry in the month after birth : not only do physiological and biological cries (e.g., a hunger cry) appear but also psychological cries such as a call cry, which demands frequent interaction and/or bodily contact with his/her caregiver (Lester, 1985 ; Tsukamoto, 1985 ; Wolff, 1969). The pattern of the cry remarkably shifts from one that is simple and rhythmic in the neonate to one that is complex and irregular (Futatsugi, 1979 ; Wolff, 1969). The cry is also a vocal expression of emotional arousal as the activities of the CNS and ANS are closely related to emotion (Fox & Davidson, 1986 ; Lester, 1985). From these physiological and anatomical indications, it is presumed that different acoustic patterns of cries represent different emotional states of the infant and also provide some information to the listener trying to infer why the infant is crying.

Comprehensive suggestions on how to proceed with the study of cry judgment are from studies of the relationship between emotion and vocal expression in the field of speech communication where is generally recognized that so-called prosody, e.g., pitch, intonation (or melody), rhythm, and tempo, plays an important role in conveying a communicator's emotion or attitude (Scherer, 1979 ; Scherer, Ladd, & Silverman, 1984 ; Williams, & Stevens, 1972). According to reviews by Scherer (1979, 1982, 1986), F_0, the variability or range of F_0, loudness, and tempo are important parameters linked to specific emotions: for instance, a high pitch level and wide pitch range, loud voice, and fast tempo, appear to characterize anger, whereas the opposite ends of these vocal dimensions, i.e., a low pitch level and narrow pitch range, downward pitch contour, soft voice, and slow tempo, appear to characterize grief/sadness. Simply put, these acoustic features are used to judge the communicator's emotional states and are thought to be also used as perceptual cues for interpretation of infant cries.

This inference is supported by a findings from our preliminary experiment (Tsukamoto, 1991), which examined the relationship between auditory impression and identification of infant cries by adults. In this experiment, we used four cry categories, hunger, anger, sleepiness, and call, and employed two groups of subjects: mothers who cared for their infants and female university students. The results showed that verbal responses to the auditory impression to cry stimuli were surprisingly similar between the mothers and university students. The most frequent verbal responses were related to the emotional states or events. On the other hand, for the results of identification by free choice, the mothers could identify hunger, sleepiness, and call cries significantly well, while the university students could identify anger and call cries as well as, or even better than, the mothers. These results suggest that the representations of the auditory impressions of the cries are attributed to an emotional dimension inherent in their infant care experiences. In other words, prosody-related acoustic features linked to emotions were possibly used as perceptual cues in representing the auditory impressions and also in judgment of the cries. Our present interest is how these acoustic features, as perceptual cues, affect the judgment of infant cries.

In discussing this issue, a method considered effective and used in the speech perception of emotional expressions (Carlson, Granstrom, & Nord, 1992; Kitahara & Tohkura, 1992; Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985), is employed to manipulate each acoustic feature characterizing the infant cries and to investigate how the manipulations affect the resulting judgment. In this study, we focused on two acoustic features, tempo and pitch, and designed two experiments to examine how tempo and pitch affect the judgment in identifying cry categories based on the infants' states. In these experiments, natural cries recorded in home situations were employed and their duration and F_0 were manipulated by using a digital signal processing method, and then cry stimuli with various tempo and pitch parameters were presented to the adult subjects to examine how these two factors affect the identification of cries.

Cry Samples

Cries of normal and healthy infants were videotaped (SONY SL-F1; SONY CCD-V8AF2) during mother-infant interactions. The observations of the infants, who ranged from 3 weeks to 12 months of age, were made in one-hour biweekly visits. The observers were several university students majoring in psychology, and one of the authors. Records were made by each observer on the behavior of the mothers and infants, and records were also made of "the infant's life rhythm" by the mothers who noted each instance of a diaper change, sucking or feeding, and sleep, as well as the awareness and physical condition of their infants the day before an observation was to take place.

In order to classify the cries, the observers evaluated them according to the infants' situations by watching repeated playbacks of the videotapes and reviewing all observations on the infant reported by their mothers. Cry category classification was done by consensus, i.e., more than 85% of the observers had to agree for a classification to be accepted. Thus, the cries were eventually classified into several categories based on what caused them to occur. We found that the cries were often produced not only by simple situations, but also for complex physical or psychological reasons and that their role in communication was functionally changed by the infant's behavioral development around six months of age (Tsukamoto, 1985; Yamada, 1982).

For this study, as samples we chose nine cries produced by simple situations before six months of age and divided them into three categories, i.e., hunger, anger, and call,. (Here, we define a call cry as a psychological cry demanding interaction and/or bodily contact with a caregiver.) These cry samples were produced by four infants ranging in age from 9 to 17 weeks. The overall duration of each sample was around 30 s.

Acoustic Properties of Samples

Prior to the perceptual experiment, it was necessary to study the acoustic features of cry samples. The samples were digitized by computer (MASSCOMP 5600) at a sampling frequency of 20 kHz (LPF=10 kHz, -96 dB/oct) and were analyzed to extract the fundamental frequency (F_0) and duration. F_0 was measured every 2.5 ms by applying an autocorrelation pitch extractor to the cry waveforms. The duration of the segmental units and pitch duration were measured at a time resolution of 2.5 ms by observing digital spectrograms and the cry waveforms A segmental unit is defined as a breath group of the cries, consisting of all the vocalizations occurring during a single respiration, and a pitch duration is defined as the duration of the vocalized parts in which clear harmonics of F_0 can be continuously observed in the spectrograms. To clarify the acoustic properties of cry samples, statistical distribution of the pitch durations and F0s for each sample are illustrated in histograms.

Figure 1 shows the pitch duration distribution of samples for each of the three categories. The pitch duration is quantized every 250 ms and its occurrences are normalized by the total number of occurrences. In this figure, the ordinate shows the pitch duration occurrences, while the abscissa shows pitch duration length. Differences in the pitch duration distributions can be found among the three categories. For the hunger samples, the histograms are somewhat skewed to a short duration. For the anger samples, the histograms for the three samples are similar and also considerably more skewed to a short duration. On the other hand, for the call samples, the pitch duration has a wider distribution and the histograms for the three samples show fewer similarities. Figure 2 shows the F_0 distribution for each of the three categories. In this figure, F_0 is quantized every 50Hz, and its occurrences are normalized by the total number of occurrences. The ordinate shows the F_0 occurrences, while the abscissa shows the F_0 value. For both the hunger and anger samples, the variances of F_0 are somewhat broad and the peak frequencies of the histograms are, to some extent, different among the three samples. On the other hand, for the call samples, the variances of F_0 and the peak frequencies are rather similar for all three samples.

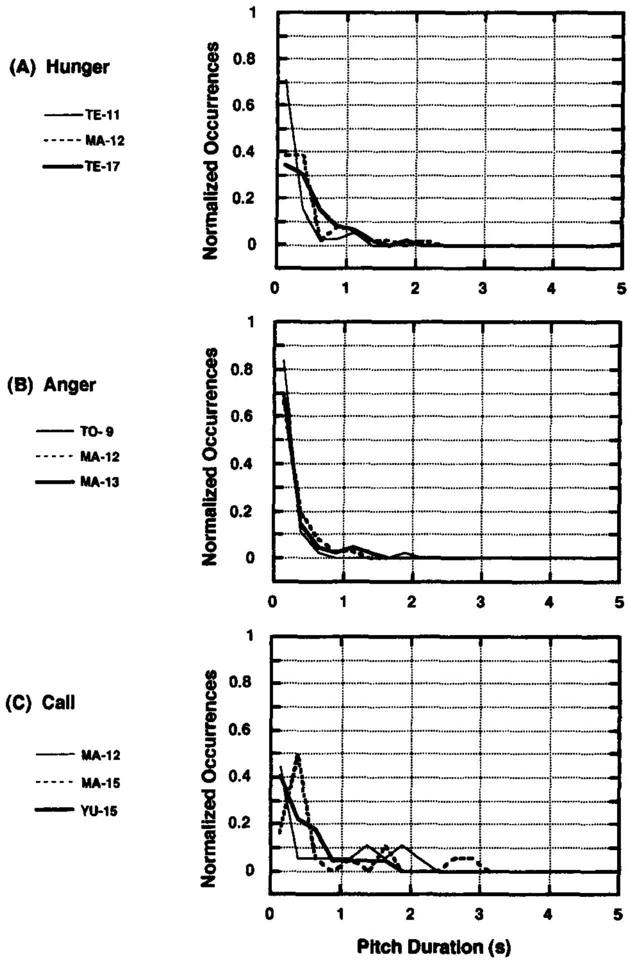

Figure 1 Pitch duration distributions for three cry categories. (A) is for three samples of hunger. (B) is for three samples of anger, and (C) is for three samples of call. Each plot has three distributions obtained from different cry samples. MA, TE, TO, and YU represent individual infants.

From these analyses of the pitch duration and F_0, it may not be easy to derive a quantitative correlation between these acoustic features and cry categories. However, similarities and differences showed within and between the three categories suggest that each cry category might have its own acoustic properties. Among these acoustic properties, the duration and F_0 are our main concerns and their role as perceptual cues for the identification of cry categories in discussed by perceptual experiments in this study.

Training

This study consisted of two experiments to examine the effects of tempo and pitch on the identification of infant cries using three categories : hunger, anger, and call. In the two experiments, we employed subjects who had had no experience in infant care. We designed a training session of cry category identification for the start of each experiment

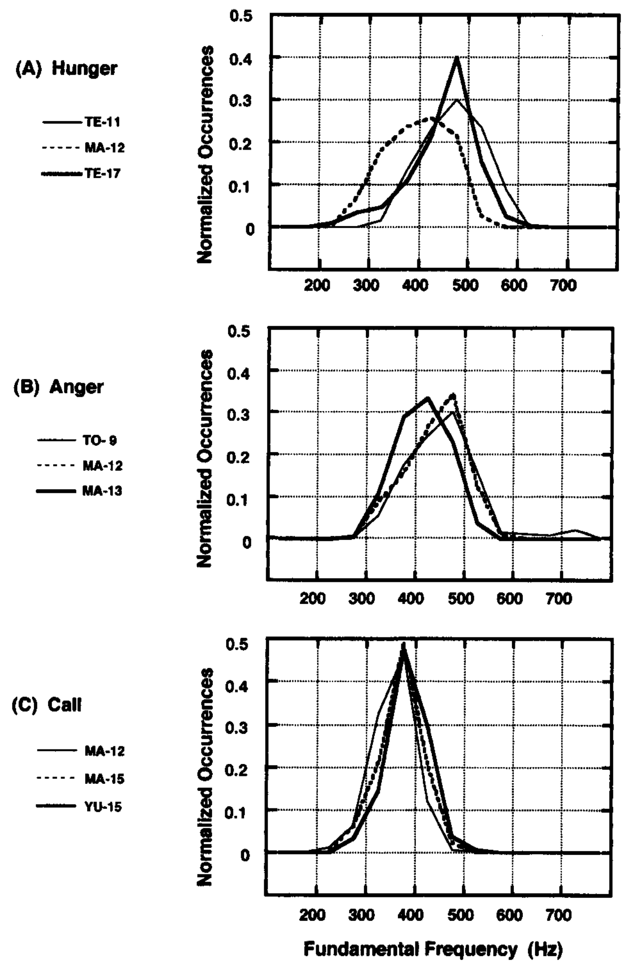

Figure 2 Fundamental frequency distributions for three cry categories. (A) is for three samples of hunger, (B) is for three samples of anger, and (C) is for three samples of call. The same samples as those used in Figure 1 have been employed.

to help the subjects learn how to categorize the cries. The subjects were expected to learn the distinctive features peculiar to each of the three categories.

To more efficiently to conduct these experiments, we examined the cry length needed for accurate identification of cry reasons in our previous experiment (Tsukamoto & Tohkura, 1990). The results showed that the perceptual units, which were correctly identified by more than 85% of the adult subjects, were from six to eight seconds' minimum duration through the three categories. As these perceptual units were assumed to contain typical acoustic features that corresponded to the cry categories, we excised portions of the perceptual units from each of the nine samples and employed them as stimuli in the training sessions and the two experiments.

Method

Stimuli

In the training session, two stimuli were used for each category. That is YU-15 weeks and MA-17 weeks for call, TE-11 weeks and TE-17 weeks for hunger, and TO-9 weeks and MA-12 weeks for anger. Cry durations ranged from 10.6 to 12.6 s.

Training procedure

The training for cry category learning consisted of four stages: (1) listening trials, (2) identification trials, (3) listening trials, and (4) the identification test. In the listening trials, the stimuli were presented twice in succession with their category labels given to the subjects prior to listening. In the identification trials, the same stimuli as those used in the listening trials were presented twice in random order (12 stimuli in total, inter-stimulus interval (ISI) = 3 s). The subjects were instructed to identify the stimuli by making a forced choice among the three categories: hunger, anger, and call cries. After the three trials, an identification test was carried out for each of the subjects. In this test, the same procedure as that in the identification trial in the training was used, except that the stimuli were presented in a different random order. Subjects who correctly identified all 12 stimuli were regarded as having who successfully learned the categorization of cries in the training.

All stimuli were recorded once on a digital audio tape (SONY DTC-1000ES) and presented to the subjects through headphones (STAX SRM-1/MK-2) in a sound booth. This training lasted about 15 min followed by about a 10-minute break and then the experiment.

Experiment 1

The purpose of this experiment was to examine the effect of tempo on the identification of cry categories by manipulating the duration of the cries.

Method

Subjects

The subjects were 36 female and 32 male university students ranging in age from 19 to 24 years. None had had experience in infant care.

Stimuli

Original stimuli. Three samples, one for each category, produced by one of the infants, but different from those used in the training, were chosen from the nine samples: call, MA-15 weeks; hunger, MA-12 weeks; and anger, MA-13 weeks. From each of these three samples, two of perceptual units were generated as two original stimuli between 6.6 and 8.1 s in duration. The acoustic features of the original stimuli are shown in Table 1.

Stimulus series. The duration of the original stimuli was changed using a signal reconstruction method based on a short-time Fourier transform (Abe, Tamura, & Kuwabara, 1989; Griffin & Lim, 1984). It is important to note that this method is capable of modifying the acoustic features independently: i.e., compressing and expanding the duration (time scale of the signals) while keeping the segmental frequency structures

Table 1 *Original Stimuli (MA) Used in Experiment 1*

Category Stimulus	Age (weeks)	Segmental Duration (s)			Fundamental Frequency (Hz)		
		Mean	Max.	Min.	Mean	Max.	Min.
Hunger							
H1	12	1.02	1.39	0.31	398.1	502.5	247.9
H2		1.23	1.74	0.62	427.7	689.1	230.1
Anger							
A1	13	1.69	2.31	0.64	427.2	557.8	285.1
A2		1.75	3.02	0.51	440.8	539.3	301.9
Call							
C1	15	1.94	5.04	0.53	353.7	430.3	215.3
C2		3.38	4.75	1.50	372.0	466.9	169.5

unchanged, and also capable of modifying F_0 while leaving the formant and temporal structures of the cries unchanged. The range of expansion and compression was set at 0.4 to 2.5 to avoid unrealistic conditions and to obtain stimuli with no perceived distortion. Furthermore, taking account of experiment efficiency, a preliminary experiment was carried out to determine the effective ranges of expansion and compression for each category. Thus, the anger stimuli were changed by expansion only in four steps (1.0 (original), 1.5, 2.0, and 2.5), the call stimuli by compression only in four steps (1.0 (original), 0.67, 0.5, and 0.4), and the hunger stimuli were changed by both compression and expansion in five steps (0.5, 0.67, 1.0 (original), 1.5, and 2.0). (Note that, for instance, compression or expansion at a ratio of 2.0 makes cry stimuli 2 times as fast or slow as the original stimuli.)

Procedure

The stimulus series was generated from two original stimuli for each category, so 26 stimuli including the original stimuli were generated in all. In addition to these 26 stimuli, another 14 stimuli (which were also excised from the same samples but from different portions) were prepared for embedding in the stimulus series as dummy stimuli. The total number of stimuli for this experiment was 40. These 40 stimuli were randomized and another 10 dummy stimuli were added at the beginning. Fifty stimuli were recorded on a digital audio tape (SONY DTC-1000ES). The ISI was set at 3 s, and a beep was presented at the beginning of each stimulus. Furthermore, three beeps were presented every 10 stimuli as a block sign. Taking into account the order effect of stimulus presentation, two tapes with different random orders were prepared and presented to counterbalanced groups of subjects. The subjects were required to identify each stimulus by making a forced choice among the three categories: hunger, anger, and call. The experiment lasted about 15 min. and was conducted in a sound booth after training. All stimuli were presented to the subjects through headphones (STAX SRM-1/MK-2).

Results and Discussion

The results were analyzed based upon the response data for 40 (25 women and 15 men) of the subjects who successfully learned the categorization of cries in their training. There were 28 subjects (11 women and 17 men) who failed to learn cry categorization. As the purpose of this experiment was to examine the effect of tempo by duration manipulation, the sex of subjects was ignored in the analysis of the results. From these data, similar response properties were observed between the two stimulus series generated from the two original stimuli for each category. Accordingly, the responses of 40 subjects for each stimulus between the two stimulus series were averaged and the relationships between the correct response rates and the variations of duration were analyzed in each category. Here, correct responses to the stimuli were defined as responses to the same category as that of the original. To clarify the relationships between the stimuli and responses, averaged responses for each stimulus are illustrated in Figure 3. In this figure, (A) is for the hunger, (B) is for the anger, and (C) is for the call stimuli. For the hunger cry, the correct response rates decreased from 38% to 18.8% when the duration was compressed and increased to 55%, which was higher than those for the original stimuli, when the duration was expanded (χ^2 (4, n=40) =36.39, p<.001). For the anger cry, when the duration was expanded, the correct response rates decreased from 66.3% to 27. 5% (χ^2(3, n=40) =77.88, p<.001). For the call cry, the correct response rates also decreased from 93.8% to 70% when the duration was compressed, although the stimuli were still predominantly identified as call cries (χ^2(3, n=40) =8.83, p<.05). These results showed that the correct responses significantly decreased in all three categories when manipulating the duration. However, the effect of duration change was different among the three categories. Moreover, for the hunger and anger stimuli, category shifts were shown between these two categories: when the duration was compressed for the hunger stimuli, more than half of the responses were identified as anger cries and when the duration was expanded for the anger stimuli, more than half of responses were identified as hunger cries. It was shown that the categories shift between the hunger and anger cries when the duration of original stimuli is changed. Taking this finding into consideration, it is indicated that the discrimination between hunger and anger cries is related to the tempo: a cry with a faster tempo tended to be identified more as anger and a cry with a slower tempo tended to be identified more as hunger. In contrast, when the duration was

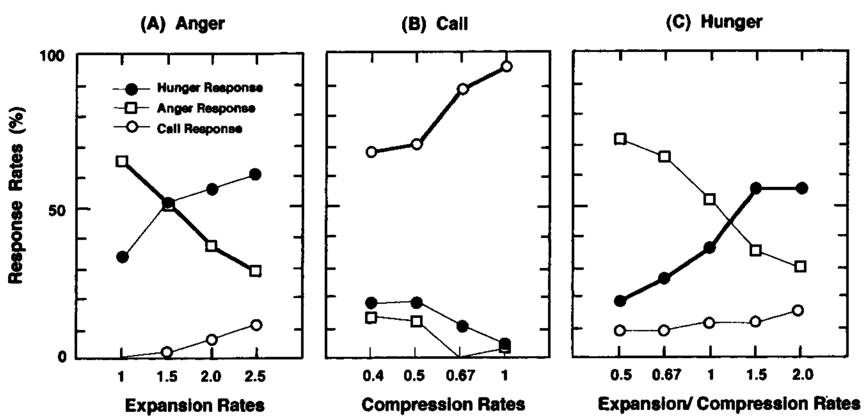

Figure 3 Relationships between response rates and the variations of duration for three cry categories: (A) is for anger, (B) is for call, and (C) is for hunger. Bold lines represent correct responses in three cases. The correct response rates significantly decreased in all three categories when changing the duration.

compressed to over 0.5, the correct response rates for the call stimuli were always higher than 70%, though to some extent they were identified as hunger (17.5%) or anger (11.5%). Accordingly, it is assumed that there might be other perceptual cues than the tempo that characterize the call cries.

Considering the general response properties for the three kinds of cries used in the experiment, the call stimuli were perceived as quite different from those of the other two categories while the hunger and anger stimuli were easily confused with each other. The results showed that the correct response rates to the original hunger stimuli in this experiment stayed as low as under 40%, though they were the same portion identified at rather high rates (over 85%) found in the previous experiment (Tsukamoto & Tohkura, 1990). One reason for this gap in the correct response rates seems to be related to the differences in the stimulus sets in the two experiments. In the previous experiment, the stimulus set consisted of a number of segment units with a variety of acoustic features. In this experiment, however, six stimuli (two for each category), which differed in various acoustic features, were used as the original stimuli. All other stimuli had segmental features that were similar to those of the original, and probably could be differentiated mainly through tempo. To avoid a perceptual bias, which may occur when using stimuli that are similar to each other, 14 dummy stimuli were embedded throughout the 26 stimuli. However, this might not be sufficient to entirely normalize the perceptual bias and reduce the subjects' excessive concentration on differences in tempo. As a result, the anger and hunger stimuli were differentiated mainly through tempo, regardless of the differences in some other acoustic features. Eventually, perceptual cues, which were used to identify the hunger cries in the previous experiment, might be less effective in this identification test. Another reason seems to be that the subjects tended to overly rely on tempo to differentiate anger and hunger cries. Accordingly, the subjects identified hunger stimuli with a faster tempo as anger cries even though they were the original stimuli. In fact, the two original hunger stimuli used in this experiment are composed ot repetitions or relatively short segments, which presumably cause perception of a faster tempo (see Table 1). Therefore, it is considered that the correct response rates for the expanded hunger stimuli were higher than for the original. Summarizing this discussion, cry tempo is a reliable cue for discriminating between hunger and anger cries. With regard to the call cry, it is assumed that there might be other perceptual cues that influence the judgement of cries.

Experiment 2

Taking account of the results in experiment 1, the purpose of this experiment was designed to examine the effect of pitch, under different temporal conditions, on the identification of cry categories by manipulating the duration and F_0 of the cries.

Method

Subjects

The subjects were 60 female and 29 male university students ranging in age from 18 to 26 years. None had had infant care experience.

Stimuli

Original stimuli. The original stimuli used in this experiment were similar to those in experiment 1 (see Table 1), though they differed in duration.

Stimulus series. From each original stimulus, a stimulus series was generated by combining the seven pitch levels and three tempi. For the tempo conditions, three kinds of duration transformations were employed: unchanged (original), compressed, and expanded. The transformation ratios chosen were different for the three categories, based on the results of the experiment 1. For compression, the transformation ratios were set to 0.5, 0.5, and 0.67 for call, hunger, and anger, respectively. For expansion, however, the transformation ratios were set to 1.5, 1.5, and 2.0 for call, hunger, and anger, respectively. As for pitch, seven levels were constructed by shifting F_0 with ratios of 1.3, 1.2, 1.1, 1.0 (original), 0.9, 0.8, and 0.7.

Procedure

There were 126 stimuli (3 tempo conditions x 7 pitch level variations x 2 stimulus series x 3 categories) in total. From the results of experiment 1, it was presumed that the subjects might tend to notice which acoustic parameters of the stimuli had been manipulated after repeatedly listening to stimuli with similar acoustic features. Therefore, to avoid this experimental distortion, 54 dummy stimuli were added to the 126 stimuli, 180 stimuli being prepared in total. These 180 stimuli were randomized and divided into two sets of 90 stimuli each. Another 10 dummy stimuli were added at the beginning of each set (100 stimuli in total). These two sets were recorded on a digital audio tape (SONY DTC-1000ES). The ISI was set at 3 seconds, and a beep was presented at the beginning of each stimulus. Furthermore, three beeps were after presented every 10 stimuli as a block sign. This procedure was repeated to prepare another tape with two sets of 100 stimuli in a different random order. These two tapes were presented to counterbalanced groups of subjects by taking account of the order of stimulus presentation. The two sets in each tape were presented to each subject group at different sessions with a short break in -between (about 5 min). The subjects were required to identify each stimulus by making a forced choice among the three categories: hunger, anger, and call. The experiment lasted about 50 min. and was conducted in a sound booth after training (see details under Training). All stimuli were presented to the subjects through headphones (STAX SRM-1/MK-2).

Results and Discussion

Effect of Fundamental Frequency

The results were analyzed based upon the response data given by 55 (38 women and 17 men) of the subjects who successfully learned cry category identification in the training session. The subjects who failed to learn numbered 34 (22 women and 12 men). To examine the effect of F_0, the correct response rates for the stimuli under three tempo conditions were analyzed. As in experiment 1, similar response properties were observed between the two stimulus series generated from the two original stimuli for each category. Therefore, all of the responses for each stimulus in the two stimulus series were averaged and graphs to clarify the relationship between the correct response rates and the shift of F_0.

Figure 4 illustrates the relationship between the stimuli and correct responses for each category. (A) is for the hunger stimuli. The correct response rates increased as F_0

increased for all three tempo conditions (from 43.7% to 60% for "unchanged", from 22.7% to 35.5% for "compressed", and from 53.6% to 70.9% for "expanded"). This tendency was most apparent and statistically significant in the case of the unchanged condition. Consequently, the F_0 dependency of the correct response rates was shown to be significant in chi-squared analysis, $\chi^2(6, n=55)=14.19$, $p<.05$. (B) shows the response rates for the anger stimuli. For the unchanged and expanded condition, the correct response rates increased significantly as F_0 increased (from 36.6% to 61.8% for

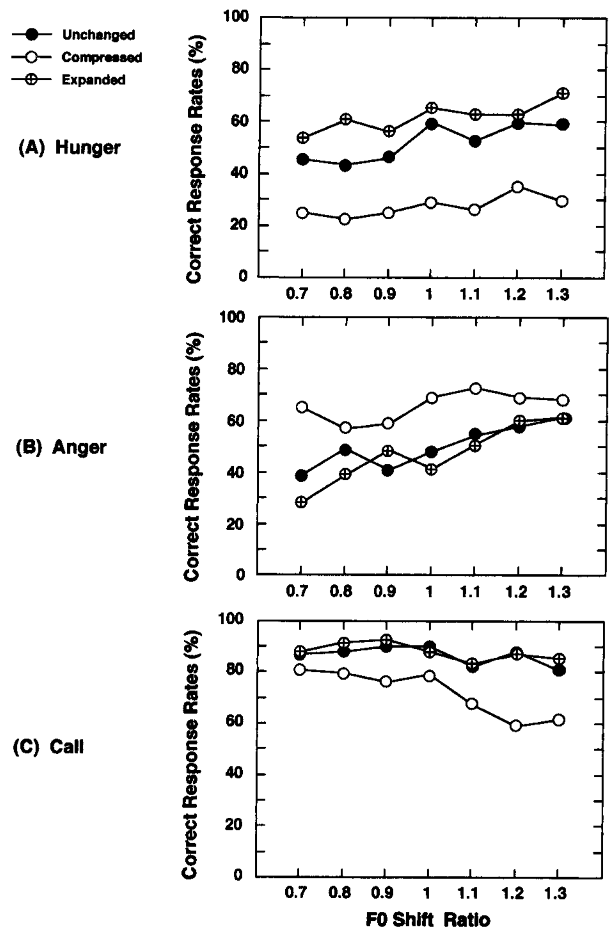

Figure 4 Relationships between correct response rates and the variations of F_0 in three tempo condition for three categories: (A) is for hunger stimuli. The correct response rates increased as F_0 increased for all three tempo conditions. (B) is for anger stimuli. The correct response rates for the unchanged and expanded condition increased as F_0 increased. For the compressed condition, the correct response rates were higher than for the other tempo conditions throughout the stimulus series, regardless of the F_0 shift. (C) is for call stimuli. The correct response rates for the unchanged and expanded condition were so high as to exceed 80% throughout the stimulus series, regardless of the F_0 shift. For the compressed condition, the correct response rates decreased at higher values of F_0.

" unchanged " and from 24.7% to 61.8% for " expanded "). For the compressed condition, however, the correct response rates were higher than for the other tempo conditions throughout the stimulus series and the effect of F_0 shift was not significant. In chi-squared analyses, the correct response rates differed significantly with respect to F_0 for the unchanged condition ($\chi^2(6, n=55) = 19.03, p<.005$), and for the expanded condition ($\chi^2(6, n=55)=37.57, p<.001$). (C) shows the correct response rates for the call stimuli. For the unchanged and expanded condition, the correct response rates show similar patterns. They were so high as to exceed 80% throughout the stimulus series, regardless of the F_0 shift. For the compressed condition, however, the correct response rates were lower than for the other tempo conditions on the whole and also significantly decreased from 80.9% to 59.1% at higher values of F_0. In a chi-squared analysis, the correct response rates of stimuli for the compressed condition were significantly different with respect to F_0, $\chi^2(6, n=55)=27.2, p<.001$.

These results showed that the F_0 shift affected, on the whole, the correct response rates in all three categories. However, the response characteristics in each category showed that the effects of varying F_0 differ among the three tempo conditions.

Relationship between pitch and tempo

Examining the relationship between the pitch and tempo, the response characteristics among the three tempo conditions were analyzed by multiple comparison tests using Ryan's procedure. Analysis results showed significant differences ($p<.05$) as follows: For hunger, the correct response rates were highest for the expanded condition followed by the unchanged and then the compressed condition. There were significant differences between the unchanged and compressed condition ($\chi^2(1, n=55)=27.53$), between the expanded and unchanged condition ($\chi^2(1, n=55)=6.6$), and between the expanded and compressed condition ($\chi^2(1, n=55)=53.94$). For anger, a different tendency was observed. The correct response rates were highest for the compressed condition followed by the unchanged and then the expanded condition. There were significant differences between the compressed and expanded condition ($\chi^2(1, n=55)=39.32$), and between the compressed and unchanged condition ($\chi^2(1, n=55)=17.93$). For the call, the correct response rates were highest for the expanded condition followed by the unchanged and then the compressed condition. There were significant differences between the expanded and compressed condition ($\chi^2(1, n=55)=18.29$), and between the unchanged and compressed condition ($\chi^2(1, n=55)=15.69$).

The error responses in each tempo condition were also analyzed to reveal the response tendency of the subjects. The results showed that the category shifts depending on the tempo conditions were observed for the compressed condition of the hunger stimuli and the unchanged and expanded condition of the anger stimuli. For example, in the hunger stimuli case illustrated in Figure 5, the F_0 shift produced a minimal effect when the stimuli had a fast tempo due to compression (A). However, as can be seen (B), the F_0 shift had a clear influence on identification when the stimuli had a slow tempo due to expansion.

It is noted that this same tendency was also shown in experiment 1. For the call, when F_0 increased under the compressed condition, the correct response rates decreased

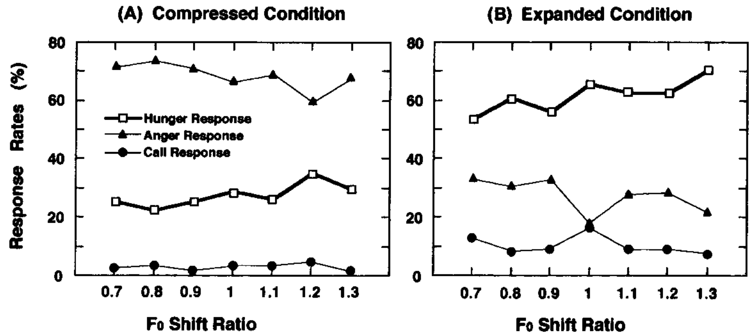

Figure 5 Response properties for hunger stimuli. Bold lines represent correct responses in two cases. The F_0 shift produced a minimal effect when the stimuli had a fast tempo due to compression (A). However, as can be seen (B), the F_0 shift had a clear influence on identification when the stimuli had a slow tempo due to expansion.

whereas the hunger responses increased (about 30%). However, there was no category shift, as can be seen in (C) in Figure 4.

These results suggest that, as far as the identification of hunger and anger cries is concerned, tempo is more influential than pitch, which might be used as a secondary cue when the acoustic features are atypical or indistinct. For the call cries, there were effects of tempo and pitch but these were not enough to cause a category shift. It is assumed that there might be other perceptual cues that are much more influential in recognizing call cries.

General Discussion

Little is known about the process of cry judgment, how the cries are perceive and then interpreted by the adult listeners. A number of studies on analysis and perception of infant cries, and even studies in the field of speech communication, suggested that the duration and the F_0 may be used as perceptual cues in judging infant cries. In this study, we focused on two acoustic features, the tempo and pitch, and examined how these acoustic features affect the category identification by using the three categories (hunger, anger, and call) of natural cries with a healthy infant when he was 12 to 15 weeks in age. In experiment 1, the effect of tempo was examined by manipulating the duration of the cries. In experiment 2, the effect of pitch by varying F_0 higher and lower in seven levels under three different tempo conditions (unchanged, compressed, and, expanded) were examined by manipulating the duration and F_0 of the cries. The results showed that differences of both the tempo and the pitch significantly affected the correct response rates in all three categories. However, the effects of pitch were different among three kinds of tempo conditions and responses of the subjects tended to depend on the tempo of stimuli. The effect of tempo by manipulating compression or expansion of the duration was quite consistent through the results of experiments 1 and 2, which indicates that tempo is more influential than pitch.

Several studies of cry perception (Bisping et al., 1990; Zeskind & Lester, 1978) found that the cries with a higher pitch were perceived as more unpleasant, urgent, or aversive and stressed that the F_0 may be the most important acoustic feature differentiated in the cries. However, these findings are related to the cries of high-risk or abnormal infants, which usually have a higher pitch (above 800 Hz, on average) than those of normal infants (ranging between about 300Hz to 500 Hz). Recently, experimental evidence related to our results mentioned above has been shown. Zeskind, Klein, and Marshall (1992) also examined the effect of duration of cries on adult listeners' perception using experimental manipulations of the cry sound. They digitally lengthened or shortened the durations of pauses and expiratory sounds in the cries and tested how these manipulations affected the perception of cries. They found that cries with pauses 50% shorter than the original were perceived to be more arousing, informative, and aversive. They also found that the monotonic effect of pause duration was enhanced in an interaction with expiration duration for perceptions of urgency. These findings indirectly support our findings that when the duration was compressed for the hunger stimuli, more than half of the responses were given as anger cries. That is to say, it can be concluded that an anger cry with fast tempo is related to the adult listener's perception of how arousing, aversive, and urgent it is.

However, we must pay attention to differences of cry stimuli used in each of our and their experiments: As the original stimuli, they used the hunger cries of a 2-day-old infant, while we used three types of cries (hunger, anger, and call) produced by an infant aged around three months. It is considered that our cry stimuli are composed of various acoustic features in general because the pattern of the cry as was described in the beginning of this paper. This comparative discussion between two kinds of experimental results, even though different stimuli with different manipulations are used in the two cases, leads us to believe that variations in the tempo clearly affected both the adult listeners' perception and their interpretations of the meaning of the cries.

Furthermore, it is important to note that the effects of the tempo and pitch were different among the three cry categories. Concerning the relationship between the acoustic features of the hunger and anger stimuli, as shown in Figures 1 and 2, they have rather similar distributions especially in F_0, which may cause these two categories to be easily confused. However, modification of cries by changing the duration and F_0 seemed to make it easier to discriminate between these two categories, assuming that by making such modification we could emphasize a distinctive perceptual feature. Here, the findings in speech perception as described by Scherer are of considerable interest (see Introduction in this study), that anger seems to be characterized by high pitch level and fast tempo, and also by Kitahara and Tohkura (1992) that temporal compression of a speech signal serves to provide neutral speech with the emotion of anger, and also that the intensity of the anger increases as F_0 becomes higher. For example, compression for the anger stimuli (i.e., with a faster tempo) might emphasize the perceptual degree of anger if the anger is characterized by a fast tempo, while expression for the hunger cries might emphasize the perceptual degree of non-anger. As far as the identification of hunger and anger cries is concerned, tempo is more influential than pitch, which might be used as a secondary cue related to the degree of intensity when the acoustic features were atypical or indistinct. Accordingly, it is possible that the subjects might identify the anger and hunger cry stimuli by using the same criterion by which they judge the emotional states of a communicator in speech perception. On the other hand, the correct response rates for the call stimuli always remained high regardless of changes to the duration and F_0, except for the stimuli whose tempo was doubled and whose pitch was 30% higher than that of the original, but nevertheless were not enough to cause any category shiat. This indicates that there might be might be much more influential perceptual cues than tempo and pitch. It can be inferred from the response characteristics of the subject that the call cries may be perceived as quite different from the other two types of cries as well as from their acoustic distributions as shown in Figures 1 and 2. That is to say the call cries have an irregular temporal pattern, a lower pitch, and a narrow pitch range that are controlled by the manner of phonation and respiration by the infants. This may be derived from the fact that the call cries are based on the infants' psychological needs (demanding interaction and/or bodily contact with a caregiver), which are generally called "amae-naki" in Japanese culture. If these cries are the same as the "faking cries" described by Wolff (1969), the call cries must have some perceptual acoustic features that identify that the infants are faking the cry.

In this study, we focused on two perceptual acoustic features, tempo and pitch, which can rather easily be changed by manipulating the duration and F_0 of the cries by using a signal reconstruction method. Although this method is capable of modifying some acoustic features independently, we should know that it would be impossible to completely manipulate all the features independently because several features are, by their nature, dependent. For example, the temporal pattern of the F_0 (i.e., so-called pitch contour) must change so as to steepen the F_0's rise or fall when the duration is compressed linearly; the opposite effects exist when the duration is expanded. Thus, the possibility has to be taken into account that the experimental results given by manipulating the duration include a concomitant effect of F_0 change. Future research is needed to examine the effects of other segmental acoustic features such as formant transitions, pitch contours, and spectral envelopes. In order to manipulate these acoustic features in a sophisticated way, however, we still need to develop a high-quality cry synthesizer based upon a cry production mechanism for manipulating many more acoustic features independently and freely.

Taking the above discussion into consideration, though it is too early to conclude that the effects of pitch and tempo obtained experimentally can be generalized for the judgment of cries, the results of this study provide meaningful suggestions and a new view to the study of cry judgment. It is safe to say that future research will deal with segmental acoustic features, such as formant transitions and pitch contours. In order to manipulate these segmental acoustic features in a sophisticated way, however, we still need to develop a high-quality cry synthesizer based upon a cry production mechanism. This issue should also be discussed in further research. Additionally, new mechanisms will be necessary for infant cry judgment, in order to study the development of communication patterns between parents and infants.

References

Abe, M., Tamura, S., & Kuwabara, H. (1989). A speech modification method by signal reconstruction using short-time Fourier transforms. *The Transactions of the Institute of Electronics, Informa-*

tion and Communication Engineers of Japan, J72-D-II, 8, 1180-1186. (In Japanese)

Bates, J. E., Freeland, C. A. B., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development*, 50, 794-802.

Bisping, R., Steingrueber, H. J., Oltmann, M., & Wenk, C. (1990). Adults' tolerance of cries: An experimental investigation of acoustic features. *Child Development, 61*, 1218-1229.

Boukydis, C. F. Z. (1985). Perception of infant crying as an interpersonal event. In B. M. Lester & C. F. Z. Boukydis (Eds.), *Infant crying : Theoretical and research perspective* (pp. 187-215). New York : Plenum Press.

Carlson, R., Granstrom, B., & Nord, L. (1992). Experiments with emotive speech-acted utterances and synthesized replicas. International Conference on Spoken Language Processing 92, October, 1-2, 671-674.

D'Odorico, L. (1984). Non-segmental features in prelinguistic communications: An analysis of some types of infant cry and non-cry vocalization. *Journal of Child Language, 11* (1), 17-27.

Fox, N. A., & Davidson, R. J. (1986). Psychophysiological measures of emotion: New directions in developmental research. In C. E. Izard & P. B. Read (Eds.), *Measuring emotions in infants and children*: vol. 2. (pp. 13-47). Cambridge University Press.

Futatsugi, T. (1979). Cry characteristics in newborn infants and developmental meanings of the infant cry. *Japanese Journal of Child Psychiatrica, 20* (3), 161-177 (In Japanese).

Golub, H. L., & Corwin, M. J. (1985). A physioacoustic model of the infant cry. In B. M. Lester & C. F. Z. Boukydis (Eds.), *Infant crying : Theoretical and research perspectives* (pp. 59-82). New York : Plenum Press.

Gustafson, G. E., & Green, J. A. (1989). On the importance of fundamental frequency and other acoustic features in cry perception and infant development. *Child Development, 60*, 772-780.

Griffin, D. W., & Lim, J. S. (1984). Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustic Speech & Signal Process, 32* (2), 236-242.

Kitahara, Y., & Tohkura, Y. (1992) Prosodic control to express emotions for man-machine speech interaction. *IEICE Transaction Fundamentals, E75-A* (2), 155-163.

Kobayashi, K., Oda, T., & Murooka, H. (1986). Characteristic pattern in time series of crying vocal early infant. *Japanese Society of Medical Electronics & Biological Electronics*, 24 (In Japanese).

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F_0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, 78 (2), 435-444.

Lester, B. M. (1978). The organization of crying in the neonate. *Journal of Pediatric Psychology, 3* (3), 122-130.

Lester, B. M. (1984). Infant crying and the development of communication. In N. A. Fox & R. J. Davidson (Eds.), *The psychobiology of affective development* (pp. 231-258). Hillsdale, New Jersey : London.

Lester, B. M. (1985). Introduction : there's more to crying than meets the ear. In B. M. Lester & C. F. Z. Boukydis (Eds.), *Infant crying : Theoretical and research perspectives* (pp. 1-27). New York : Plenum Press.

Lester, B. M., & Zeskind, P. S. (1982). A biobehavioral perspective on crying in early infancy. In H. E. Fitzgerald, B. M. Lester & M. W. Yogman (Eds.), *Theory and research in behavioral pediatrics* (pp. 133-180), Plenum Press: New York.

Lounsbury, M. L., & Bates, J. E. (1982). The cries of infants of differing levels of perceived temperamental difficultness: Acoustic properties and effects on listeners. *Child Development, 53*, 677-686.

Michelsson, K., & Wasz-Höckert, O. (1980). The value of cry analysis in neonatology and early infancy. In T. Murry & J. Murry (Eds.), *Infant communication : Cry and early speech* (pp. 152-182). Houston : College-Hill Press.

Muller, E., Hollien, H., & Murry, T. (1974). Perceptual responses to infant crying: Identification of cry types. *Journal of Child Language, 1,* 89-95.

Murray, A. D. (1979). Infant crying as an elicitor of parental behavior: An examination of two models. *Psychological Bulletin, 86* (1), 191-215.

Murry, T., Hoit-Dalgaard, J., & Gracco, V. L. (1983). Infant vocalization: A longitudinal study of acoustic and temporal parameters. *Folia Phoniatrica,* 35, 245-253.

Rosenhouse, J. (1977). A preliminary report: An analysis of some types of a baby's cries. *Journal of Phonetics, 5,* 299-312.

Scherer, K. R. (1979). Nonlinguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions in personality and psychopathology* (pp. 495-529). New York: Plenum Press.

Scherer, K. R. (1982). The assessment of vocal expression in infants and children. In C. E. Izard & L. M. Dougherty (Eds.), *Measuring emotions in infants and children* (pp. 127-163). Cambridge University Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99* (2), 143-165.

Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76* (5), 1346-1356.

Sherman, M. (1927). The differentiation of emotional responses in infants: 2. The ability of observers to judge the emotional characteristics of the crying of infant, and of the voice of an adult. *Journal of Comparative Psychology, 7,* 335-351.

Tsukamoto, T. (1985). Development and mechanism of the infant cry. *Annual Reports of Konan Women's University, 4,* 1-20 (In Japanese).

Tsukamoto, T. (1991). Acoustical impressions of infant cries and the effect of infant care experience on cry judgment. Proceedings of the 33 Japanese Association of Educational Psychology, (In Japanese).

Tsukamoto, T., & Katagiri, S. (1988). The quantitative acoustic feature analysis of the infant cry [English summary] . Proceedings of the 52nd Japanese Psychological Association, 52, 532 (In Japanese).

Tsukamoto, T., & Tohkura, Y. (1990). Perceptual units of the infant cry. *Early Child Development and Care, 65,* 167-178.

Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T., & Valanne, E. (1968). *The infant cry: A spectrographic and auditory analysis.* Clinics in Developmental Medicine 29 (pp. 1-42). London: Spastics International Medical Publications Heinemann.

Wiesenfeld, A. R., Malatesta, C. Z., & DeLoach, L. L. (1981). Differential parental response to familiar and unfamiliar infant distress signals. *Infant Behavior and Development, 4,* 281-295.

Williams, C. E., & Stevens, K. N. (1972). Acquaintance and accuracy of vocal communication of emotions. *Journal of Communication, 22,* 300-305.

Wolff, P. H. (1967). The role of biological rhythms in early psychological development. Bulletin of the *Menninger Clinic, 31* (4), 197-218.

Wolff, P. H. (1969). The natural history of crying and other infant vocalization. In Foss, B. M. (Ed.), *The determinants of infant behavior* Vol. 4. (pp. 81-109), London: Methuen.

Yamada, Y. (1982). Expressions of request and rejection and the emergence of self in infancy. The *Japanese Journal of Educational Psychology, 30* (2), 128-138.

Zeskind, P. S. (1985). A developmental perspective of infant crying. In B. M. Lester, and C. F. Z. Boukydis (Eds.), *Infant crying: Theoretical and research perspectives* (pp. 159-185), New York: Plenum Press.

Zeskind, P. S., & Lester, B. M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. *Child Development, 49,* 580-589.

Zeskind, P. S., & Marshall, T. R. (1988) The relation between variations in pitch and maternal perceptions of infant crying. *Child Development, 59,* 193-196.

Zeskind, P. S., Klein, L., & Marshall, T. R. (1992). Experimental modification of relative durations of pauses and expiratory sounds in infant cries alters adult's perceptions. *Developmental Psychology, 28* (6), 1153-1162.