Perceptual differences in infant cries revealed by modifications of acoustic features

Athanassios Protopapas[a] and Peter D. Eimas
Department of Cognitive & Linguistic Sciences, Brown University, Providence, Rhode Island 02912

(Received 22 July 1996; revised 12 June 1997; accepted 30 June 1997)

Previous studies of infant cry acoustics and their perceptual significance have remained inconclusive as to the graded nature of cry production and perception and to the exact role and importance of particular acoustic features. In this study, a set of infant cries were digitally analyzed and resynthesized to form natural-sounding cries with varying fundamental frequency (F_0), degrees of jitter (period to period variations in F_0), and rise time (time for F_0 to reach its maximum value). In a perceptual rating task, higher-F_0 cries as well as cries with larger amounts of jitter tended to be given more negative ratings than were lower-F_0 cries and cries with less jitter, respectively. The perceptual ratings of the rise time manipulations were inconclusive. This study demonstrated a perceptual effect of F_0 and jitter independently of other parameters, consistent with current notions of infant cry gradedness. It was also shown that digital signal processing techniques can be fruitfully applied to infant cry research. © *1997 Acoustical Society of America.* [S0001-4966(97)05311-3]

PACS numbers: 43.71.Bp, 43.66.Lj, 43.71.Ft, 43.72.Ja [WS]

INTRODUCTION

The infant cry is a signal of considerable communicative and adaptive value. It is an effective means of conveying information about the infant's state even in the absence of visual contact: Differences in the sound alone can lead to differences in listeners' perception of the needs or state of the crying infant. It has been a subject of debate for a long time whether the perceptually salient acoustic differences affect caretakers' responses in a graded or discrete nature, or perhaps a combination of the two. In other words, do graded differences in infant cry signals along acoustic continua convey to caretakers graded information about an infant's state and needs, or do cries fall into functional categories with distinct constellations of acoustic features? This study was undertaken to determine the extent to which a few actually graded acoustic features function in a perceptually graded manner.

Many acoustic features have been examined as candidates for conveying information about the infant's state, including fundamental frequency of phonation (F_0), duration, and spectral characteristics of the cry. Cries with higher mean or maximum F_0 have been found to sound more aversive on a number of different perceptual scales (e.g., Zeskind and Marshall, 1988) and to ''elicit less optimal responses'' (Frodi and Senchak, 1990). The formant structure of the cries (Johnston and O'Shaughnessy, 1988), the segmental duration or ''tempo'' (Tsukamoto and Tohkura, 1992), and the duration of pauses and expiratory sounds (Zeskind *et al.*, 1992; Bryan and Newman, 1988) have also been found to affect perceptual ratings. Boukydis (1985) has provided a review of earlier findings.

Consistent with current notions of infant crying as a graded signal (Zeskind *et al.*, 1985; Zeskind *et al.*, 1993; Gustafson and Harris, 1990), Porter *et al.* (1986) found that durational as well as F_0-related characteristics are correlated with the severity of the procedure that the infant was undergoing at the time and that adult listeners were able to correctly assess the urgency of a cry. In addition, acoustic features have been found to correlate with situationally defined cry ''types,'' which presumably reflect the infant's state. For example, Fuller and Horii (1986, 1988) found that the F_0 and the energy distribution (tenseness) of cries differs between ''pain-induced'' and ''fussy'' or ''hungry'' cries.

The notion of gradedness is not incompatible with the use of situationally defined cry types. Cries within a category (type) may vary acoustically and perceptually along continuous scales. Furthermore, sets of cries with similar acoustic or perceptual characteristics may be taken to constitute a cry ''type,'' if, for example, they occur in similar eliciting conditions. Thus ''cry types'' may refer to clusters of cries around particular points on graded acoustic or perceptual continua. Wolff (1969), for example, defined the ''mad'' cry as a variation of the basic pattern of ''hunger'' cry and also noted that ''pain'' cries settle down to the same basic rhythmical pattern. The notion of gradedness clearly underlies Wolff's observations even though situationally defined cry types are named.

Listeners can identify the cause of a cry, although not as accurately as they can assess its urgency (e.g., Gustafson and Harris, 1990; Papoušek, 1989). It is not clear, however, how each acoustic feature affects listeners' perception of the cry, and which features, singly or in combination, are the more salient components for perception. The F_0, for example, is generally considered to be a perceptually salient feature of infant crying. However, its precise role and significance remain to be determined. The correlations between F_0 and perceptual ratings found by Zeskind and Marshall (1988), Zeskind and Collins (1987), and others cannot be interpreted with confidence because the acoustic context (i.e., the overall ''quality'' of the cries) may covary with F_0.

---

[a] Present address: Scientific Learning Corp., 1995 University Ave., Ste. 400, Berkeley, CA 94704-1074; Electronic mail: protopap@scilearn.com

Gustafson and Green (1989) did not find any significant correlations between F_0 measures and perceptual ratings but noted that ‘‘some of the correlations...[were] of a magnitude that would yield statistical significance with a larger sample’’ (p. 778). Based on strong correlations found between the perceptual ratings and other acoustic measures, such as cry duration, amount of dysphonation, and energy in low and high frequencies, they suggested that variations in F_0 may not be very important to cry ratings when the range of variation in F_0 is small. It is thus necessary to investigate small systematic variations in F_0 in a controlled acoustic context, as was done in the present study.

Studies that are based on acoustic analyses of natural cries and the correlation of their acoustic features with either perceptual ratings or physiological measures suffer from lack of controlled acoustic conditions, because the effects of single acoustic parameters (and of two or more parameters together) cannot be isolated and identified in natural cries. Given the nature of the infants' cry production system, it is certain that many closely coupled parts are interacting when a cry is produced, and that the resulting sound is affected in many ways by single events. For example, increased tension of the laryngeal muscles stiffens the vocal folds, thus resulting in more energy in higher frequencies (affecting timbre) and in higher F_0 (affecting pitch). It is therefore impossible in principle to completely separate acoustic features and their consequent perceptual effects in natural cries.

Studying the effects of specific features, individually and together, is the only way to identify the acoustic characteristics and their combinations that have perceptual effects as well as to determine whether and under which circumstances cries may be perceived as acoustic gestalts, i.e., with perceptual properties not directly attributable to their individual acoustic components. One way to accomplish this goal is to perform experiments with cries that are identical to one another in all respects but the single acoustic feature of interest. In order to create a set of stimuli that meet this condition, signal processing techniques must be used to alter a particular acoustic feature in a controlled way, while holding other features constant as much as possible. This was not feasible until recently, because of the technical demands of such a process.

Recognizing the merit of such an approach, Okada *et al.* (1987) used partially altered stimuli to investigate the perceptual effects of the ratio of voiced phonation to unvoiced phonation. More recently, Bisping *et al.* (1990) created resynthesized cries to control the F_0 and the melodic structure of the cry using an interpolation technique that scaled the F_0 and formant frequencies without affecting duration. (Some important methodological concerns regarding the latter study are discussed in the introduction to experiment 1 below.) Along the same lines, Zeskind *et al.* (1992) investigated the perceptual effects of variations in the temporal structure of cry bouts by digitally altering the duration of pauses and expiratory sounds.

It is becoming obvious that we need to retest the important findings from earlier studies using acoustically controlled procedures. In this study we used the linear predictive coding (LPC) method to decompose infant cries into sets of parameters, which could be independently manipulated and then used to resynthesize artificial cries with precisely controlled acoustic differences in F_0-related parameters. LPC is a well-tested and widely applied method in speech research whose behavior and requirements are well understood. Although far from perfect, for reasons discussed in the final section, LPC can be fruitfully applied to infant cry research if care is taken in the selection of processing parameters and in the kinds and ranges of acoustic modifications that are applied.

Besides average pitch, the amount of short-time fluctuation of F_0, referred to as ‘‘jitter,’’ is an important acoustic characteristic of infant crying. Jitter is always present in adult speech (Lieberman, 1961) and has been found to be linked to ‘‘stressor-provoked anxiety’’ (Fuller *et al.*, 1992) and to the emotional condition of the speaker (Lieberman and Michaels, 1962). It is perceptible in artificial signals, even in small amounts (Pollack, 1968; Rosenberg, 1966), and it is of potential diagnostic value in cases of laryngeal pathologies in adults (Lieberman, 1963; Perkins, 1985; Jacobson, 1994) and in infants (Hirschberg, 1990), as well as in cases of central nervous system disorders in infants (Mende *et al.*, 1990; Lüdge and Rothgänger, 1990). Fuller and Horii (1986) did not find any differences in amount of jitter between ‘‘different types of infant vocalizations’’ and considered jitter an ‘‘inaudible variable.’’ Fitch *et al.* (1992), however, found a correlation between the intensity of the cry-eliciting stimulus and jitter, perhaps due to a more appropriate analysis and higher temporal resolution [see Titze *et al.* (1987) for a discussion of temporal accuracy requirements in jitter measurements]. If jitter is indeed reliably correlated with the infant's level of stress, it is reasonable to assume that adult listeners might make use of it in their judgments about infant cries.

We were also interested in the perceptual significance of the time from the onset of phonation until the pitch reaches its maximal value, known with regard to infant cries as rise time, in that it appears to be strongly correlated with the infants' level of distress. For example, Wasz-Höckert *et al.* (1968) reported that a falling-melody form was characteristic of pain cries, and a rising–falling melody form was characteristic of hunger cries, which are presumably less aversive, at least at the beginning of the cry episode (Zeskind *et al.*, 1985; Rothgänger *et al.*, 1990). Given that a cry rarely starts off at its maximum pitch, usually taking some time to reach it, a ‘‘falling melody’’ should be interpreted as rapidly rising at first (i.e., having a short rise time) and then gradually falling. Therefore, it seems reasonable to expect that cries with faster rise times correspond to higher levels of arousal. Porter *et al.* (1986) provided strong support for this association in their study of cries from infants undergoing circumcision. Specifically, they found that moderately invasive procedures evoked rising–falling pitch pattern cries, whereas invasive surgical steps elicited cries ‘‘with pitch rapidly rising to high frequencies at the onset of voicing.’’ Rise time is thus in a similar state as F_0 in that its perceptual correlates, as identified in past studies, need to be rigorously investigated in controlled acoustic contexts.

A major contribution of the study of Porter *et al.* (1986)

TABLE I. Summary of cry characteristics of the eight natural cries that were analyzed and resynthesized.

Cry	Mean F_0 (Hz)	Max F_0 (Hz)	Duration (ms)	Risetime (ms)	Risetime (%)
1	435	500	907	505	56
2	425	476	1061	337	32
3	517	541	560	171	31
4	429	488	1983	558	28
5	455	500	2278	202	9
6	473	540	1673	232	14
7	579	625	996	272	27
8	596	667	1842	242	13

was that it provided clear evidence for the gradedness of infant crying, in both production and perceptual terms. They showed that the physical discomfort caused by the surgical procedures gradually affected acoustic characteristics and that these graded acoustic differences gave rise to graded perceptual differences. The issue of gradedness is a complex one, and many different lines of evidence are necessary before an appropriate definition of the term ‘‘graded’’ can be given. For the purposes of the present study the notion of gradedness is understood as follows: There is at least one continuum of discomfort that underlies the continua of acoustic features that can be perceived and ranked in a continuous manner along a number of perceptual scales. This study was designed to investigate whether this notion of gradedness applies to infant cry perception, addressing the perceptual role of gradations in individual acoustic features. The features that were investigated included fundamental frequency (F_0), whose precise role has remained controversial after much studying, jitter, whose role has yet to be explored but is potentially very significant, and rise time, which is so hard to ignore in a cry and yet remains relatively unresearched.

I. CONSTRUCTION OF STIMULI

Because of time constraints on the duration of experimental sessions, we could not use a large number of natural cries. As a consequence, in order to assess the generality of our method and perceptual findings, it was necessary to select a small but diverse set of cries for manipulation. We thus selected eight cries (individual cry expirations) from a set of 81 for their perceptual diversity, as indicated by a previous pilot experiment, and for their recorded quality and typicality of basic acoustic parameters. All cries were recorded with a cassette recorder from healthy, full-term infants at a hospital when they were between three and four weeks old. Cries 1 through 4 were spontaneous vocalizations of the infants prior to a medical examination. Cries 5 through 8 were elicited by the removal of a cardiac electrode, a moderately painful stimulus. Each cry that we used was produced by a different infant. Some acoustical characteristics of the eight cries are listed in Table I.

To create the resynthesized cry stimuli, each natural cry was low-pass filtered at 9.8 kHz, digitized at 12 bits with 20-kHz sampling rate,[1] upsampled to 80 kHz with quadratic interpolation for increased temporal resolution, and processed by means of a fortieth-order linear predictive coding (LPC) analysis using the autocorrelation method with Durbin’s recursive algorithm for solving the LPC equations (Rabiner and Schafer, 1978, p. 411) on 10-ms frames with 50% overlap. A pitch tracking program estimated the pitch[2] at 20-ms intervals using an autocorrelation-based algorithm after low-pass filtering at 800 Hz using an odd-length zero-phase FIR filter. The pitch estimates were visually inspected against the plotted waveform and any erroneous values were manually adjusted. The LPC reflection (PARCOR) coefficients and pitch values were stored to be independently manipulated and then used by the synthesis program. Synthesis was done by direct implementation of the recursive LPC filter, driven by impulses scaled by the LPC residual energy. The PARCOR coefficients and the pitch and gain parameters were updated at the beginning of each pitch period by logarithmically interpolating their stored values. The final resynthesized cries were passed through a 89-tap low-pass FIR filter with a cutoff frequency of 9.5 kHz and were downsampled to 20 kHz before being presented to the subjects.

There were four types of acoustic manipulation, henceforth termed ‘‘manipulation conditions.’’ In the first, termed simply ‘‘pitch’’, each cry was resynthesized with each pitch period scaled by a constant factor throughout the cry. There were five levels of pitch manipulation, with pitch periods equal to 90%, 95%, 100%, 105%, and 110% relative to those of the original cry.[3] The resynthesized cry with pitch periods equal to those of the original cry (scale value 100%) and the minimum amount of jitter is termed the ‘‘neutral’’ cry.

In the second manipulation condition, termed ‘‘jitter,’’ cries were resynthesized with pitch periods initially equal to those of the original cry (i.e., no scaling), but with increasing amounts of jitter across cries. Jitter was added by introducing random perturbations to the pitch estimates. There were three levels of jitter manipulation, with perturbation ranges $\pm 45\ \mu$s, $\pm 70\ \mu$s, and $\pm 90\ \mu$s. The neutral cry was treated as a fourth (baseline) level of jitter manipulation in the analysis.

The remaining two manipulation conditions involved shifting the point when the cry reached its maximum pitch. One can define that time either relative to the length of the entire cry (as a percentage) or as an absolute amount of time (in ms). We implemented both in order to investigate which would provide a better predictor of the perceptual differences. They were termed ‘‘risetime-P’’ and ‘‘risetime-T,’’ respectively. There were three levels of rise time manipulation in each of the two conditions: In risetime-P the peak was set at the point of 10%, 25%, and 50% of the length of the cry. In risetime-T the peak was set at 100, 250, and 500 ms from the onset of phonation. In order to shift the max-F_0 point, the LPC, pitch, and energy parameters were interpolated differentially around the original F_0 peak, thus ‘‘stretching’’ time on one side and ‘‘squeezing’’ it on the other.

The range of manipulation was always well within the normal range for all the parameters we altered. In this manner the acoustic changes could be treated as gradual and not as changing the quality of the cry. In particular, F_0 variations were kept small, so that the claim of Gustafson and Green

(1989) could be directly addressed and evaluated. Moreover, jitter variation was not so large as to create cries that would sound ‘‘dysphonated.’’ Our belief is that testing with more extreme parameter values might have produced results difficult to interpret if the perceived cries were qualitatively removed from the ends of their graded nature.

II. EXPERIMENT 1

Before conducting the main experiment, it was necessary to test the resynthesized stimuli for naturalness. In a sense, this is more of a pretest for experiment 2 than an independent experiment. However, there are two reasons to present it independently. First, it is of interest with respect to the methodology whether our method of creating resynthesized stimuli results in tokens natural enough to be useful for perceptual experiments. A thorough investigation of the perceptual properties of our stimuli will be of use to researchers in search for a good method with which to acoustically modify infant cries. Second, we deemed appropriate to emphasize the issue of stimulus naturalness because it has been overlooked in previous research, leading to perceptual findings whose interpretation is unclear.

Previous studies using acoustically altered cries have not directly addressed the question of the ‘‘naturalness’’ of the stimuli. Certainly, the experimenters have found the resynthesized cries to sound natural, but it is not clear whether naive listeners could have detected aspects of the stimuli that did not quite sound natural. The study of Bisping *et al.* (1990), where the spectral envelope, and thus the formants, was scaled along with F_0, illustrates the point. It is well known that F_0 may vary considerably, because it is mainly the result of the interactions of subglottal pressure and laryngeal muscular tension. In contrast, the range of formant frequencies is determined by the size and flexibility of the vocal tract and it is quite restricted for very young infants, whose larynges have not yet descended to their mature positions (Lieberman, 1980, 1985). Halving or doubling the formant frequencies of an infant cry is effectively similar acoustically to doubling or halving, respectively, the length of the infant’s vocal tract and the resulting cries may sound unnatural. Moreover, such dramatic alterations of F_0 may cause different interpretations of the cries if, for example, they sound as if produced by abnormal infants or if they change the perceived mode of phonation (e.g., from phonated to hyperphonated). Unfortunately, Bisping *et al.* (1990) did not discuss these issues.

In our view, caution must be used in interpreting results where naturalness was not directly measured. To counter possible concerns of this kind regarding our experiment 2, we conducted the present experiment to justify the use of our computer-generated cries. Subjects were asked to decide whether each stimulus was natural or synthetic. Given an adequate procedure for cry generation, we expected to find a lack of discriminability between natural and synthetic cries.

A. Method

1. Subjects

The subjects were 11 graduate students (6 women and 5 men) at Brown University, ranging in age from 23 to 38 years, who volunteered their participation. Five of the women reported having experience in childcare. One man was the father of a 2-year-old.

TABLE II. Frequencies and percentages of responses for natural and synthetic stimuli in experiment 1.

	Response			
	Natural		Synthetic	
Stimulus	*N*	%	*N*	%
Natural	282	55	232	45
Synthetic	448	43	602	57

2. Stimuli

There were two sets of stimuli in this experiment, a set of natural cries and a set of resynthesized cries. We used all 8 natural cries that had been analyzed and 6 resynthesized variants of each natural cry (for a total of 48 different resynthesized cries). Each natural cry was presented 3 times (in random positions) in each session, making for a total of 24 natural cry trials, in order to bring the number of natural cries closer to that of the synthetic cries ($n=48$).

Besides time constraints, the number of resynthesized cries used in this experiment had to be kept small because we did not want to bias our subjects by using a large number of resynthesized cries and only eight natural cries. The same six variants were chosen from each of the eight series: The neutral (baseline jitter and 100% pitch), the maximum jitter ($\pm 90\ \mu$s), the minimum and maximum pitch (90% and 110%), and the slowest and fastest rise time (50% and 100 ms). These covered the entire range of manipulations and by testing with the most extreme levels we believed we were more likely to include sounds that might be judged to sound artificial. If these cries were found to sound natural, we could safely conclude that the remaining cries, being the products of less extreme acoustic manipulations, would also sound natural.

3. Procedure

Subjects listened to one cry at a time and were asked to decide whether it was a natural or synthetic cry (forced choice task). The cries were played over a loudspeaker and the subjects responded by pressing a button on a response box. The entire set of stimuli was presented twice in different random orders, making for a total of 144 trials administered in each session, including 48 natural and 96 synthetic. Subjects had 3 s to respond and the next stimulus occurred 3 s after the response.

B. Results and discussion

The total number of responses in each condition is shown in Table II. Twenty responses (1.3%, including 14 in natural trials and 6 in synthetic trials) were not recorded because of response times greater than 3 s. From the percentages of hits and false alarms we computed for each cry a d' score, which is a z-score discrimination index not affected by response biases (Macmillan and Creelman, 1991). When the distributions underlying the subjects' responses are normal and their variances equal, the d' score is the distance between the distribution means of the two stimulus categories, expressed in standard deviation units. For each resynthesized cry only the responses to the corresponding natural cry were used in the calculation of the d' score.

The individual d' scores for each cry ranged between -0.47 and 1.20. There were four cries with an individual d' greater than 1.00, and seven with d' between 0.75 and 1.00. Of these 11 cries 5 were in one cry series, that of cry 6. Consequently, all stimuli from the cry 6 series were excluded from further experiments. Although there remained one cry with a $d'=1.03$, we chose to include it in experiment 2, because the amount by which its d' score exceeded a commonly accepted cutoff point of discriminability (1.0) was not large enough to justify rejection of this whole series of cries in our view. It was not feasible to discard only that particular stimulus, because the design of experiment 2 required that cries and levels be completely crossed. The 42 remaining synthetic cries were not nearly discriminable from the natural ones (d' well below 0.75, with six exceptions between 0.76 and 1.03).

The individual d' scores for each subject (including each subject's responses to all cries) ranged from -0.22 to 1.02, indicating that large differences exist between individuals in what is taken to sound like a ''natural'' infant cry. Interestingly, it was the two males with the least experience with infants that discriminated best between the natural and resynthesized cries (d' scores of 1.02 and 0.94), whereas the two females who reported a lot of experience in babysitting and the one subject who was a father showed very poor discrimination (d' scores between 0.23 and 0.30). This experiment was not designed to investigate differences related to experience with natural infant cries and interpretation of *post hoc* observations is of questionable validity. This individual variation notwithstanding, we conclude that our stimuli retained the most important acoustic properties of baby cries, since they sounded natural to the experienced subjects.

In summary, the results indicated that subjects were unable to discriminate reliably between the natural and the synthetic cries. All subjects reported that they found the task extremely difficult and none were confident about their responses. They said that almost all cries sounded more or less natural. On the basis of these results, we concluded that it is possible to synthesize artificial infant cries that sound natural using the LPC method. Although some cries might present problems, perhaps due to undesirable interactions between formant frequencies and the fundamental and to spectral properties that LPC cannot always handle appropriately (e.g., nasality), our results justify the use of synthetic cries in cry perception experiments and allow experimental conclusions to be extended to natural cry perception.

III. EXPERIMENT 2

The purpose of this experiment was to assess the perceptual effects of small variations in fundamental frequency, jitter, and rise time. In accord with earlier findings, we expected that higher fundamental frequencies would lead to higher ''negative'' ratings of the cries. We also expected higher negative ratings for cries with more jitter and for cries with shorter rise times.

The subjects were asked to rate the cries on four seven-point perceptual scales, called ''urgency,'' ''distress,'' ''health,'' and ''feeling.'' In the past, some researchers have used as many as fifty scales (e.g., Brennan and Kirkland, 1982); however, Zeskind and Marshall (1988) and Gustafson and Green (1989) showed that almost all those scales are highly interrelated. Zeskind and Lester (1978) suggested that there is a basic dimension of ''aversiveness,'' along which most cries differentiate, and an extra dimension of ''sickness'' or ''abnormality,'' which has an effect only with special cases of cries. We thus selected the four scales that we believed cover the spectrum of salient perceptual features, namely, severity or urgency, the inferred emotional state of the infant (distress), the possible abnormal aspect of the cry (health), and the aversive character of the cry (feeling).

A. Method

1. Subjects

The subjects were 20 undergraduate students at Brown University, including 8 men and 12 women ranging in age from 19 to 23, and were paid for their participation. None of the subjects were parents, but three of the men and eight of the women reported some experience in childcare (babysitting).

2. Stimuli

The total number of stimuli in this experiment was 105, 98 of which were resynthesized and 7 were natural cries. All 14 synthetic versions of each of the 7 remaining natural cries were used, i.e., there was a neutral version, 4 with altered pitch and baseline jitter, 3 with increased amounts of jitter, and 6 with altered rise time (3 by absolute time and 3 by proportion). The seven natural cries that had been originally analyzed were used in order for their ratings to be compared with those of the corresponding resynthesized cries. The ratings of the natural cries were excluded from the analyses for acoustic manipulation effects and were only used to measure the level of perceptual overlap between them and the resynthesized cries.

3. Procedure

Subjects heard one cry per trial and responded by pressing one of seven buttons arranged along an arc on a specially designed response box. The responses were automatically recorded in an IBM AT computer. Each stimulus was presented twice during each rating session and the mean of the two responses was used for the analysis. The order of the stimuli was randomized before each rating session. Subjects had 3 s to respond and the next stimulus occurred 2 s after the response.

The instructions given to the subjects were to rate the cries on a scale from one to seven according to how urgent they sounded, how distressing they sounded, how sick the infant sounded, and how angry or sad the cry made them feel; the endpoints of the scales were anchored by the words

TABLE III. Correlation of ratings of natural cries with ratings of corresponding neutral resynthesized cries and slope of linear regression. All correlations are significant to $p < 0.0001$.

Rating scale	Correlation coefficient	Linear regression slope
Urgency	0.994	0.890
Distress	0.981	0.887
Health	0.992	0.951
Feeling	0.991	0.879

''most'' and ''least'' before the words ''urgent,'' ''distressing,'' ''sick,'' and ''feeling angry or sad,'' respectively, and were indicated next to the endpoint buttons throughout each rating session.

The stimuli were presented over a single loudspeaker positioned to face the subject directly. The sound level was adjusted to a comfortable level and remained unchanged throughout the experiment. Each subject rated all cries on one scale before proceeding to the next scale. This way interference between rating scales was minimized and, importantly, subjects did not have to remember the cry for a long time while indicating their responses to several rating scales. The presentation order of the four scales and the direction (polarity) of the rating scale were balanced across subjects using a latin square design. The experiment was divided into two parts, each lasting about one half hour.

B. Results

In order to compare the ratings of the natural cries to those of the resynthesized ones across the 20 listeners, we calculated for each of the 4 rating scales the correlation coefficient between the mean ratings of the neutral resynthesized cries and the mean ratings of the corresponding natural cries, which were included for just this purpose. Table III shows the resulting correlation coefficients and linear regression slopes. Although the neutral cries may not always be the closest to the natural ones,[4] either acoustically or perceptually, and the analysis–synthesis procedure tended to decrease the perceptual range slightly, across subjects the order of the neutral cries, as given by the perceptual ratings, was always the same as the order of the natural cries and there was in fact a very close correspondence between the ratings of the natural and the neutral stimuli in all conditions.

The effects of the four manipulation conditions in the four rating scales are shown in Fig. 1. In brief, scale ratings tend to increase, i.e., become more negative, as F_0 increases and as the amount of jitter increases, in agreement with initial expectations. Effects of the two rise time manipulations were very small, although risetime-P followed the expected pattern, i.e., toward lower (less negative) ratings for longer rise times. Risetime-T only followed the pattern for the two shorter rising times, and that was not the case for the health ratings.

Two-way, within-subjects analyses of variance were performed, separately for each of the four rating scales and for each of the four manipulation conditions. The two factors, treated as fixed-effects variables, were cry item (the seven natural cries on which the resynthesized cries were based, henceforth termed plainly ''cry'') and manipulation level of the acoustic variable (five levels for pitch, four for jitter, and three for each of the risetime conditions). Thus for each of the four rating scales we performed four ANOVAs: Cry(7)×Pitch(5), Cry(7)×Jitter(4), Cry(7)×Risetime-P(3), and Cry(7)×Risetime-T(3).[5]

The effect of pitch was significant for each rating scale at the 0.005 level or less, and the effect of jitter was significant for all rating scales except urgency at the 0.04 level or less. The exact F and p values are shown in Table IV, as calculated with the Huynh–Feldt method for adjusting the degrees of freedom according to the $\tilde{\varepsilon}$ estimate of the population treatment-difference variances (Maxwell and Delaney, 1990, p. 477). The adjusted degrees of freedom are fractional because there is no need to round them to the closest integer, given the use of computer programs to calculate the exact probability of occurrence for any F value with any possible adjusted degrees of freedom. Risetime-P never had a significant effect, whereas risetime-T produced significant effects, but only for the ratings of distress [$F(2.0,38.0)=3.33$, $p=0.047$] and feeling [$F(2.0,38.0)=6.73$, $p=0.003$]. There was always a significant effect of cry at the 0.0001 level for all manipulation conditions and rating scales, which shows that our selection of cries was diverse, as intended.

There was also, in three cases, a significant interaction of cry with pitch [urgency: $F(17.0,323.7)=2.27$, $p=0.003$; distress: $F(17.0,323.8)=2.10$, $p=0.007$; feeling: $F(16.8,318.2)=2.37$, $p=0.002$], and, in two cases, of cry with rise time [risetime-P, health: $F(11.9,225.5)=2.14$, $p=0.016$; risetime-T, urgency: $F(11.4,216.2)=1.89$, $p=0.040$]. These findings indicate that the effects of the pitch and rise time manipulations depended on the original characteristics of each cry and thus were not uniform across all cries. The interaction of cry with pitch is illustrated in Fig. 2, which shows the effects of pitch on each cry in the distress condition. It is clear that the effect of changing F_0 is not the same for all cries but, except for cry 2, it is more pronounced and more in line with expectations for cries that received the highest overall ratings, i.e., the ''pain cries.'' Note also the rating consistency between subjects, evidenced by the small standard errors, and the use of the entire rating scale, indicating that our stimuli covered an extensive perceptual range.

In order to further investigate the effect of the pitch manipulation and its interaction with cry we performed a trend analysis on the ratings for each rating scale. In all four rating scales there was a significant linear trend in the ratings with higher, i.e., more negative, ratings with higher pitch levels [urgency: $F(1,19)=42.76$, $p<0.0001$; distress: $F(1,19)=67.10$, $p<0.0001$; health: $F(1,19)=11.84$, $p=0.0027$; feeling: $F(1,19)=55.00$, $p<0.0001$]. This linear trend interacted significantly with cry in all but the health ratings [urgency: $F(6,114)=6.72$, $p<0.0001$; distress: $F(6,114)=4.65$, $p=0.0003$; feeling: $F(4.85,92.15)=5.41$, $p=0.0002$], indicating that the slopes of the rating curves were different for different cries. Linear analyses were performed for each cry separately, whereby the effects of pitch were found to be significantly linear (after the Bonferonni adjustment for *post hoc* contrasts) for the urgency ratings of

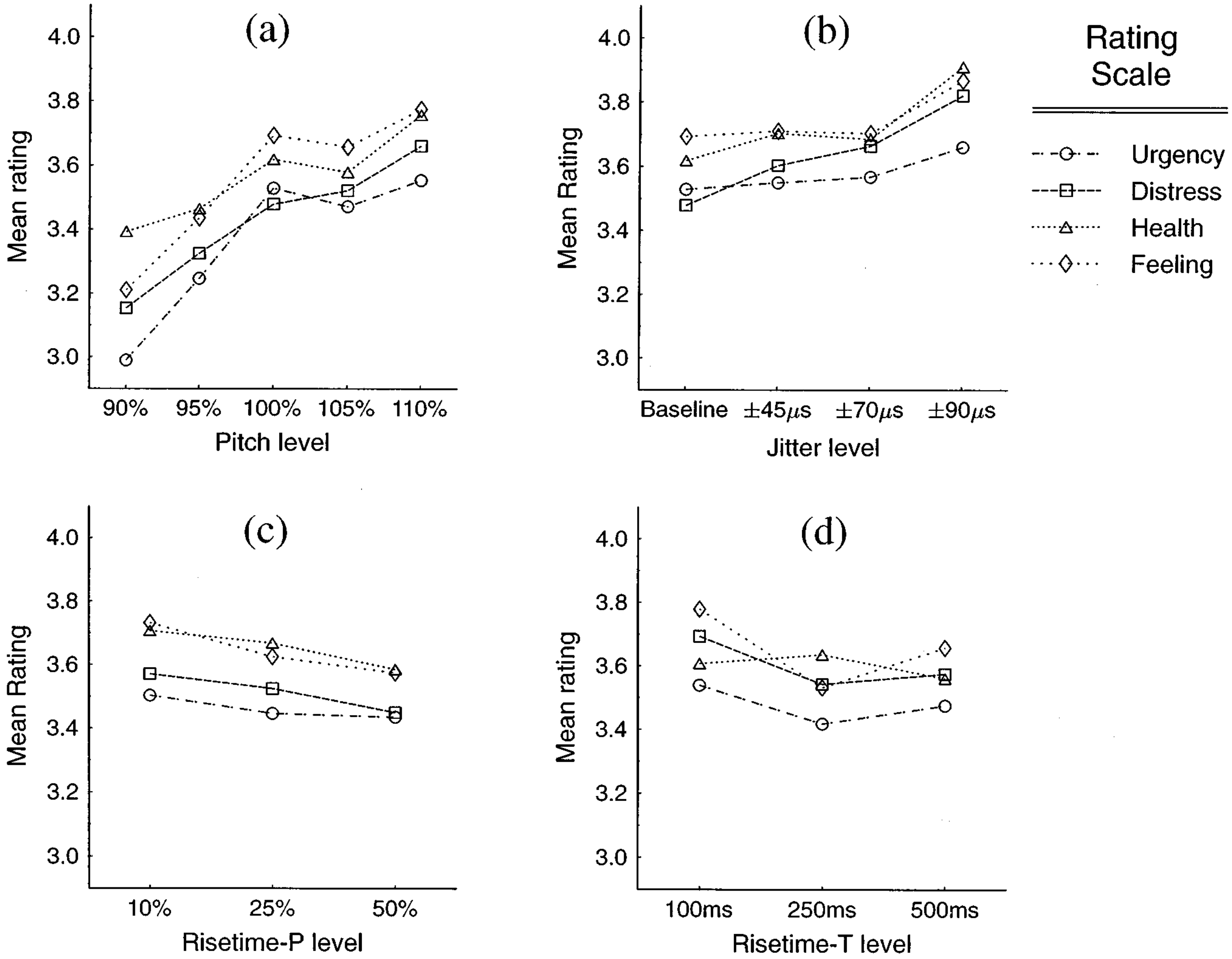

FIG. 1. Variation of ratings (averaged across cries) with level of acoustic manipulation in experiment 2 in all four rating scales. (a) Pitch (relative to original pitch of natural cry); (b) jitter (pitch period perturbation range); (c) risetime-P (cry portion before maximum pitch); and (d) risetime-T (time length before maximum pitch). The rating scale was always from 1 to 7, with 7 corresponding to the "negative," i.e., most aversive, end of each scale.

cries 2, 5, 7, and 8, the distress ratings of cries 2, 5, and 8, and the feeling ratings of cries 4, 5, 7, and 8. Analysis of the health ratings was not performed for each cry separately since we found no significant interaction of the linear trend of health ratings with cry $[F(6,114)=1.14,\ p=0.3410]$. In other words, the effects of the pitch manipulation vary with cry and with rating scale. It appears that the cries that received the highest overall ratings, in particular the pain cries, tended to be more affected, but not without exceptions.

TABLE IV. Results of the significance tests for the main effects of acoustic manipulations of pitch and jitter in experiment 2 (with Huynh–Feldt adjustment).

Acoustic manipulation	Rating scale	Fixed-effects analysis		
		F	adj. df	p
Pitch	Urgency	19.67	3.7,69.7	<0.0001
	Distress	15.94	4.0,76.0	<0.0001
	Health	4.87	3.5,66.6	0.0026
	Feeling	18.30	4.0,76.0	<0.0001
Jitter	Urgency	1.21	2.5,46.7	0.3134
	Distress	7.88	2.6,50.2	0.0004
	Health	4.14	2.5,47.6	0.0152
	Feeling	3.14	2.9,55.2	0.0337

It is not surprising that the ratings to the pain cries (cries 5, 7, and 8) were higher than the ratings to the spontaneous cries (cries 1 through 4). Examination of Fig. 2 and Table I reveals that three of the four cries that received the highest overall ratings were the ones with the longest absolute duration (cries 5, 4, and 8, in order of decreasing duration). Two of these were also the cries for which the pitch manipulation always resulted in significant perceptual differences (cries 5 and 8). Since the duration factor was not explicitly manipulated or controlled for in the design of this experiment, it is not possible to draw any conclusions regarding possible perceptual effects of duration or of an interaction between duration and F_0. It is plausible that pitch differences in longer cries are perceptually more salient than comparable pitch differences in shorter cries (cf. Tsukamoto and Tohkura, 1990, on increased category identification rates with longer cries). Because there are many covarying factors in natural infant cries it is impossible to control simultaneously for more than

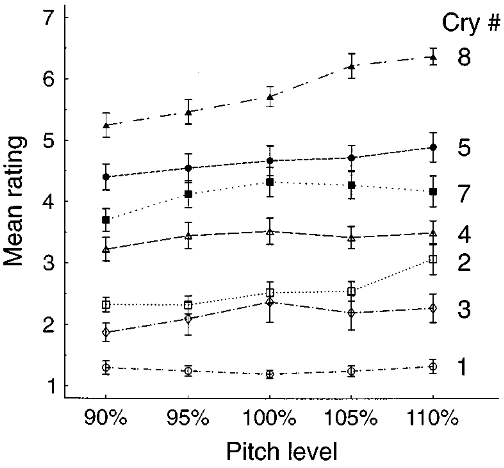

FIG. 2. Variation of ratings of individual cries with level of acoustic manipulation of pitch (relative to original pitch of natural cry) in experiment 2 in the distress rating scale. The errors bars show standard errors.

one or two of them. Another study is needed to investigate the effects of duration manipulations.

There was no significant interaction of cry with jitter in any of the rating scales, indicating that the perceptual effects of jitter were uniform across cries. A trend analysis indicated that the significant effect of the jitter manipulation on the distress, health, and feeling ratings was due to the statistically significant linear trend towards higher ratings for cries with more jitter [distress: $F(1,19)=18.87$, $p=0.0003$; health: $F(1,19)=5.93$, $p=0.0249$; feeling: $F(1,19)=6.31$, $p=0.0212$]. Although the main effect of jitter did not interact with cry in any rating scale, in the linear analysis the linear trend interacted with cry for the distress ratings only [$F(5.52,104.97)=3.16$, $p=0.0084$], indicating that the linear effect was not uniform across cries for this rating scale. Separate linear analyses of the distress ratings for each cry (with the Bonferonni adjustment) showed a significant linear trend for the ratings of cries 2, 7, and 8.

Surprisingly, the effects of risetime-P and risetime-T were very small. In addition, the effects of risetime-P were never statistically significant and the effects of risetime-T failed to show the expected trend. No clear global effect was seen in the risetime-P ratings, but several cries showed a small trend one way or the other. For some cries shorter rise time led to higher ratings, whereas for others longer rise times led to higher ratings. Similarly, the effects of risetime-T failed to show a uniform pattern. The fact that the effect of risetime-T was statistically significant in the feeling and distress conditions does not lend itself to a meaningful interpretation, in that the rating function for risetime-T is nonmonotonic. No significant linear trends were found for the ratings in the rise time manipulations.

Given these main effects and interactions of the various manipulations, it is reasonable to assess the extent to which any conclusions might be generalizable to infant cry perception in general. Treatment of the cry factor as a fixed-effects factor in the ANOVA only permits an inference of significance of reliable effects for the particular items (the natural original cries) that were used. Because the cries were chosen so as to cover a wide perceptual range, and because their acoustic characteristics are typical of clearly phonated yet quite diverse cries within that class, we can assume that they constitute a representative sample of all or at least of the most clearly phonated cries of comparable duration. We can then test the significance of the effects of the acoustic manipulations with cry considered a random factor. To do this, we used the method of Quasi-F (Myers, 1979, pp. 188–192) to calculate the appropriate denominator terms and degrees of freedom. The results are shown in Table V. Note that the effect of the pitch manipulation is significant in all conditions, and that of the jitter manipulation is significant in two of the four conditions, namely distress and health. It would then seem that the effects of F_0 are generalizable across cries for all rating scales, but only for ratings of distress and health in the case of jitter. Inasmuch as the results of the fixed-effects analyses for rise time were inconclusive, there was little point in performing this analysis for the risetime-P and risetime-T conditions.

TABLE V. Results of the significance tests for the main effects of acoustic manipulations of pitch and jitter in experiment 2 with cry treated as a random factor.

Acoustic manipulation	Rating scale	Mixed-model analysis		
		Quasi-F	adj. df	p
Pitch	Urgency	9.48	4,26	<0.001
	Distress	7.59	4,22	<0.001
	Health	4.70	4,23	<0.01
	Feeling	7.42	4,21	<0.001
Jitter	Urgency	0.93	3,16	>0.25
	Distress	7.19	3,20	<0.005
	Health	4.15	3,21	<0.025
	Feeling	2.53	3,9	>0.10

IV. DISCUSSION

Fundamental frequency has long been hypothesized to affect perception of infant cries, both in a graded manner (as defined in the introduction) and in a discontinuous manner (as defined by the modes of infant cry phonation). Here we confirmed the graded effect with a manipulation in a constant context, the causal nature of which constitutes an improvement in methodology over that found in correlational studies. In addition, we have shown that jitter has similar (if smaller) perceptual effects that have not been emphasized in the literature and need to be further investigated. Finally, we have been unable to provide evidence for perceptual effects of rise time, although descriptions of cries in the literature might have led one to expect such effects. In the following we elaborate on the details of these findings and on their significance, in the context of the methodology we have used.

A. Gradedness

The main issue motivating this study was whether graded differences in particular acoustic features in a constant acoustic context would give rise to graded perceptual responses. The linear relation found between cry ratings differing only to a small degree in mean F_0 or jitter unequivocally supports the notion of gradedness in infant cry perception, inasmuch as a range of acoustic variability was shown to map onto a range of perceptual variability. This notion of gradedness has also been extended to temporal features by Zeskind *et al.* (1992, 1993), who found that cry bouts with shorter expirations and with shorter pauses between expirations were given higher negative ratings.

However, nongraded aspects of infant cry perception have also been reported and must temper our conclusions on the pervasiveness of graded infant cry perception. For example, Wiedenmann and Todt (1990) asked listeners to respond to recorded infants' cries when they felt ''emphasis being expressed in the vocalizations'' or when they would ''get up to soothe the infant in a real life situation'' (p. 181). They found that their subjects' responses, indicating high cry aversiveness, were clustered at particular points in the cries and were not uniformly distributed. It would be hard to ascribe that finding to graded perception; a more likely explanation is that a more or less discrete acoustic parameter, e.g., onset of dysphonation or ''biphonation,'' was causing the abrupt changes in behavioral response (cf. Gustafson and Green, 1989, regarding separate treatment of phonation modes). Furthermore, the issue of gradedness cannot be examined independently of physiological and social factors that are active in the determination of the cry signal (Lester and Boukydis, 1992).

B. Fundamental frequency

The present findings on the perceptual effects of F_0 are in general agreement with those in the literature, showing a shift to more negative ratings as the mean F_0 increases. The effect is highly significant, and although it may seem small, it should be borne in mind that pitch periods were altered by at most plus or minus 10%. The fact that the effect of the pitch manipulation was not uniform across all cries indicates that there are other acoustic attributes, as yet unknown, that take precedence in determining the overall quality of the cry, or that counter or neutralize the F_0 manipulation. Also, in our manipulation, although the highest-pitched version of one cry was often higher pitched than the lowest-pitched version of another cry, their perceptual ratings rarely crossed, i.e., the effect of changes in F_0 was relative to the particular cry and not to the absolute value of F_0. As illustrated in Fig. 2, the ratings of different cries were well separated, each cry series occupying a small region within the available rating space.

One reason that the ratings of each series of cries tended to cluster together might be that subjects tried to be consistent by remembering the cries. Some subjects reported that they thought they heard only a few cries, and were trying to give the same rating to the ''same'' cry. Admittedly, cries that only differ in F_0 by 5% sound very similar. In fact, unless heard in succession, they might well sound identical and, consequently, if subjects were trying to be consistent, the effect would be a flattening of the rating curves. Therefore, the criticism of Zeskind and Huntington (1984) that within-subject designs may accentuate unreliable perceptual differences would not apply to our study because of the perceptual similarity between the items and the tendency of listeners to be consistent. The fact that the effect of pitch is nonetheless clear for a number of cries and highly significant demonstrates that it is indeed robust.

As noted, this finding corroborates the findings from correlational studies that higher-pitched cries are perceived as being more aversive. Furthermore, by using small F_0 ranges, we can address the point brought up by Gustafson and Green (1989) about the perceptual importance of F_0 being small when F_0 does not vary much. The overall subjective quality of each cry series did not change between versions with different mean pitch because the alterations were very small, so it cannot be the case that the perceptual differences we found were due to perception of resynthesized cries differing only in pitch as belonging to distinct cry types (i.e., cries that distinctly differ in their perceived quality). Furthermore, acoustic variables that may, in natural cries, be intrinsically coupled with F_0 (such as spectral energy distribution) were kept constant, to the extent possible. Thus the observed effect shows clearly that F_0 can have a direct gradual perceptual effect, independently of other parameters, albeit not for every cry.

Although it has long been accepted that F_0 is perceptually salient, it was not previously possible to distinguish between the possibility that other parameters, related to F_0, were responsible for the observed correlations, and the possibility that F_0 was directly perceived as a cue to cry aversiveness. Our study shows that small changes in F_0 alone can give rise to significant perceptual differences, although there are probably other acoustic parameters that are as important or perhaps even more important in defining the overall aversive quality of a cry. We are therefore providing direct support both for Gustafson and Green's contention that F_0 is not the most important acoustic determinant of perceived cry aversiveness and for their suggestion that the perceptual effects of F_0 might be statistically significant in a larger (or, as in our case, better controlled) cry sample.

C. Jitter

Similar conclusions with regard to gradedness may be drawn regarding the effect of jitter. Cries only differing by a small amount of jitter are difficult to discriminate, and thus subjects' striving for consistency would again have worked against our hypothesis and may have been the cause of the reduction in reliable effects when cries were taken to be a random variable. The trend for higher (more negative) ratings to cries with more jitter serves to strengthen a point of evolutionary adaptation of the mother–infant communication system. If jitter is indeed correlated with particular disorders of the nervous system or other indices of need for special treatment, our auditory system may have naturally evolved to incorporate jitter into the set of perceptually salient acoustic features of infant crying. Given recent findings on the potential value of jitter analysis for diagnostic purposes (Hirschberg, 1990; Mende *et al.*, 1990) and on the correlation of F_0 variability with the short-term affective state and with the long-term dispositions of infants (Fitch *et al.*, 1992), there are many issues related to jitter that are worth investigating. It remains to investigate, for example, what kinds of jitter are present in natural cries and how each kind correlates with production physiology and with perceptual judgements using our current methodology.

Notably, we discovered during pilot experiments that jitter failed to show a perceptual effect when the stimuli were presented through headphones. In fact, subjects were unable to discriminate cries with different amounts of jitter when they heard them through headphones, but could perform the task relatively easily when the same stimuli were presented over a loudspeaker. This finding is not surprising in the light of similar findings by Wilde *et al.* (1986) on the perception of jitter in synthesized speech. On the basis of those results, it is now becoming common practice by researchers who use synthetic voice stimuli to prefer loudspeaker presentation to headphone presentation (e.g., Hillenbrand, 1988; Milenkovic, 1993). Wilde *et al.* (1986) attributed this perceptual effect to reverberation, which is present in free-field presentation (i.e., speakers) but not through headphones. Little is known about other possible differences between the two modes of presentation, but this study shows that jitter plays a role in cry perception and appears to be affected by presentation mode. We thus suggest that cry perception experiments involving synthetic stimuli should be performed using loudspeakers rather that headphones, particularly if jitter perception is of interest.

D. Rise time

As for the effects of rise time, we were surprised to find that they were small and in most cases not statistically significant. Although informal listening showed that cries differing in rise time sounded quite distinct, it appears that this kind of difference carries little or inconsistent information about the infant's state. We suggest that rise time might only be correlated with the indicators of cry aversiveness and not a determinant of aversiveness by itself. Certainly, further research is necessary to resolve the issue.

E. Stimulus naturalness and LPC

Finally, we briefly discuss the merits of our resynthesis method for perceptual investigations. The correlation of the ratings of the original natural cries with the ratings of the resynthesized cries, in conjunction with the results of experiment 1, shows that resynthesized cries retain the acoustic properties of the natural cries of infants. They capture well the aversive character of the cry in that the ordering of cries by human listeners in all four rating scales is preserved in the resynthesized cries. Consequently, an acoustic feature that affects perception of resynthesized cries toward higher or lower ratings can be reasonably assumed to affect perception of natural cries in a similar manner and we feel confident in concluding that the trends in rating differences caused by our acoustic manipulations would be nearly the same with natural cries (were it possible to obtain cries with comparable acoustic properties). The fact that our synthetic stimuli were nearly indiscriminable from natural cries (experiment 1) serves to strengthen this conclusion.

The naturalness of synthetic stimuli is an important aspect of perceptual studies, often overlooked in previous research. The use of synthesized cries is not a panacea for investigations involving acoustic features; a host of complications may interfere with the experimental manipulations. As already discussed, major problems can arise from allowing the formant frequencies to be scaled along with F_0. Nevertheless, within the range of parameters we have studied, LPC has proved to be an adequate signal processing method for constructing artificial cry stimuli (although, because of the noisy character of the infants' glottal excitation of the vocal tract and of the nasality of most infant vocalizations, the physical meaning of the LPC parameters is not as clear as it sometimes is in speech modeling). The ability to separate the quasi-periodic excitation component from the spectral shaping component according to the source-filter theory of speech (Fant, 1960; Lieberman and Blumstein, 1988) and cry production (Golub and Corwin, 1985) is important in that it allows almost independent control of many acoustic features, and our study shows that the applications of LPC can be fruitfully extended into infant cry processing as it already has in speech as well as in animal vocalization research.

However, it must be pointed out that LPC is probably not the optimal method for studies like this. The dependency of LPC spectra on F_0 and the inherent inability of LPC to model correctly the nasality often present in infant cries make selection of cry stimuli and estimation of processing parameters an involved trial-and-error process. Keeping the extent of the acoustical modifications small is critical to retain the naturalness of the resulting stimuli. In light of more recent advances in digital signal processing for speech, it may be fruitful to explore the potential of time-domain pitch-synchronous overlap/add (PSOLA) methods (Moulines and Laroche, 1995; Veldhuis and He, 1996) for F_0 modification of infant cries. As with LPC and any other speech processing methods, one must experimentally establish whether successful application to infant cry signals is possible, because of the many acoustic differences between cries and speech. In particular, the high F_0 of infant cries may present problems for PSOLA methods since, in a recent psychoacoustical evaluation of PSOLA with single-formant synthetic stimuli, Kortekaas and Kohlrausch (1997) concluded that ''distortions introduced to signals with higher fundamental frequencies are expected to be more easily detectable'' (p. 2211).

V. CONCLUSION

In summary, we have found significant perceptual effects of the fundamental frequency of the cries, or pitch, and of its short-time perturbations, known as jitter. These findings support the notion of infant crying being a graded signal in the sense that gradual acoustic variation gives rise to gradually differing perceptual ratings. Our findings on the role of rise time were inconclusive. With respect to our methodology, we have showed that a digital signal processing technique can be successfully applied to infant cry perception research by providing acoustically controlled artificial cry stimuli that sound natural. By using this method it is possible to investigate the role not only of individual acoustic features but also of their various combinations, and to examine the relationships between the mechanisms of perception and production. Further research of this nature should improve our understanding of the communicative function of infant crying, with important implications for developmental research and parenting.

ACKNOWLEDGMENTS

This study was supported in part by Grant No. HD 05331 from the National Institute of Child Health and Human Development to PDE. Treatment of human subjects in the experiments described herein was in accordance with the ethical standards of the APA. The authors thank C. F. Zachariah Boukydis and three anonymous reviewers for comments on the manuscript. The infant cries that were used were originally recorded for research projects by Barry M. Lester, Brown University, E. P. Bradley Hospital, and Women and Infants Hospital, Providence, RI.

[1] The maximum amplitude of each cry was individually adjusted during digital sampling to cover the available dynamic range.

[2] The terms ''pitch'' and ''fundamental frequency'' are used interchangeably.

[3] A small amount of jitter was added to each period after scaling because some synthetic stimuli do not sound natural without jitter. The minimum amount necessary for each cry to sound natural was found by trial and error, and ranged between $\pm 15\ \mu$s and $\pm 30\ \mu$s. The same amount was added in all variants of each cry in the pitch and risetime manipulations.

[4] The neutral cries are not always the closest to the natural ones because the natural cries often contained more jitter than that in the neutral synthetic ones, whose pitch contour was smoothed to minimize the jitter content. Therefore it was expected that neutral cries would be given on average lower ratings than the corresponding natural cries if jitter had the expected perceptual effect. An additional reason concerns the fact that the spectral flattening of the most aversive sounding cries that is caused by noise excitation produced at the infant's glottis cannot be perfectly accommodated in the LPC model with periodic excitation.

[5] Separate analyses for the between-subjects factors of sex and childcare experience (with cry and level being within-subjects factors) showed no effect of experience [$F(2,17)<2$, $p>0.17$], but a significant effect of sex [$F(1,18)>4.41$, $p<0.05$], with men giving higher ratings than women in all conditions.

Bisping, R., Steingrüber, H. J., Oltman, M., and Wenk, C. (**1990**). ''Adults' tolerance of cries: An experimental investigation of acoustic features,'' Child Dev. **61**, 1218–1229.

Boukydis, C. F. Z. (**1985**). ''Perception of infant crying as an interpersonal event,'' in *Infant Crying*, edited by B. M. Lester and C. F. Z. Boukydis (Plenum, New York), pp. 187–215.

Brennan, M., and Kirkland, J. (**1982**). ''Classification of infant cries using descriptive scales,'' Inf. Behav. Dev. **5**, 341–346.

Bryan, Y. E., and Newman, J. D. (**1988**). ''Influence of infant cry structure on the heart rate of the listener,'' in *The Physiological Control of Mammalian Vocalizations*, edited by J. D. Newman (Plenum, New York).

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fitch, T., Bryan, Y., Huffman, L., Lester, B. M., and Newman, J. (**1992**). ''Pitch perturbation in human infant cries,'' Paper presented at the IVth International Workshop on Infant Cry Research (Munich).

Frodi, A. M., and Senchak, M. (**1990**). ''Verbal and behavioral responsiveness to the cries of atypical infants,'' Child Dev. **61**, 76–84.

Fuller, B. F., and Horii, Y. (**1986**). ''Differences in fundamental frequency, jitter, and shimmer among four types of infant vocalizations,'' J. Comm. Disord. **19**, 441–447.

Fuller, B. F., and Horii, Y. (**1988**). ''Spectral energy distribution in four types of infant vocalizations,'' J. Comm. Disord. **21**, 251–261.

Fuller, B. F., Horii, Y., and Conner, D. A. (**1992**). ''Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety,'' Res. Nursing Health **15**, 379–389.

Golub, H. L., and Corwin, M. J. (**1985**). ''A physioacoustic model of the infant cry,'' in *Infant Crying*, edited by B. M. Lester and C. F. Z. Boukydis (Plenum, New York), pp. 59–81.

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

Gustafson, G. E., and Harris, K. L. (**1990**). ''Women's responses to young infants' cries,'' Dev. Psychol. **26**, 144–152.

Hillenbrand, J. (**1988**). ''Perception of aperiodicities in synthetically generated voices,'' J. Acoust. Soc. Am. **83**, 2361–2371.

Hirschberg, J. (**1990**). ''The value of the acoustic analysis of pathological infant cry and breathing noise in everyday practice,'' Early Child Dev. Care **65**, 57–69.

Jacobson, B. H. (**1994**). ''Objective voice analysis: the clinical voice laboratory,'' in *Vocal Arts Medicine: The Care and Prevention of Professional Voice Disorders*, edited by M. S. Benninger, B. H. Jacobson, and A. F. Johnson (Thieme Medical, New York), pp. 135–152.

Johnston, C. C., and O'Shaughnessy, D. (**1988**). ''Acoustical attributes of infant pain cries: Discriminating features,'' in *Proceedings of the Vth World Congress on Pain*, edited by R. Dubner, G. F. Gebhart, and M. R. Bond (Elsevier, Amsterdam).

Kortekaas, R. W. L., and Kohlrausch, A. (**1997**). ''Pyschoacoustical evaluation of the pitch-synchronous overlap-add speech-waveform manipulation technique using single-formant stimuli,'' J. Acoust. Soc. Am. **101**, 2202–2213.

Lester, B. M., and Boukydis, C. F. Z. (**1992**). ''No language but a cry,'' in *Nonverbal Vocal Communication, Comparative and Developmental Approaches*, edited by H. Papoušek, U. Jürgens, and M. Papoušek (Cambridge U.P., Cambridge, UK).

Lieberman, P. (**1961**). ''Perturbations in vocal pitch,'' J. Acoust. Soc. Am. **33**, 597–603.

Lieberman, P. (**1963**). ''Some acoustic measures of the fundamental periodicity of normal and pathologic larynges,'' J. Acoust. Soc. Am. **35**, 344–353.

Lieberman, P. (**1980**). ''On the development of vowel production in young children,'' in *Child Phonology: Perception and Production*, edited by G. Yeni-Komshian and J. Kavanaugh (Academic, New York).

Lieberman, P. (**1985**). ''The physiology of cry and speech in relation to linguistic behavior,'' in *Infant Crying*, edited by B. M. Lester and C. F. Z. Boukydis (Plenum, New York), pp. 29–57.

Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge U.P., Cambridge, UK).

Lieberman, P., and Michaels, S. B. (**1962**). ''Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech,'' J. Acoust. Soc. Am. **32**, 922–927.

Lüdge, W., and Rothgänger, H. (**1990**). ''Early diagnosis of CNS disturbance from computer analysis of infant cries: A new method of fundamental frequency jitter computation with high resolution in frequency and time,'' Early Child Dev. Care **65**, 83–90.

Macmillan, N. A., and Creelman, D. C. (**1991**). *Detection Theory: A User's Guide* (Cambridge U.P., Cambridge, UK).

Maxwell, S. E., and Delaney, H. D. (**1990**). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (Wadsworth, Belmont, CA).

Mende, W. E., Wermke, K., Schindler, S., Wilzopolski, K., and Höck, S. (**1990**). ''Variability of the cry melody and the melody spectrum as indicators for certain CNS disorders,'' Early Child Dev. Care **65**, 95–107.

Milenkovic, P. H. (**1993**). ''Voice source model for continuous control of pitch period,'' J. Acoust. Soc. Am. **93**, 1087–1096.

Moulines, E., and Laroche, J. (**1995**). ''Non-parametric techniques for pitch-scale and time-scale modification of speech,'' Speech Commun. **16**, 175–205.

Myers, J. L. (**1979**). *Fundamentals of Experimental Design* (Allyn and Bacon, Boston, MA), 3rd ed.

Okada, H., Murai, N., and Adachi, T. (**1987**). ''VP ratio of the infant cry sounds and its influences on their perception,'' Tohoku Psychologica Folia **46**, 36–42.

Papoušek, M. (**1989**). ''Determinants of responsiveness to infant vocal expression of emotional state,'' Inf. Behav. Dev. **12**, 507–524.

Perkins, W. H. (**1985**). ''Assessment and treatment of voice disorders: State of the art,'' in *Speech Disorders in Adults*, edited by J. Costello (College-Hill, San Diego, CA), pp. 79–111.

Pollack, I. (**1968**). ''Detection and relative discrimination of auditory 'jitter','' J. Acoust. Soc. Am. **43**, 308–315.

Porter, F. L., Miller, R. H., and Marshall, R. E. (**1986**). ''Neonatal pain cries: Effect of circumcision on acoustic features and perceived urgency,'' Child Dev. **57**, 790–802.

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).

Rosenberg, A. E. (**1966**). ''Pitch discrimination of jittered pulse trains,'' J. Acoust. Soc. Am. **39**, 920–928.

Rothgänger, H., Lüdge, W., and Grauel, E. L. (**1990**). ''Jitter-index of the fundamental frequency infant cry as a possible diagnostic tool to predict future development problems,'' Early Child Dev. Care **65**, 145–152.

Titze, I. R., Horii, Y., and Scherer, R. C. (**1987**). ''Some technical considerations in voice perturbation measurements,'' J. Speech Hear. Res. **30**, 252–260.

Tsukamoto, T., and Tohkura, Y. (**1990**). ''Perceptual units of the infant cry,'' Early Child Dev. Care **65**, 167–178.

Tsukamoto, T., and Tohkura, Y. (**1992**). ''Tempo as a perceptual cue for judgement of infant cries,'' Percept. Motor Skills **74**, 258.

Veldhuis, R., and He, H. (**1996**). ''Time-scale and pitch modifications of speech signals and resynthesis from the discrete short-time Fourier transform,'' Speech Commun. **18**, 257–279.

Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T., and Valanne, E. (**1968**). ''The infant cry: A spectrographic and auditory analysis,'' Vol. 29 of *Clinics in Developmental Medicine* (Spastics International Medical).

Wiedenmann, G., and Todt, D. (**1990**). ''Discrete responses of adult subjects during exposure to infant cry sequences,'' Early Child Dev. Care **65**, 179–188.

Wilde, M. D., Martens, W. L., Hillenbrand, J. M., and Jones, D. R. (**1986**). ''Externalization mediates changes in the perceived roughness of sound signals with jittered fundamental frequency,'' in *International Computer Music Conference* (Computer Music Association, San Francisco, CA).

Wolff, P. H. (**1969**). ''The natural history of crying and other vocalizations in early infancy,'' in *Determinants of Infant Behavior IV*, edited by M. Foss (Methuen, London).

Zeskind, P. S., and Huntington, L. (**1984**). ''The effects of within-group and between-group methodologies in the study of perception of infant crying,'' Child Dev. **55**, 1658–1665.

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

Zeskind, P. S., and Marshall, T. R. (**1988**). ''The relation between variations in pitch and maternal perceptions of infant crying,'' Child Dev. **59**, 193–196.

Zeskind, P. S., and Collins, V. (**1987**). ''Pitch of infant crying and caregiver responses in a natural setting,'' Inf. Behav. Dev. **10**, 501–504.

Zeskind, P. S., Klein, L., and Marshall, T. R. (**1992**). ''Adults' perceptions of experimental modifications of durations of pauses and expiratory sounds in infant crying,'' Develop. Psychol. **28**, 1153–1162.

Zeskind, P. S., Wilhite, A., and Marshall, T. R. (**1993**). ''Infant cry as a graded signal: Experimental modifications of durations of pauses and expiratory sounds alter mothers' perceptions,'' Early Dev. Parenting **2**, 99–105.

Zeskind, P. S., Sale, J., Maio, M. L., Huntington, L., and Weiseman, J. R. (**1985**). ''Adult perceptions of pain and hunger cries: A synchrony of arousal,'' Child Dev. **56**, 549–554.