An assessment of the perceptual role of individual acoustic features of infant cries

by
Athanassios Protopapas
B.Sc., University of Patras, 1991

Thesis
Submitted in partial fulfillment of the requirements for the
Degree of Master of Science in the Department of
Cognitive & Linguistic Sciences at Brown University

May 1993

** CONDENSED TYPEFACE – TWO COLUMN VERSION **

Contents

Acknowledgements

I wish to express my gratitude to Dr. Peter Eimas for directing my work, and for helping me clarify many methodological and experimental design issues. This study would certainly have not been possible without his expert guidance. I want to thank Dr. Barry Lester, who commented on an earlier manuscript, and whose experience in infant cry research has been an invaluable resource. I had the pleasure of participating in a number of enlightening discussions with Dr. Barry Lester and Dr. Zachariah Boukydis, which contributed a great deal in improving my way of thinking about the important issues. My thanks to Dr. James Morgan, who also read and commented on an earlier draft of this study. Finally, I wish to express my deepest appreciation for my colleague Tecumseh Fitch, whose work inspired this study, and who has always been available for help and for the most constructive of criticisms.

Introduction

Infant crying has long been a field of study. The communicative and diagnostic values of infant cries are now recognized. Acoustic analyses and perception experiments have been performed in an attempt to "decode" the message of the cry. However, it was not possible, until recently, to isolate individual acoustic features of the cry signals and their respective significance. Given major technological advances over the past decade, sophisticated signal processing techniques are now feasible and relatively easily implemented in modest equipment.

The purpose of this study was to assess the role of various acoustic parameters on the perception of infant cries by adults using computer generated cries. Based on a number of real cries, LPC analysis was used to create a set of parameters for each cry, which were then altered in a controlled way to synthesize artificial cries. The synthesized cries were rated by adult listeners on a number of scales that have been shown to differentiate the perceptual dimensions of infant cries. The correlation of the individual acoustic features with perceived qualities provides a reliable assessment of their perceptual role.

The theoretical questions that are central to my thesis concern the perceptual differentiation of infant cries according to acoustic features and the specific messages such differentiations convey. A closely related aspect, which is beyond the scope of this present study, is whether specific variations of the cries in terms of acoustical features can be reliably related to corresponding states or needs of the infants. I am concerned with the meaning that the cries have for the perceiving adults, not with the meaning they might have for the emitting infant, although evolutionary considerations necessitate to argue for a strong correlation between the two.

It is important to make the distinction between the different stages and the causal relations involved in the cry-production and cry-perception process: First, there is the internal or external stimulus that interacts with the infant's state and is the (major) cause of the cry. Then there is the neural and motor coordinated activity that produces the cry (sound)[1], which has certain acoustic features. These features may or may not be related to the cry-eliciting stimulus, to the infants prior state and to the infants general (clinical or other) condition, dispositions, etc. On hearing the cry, adults experience a certain emotion, often compelling, related to an inferred condition of the infant, whereby they often act upon, according to a number of factors, also outside the scope of this study.

Assessment of the importance of particular aspects of the cry acoustics should be a significant contribution to the general understanding of the function of infant crying, because it will improve our understanding of the interaction between the human infant and its caregivers. The research should also aid discovering the relation between auditory perception and emotions, if a correlation between acoustic features in perception and activation of emotion is found. Furthermore, studies should be prompted to examine the cry production of perceptually significant acoustic features, relating them to the state of the infant or to particular stimuli, contributing to the clinical value of infant cry research and improving the diagnostic potential of cry analysis.

With respect to methodology, this study addresses a number of issues, mainly concerning the validity of using computer processed cries for cry perception experiments. I show experimentally that the re-synthesized cries sound natural to adult human subjects, therefore their use in such studies is justified. I also discuss some problems that one encounters when trying to create such artifi-

---

[1] This sound signal can be captured on storage media independently of other reactions or events, and it contains all the information I am concerned with.

cial stimuli, and suggest solutions for the appropriate types of processing. I raise technical issues, including the problem of free-field vs. headphone presentation, which has attracted some attention in other domains that use synthetic audio stimuli, and I demonstrate how an inappropriate choice might seriously affect the outcome of the experiment.

The first chapter begins with a brief overview of the infant cry literature, noting some of the theories that have been proposed for the production, physiology, developmental course, and for the social and medical utility of the cry. I discuss the cry acoustics, which have been extensively described in the past, placing emphasis on the features in which I am primarily interested. In addition, I review the cry perception literature, where my choices for which acoustic features to examine are justified, and conclude with a closer look at the research on the perceptual role of these features.

The second chapter contains a discussion of the digital signal processing techniques I use and the reasons for their choice, followed by the description of my processing scheme. There is also a discussion of possible improvements for a more automated procedure. The third chapter describes the experiments on the perception of the cries and discusses the significance of the experimental data.

The infant cry

The reasons for crying

In order to assess the perceptual function of infant crying as completely as possible, one should be concerned with the causes of the cry. The communicative or other values of the cry can be understood only with respect to the physiological reasons which give rise to cry behavior.

The psychological and physiological reasons for crying in infants have long been debated. The complexity of the issues involved is now realized (Lester, 1985). Although it is commonly argued that crying is usually an expression of discomfort or distress (e.g., Dunn, 1977), the functional role of crying is still unclear, and a number of theories have been proposed. Since crying is one of the infants' few ways of communicating its needs, any attempt to explain it should consider the evolutionary advantages that drove the development of crying behavior.

The most well documented cause of crying is physiological pain. Crying has often been considered as an indication of pain (Anand and Hickey, 1987; Johnston and Strada, 1986) and it has been suggested that the cry can be used for evaluating the actual degree of pain in the infant (Beyer and Wells, 1989; Levine and Gordon, 1982; Fuller et al., 1989). However, Owens (1984) has pointed out that crying is not an adequate measure of the degree of pain; other measures as well as context need to be taken into account.

In fact, it is not yet clear what the meaning (both neurologically and psychologically) of pain in infants is (Owens, 1984). Nevertheless, the "pain cry" has been extensively described in the literature, often as a distinct "cry type"[2]. With the exception of the oxymoronic "pleasure cry", which occasionally appears in the literature (Wasz-Höckert et al., 1968), almost all documented cries result from physical discomfort. However, in the cases of "colicky" infants, it seems that crying might be the expression of a heightened sympathetic/low parasympathetic reaction in the hypothalamic and limbic circuits, which triggers the vagal complex and results in gastrointestinal activity as well as in excessive crying (Lester et al., 1992). In other words, crying might in some cases be a result of neural "mistuning", in addition to being an indicator of physical discomfort or need. Developmental factors must also be taken into account, because crying changes with age (Fisichelli et al., 1974; Colton et al., 1985; Zeskind, 1985).

Lester (1984) has proposed a biosocial model of infant crying, where the physiology of the motor parts and the neural parts is integrated with developmental factors in a social context . In this model, the interactions between the infant and the caregiving environment are considered. It is argued that the feedback from the environment given to the infant plays an important role not only in the general developmental course, but also in the crying behavior *per se.* Crying is also examined as a facilitator of social interaction that can promote growth and development. The model takes into account the gestational age of the infant, measures of cry acoustics, home environment factors, and mother-infant interaction factors, scored on categories such as body position, head position, facial expression, vocalizations etc. The predictive output of the model is correlated ($r^2 = 0.72$) with Bayley Mental Developmental Index scores at age 18 months.

---

[2] discussed in detail under "cry types"

The production of the cry

In order to evaluate the different acoustic parameters of the infant cry, it is also important to understand how the cry is produced. Naturally, all the systems involved in the cry production can influence the acoustics of the cry, and perhaps the perceptual impression it can make. In other words, the perceptual meaning of the cry is affected not only by the infant's state which is the main reason for the cry, but also by the systems that interfere with or participate in the emission of the sound.

Although there is now some general understanding of the cry physiology, there are still many unanswered questions, mainly as to the neural mechanisms that underlie and result in the actual crying behavior. Crying involves delicate coordination of a number of muscles and neural events, including the laryngeal, supralaryngeal and respiratory muscle groups and the vagal neural complex. It is generally accepted that the cry of the newborn is involuntary and reflexive (symptomatic). Control of cry production mechanisms develops with age, so that infants can vocalize voluntarily after a few months, presumably to convey specific messages or to signal their condition to their caregivers. The sophistication of the vocal signs of infants gradually increases to reach symbolic-linguistic level after several months (Nakazima, 1980).

The physiological and motor processes, in terms of respiration and articulation, that produce the cry sound have been described from the perspective of a pre-linguistic (nonverbal) tuning of vocalization (Lieberman, 1985). Golub and Corwin (1985) proposed a "physioacoustic model" of crying in infants, which emphasizes the neural mechanisms that control the motor behavior, but their 3-component approach has been challenged on the basis of more recent statistical analyses (Green and Gustafson, 1990). We probably have to integrate information from both the motor and the neural approach to fully understand the physiology of crying.

The information in the cry

As a result of the neural and muscular activity that produce the cry, in conjunction with the state that caused it, the resulting sound may convey specific messages, which reflect the situations in all systems involved. Such messages have been found to be multi-dimensional, and of great potential value to the infant's caregivers. They can be classified in three generic categories:

Predictive An interesting aspect of the infant cry sound is its potential predictive value in terms of developmental outcome, as it is measured by mental (cognitive) and motor indices, such as the Bayley Scales of Infant Development, the McCarthy General Cognitive Index, the Psychomotor Developmental Index, the Mental Developmental Index etc. A positive correlation has been shown to exist between acoustic features of the cry signal (e.g. mean fundamental frequency and its variations) and scores in tests taken several months later (Lester, 1987; Colton et al., 1985; Grauel et al., 1990; Fitch et al., 1992), or subsequent developmental problems, including developmental delays and severe mental retardation. However, because of the small samples and few number of studies, these results are not yet very robust.

Diagnostic Since a cry sound may reflect some of the anomalies in all the cry-production stages, it is expected that some aspects of the infant cry sound can be directly related to neurological deficits or damage, as well as to genetic anomalies and respiratory problems. This has been demonstrated for a number of deficits and abnormalities in the infants' neural and respiratory systems (Fisichelli et al., 1966; Frodi and Senchak, 1990; Lüdge and Rothgänger, 1990; Mende et al., 1990), as well as in the physiological system of cry/speech production (Hirschberg (1990) showed an acoustic manifestation of pathologic larynx cases in infants' cries). Lester et al. showed that cocaine exposure in utero appears to affect the acoustic characteristics of the newborn infant cry, although it is not clear whether this is directly connected with any future developmental problems.

Other researchers (Wasz-Höckert et al., 1968; Golub and Corwin, 1982; Wasz-Höckert et al., 1985) have described the differences between the cries of problematic (abnormal) infants and those of normal infants. In an interesting evolutionary perspective, these cries are usually "incorrectly" perceived, in the sense that, because of their abnormal quality, they do not elicit the appropriate actions from caregivers (Zeskind, 1980), thus jeopardizing the infant's survival.

Cries of preterm infants have also been reported to differ from those of full-term infants (Michelsson et al., 1982), and cries of under- or overweight infants have been found to differ from those of average weight infants (Zeskind, 1981; Zeskind and Lester, 1981). Some studies suggest, however, that it is the risk status[3] of the infant rather than the gestational age at birth that accounts for the difference (Friedman et al., 1982). This possibility is even more interesting, because, if adequately supported, it could enable pediatricians to perform reliable assessments of the risk status of infants based on objective measurements of acoustic features, and take appropriate preventive action.

Communicative Besides being an invaluable tool for the pediatrician, the cry is believed to be primarily a means of communication for the infant. Particularly in modern western societies, where infants are physically separated from their mothers immediately after birth, and excessive contact is often discouraged, the infants, being physically isolated in strollers or playpens, have no alternative other than vocalizations, in order to signal their needs. In many other cultures (e.g. most primitive ones as well as modern African and east Asian), infants are in constant physical contact with the mother (i.e., always carried) for a very long period in life (Konner, 1972; Mead and Newton, 1967). In such cultures infants primarily utilize other means to communicate, mostly subtle motor signals, and cry is considered an "emergency signal"[4]. Still, even as a signal, it is far from simple, and particular variations may convey different messages.

A number of studies have demonstrated the great communicative values of the infant cry signal. Although there is a high degree of individuality in the cries of different infants (Gustafson et al., 1984), so that parents or primary caregivers can tell the cries of their own babies from the cries of others (Green and Gustafson, 1983; Cismaresco and Montagner, 1990), enough features are similar between infants to provide sufficient cues for evaluating the infant's condition. For example, Porter (1986) found that the judgment of adult subjects on the severity of the infant's condition based on the recorded cry alone accurately matched the invasiveness of the procedure the infant was undergoing at the time. Other studies have investigated the differences between situationally defined cry types (Wasz-Höckert et al., 1968; Brennan and Kirkland, 1983; Johnston and O'Shaughnessy, 1988; Fuller et al., 1989; Fuller, 1991).

The potential of infant crying, or other types of vocalizations, to directly elicit or affect social interaction is very significant, as studies of adults' responses to cries and smiles of infants have shown (Frodi et al., 1978). Non-cry vocalizations are also powerful in eliciting a consistent response, as Shimura et al. (1992) have demonstrated. Lester's (1984) model of infant crying also considers this aspect of the communicative power of crying. More recently, Lester et al. (1993) argued that the goodness of fit between the characteristics of infants' cries and their mothers' perception of the cries is a good predictor of developmental outcome measures, indicating the importance of crying as an efficient means of communication for proper development.

Naturally, validation of the communicative value of the cry relies upon the correlation of the cry characteristics with specific situations or intended messages from the infant's point of view, as well as upon the ability of the caregivers to make perceptual distinctions based on it. The assumption that such perceptual distinctions can be made, which is critical for this study, is well supported in the literature, and I return to this later, when I talk about cry perception.

The cry types

Many researchers have relied upon the existence of distinct "cry types" in their studies. Whether such types actually exist is still under debate. It is common, in the infant cry literature, to define and use particular cry types, usually in a situational context, i.e. with respect to temporally proximal stimuli or to a condition inferred from other observations. For example, pain cries have been defined as the vocalizations (of previously quiet infants) following presumably painful stimuli, such as snapping the sole of the foot with a rubber band, pinching the arm, pulling hair, snapping the ear, as well as some medical procedures, including heel-stick procedures, routine immunizations, and invasive surgical procedures (Wasz-Höckert et al., 1968; Fuller and Horii, 1986; Fuller and Horii, 1988; Fuller et al., 1989; Fuller, 1991; Golub and Corwin, 1982; Michelsson et al., 1982; Zeskind and Lester, 1981; Porter et al., 1986; Johnston and O'Shaughnessy, 1988). Hunger cries have been defined as the spontaneous vocalizations of presumably otherwise comfortable infants that occur some time after the normal feeding time and cease

---

[3] Risk status is defined in pediatrics with respect to a number of obstetric conditions, including the social-economic and health status of the mother as well as parturitional factors (Zeskind, 1980).

[4] Quite interestingly, it has been observed in foraging people's communities that hunger cries attract only the infant's primary caregivers, whereas pain cries may elicit an immediate response from the entire village (Konner, 1972)

when feeding begins (Wasz-Höckert et al., 1968; Fuller and Horii, 1986; Fuller and Horii, 1988), or as the vocalizations occurring when feeding is suddenly interrupted after a few seconds and terminating when feeding is resumed (Müller et al., 1974; Murry et al., 1977).

Although pain and hunger are the most frequently designated "types", other types have been defined in an attempt to judge the possible differentiation of acoustic or perceptual features as they relate to the cry-eliciting situation. Indeed, a significant correlation has been found in some studies between the assumed cry type and specific acoustic features, such as maximum, minimum, mean, and range of fundamental frequency, melody, spectral energy distribution, duration, position and bandwidth of formants, etc. (Wasz-Höckert et al., 1968; Wolff, 1969; Freeburg and Lippman, 1986; Fuller and Horii, 1986; Fuller and Horii, 1988; Johnston and O'Shaughnessy, 1988; Fuller, 1991) Based on such observations, acoustic predictors have been developed for automatic classification of cries, often with good results, although this depends heavily on the definitions of types and the cry samples used (Wasz-Höckert et al., 1968; Cohen and Zmora, 1984; Fuller, 1991). Lundh (1986) proposed a baby-alarm for hearing-impaired parents, which alerts the parent only when the cry is considered to be the result of an aversive condition, (basing the judgement on the tenseness of the sound,) making a distinction between "happy", "weeping", and "distressed" cry types.

On the other hand, there are a number of studies that consider the cries as graded signals, and ask subjects to rate them on descriptive perceptual polar scales, such as distressing/non distressing, arousing/soothing, urgent/not urgent, sick/healthy, etc. (Brennan and Kirkland, 1982; Zeskind, 1983a; Zeskind and Huntington, 1984; Zeskind et al., 1985; Zeskind and Marshall, 1988; Brennan and Kirkland, 1983; Brennan and Kirkland, 1985; Porter et al., 1986; Green et al., 1987; Lester et al., 1989) or on pairwise similarity, i.e. how similar or dissimilar do two cries sound (Green et al., 1987). Subsequent analyses in these studies reveal a number of dimensions along which the cries vary either acoustically or perceptually or both. Unfortunately, these dimensions are usually reported in terms of statistical factors and their respective loads on the different acoustic features, and they are not very illuminating.

The theoretical approach to the subject, I believe, clearly favors the latter idea, namely, that cries are graded and not categorical signals. All models that have been recently proposed (Murray, 1979; Lieberman, 1985; Golub and Corwin, 1985; Lester and Zeskind, 1982; Lester, 1984; Fuller, 1991), irrespectively of their particular subdomain, consider the infant cry as a graded signal, indicating either some kind of Central Nervous System level of arousal or some other stress-arousal type gradation. Gustafson and Harris (1990) compared the two approaches and found that subjects were generally unable to categorize cries correctly by eliciting situation, yet showed very good discrimination along a perceptual continuum of discomfort. Similarly, Papoušek (1989) concluded that "gradedness may be more relevant". It is generally accepted now that infant cries are not categorical signals. Therefore, the question arises, whether one is justified to talk about situationally defined cries, or even, about cry types.

An important point against the use of situationally defined cry types is that the acoustics of cries change considerably over time (Zeskind et al., 1985; Porter et al., 1986). For example, pain cries become less aversive as the effect of the painful stimulus wears off, while hunger cries become more aversive as the hunger feeling presumably becomes more intense, and, over time, they start to resemble pain cries (Zeskind et al., 1985). Supporting this idea, Gustafson and Harris (1990) found that initial parts of the cry are better discriminated than intermediate (later) parts. To this point, Johnston and O'Shaughnessy (1988) have suggested that pain might be a *more extreme* rather than a *different kind of* arousal. Rothgänger et al. (1990) arrived more recently at the same conclusion.

It should be noted, however, that gradation of cries over some range along a number of dimensions does not mean that one cannot define categories, nor that there are not cry types that relate certain acoustic characteristics to specific stimuli or eliciting situations. Indeed, for diagnostic, communicative, or research purposes, we might want to define unique cry categories within a narrow range of variability of some acoustic feature or features. Furthermore, if there is a significant correlation between a range along certain acoustic or perceptual features and some cry eliciting stimulus, it might also be useful to consider defining a cry type pertaining to that stimulus.

For example, since hunger is usually less intense than pain (at least initially), it is expected that crying as a result of hunger will in general be different from crying as a result of pain. As long as we keep in mind that the difference is mostly quantitative, rather than qualitative, we may define a "hunger cry" type and a "pain cry" type for identification purposes. Still, it is important to know that the categories may overlap, and that a categorization scheme which fits a given situation might be inappropriate in another.

In this light, I prefer to attempt to reveal differences along perceptual dimensions, rather than in the responses of categorization or identification. Type definitions in the literature are vague and variable and do not lend themselves to useful interpretations. The acoustic differences in the construction of the stimuli, and the experimental method in this study will reflect this point of view on the subject of cry types: graded stimuli (cries) and graded responses (perception).

The cry acoustics

The auditory properties of the infant cries were not possible to quantify until the invention of the spectrograph. Prior to the breakthrough of spectrographic analysis, studies would report findings in terms of musical tones (Gardiner, 1838) or percentages of vowels and consonants (Fisichelli et al., 1966). The widespread availability of the spectrograph and the appropriate mathematical formulations for sound in terms of the frequency domain, enabled researchers to examine the "voiceprint" of the cry sound in great detail.

The acoustic features of infant cries have been described by a number of researchers (Wolff, 1967; Wasz-Höckert et al., 1968; Wolff, 1969; Michelsson, 1980; Michelsson et al., 1982; Michelsson et al., 1983; Murry et al., 1977; Zeskind, 1983b; Thodén and Koivisto, 1980; Thodén et al., 1985). Although the spectrograph has been an invaluable tool for cry analysis, modern methods utilizing the available technology in signal processing enabled researchers to extract quantitative measures of the various parameters and study the acoustics in more detail (Fuller and Horii, 1986; Fuller and Horii, 1988; Johnston and O'Shaughnessy, 1988; Golub and Corwin, 1985).

In summary, the infant cry has the following characteristics:

- With the exception of newborns, where unvoiced cries occur more frequently, voiced cries are the norm. Three modes of phonation have been described in the literature:

  Phonation, when the vocal cords are fully vibrating quasi-periodically, at a frequency roughly between 250Hz and 700Hz.

  Dysphonation, when turbulence noise is generated at the vocal folds and is modulated by vocal fold vibration.

  Hyperphonation, when the frequency of oscillation of the vocal cords shifts to 1000–2000 Hz in a "falsetto"-like vibration pattern.

- The duration of individual cries (from the onset of crying to the end of phonation before inspiration) averages at about 2.5 sec. The cries and the intervals between them are shorter when the infant is apparently in more severe pain.

- The melodic form (F_0 contour) of the cries can be falling, rising/falling, falling/rising or flat. The most common melody type is rising/falling.

- In normal phonation the cry harmonics are distinct and clear. When the cry is dysphonated, however, the harmonics become obscured by noise and, occasionally, disappear.

- The energy of a cry is usually higher at lower frequencies (below 1 kHz). The spectral content above 1 kHz becomes stronger when the infant is apparently in severe pain.

- Occasionally, the cry becomes extremely low-pitched (vocal fry). This occurs at the onset or, more frequently, at the end of a phonation.

- Sometimes parallel lines appear in the spectrogram between the fundamental and its harmonics (double harmonic break). Double series of fundamental frequencies (biphonation) are also possible.

- The amplitude and fundamental frequency of the cry do not remain constant, even between adjacent periods. Their perturbations[5] have the effect of flattening the spectrum making the cry more "noisy".

With such a host of possible acoustic parameters to be analyzed, the task of researchers has become very demanding, and one might wonder whether all the different types of features are really necessary for a good understanding of crying

---

[5] The amplitude fluctuation between adjacent periods of the waveform is called shimmer; the fluctuation of period length is called jitter.

and reliable judgements concerning clinical, developmental, or communication issues. Green and Gustafson (1990) have suggested that there is actually a lot of redundancy in the various acoustic features, so that only four "composite variables" are in fact independent, and therefore necessary for a complete description of a cry. These feature clusters, identified with Multidimensional Scaling (MDS) statistical methods, could provide the basis for a better understanding of the cry mechanisms, in accord with some appropriate theory which has to be developed. However, they do not recommend that the individual acoustic variables be altogether dismissed, because some of them have been shown to be of great diagnostic value in terms of critical values.

The impact of the cry

In order to understand the communicative function of the infant cry, it is also important to understand the mechanisms that elicit the appropriate reactions from the caregivers. However, it is not yet clear how cries affect listeners. Studies have demonstrated that exposure to the infant cry affects the heart rate of the listener (Bryan and Newman, 1988), elicits autonomic arousal in terms of skin potential (Boukydis, 1980), and causes various verbal and behavioral responses (Zeskind, 1980; Frodi and Senchak, 1990), with differential effects at different stages of the maternal cycle (Bleichfeld and Moely, 1984). Of particular interest in these studies have been the cries of atypical infants, because those infants have greater need of attention. Frodi and Lamb (1978) found that the cries of premature infants elicit greater autonomic arousal. Similarly, Zeskind (1987) found that there is greater change in arousal in response to crying of high-risk infants.

Based on the current views on the communicative role of the infant cry, different models for its function have been proposed. Tied with the notion of distinct cry types was the releaser model, which suggested an inherent, physiological response from the listener that was triggered by the cry, and was different for different cry types. This model has been criticized on the basis of more recent evidence (Murray, 1979; Murray, 1985).

The emotion activator model asserts that the cry sound affects the emotional condition of the listener and thereby indirectly causes a reaction. This reaction can be explained with the assumption of either an altruistic (to soothe the suffering infant) or an egoistic (to cease the source of auditory annoyance) motive. Michelsson et al. (1990) found that crying aroused feelings of tenderness or irritation. Either way, an action is taken by the caregiver that aims to stop the crying by eliminating its cause.

In cases of excessive or abnormal crying or nonoptimal social or other parenting situations, the corrective action may sometimes be against the infant in the form of physical abuse (Frodi, 1985). In less nonoptimal cases, caregivers have been reported to avoid the crying infants (Bennett, 1971), thus avoiding exposure to the annoying stimulus[6]. (In fact, Frodi and Lamb (1980) found that child abusers' responses to infants' smiles was similar to their responses to infants' cries, negative in both cases.) This is particularly true in the cases of brain-damaged or asphyxiated infants whose cries have been reported to be more aversive (Boukydis, 1985).

It seems that, although the egoistic motive hypothesis fails to account for most of the caregiving behavior, there is a threshold of tolerance which, when exceeded, gives way to negative behavior. Many studies have shown the compelling effects of the infant cry to adult listeners; more research is necessary to reveal the underlying processes. Knowing which are the relevant acoustic parameters and how they affect perception might improve our understanding of the function of crying.

The perception of the cry

Of critical importance to my thesis is whether adults can differentiate between cries and "read" specific signals the infant might be emitting. Some early studies (Sherman, 1927; Aldrich et al., 1945) failed to support this idea, but they have been strongly criticized for their methodology. More recently, researchers have investigated the ability of adults to categorize cries according to their respective eliciting situations; with one exception, where ill-defined categories were used (Müller et al., 1974), such studies had positive results and demonstrated the effects of sex, training, experience in childcare, and interactions of these on the ability to accurately identify and label cries (Wasz-Höckert et al., 1968; Freeburg and Lippman, 1986; Gladding, 1979). Tsukamoto and Tohkura (1990) found similar effects using a procedure which included a learning session prior to the cry categorization task.

---

[6] Avoidance behavior has also been considered to be a measure of the aversiveness of the cry (Bisping et al., 1990).

Unfortunately, these results are not directly comparable, because of the different methods and stimuli used. The only robust conclusion that can be drawn is that adult listeners can indeed distinguish infant cries and make some kinds of judgments about the situation of the crying infant, although this depends heavily on the kinds of cries used, and there is a considerable overlap between identified categories. Gustafson and Harris (1990) found recently that mothers and nonmothers alike could not perceive "correctly" a cry as originating from pain or hunger, and this result bears on the long-term temporal trends of the cries, as previously discussed.

Cry rating experiments, on the other hand, have utilized bipolar perceptual scales for subjects to rate the cries and have revealed the perceptual dimensions of the cries and how they are affected by age, sex, parity[7], childcare experience, cross-cultural differences, and some of their interactions (Brennan and Kirkland, 1982; Brennan and Kirkland, 1983; Zeskind, 1983a; Brennan and Kirkland, 1985; Porter et al., 1986; Green et al., 1987; Gustafson and Green, 1989; Lester et al., 1989). These results also fail to agree on some issues such as the effect of sex or parity (parents vs. nonparents). The differences in the underlying assumptions and methodologies between experimenters could probably account for such discrepancies. Interestingly enough, the cries of preterm infants and of high-risk infants have been found to consistently differ perceptually from those of term and low-risk infants, respectively (Zeskind and Lester, 1978; Friedman et al., 1982; Zeskind, 1983a).

More specifically, some interesting findings about infant cry perception that illustrate the complexity of the matter include the following: Teenage mothers have been found to be "less accurate than adult mothers in perceiving the characteristics of their infants" (Lester et al., 1989). It has been suggested that "parents are integrating frequency information from across the cry, whereas nonparents focus on information at discrete points in the cry" (Green et al., 1987). Along the same lines, Adachi et al. (1985) found that mothers tend to use different acoustic cues for different rating scales, whereas nonmothers used the same cues for all ratings. Zeskind (1983a) reported cultural differences in the perception of the cries of the infant at risk among multiparous mothers, and Zeskind (1980) observed "more 'affective' responses" in order "to terminate a cry that is more distressing and is coming from an infant who is crying for a more 'urgent' reason and more 'tender and caring' responses to an infant who sounds sick". Kirkland (1990) recently found that the perception of cries is context-sensitive, and that parents' disposition towards their competence to manage the infant directly affects their perception of the cry.

It has been found that "low-risk cries are perceived along a single factor representing the unpleasant quality of the cry" and " a second factor emerges in the perception of the high-risk-infant cry that conveys the 'sick' and 'urgent' nature of the cry sound" (Zeskind and Lester, 1978), and that "different segments of a cry in response to a single stimulus may convey different messages to the caregiving environment", because "a 'pain' or 'hunger' cry is a dynamic signal with varied perceptual meanings" (Zeskind et al., 1985). Also, that "cries of preterms who are ostensibly healthy at expected date of birth can reflect the medical risk associated with their neonatal condition" (Friedman et al., 1982). Boukydis and Burgess (1982) found differences in the perception of cries from temperamentally different infants, which, they suggested, might lead to predictable differences in caretakers' interpretation of and reaction to the cries. Finally, Papoušek (1989) reported differences in the ability to decode differential information in infants' cries between subject groups with different amounts and kinds of experience, so that joy sounds are often misread as discomfort sounds, except by parents of same age infants.

Concerning the methodology of studying infant cry perception, studies have cautioned researchers, among other things, about the effects of typicality of the experimental cries and of composition of the stimulus tapes[8] on discriminability (Freeburg and Lippman, 1986), and about the effects of subject training in cry identification (Gladding, 1979). With respect to the presentation design, Zeskind and Huntington (1984) found that "within-group methods of cry presentation accentuate the perceptual distance among cry types and may actually create many reliable differences that would not be found in between-group comparisons". Furthermore, Bisping et al. showed that information given to the subjects about the infants health status influences their responses.

As already discussed, it is reasonable to assume, that a specific situation (e.g. frustration from being restricted, or startle from a loud noise, or pain

---

[7] parents vs. nonparents and primiparous vs. multiparous.

[8] Single cries vs. naturally occurring sequences

from a heel-stick) will often result in similar levels of arousal and thus to similar-sounding cries. Although it would not be possible to distinguish between two cries that are the result of different stimuli, when they cause the same amount of distress, it should be possible to distinguish cries provoked by mild stimuli from those provoked by intense stimuli. Such an approach might well explain both the positive and negative results that have been at times reported on cry perception. It might also help to put both the categorization and the rating studies in a common perspective.

Caution must be exercised, however, not to mistake cry perception as being unidimensional, i.e., on a level-of-arousal or any other scale. Some studies have found the perceptual space of infant cries to be at least three- or four-dimensional (Brennan and Kirkland, 1985; Green et al., 1987). Because of the large number of factors that can influence the acoustics of cries and, subsequently, cry perception, and because there are many audible acoustic parameters that can be altered in many ways, it is likely that there is more than one perceptual dimension. The dimensions uncovered from the statistical analysis of the experimental data are not very enlightening or intuitive (arbitrarily named as "Affect"—"Potency"—"Evaluation" (Brennan and Kirkland, 1985; Brennan and Kirkland, 1983) or merely as first, second, etc. dimension (Green et al., 1987)), but the calculation of the load distribution of specific rating scales on each dimension might give some clues about the meaning of such dimensions.

The relation of acoustics to perception

If listeners can differentiate cries on the basis of the acoustic signal alone, then it follows that there is adequate information in the signal to justify their judgement. Wiedenmann and Todt (1990) argued, on the basis of their direct-on-line-response (DOL) data, that there must be something in the acoustics of the cries that prompts reaction, otherwise their subjects' responses would not have been concentrated at particular moments of the cries, but they would have been uniformly distributed along the whole cry. Many studies have attempted to isolate the acoustic correlates of cry perception, mainly by statistical analysis of the cry properties that are correlated with the listeners' ratings.

Some acoustical features that have been reliably established as playing an important role in cry perception are the maximum, minimum, average, and range of fundamental frequency, the melodic form and the noise content of the cry, and the duration of the individual expirations and inhalations. Many such studies were reviewed by Boukydis (1985). Specifically, Bisping et al. (1990) found that "a higher fundamental frequency in natural cries is perceived as more aversive than is a lower fundamental frequency" and that "the aversiveness of high frequency cries is reduced in the presence of a more complex spectrum". Frodi and Senchak (1990) found that "infants with unusually high-pitched cries tend to elicit less optimal responses from adult listeners". Zeskind and Marshall (1988) found that "increases in the peak, as well as the mean and variability of the fundamental frequency of crying were reliably related to increases in how urgent, distressing, arousing, and sick the cries were perceived to be by multiparous mothers" . Gustafson and Green (1989) reported that "the acoustic features most strongly related to the adults' ratings were: the duration of the cry, the amount of dysphonation, and the amount of energy in low and high frequencies" and that "it appears that all the ratings may be measures of a single underlying dimension" . There is also some evidence for the importance of the formant structure (Johnston and O'Shaughnessy, 1988), the segmental duration, or "tempo" (Tsukamoto and Tohkura, 1992), and the VP ratio[9] of the cries (Okada et al., 1987; Adachi et al., 1985).

Many studies (mentioned in previous sections) have attempted to correlate the acoustics of the cry to the infant's situation, rather than to their perceptual effects. Arguably, it might be more important to judge an infant's condition from the acoustic features of its cries. Unfortunately, in most cases, one cannot precisely define the infant's condition, but rather, must infer it from other external observations. Moreover, because of the great variability in the cry features, and because of the differences between infants, it is harder to extract any meaningful data from such analyses, especially without any established standards for the cry-eliciting stimuli and for the measurements.

Of the studies that have attempted to discover the correlations between acoustics and perception, a few have attempted to alter the acoustics of natural cries by using computer synthesized or simply modified stimuli for perceptual experiments. Specifically, the Japanese group that reported that

---

[9] Voiced Phonation ratio: the ratio of the frames of voiced phonation to the total frames of phonation (Okada et al., 1987).

the VP ratio of cries has a strong effect on the subjects' perceptual ratings (Okada et al., 1987) used the LPC method to create their stimuli. They used the residual error signal to synthesize most parts of their signals, which means that they recovered exactly the original analyzed cry in those parts, and substituted noise for the excitation signal for some parts of the cries, in order to alter the VP ratio. They reported that their stimuli sounded very natural, although they did not test this claim experimentally.

Bisping et al. (1990) used synthesized cries to alter the fundamental frequency and the melodic structure of the cry. Specifically, they used an interpolation procedure to create the artificial cries with a precisely controlled difference in fundamental frequency. Differential interpolation within a cry resulted in alterations of the melodic form. Superimposing an increased-frequency cry rendered a biphonated cry. Not surprisingly, significant perceptual differences were found for all three variables tested. Unfortunately, there are many technical issues involved in such a manipulation, and although they reported that their stimuli sounded natural, it is clear that the values of fundamental frequency, and, more importantly, of formant frequencies that they used were not within the normal range of infant cries. The major problem arises from allowing the formant frequencies to be scaled along with the fundamental, which is unavoidable when simple interpolation of the waveform is the method of choice. It is well known that the fundamental frequency may vary within an extensive range, because it is mainly the result of the interactions of subglottal (lung) pressure and laryngeal muscular tension. The range of formant frequencies, however, is determined by the size and flexibility of the vocal tract, and for very young infants, whose larynges have not yet descended to their mature positions, it is quite restricted (Lieberman, 1980; Lieberman, 1985). Consequently, doubling or halving the formant frequencies of an infant cry is effectively similar to doubling or halving, respectively, the length of the infant's vocal tract, and it is unclear to me how such a manipulation could produce naturally-sounding infant cries.

Furthermore, because of the simple signal processing method that was used in that study, it was impossible to completely isolate any single acoustic parameter. Altering the fundamental frequency also affected the whole spectrum of the cry, including the formants (position and bandwidth) and the energy distribution in specific bands, all of which have been connected with perceptual differences, as already discussed. The advantage of the Linear Predictive Coding Method I used is that individual parameters can be independently manipulated. This way, the effects of any single feature can be identified and the effects of feature interactions can be separately controlled. It is also possible to investigate the role of spectral properties that are not directly related to the fundamental frequency, as well as other features, such as noise content and short-time fluctuations.

The potential value of perturbations

Because of the effects of stress/arousal situations on the respiration and muscle tension, it was expected that such situations would be reflected in phonation and articulation in adult speech. Sherer (1981) justified theoretically the existence of vocal indicators of stress or other emotional conditions. . Some studies even aimed at practical applications, such as in determining pilot stress through voice analysis (Kuroda et al., 1976).

The ability of listeners to perceive differences and actually identify emotional conditions from speech samples has been documented (e.g., Lieberman and Michaels, 1962). Williams and Stevens (1972) described some effects of the emotional state on the speech signal parameters. Sherer (1986) proposed a model which addressed the situation from a physiological point of view, including a list of emotional states and their effects on vocal features. He and others have considered the fluctuations and perturbations in the fundamental frequency and the amplitude of the speech (tremor, jitter, and shimmer, respectively) to be likely indicators of vocal stress.

Voice tremor has been positively linked to stress in adults (Brenner et al., 1979; Inbar and Eden, 1976; Smith, 1977) and also in children (Wiggins et al., 1975). Jitter and shimmer have also been fairly well studied in adult speech [10]. It turns out that vocalization signals naturally exhibit a great amount of short-time variation (Lieberman, 1961). Such variation is of great interest, because it has been linked to perceptual features, such as hoarseness and nasality of speech in adults (Muta et al., 1988; Wolfe et al., 1991) and in children (Zajac and Linville, 1989). Jitter also has potential diagnostic value for some pathological cases of the larynx in adults (Lieberman, 1963) and in infants (Hirschberg, 1990) and for Central Nervous Sys-

---

[10] Wong et al. (1991) proposed a vocal fold model which accounts for perturbations.

tem disorders in infants (Mende et al., 1990; Lüdge and Rothgänger, 1990). Furthermore, and perhaps more interestingly, it might be related to the emotional content of speech (via the emotional condition of the speaker) (Lieberman and Michaels, 1962) and to the general level of arousal in the Central Nervous System. A number of methods have been proposed for measuring jitter (Titze et al., 1987; Milenkovic, 1987; Hollien et al., 1973; Deem et al., 1989).

Given that according to the current models of infant crying, the cry is a graded signal that indicates the infant's state of arousal, and that human listeners are able to consistently rate cries along perceptual continua, it is only reasonable to assume that there must be specific cues in the cry signal that vary according to level of stress (or distress), which can be reliably picked by human listeners and/or computer signal processing software (because some automatic classifiers have been relatively successful). In the light of the observations on pitch perturbations and their relation to stress-arousal situations in adults and children, we might expect to find similar correlations of the jitter in infant vocalizations with the infants' stress-arousal situations. Such an acoustic cue might be naturally utilized by human listeners when they make judgements about infant cries; or computer programs could be written to perform specific functions based on some related acoustic measure.

Recently, a correlation was found between stimulus intensity and jitter in the infant's cry (Fitch et al., 1992). Although Fuller and Horii (1986) found no differences in jitter between "different types of infant vocalizations", Fitch et al. pointed out that it might be that the temporal resolution of their analysis was inadequate[11]. It could also be that their formula for calculating jitter or the segments that were selected for the analysis were not appropriate. As has been often noted in the literature, there are many ways of defining and calculating jitter (Pinto and Titze, 1990) and results from different studies are not directly comparable (Karnell et al., 1991). Furthermore, jitter was considered in that study to be "an inaudible variable", although it is has been demonstrated that jitter is audible, and that the threshold for detecting it is very low (Pollack, 1968; Rosenberg, 1966).

Conclusion

I have briefly reviewed the infant cry literature, examining the various stages of crying. The current state of research on why and how the infants cry, what kind of information might be contained in the cry, and how the cry affects and is perceived by adults was presented, all in relation to the cry acoustics, which are of major interest for this study. For a valid assessment of the value of any acoustic parameter in a more broad context, it is critical to understand all the stages involved, and how each one might influence the acoustics.

The most well documented acoustic feature with respect to its perceptual effects in adults is the mean fundamental frequency of the cry. The most unlooked at, yet very promising because of other research in related fields, is jitter (perhaps perturbations of F_0 in general). Another very interesting feature for which very few data exist is the long-term variation of the fundamental frequency (melodic contour), and in particular the time it takes for the F_0 to reach its maximum, known as rise time. I performed my experiments with these three parameters. This way I had a measure of the validity of my method, in relation to the existing literature (by comparing my results on the effects of fundamental frequency to what is more or less known), and, at the same time, I was able to make a novel contribution to our understanding of cry perception. Future experiments might be aimed at assessing the role of the spectral structure (including formant frequencies and tenseness), and durational aspects, which are somewhat documented, although their perceptual significance is rather vaguely described.

The signal processing

In order to be able to precisely control individual acoustic features of cry signals, a parametric description of the sound is necessary. As a result of extensive research in signal processing, sophisticated methods have been developed for parametric representations of sound and speech in particular. Although none of these methods have been applied to infant cry signals before[12], the similarity of the physiological basis of crying to that of speech, in terms of modes of excitation and resonating chambers, allows us to expect that models derived for analysis and representation of speech

---

[11] This is in agreement with the conclusion of Titze, Horii and Scherer (1987) about temporal accuracy in jitter measurements.

[12] Okada et al. (1987) used LPC analysis to process their cries, but they kept their manipulations to a minimum.

might apply to crying as well. This assumption can be marginally justified by the theoretical description of the methods, but must ultimately be experimentally verified.

Digital speech processing

The existing speech processing algorithms can be divided into those that are mathematical descriptions of the speech waveform, possibly taking some advantage of its particular characteristics, such as quasi-periodicity or slow rate of change, and those that are based on models of speech production or communication. Another distinction is usually made between time-domain algorithms, i.e., formulations with time being the independent variable, and frequency-domain algorithms, which have frequency as the independent variable. An introduction in speech signal processing can be found in Saito and Nakata (1985). Rabiner and Schafer (1978) offer a more thorough and detailed coverage of the field.

All algorithms and methods mentioned here are described in the digital domain (as opposed to analog). Although some can be implemented in analog designs, the more sophisticated methods can be implemented only in microcomputer applications. Before any digital processing can take place, the analog sound signal must be sampled and quantized, thus converted into computer readable form. Sampling and quantization are now easily implemented with off-the-shelf products and will not be of concern in this study. It will be assumed that the conversion to digital is optimal, i.e., proper filtering etc. has been correctly performed.

The simplest methods for digital signal processing were developed with radio or telephone communications in mind, and only aim to compress the signal, so that it takes less space to store, and less time and bandwidth to transmit. Such methods usually work on any signal and do not exploit the special characteristics of the speech waveform. One step up in complexity, speech analysis schemes such as filter-bank designs, phase vocoders, and channel vocoders, take into account the bandwidth and the quasi-periodic nature of speech to create frequency-domain representations. More sophisticated, the Linear Predictive Coding method is based on modeling of the human vocal tract. Finally, phonemic methods, which code the phonemes contained in speech segments, are only concerned with conveying the lexical information of the speech signal and are successfully used only for speech synthesis applications.

The choice of LPC

Of the many existing algorithms and formulations, I chose to use the Linear Predictive Coding (LPC) method for my study (Makhoul, 1975; Markel and Gray, 1976). The factors taken into account for the choice of LPC over more intuitive frequency domain representations[13] were basically two: applicability and ease of use. According to the rationale of this study, I needed an analysis method that would provide independent measures of the different acoustic parameters of interest and that would allow each of them to be controlled independently from the others during re-synthesis. The acoustic parameters I was interested in were pitch[14], jitter, and rise time, all of which have to do with the fundamental frequency (F_0) of the signal. It was therefore of primary importance that F_0 be a self-contained part of the analysis, with as few interactions with the other parameters as possible[15].

There are many methodological advantages to the use of LPC. It offers complete separation of pitch and spectral envelope, so that each can be independently manipulated. LPC coefficients can be interpolated without creating instabilities in the filter[16], so that changes can be introduced easily and without problems. It is possible to calculate many sets of parameters from the LPC coefficients, including formant frequencies, filter poles, and vocal tract area function. Thus, it is possible to alter cries directly at any level we wish, or to create sets of parameters with intuitive meanings (for example, the vocal tract area function, which approximates the shape of the vocal tract during phonation, or the log power spectrum, which shows the distribution of energy in the different frequencies). The power spectrum can be computed directly from the filter coefficients (independently of pitch) and manipulated by altering peaks and

---

[13] LPC is considered to be a time-domain representation, because it models the speech waveform itself rather than its spectrum (Atal and Hanauer, 1971).

[14] In the digital signal processing literature, “pitch” refers to the frequency of lowest-order periodicity of the signal, which, for speech and similar signals, is the fundamental frequency of phonation. I adopt this terminology in this chapter.

[15] Naturally, it is impossible not to have any interactions between F_0 and the formant frequencies, because F_0 determines the exact position of the formants within a narrow range specified by the spectral envelope, but this can be safely ignored in most cases.

[16] Actually, the PARCOR coefficients are interpolated, but they are equivalent to the filter coefficients and, often, a by-product of the analysis.

bandwidths. Finally, LPC is relatively simple to program and test, the procedures are well documented, and many algorithms have been developed that exploit the special nature of LPC formulations to yield fast solutions.

Besides, LPC is the only digital signal processing technique available today which is tractable and implementable without the need for unrealistically long computation times and awkward restrictions. It is a well tested, widely applied method, and researchers have acquired extensive experience with the behavior and the requirements of the algorithms. Other methods for speech analysis might have advantages in special areas. For example, the time-dependent discrete Fourier transform representation provides a more accurate spectral description, but does not separate pitch from the spectral envelope. Channel vocoding lends itself to spectral modifications with simple operations, but its spectral resolution is very crude. Homomorphic processing retains more spectral information while still separating excitation source from filter, but it suffers from similar problems as LPC, and is more complex mathematically. In general, LPC seemed the best candidate for the kind of signal processing required for this study. However, it remained to be seen whether it would be appropriate.

Potential problems with LPC

With respect to the applicability of the LPC method to infant crying, there might be problems arising from the high pitch of infant cries, from the phonation modes that contain both a periodic component and noise, and from the nasal coupling in many cries.

LPC has been very successful in applications involving adult male speech. It is not as good with female speech, because of the higher pitch of female speech. When the pitch period is so short that the impulse response of the LPC filter does not have time to decay sufficiently before the next excitation pulse arrives, the interactions between periods cause distortions in the output signal. However, in many cases, this seems more apparent in the mathematics than to the ear. Although infant cries are much higher-pitched than adult male speech, if a filter of high enough order is used, they can probably sound natural enough with minimal deterioration at higher frequencies. Fortunately, the particular spectral content of cries is not of primary interest in this study. As long as the resynthesized cries that will be compared perceptually sound natural and have identical spectral envelopes with one another, small differences from the original cries are of no concern.

With respect to the excitation source of the linear model, there could be a problem with cries that have both a periodic and an aperiodic excitational component. The mathematical formulation of LPC does not allow for simultaneous excitation from both sources[17]. My experimentation with combinations of periodic and aperiodic excitation did not yield perceptually acceptable results. Therefore, the only way to compensate for the lack of spectral flattening that characterizes noisy excitation is to include more poles in the LPC filter. Because of the limited bandwidth of the signal[18], there is a limit to the effective improvement that can be achieved this way.

As far as nasality is concerned, although nasal coupling is known to introduce zeros to the transfer function of the vocal tract, it has been shown mathematically that the effect of zeros can be simulated by increasing the number of poles. Thus, using a higher order filter should practically eliminate this potential source of problems. However, the LPC parameters then lose their clear physical meaning, because the model from which they were derived is an all-pole model assuming no nasal coupling. Thus, in the case of nasal sounds, LPC is not a model of the vocal tract, but a model of the particular waveform of the signal, and transformations to higher-level parameters, such as the vocal tract cross-sectional area function coefficients, are meaningless.

The processing method and organization

Given a decision to use the LPC method for analysis and resynthesis of the cries, it is still necessary to implement the appropriate algorithms with the particulars of the task in mind, adjusting the various parameters to the characteristics of the infant cry signals. Because the goal of the processing is to create series of cries identical in all but one respect, it makes sense to store the results of the analysis in a way that allows for their easy inspection and manipulation. This way it will also be possible to eliminate any errors, which is of critical

---

[17] The excitation source driving the LPC filter can produce either white noise or periodic unitary impulse trains, but not both at the same time.

[18] The Nyquist theorem states that frequencies up to half the sampling frequency can be accurately represented in the digital signal.

importance, because the resynthesized cries should sound perfectly natural.

In order to accurately control the amount of jitter put into the synthesis of the stimuli, it is necessary to have an adequate temporal resolution. This can be achieved either with high sampling frequency, or with glottal pulse modeling (Milenkovic, 1992). Because we don't have a model of the infants' glottal excitation (and because the latter method is substantially more complicated), I decided, for the second set of experiments, to upsample (with interpolation) the cries, effectively increasing the temporal resolution. I used quadratic interpolation because it results in much smoother waveform than linear interpolation does and because it retains the length and the structure of the pitch periods more accurately. With these general issues in mind, I now discuss the specifics of my digital processing organization, with emphasis on the problems that arise and the way they can be solved.

The LPC analysis

The first step in the analysis is to decide which algorithms to use, and how to set the working parameters. To determine the optimal solution, one must take into account the particular characteristics of the problem at hand. In this case the most important factor was the quality of the output. In order to make the re-synthesized cries sound as natural as possible, the LPC filter should be of sufficiently high order. In most speech applications, when the sampling frequency is around 10KHz, an order of ten or twelve produces good results for adult male speech, because there are enough poles to determine three or four formants (two poles each) for the vocal tract area function and account for the lip radiation and glottal source effects in spectral shaping. A general rule of thumb for LPC speech processing systems is to use one pole per Kilohertz of sampling frequency, plus two. This results in a close enough spacing of the poles, so that there are enough to accurately model the entire spectrum. Because I used a sampling frequency of 20KHz, an order of 24 should be sufficient. However, when I needed to use higher effective frequencies to increase the temporal resolution for the jitter manipulations, going as high as 80KHz, the 24 poles were not enough to cover the entire 40KHz range, even though there is very little energy above 10KHz. By trial and error, I found that 40 poles were necessary for good results.

Close examination of the infant cry waveforms revealed the diversity of their characteristics in terms of periodicity. With pitch periods ranging from less than 1.5 msec to more than 4 msec, going up to 25 msec in segments with vocal fry, and ranging from clearly periodic, stable, waveforms to almost pure noise, it was hard to identify any useful average values. Although for the most part cry waveforms were changing very gradually, there were some segments with rapid changes in the waveform characteristics. To obtain an accurate parametric representation of the fast-changing segments, a small window length was necessary. On the other hand, the window should be large enough to contain at least two or three pitch periods at all times, otherwise the periodicity information would not be correctly extracted and the calculated spectrum would be wrong. Furthermore, for an acceptable pitch detection accuracy[19], three to five pitch periods should be able to fit in the analysis window. After experimentation with different settings, I decided to use different window lengths for the pitch and the LPC analysis. The window length for the LPC analysis was set at 10 msec and the frame rate was set at 5 msec. With a 50% overlapping between successive frames, most of the rapid variation in the waveform can be captured.

The method I used for LPC analysis was the autocorrelation method using the Durbin recursive algorithm for solving the LPC equations[20] (Rabiner and Schafer, 1978). This method is very robust, because it is guaranteed to yield a stable filter, and very fast, because it exploits the Toeplitz nature of the system matrix. In order for this algorithm to work well, double or extended precision arithmetic is recommended. This is not a problem when doing off-line analysis on a personal computer or Unix workstation.

The calculation of the residual energy, which was to be used as the gain factor of the filter during synthesis, although straightforward in theory, was in some cases problematic. When the waveform was very smooth and there was a resonance at a multiple of F_0, the predictor coefficients could be calculated with high accuracy and the actual residual energy was much smaller than calculated, because the filter was not adequately damped, and energy could accumulate from previous periods.

---

[19] The pitch detection analysis and the LPC coefficient analysis are practically independent, although they must be done on the same speech segments.

[20] A comprehensive analysis of the LPC mathematics is contained in Strobach (1990).

In such cases, the amplitude of the re-synthesized waveform would keep increasing and when it exceeded the maximum allowable value it would get clipped. The clipping was clearly audible in the synthetic cries. To eliminate it, the energy values were manually reduced as much as it was necessary to avoid clipping, with special care taken not to affect the output amplitude level in any abrupt way. Furthermore, I included a smooth ceiling function in the synthesis program to be able to adjust the maximum energy without manually editing the file each time. All synthesized cries were checked after synthesis to make certain that no clipping had occurred. The energy values were also trimmed at the initial and final segments, so that there would not be any noise preceding or following resynthesised cries.

Pitch analysis

The method used for pitch detection was the modified autocorrelation function with center clipping and coefficient scaling (Hess, 1983; Tsakalos and Zigouris, 1991), which gives good results for spoken vowels when the window contains a few pitch periods. However, infant cry waveforms often contain a large amount of noise, and pitch periods become obscure. Especially in dysphonated segments, automatic pitch tracking is all but impossible and voicing classifiers incorrectly output "unvoiced". The problem is that the excitation source is actually noise, which is modulated by glottal pulses, so it has some quasi-periodic component that is clearly perceived as pitch, although the spectrum is flattened and the waveform appears random. Other major problems in pitch tracking include the wide range of variability of the fundamental frequency, which makes it difficult or impossible to use "smart" peak-picking algorithms, and the interactions of the fundamental frequency with formants, which often resulted in disproportionately pronounced harmonics that confused the pitch tracking routine.

A partial solution to most of the above problems is to low-pass filter the signal, so as to eliminate as much spectral information above the higest possible F_0 as possible. I used an odd-number window length FIR filter with a cutoff frequency (-3dB) at 800Hz and a rate of descent so that the log amplitude of the transfer function reached -60dB at 1200Hz. The filter was designed using the Kaiser window method (Oppenheim and Schafer, 1989) and had zero phase response, so the waveform peaks would remain at their relative positions. Application of this filter removed most of the noise and the effects of higher-order harmonics, thus increasing significantly the efficiency of the autocorrelation-based pitch detection algorithm. Because of the ambiguity in the autocorrelation function when more than one peak were good candidates for the pitch estimate, I used the autocorrelation of the autocorrelation function of the cry waveform to enhance the global periodicity. Finally, I incorporated in the algorithm a restriction on the maximum allowable change of pitch between successive processing frames.

I also tried inverse filter tracking and various spectral algorithms (Hess, 1983), with and without prior filtering, as well as center clipping, with no improvement. Because of these problems, following pitch tracking, all results were visually inspected against the waveforms, and the problematic sections were analyzed manually and corrected when necessary, so that there would not be any periods with pitch errors. The pitch contour was post-processed with nonlinear (median) smoothing through a window three frames long, so as to eliminate any short-term perturbations in F_0 that might interfere with accurate jitter synthesis.

The synthesis

Synthesis was performed directly with the LPC filter from the α_i filter coefficients (Markel and Gray, 1976). A pitch-synchronous strategy was used, with all parameters being updated only at the beginning of pitch periods. Every time a new pitch period started, the pitch, energy, and filter coefficients were interpolated, in order to avoid abrupt changes in the resulting waveform. The pitch and energy values were calculated via logarithmic interpolation. The filter coefficients cannot be directly interpolated, because intermediate value sets might result in unstable filters. Thus, the equivalent set of reflection coefficients[21] were used, which were calculated as by-products of the Durbin algorithm and stored instead of the filter coefficients. These can be directly interpolated with all intermediate value sets giving stable configurations. The interpolated reflection coefficients were then converted to the filter coefficients prior to resynthesis.

Whenever a new pitch period was about to begin, the period length was changed according to the particular synthesis condition so that cries with

---

[21] The terms reflection coefficients, partial correlation coefficients, and PARCOR coefficients are used interchangeably

uniformly higher or lower pitch or with perturbed pitch could be easily produced. The method for producing jittered cries was to add or subtract a random number to or from each successive pitch value, and then use the result for the pitch period. The random numbers used were uniformly distributed[22]. The range of perturbation was determined by a constant at the beginning of the program as the maximum allowable number of sample periods to be added to or subtracted from each pitch period. For the cries that were used in the second set of experiments, the direction of the deviation was forced to change sign every period, so as to approximate the double harmonic break effect often observed in infant cry spectra. When synthesizing cries with higher or lower pitch, the actual pitch value was multiplied by a constant factor (the same for the whole cry) and the result was used for the pitch period.

To manipulate the rise time, it was necessary to know the maximum pitch and the time point at which it occurred. With this information all the parameters (pitch, filter, and energy) were appropriately interpolated (in parallel, so that their combinations would remain identical) so that the point of maximum pitch was moved to the desired position. This process effectively "stretched" the analysis parameters on one side of the peak and "squeezed" them on the other, creating a new set of parameters which defined a cry of equal length and of the same sound quality. The choice of points to move the maximum pitch to could be done in at least two ways: By defining rise time either as a percentage of the total time of phonation, or by an absolute amount of time. Because there was no reason to justify the choice of either method, I used them both, thus producing two different rise time variants for each original cry.

In summary, the processing was organized as follows: First, the cry was piecewise interpolated with a second-order polynomial, and for each sample point in the original, the quadratic curve was evaluated at four points, effectively multiplying the sampling frequency by four. Following upsampling, a pitch estimator processed the waveform, creating a series of pitch estimates which were stored on disk. Subsequently, these pitch estimates were visually inspected against the plotted waveform of the cry, and the erroneous values were manually adjusted. Then an LPC analysis was performed, saving the results (combined with the pitch estimates) on disk. For a series of cry stimuli to be created, the LPC synthesis program was run, using these files, with different parameter settings each time, which controlled the type of period manipulation. Finally the resulting waveforms were low-pass filtered and downsampled back to 20KHz, by simple 4:1 decimation. For the first series of experiments, the upsampling and downsampling steps were omitted and all processing was done at 20KHz.

Results and conclusion

The resulting re-synthesized cries sounded identical in quality to the original (natural) cries when the latter contained only clear voiced phonation. They usually differed slightly in intensity, because of the energy accumulation effect I discussed earlier, which was not uniform. Dysphonation was crudely approximated, i.e., there was some spectral flattening in the resynthesised cries, but generally they sounded smoothed, without the harsh quality of natural dysphonation. It remains unclear how one could go about to produce dysphonated cries with LPC. An extension to the voice production model will be necessary, which will allow excitation from combined noise and pulses.

The order of the filter was not critical. Experimentation with lower-order filters showed no audible differences[23] with lower order filters (as low as 12 for the 20KHz cries, and 35 for the 80KHz cries). Therefore, there were more than enough poles to improve and refine the synthetic waveform as much as the model could allow. The window size was probably more important in determining the upper limit in output quality. Perhaps an LPC analysis with adaptive window could better account for both the long pitch periods and the sudden changes. Also, the pitch detection method could use an adaptive window and more rules about what to anticipate and how to handle the most difficult cases, for more accurate estimations throughout the cries.

Because of interaction effects of the fundamental frequency with the formants, it is not advisable to alter the fundamental very dramatically. The fundamental frequency of infant cries is often so high that the spectral envelope is specified by very few points. If it is shifted too far, there is selective amplification at particular frequencies different from

---

[22] For adult voices, it is known that the distribution of the jitter-type fluctuations is of the form $1/f$ (Hillenbrand, 1988), but there are no data on infant cries.

[23] This was a personal, subjective observation, and was not experimentally tested.

those of the original, that change the quality (timbre) as well as the loudness of the cry. I found that a change of up to about ten percent does not have any adverse effects of this kind and can be safely induced.

Overall, although the LPC model is not theoretically designed for infant vocalizations, the results seem quite satisfactory (in those cases in which cry production resembles speech production). Speech production (Flanagan, 1972; Lieberman and Blumstein, 1988) and cry production (Golub and Corwin, 1985; Lieberman, 1985) have many physiological similarities. It remains to be experimentally verified that LPC is a valid method for infant cry manipulations.

The Experiments

This chapter describes two series of experiments The first four experiments were run during the summer of 1992, using headphones to present the stimuli to the subjects. The first experiment was done with natural cries, and Experiments 2, 3, and 4 used the 20KHz synthetic cries. Experiments 5 and 6 were conducted in the spring of 1993, using cries processed at 80KHz, and the stimuli were presented through a loudspeaker. The reason for the change was that, in the meantime, it came to my attention that there are differences in jitter perception between headphone and loudspeaker presentation. Wilde et al. (1986) found that jittered synthetic vocal stimuli sounded rougher when heard over loudspeakers than when heard through headphones. Although no published replication of that finding has come to my attention, researchers who use synthetic voice stimuli now commonly perform their experiments using loudspeakers for this very reason (Hillenbrand, 1988; Milenkovic, 1992). A plausible hypothesis to explain this finding by Wilde et al. pertains to the effects of reverb in the case of free-field presentation.

Before it is possible to use synthetic[24] cries to assess the role of acoustic parameters, it is important to establish that the synthetic cries are appropriate stimuli. Also, it is important to verify the ability of subjects to rate cries consistently with respect to perceptual features. Therefore, the experimental plan was designed as follows: Experiment 1 was a preliminary cry rating task, which was carried out in order to demonstrate the rating consistencies one should expect, and to aid in the design of the other experiments. Experiment 2 was a discrimination experiment, wherein the computer-generated stimuli were tested for naturalness against natural cries. In Experiment 3, I tested the subjects' sensitivity to jitter with a standard A-X pair-comparison procedure. Then Experiment 4 addressed directly the perceptual differences in cries differing only in one acoustical feature with a rating task in an attempt to reveal underlying correlations between acoustics and perception.

Following extensive revision of the signal processing methods, including the use of higher effective sampling rate for improved jitter control, and with loudspeaker presentation of stimuli, I repeated the procedure of Experiments 2 and 4 in Experiments 5 and 6 respectively. Because Experiments 4 and 6 address the same issues, and because the latter constitutes an improvement of the former, I give only a summary of the results of Experiment 4. The final discussion will take into account all findings, and I try to draw some general conclusions and present my ideas about open issues to be addressed in future research.

The original (natural) cries that were used in this study were provided by Mr. T. Fitch, of the Department of Cognitive and Linguistic Sciences at Brown University, originally recorded for research projects by Dr. Barry Lester of Brown University, E. P. Bradley Hospital, and Women and Infants Hospital in Providence, RI. The cries were provided on a diskette in a computer-readable format (for IBM PC computers), already sampled at 20 KHz and quantized at 12 bits, after appropriate anti-aliasing filtering. In all experiments, the stimuli were presented using an IBM AT personal computer connected to a D/A board and high-quality audio amplifier. When headphones were used, the stimuli were presented monaurally.

Experiment 1: Preliminary ratings

The purpose of this experiment was threefold. First, I wanted to see whether subjects consistently rate infant cries according to their severity and whether ratings from different subjects are similar. Secondly, I wanted to obtain a measure and an ordering for the various cries, in terms of severity and of rating consistency between subjects to aid in selecting my stimuli for the next experiments. Thirdly, I was interested in finding out whether

---

[24] The term synthetic cries will be used to denote cries generated by a computer based on parametric descriptions (pitch and LPC coefficients) of natural cries, i.e. cries recorded from infants.

there would be any perceptual distinctions among the cries according to the presumed infant's condition during recording that could serve as a pilot for acoustic analyses and as an indication for parameters to investigate.

Method

Subjects The subjects were five male graduate students in the Department of Cognitive and Linguistic Sciences at Brown University. None of the subjects were parents and only one reported having some childcare experience. The mean age was 28 years.

Stimuli Eighty one cries were used, forty of which were labeled as "pain" and forty as "nonpain", according to the recording conditions. The "pain" cries were recorded after the removal of a cardiac electrode, a stimulus of relatively moderate intensity. The "non-pain" cries were spontaneous vocalizations of the infants. One cry was not labeled. The durations of the cries ranged from 540 to 6300 msec. The cries were recorded from ten healthy, full term infants; some were recorded at age one week, some at age one month. Each sampled cry consisted of one expirational vocalization, followed in a few cases by an inspiratory phase. From each entire crying event (lasting a few minutes) originally recorded on tape, two cries ("syllables") had been selected and sampled, so the cries could be arranged into ordered pairs of cries originating from the same crying event. Overall, the available set of cries covered a broad perceptual spectrum, possibly corresponding to a diverse set of infant "moods".

Procedure The subjects heard the cries monaurally through headphones, and entered their responses by means of a four-button response box. The sound level was adjusted at a comfortable level and maintained unchanged for all subjects. The subjects were asked to rate the cries on the basis of severity on a scale from one to five. After each cry was played, the subjects were required to press a button to indicate a point in the range from "best-sounding" to "worst-sounding". The middle point of the scale (3) was indicated by not pressing any of the four buttons. The subjects were instructed to use any perceptual dimension they felt appropriate. The cries were presented in a random order, and each was rated twice, to test within-subject consistency. After the rating experiment each subject was asked about the basis for his rating scale.

Results

The cries were consistently rated by all subjects, and the entire range from one to five was used by all of them. The distribution of the ratings is shown in Table 1, and in Figure 1 (left graph), separately for the pain and nonpain conditions, and overall. The average standard deviation of ratings of the same cry was 0.7 between subjects, ranging from 0.5 to 0.6 within subjects. The average rating was 3.5 ± 1.3 for the pain cries and 2.6 ± 1.1 for the nonpain cries. The difference between the two cry groups was statistically significant at the 0.001 level ($t = 26.83$, $df = 4$).

Cry Group	Rating (1 is best sounding)				
	1	2	3	4	5
Pain	31	70	77	115	107
Nonpain	75	116	119	76	14
All	106	186	196	195	127

Table 1: Distribution of ratings by cry category

Discussion

The results of this preliminary rating suggest that the cries are perceptually distinct from each other, and can be consistently rated by different subjects, even when the subjects have no experience with infants. The subjects reported rating the cries based on "how urgently the infant sounded to need help" (2 subjects), "how much in discomfort the infant sounded to be" (2 subjects), or "how bad the cry made them feel" (1 subject). Because the between-subjects consistency was very high it appears that there is a single underlying dimension which is expressed in all the wordings used.

The results indicate that the pain cries were rated significantly worse-sounding than the nonpain cries, although there is not a clear distinction in accordance with the assumed pain status. The high degree of overlapping between the ratings of pain cries and ratings of nonpain cries could be explained by the fact that the painful stimulus was not very intense and that the nonpain condition was not precisely controlled. Moreover, the first and second syllable of the same cry were often given very different ratings. Although the difference between the first- and the second-syllable groups just failed to reach significance at the 0.05 level (ANOVA Pain×Syllable, $F(1, 796) = 3.78$, $p = 0.0518$ for Syllable), the trend appears to be in agreement with the literature (Zeskind et al.,

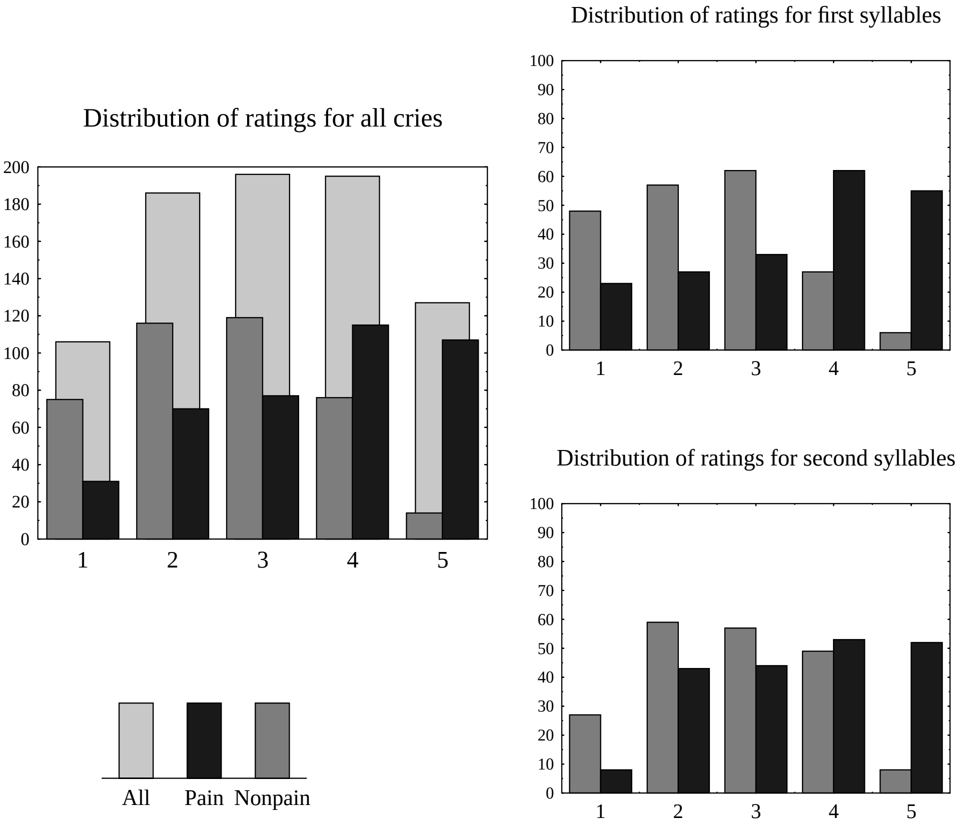


Figure 1: Distribution of ratings in the first experiment

Rating (1 is best sounding)	First syllables		Second Syllables		All	
	Pain	Nonpain	Pain	NonPain	Pain	NonPain
1	23	48	8	27	31	75
2	27	57	43	59	70	116
3	33	62	44	57	77	119
4	62	27	53	49	115	76
5	55	6	52	8	107	14

Table 2: Distribution of ratings for the first and second syllables

1985) that the second syllable of a pain cry tended to be rated lower than the corresponding first syllable, whereas the second syllable of a nonpain cry tended to be rated higher than the corresponding first syllable. This is illustrated in Table 2 and in Figure 1. Perhaps with better controlled stimulus conditions such findings could have been more pronounced.

Experiment 2: Synthetic vs. Natural cries

This experiment was done in order to justify the use of computer generated cries in cry perception experiments. I expected that the synthetic cries would sound natural, although it would not be necessary for them to sound identical to their original counterparts. In order to assess the extent to which synthetic cries resemble real cries, both a forced choice and a same-different discrimination task were performed.

Method

Subjects The subjects were six males and five female students at Brown University, ranging in age from 20 to 37 years. All females and one male reported having experience in childcare. None of the subjects were parents. All subjects volunteered their participation.

Stimuli Eight cries from the eighty one available from the previous experiment were chosen for analysis and resynthesis, with carefully selected to span the whole range of ratings from the first experiment, so as to cover as broad a spectrum as possible. The selected cries ranged in mean rating from 1.0 ± 0.0 to 4.8 ± 0.4 and in duration from 625 msec to 2459 msec. Their characteristics are summarized in Table 3.

The stimuli were synthesized with the method described in Chapter 2. There were two manipulation conditions, Pitch and Jitter. In the Pitch condition, there were five levels, P1, P2, N, P3, and P4, which corresponded to F_0 approximately equal to 90%, 95%, 100%, 105%, and 110% that of the original cry respectively. For example, with Cry1, P2 was synthesized using the parameters derived from analysis of Cry 1, but with multiplying each pitch period by 1.05 before using it. This way, the alterations were restricted to ±10%, so as to keep unwanted interaction effects with formants to a minimum. In the Jitter condition, there were four levels, N, J1, J2, and J3, which corresponded to range of period deviations 0μs, $\pm 25\mu$s, $\pm 50\mu$s, $\pm 75\mu$s, or 0, ±0.5, ±1.0, and ±1.5 times the sampling period, respectively. These times correspond to jitter 0%, 2%, 4%, and 6%, for a mean F_0 of 400Hz, and to jitter 0%, 3.33%, 6.66%, and 10% for a mean F_0 of 666Hz, respectively. For example, with Cry2, J2 was synthesized using the parameters derived from analysis of Cry 2, but with adding a random number between -1 and 1 to each pitch period before using it, independently of the original length of the period. This way, the percent jitter was higher in the higher-pitched portions of the cry, in agreement with previous findings about the relationship between F_0 and jitter. Note that the N cry is the same in both conditions, and was only synthesized once. Consequently, there were eight synthetic cry stimuli in all for each of the eight natural cries, with varying amounts of pitch and jitter, completely separated (the pitch and jitter manipulations were never crossed). Also note that the cries in the pitch condition contained little or no jitter (the minimum due long-term F_0 variability) because of the F_0 contour smoothing that was done after the analysis.

The amount of jitter actually contained in the synthesized cries was measured after synthesis, and was found to correlate well with the expected values. Furthermore, it was found to be well within the range of jitter in natural cries[25], the measures

---

[25] See Fitch (1992) for measurements of jitter in infant cries.

Cry #	Mean F_0 (Hz)	Max F_0 (Hz)	Duration (msec)	Rating
1	435.4	500.0	1062.7	1.0 ± 0.0
2	425.0	487.8	1216.3	1.7 ± 0.8
3	516.8	540.5	625.3	2.5 ± 1.2
4	428.8	487.8	2240.3	2.9 ± 0.9
5	455.1	526.3	2459.1	3.4 ± 0.7
6	465.4	540.5	1677.1	3.9 ± 0.7
7	578.8	625.0	1133.8	4.3 ± 0.8
8	595.6	666.7	2009.8	4.8 ± 0.4

Table 3: Summary of cry characteristics

spanning at least half the range of those of natural cries. Thus it is reasonable to assume that the synthetic cries constitute a representative, if somewhat conservative, set of stimuli for jitter experiments.

Procedure There were two parts to this experiment. In Part 1 subjects listened to one cry at a time, and were asked to decide whether it was a natural or a synthetic one (i.e., a forced choice task). In Part 2 subjects listened to pairs of cries, and decided whether the two cries were the same or different (A-X two-interval same-different design). The order of the two parts was randomly chosen for each subject. There were some practice trials before each part, so that subjects could familiarize themselves with the procedure. The cries were played monaurally through headphones to the subjects, who responded by pressing a key on a four-button response box.

In the forced-choice task (Part 1) there were seventy cries in all, each repeated twice, for a total of one hundred and forty trials. (Each trial consisted of a single cry.) The trials were presented in random order, different for each subject. Of the 70 cries, 48 were synthetic, and the remaining 22 were natural. The natural cries included those that were used to create the synthetic ones, plus fourteen more, selected at random from the remaining 73 from the first experiment. The synthetic cries that were used in this experiment included the neutral version N00 (zero jitter, pitch equal to the original) the three jitter versions, and two of the four altered-pitch versions, P1 and P4. The reason that not all the synthetic cries were included was to reduce the amount of time that was necessary to run the experiment in order to prevent subject fatigue. Also, to keep the number of natural cries relatively close to the number of synthetic cries for the statistical analysis. After all, there were not any qualitative differences between cries differing in mean F_0 in the pitch condition.

In the A-X design (Part 2) the eight natural cries (N) and the corresponding neutral (N00) synthetic ones (S) were paired and presented in all four possible combinations for each pair (N-N, N-S, S-N, S-S), resulting in a total of thirty-two trials. Each pair was played once to each subject and the order of the pairs was randomly varied. The equipment used for this part was the same as for the first part.

Results

The results from the two tasks are summarized in Tables 4 and 5. The ability of subjects to correctly identify a natural cry was significantly better than chance ($P(\text{"N"}|\text{N}) = 0.66$, $t = 5.37$, $p < 0.01$), whereas the ability to correctly identify a synthetic cry was not significantly different from chance ($P(\text{"S"}|\text{S}) = 0.54$, $t = 0.92$, $p > 0.2$). Subjects were significantly better in correctly identifying the natural cries than in correctly identifying the synthetic cries ($t = 2.40$, $p < 0.05$), but they were not able to consistently tell them apart in the forced choice task (d'=0.51).

They were able to tell them apart, however, in the same-different task. The corresponding d' score was 1.17. The *true* d' score, which takes into account the peculiarities of the A-X same-different experimental design and is corrected for response bias (Kaplan et al., 1978) was 2.21, which again shows the difference in performance in identifying the different cries.

No significant differences were found between males and females. The d' scores for individual subjects ranged from -0.63 to 1.36 for the forced-choice task and from 1.05 to 4.33 (true d') for the same-different task. There were also no significant differences between the naturalness of different types of cries (jittered vs. altered-pitch etc.)

Actual	Response				d' score
	Natural		Synthetic		
	N	Probability	N	Probability	
Natural	320	**0.66**	162	**0.34**	0.51
Synthetic	480	**0.46**	568	**0.54**	

Table 4: Natural vs. Synthetic cries — Forced choice task

Actual	Response				d' score (True d')
	Same		Different		
	N	Probability	N	Probability	
Same	149	**0.85**	27	**0.15**	1.17
Different	80	**0.45**	96	**0.55**	(2.21)

Table 5: Natural vs. Synthetic cries — Same-different discrimination task

Discussion

From the above results, we can conclude that even though the synthetic cries do not sound identical to their natural counterparts when played in temporal succession, subjects cannot reliably distinguish between the two in terms of naturaleness. Therefore, it is reasonable to conclude that the synthetic cries do indeed sound natural. Subjects reported after the experiment that the task was very difficult, and that any differences they had picked in the same-different task were in terms of loudness. Indeed, due to the peculiarities in the calculation and interpretation of the residual energy in the LPC analysis, it is very hard to make a synthetic cry having exactly the same intensity as the original. In fact, there was another difference, in that the synthetic cries always had a much smoother F_0 contour than the natural ones, and perhaps this helped the subjects realize that there was a difference.

In the forced-choice task, all subjects reported that they could not tell which cries were natural, and the data show that some of the cries that were consistently rated as being "synthetic" were, in fact, natural. Synthetic cries were rated as synthetic significantly more often than natural cries are, but not significantly more often than chance. This, along with the fact that subjects are generally unable to discriminate the natural from the synthetic cries, (as the d' score from the forced-choice task indicates,) leads to the conclusion that synthetic cries can be used in place of natural cries in perception experiments.

Experiment 3: Jitter discrimination test

This experiment investigated the discriminability of cries with different amounts of jitter, independently of the perceptual effect a distinction might have. I was concerned about the ability of adults to perceive jitter in natural signals[26] (specifically, infant cries) and to discriminate the signals according to it. An A-X two-interval same-different discrimination task was used to reveal possible perceptual differences. I expected that subjects would discriminate cries with different amounts of jitter. I also expected that subjects would discriminate more readily between cries with larger differences in jitter level than between cries with similar amounts of jitter.

It was not necessary to do a similar experiment with pitch differences, because there is strong evidence in the literature that pitch differences in infant cries are perceptually salient. On the contrary, there is no evidence on jitter perception in infant cries.

Method

Subjects The same six male and five female students were used, this experiment occurring in the same sessions with Experiment 2.

Stimuli The stimuli used for this experiment were pairs of synthetic cries with different amounts of jitter, synthesized as described in the previous experiment. All three jitter versions (J1, J2, and J3) and the neutral version (N) of four cries were

---

[26] Natural signals in this case refers to signals occurring in nature, e.g. speech, crying, etc.

used. These four versions were presented in pairs differing by zero, one, or two jitter levels. These pairs were N–N, J1–J1, J2–J2, J3–J3, N–J1, J1–N, N–J2, J2–N, J1–J2, J2–J1, J1–J3, J3–J1, for a total of 48 stimuli.

Procedure The procedure was identical to the procedure of the previous A-X same-different experiment. The 48 cry pairs where presented in random order and the subjects were asked to decide whether the two cries were the same or different. The instructions given to the subjects were identical to the instructions for Part 2 of Experiment 2. The equipment used was the same as that for the previous experiments.

Results

The results are summarized in Table 6. The true d' score for all subjects and all conditions was 0.59, ranging from negative to 0.95 for the different conditions (see Table 7). The true d' scores ranged from -0.82 to 2.0 for individual subjects. There was a difference between the true d' for males ($d' = 1.15$) and for females ($d' < 0$), but it was because males were more accurate in correctly identifying the "same"-pairs ($P_{males}(\text{“S”}|S) = 0.84$, $P_{females}(\text{“S”}|S) = 0.76$) and not because they could correctly discriminate cries with different amounts of jitter ($P_{males}(\text{“D”}|D) = 0.29$, $P_{females}(\text{“D”}|D) = 0.28$), therefore it is not of interest whether it is significant or not.

Discussion

These results lead us to conclude that subjects are unable to discriminate between cries with different levels of jitter when the difference in jitter was less than 3.3%. Subjects characterized the task as "extremely hard" and few could give any indication as to how the cries that they rated "different" sounded different. Two subjects reported differences in the melody of the cries (which was not correct, because the cries were always identical except for the amount of jitter they contained), and one subject reported some cries being "harsher" than others, but the true d' of these three subjects was negative. The data thus suggest that either jitter in infant cries is not perceived, or that the experimental procedure was for some reason not sensitive enough to reveal the effects that I anticipated.

The former conclusion is in disagreement with previous findings about the perceptibility of jitter, which have demonstrated a robust effect of jitter in perception of pulse trains (Pollack, 1968) and in perception of synthetically generated voices (Hillenbrand, 1988). It is also inconsistent with the findings of Lieberman and Michaels (1962), which suggested that variations in fundamental frequency of phonation may play a role in identifying the emotional condition of the listener. The possibility of insensitivity of the experimental procedure will be discussed in detail in Experiments 4 and 6.

Experiment 4: Varying pitch and jitter

The purpose of this experiment was to assess directly the role of pitch and jitter in the perception of infant cries. For this purpose, subjects were presented with cries, some of which differed in only one of these features, and all the other acoustic parameters were kept constant, and had to rate them on perceptual scales. If the subjects' ratings of cries differing in only one feature were significantly different, it would be evident that the differing feature is perceptually salient. I thus expected that the subjects' perception of the cries would change when either pitch or amount of jitter were different. Although the results of Experiment 3 in a way precluded any perceptual role for jitter, it is preferred to have a common way of evaluating the significance of the different acoustic parameters.

Method

Subjects The subjects were 4 males and 6 females, ranging in age from 22 to 38. One female and two males reported having no experience in childcare. Of the others, one male was a father and one female had worked in a children's hospital. All subjects were either college or graduate students or recent graduates. Informed consent was obtained from all subjects prior to the experiment. Each subject was paid $8.00 per hour for his or her participation.

Stimuli The stimuli for this experiment were sixty-four synthetic cries, divided into eight cry groups (according to the original natural cry from which the data for the synthesis were obtained) with eight versions each. Each group consisted of one cry with zero jitter and with pitch periods equal to those of the natural cry (N), three cries with different amounts of jitter (J1, J2, and J3), and four cries with pitch periods differing from those of the original by different amounts (P1, P2,

Actual	Response				d' score (True d')
	Same		Different		
	N	Probability	N	Probability	
Same	134	**0.76**	42	**0.24**	0.13
Different	250	**0.72**	99	**0.28**	(0.59)

Table 6: Frequencies, probabilities, and d' score from the jitter discrimination experiment

Actual	Response				True d'
	Same		Different		
	N	Probability	N	Probability	
N–J1 (Different)	62	**0.71**	25	**0.29**	0.51
N–J2 (Different)	56	**0.64**	31	**0.36**	0.95
J1–J2 (Different)	67	**0.77**	20	**0.23**	< 0
J1–J3 (Different)	65	**0.75**	22	**0.25**	0.51
N–N (Same)	32	**0.73**	12	**0.27**	
J1–J1 (Same)	33	**0.75**	11	**0.25**	
J2–J2 (Same)	33	**0.75**	11	**0.25**	
J3–J3 (Same)	36	**0.82**	8	**0.18**	

N: No jitter, J1: $\pm 25\mu sec$, J2: $\pm 50\mu sec$, J3: $\pm 75\mu sec$

Table 7: Frequencies, probabilities, and d' scores by pair group condition

P3, and P4). See the description under "Stimuli" for Experiment 2 for more details on synthetic cries.

Procedure The procedure was similar to that of the first experiment, in that subjects heard one cry at a time and rated it on a scale on the basis of some perceptual feature. Each cry was presented and rated twice for each of the four scales, not in successive trials. A scale from one to seven was used rather than the scale of one to five used in Experiment 1, to allow more subtle perceptual differences to be revealed. A scale with more subdivisions can also help reduce the load on the endpoints. However, a scale with too many subdivisions will be meaningless and misleading. In accordance with the literature, a seven-point scale was chosen as most appropriate.

The subjects were required to rate the cries on the basis of four different perceptual scales in four repetitions of the experimental cries (i.e., the subjects rated all the cries on one scale before proceeding to the next scale). These scales were named "Urgency", "Distress", "Health", and "Feeling", and the instructions given to the subjects were to rate the cries on a scale from one to seven according to how urgent it sounded, how distressing they sounded, how sick the infant sounded to be, and how angry or sad the cry made them feel, respectively. Of the many scales that have been suggested in the past, I chose to use these in order to cover the spectrum of salient perceptual features in as full a way as possible, while using as few scales as possible (to reduce the complexity of and time required for the experiment). Although some researchers have used as many as fifty scales for the subjects to rate the cries (Brennan and Kirkland, 1982), it has been shown that almost all those scales are highly interrelated (Zeskind and Marshall, 1988). 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. The four scales that were selected for this experiment cover the inferred severity of the case (urgency), the obnoxious character of the cry and its impact on the listener (feeling of the listener), the assumed emotional state of the infant (distress) and the possible abnormal aspect of the cry (health status). The scales of urgency and listener feeling also might give an indication as to what kind of action the listener would take if he/she were the caregiver.

In order to reduce subject fatigue the experiment was divided into two separate sessions, in each of which the cries were rated for two perceptual features. The order of the four items and the polarity of the scales were randomly varied between subjects. The order of presentation of the stimuli was randomized each time. The subjects entered their responses by pressing a key on a computer key-board.

Results

The ratings of the N cries (in all instruction conditions) were correlated with the ratings of their natural counterparts in Experiment 1 (mean $r = 0.92$). This result again shows that the qualitative character of the original cries was retained in the synthetic versions, further justifying the use of synthetic cries for cry perception experiments.

Tables 8 and 9 summarize the results in the four instruction conditions, for the effects of pitch and jitter, respectively. These results are graphically demonstrated in Figure 2. Note that the y-axis is adjusted to show the magnitude of the effects more clearly.

The data were analyzed separately for pitch and for jitter. The ratings of the cries N, P1, P2, P3, and P4 were analyzed for effects of pitch, cry, and instruction. The ratings of the cries N, J1, J2, and J3 were analyzed for effects of jitter, cry, and instruction. The data were first analyzed separately for each instruction condition. Analyses of variance (one for each instruction condition) were performed in each of the two cases (pitch and jitter), using a model with two fixed within-subjects factors completely crossed.

The analysis shows that pitch is significant in all four conditions. Figure 2 (top) shows the effects of pitch in all instruction conditions. Note that the scale is adjusted to show the effects mor clearly. The corresponding F and p values are $F(4,36) = 12.28$, $p < 0.00005$ in the urgency condition, $F(4,36) = 23.05$, $p < 0.00005$ in the distress condition, $F(4,36) = 4.40$, $p < 0.006$ in the health condition, and $F(4,36) = 9.27$, $p < 0.00005$ in the feeling condition. This means that subjects rated cries with higher pitch as more urgent, more distressing, coming from a sicker infant, and making them feel more angry or sad. Figures 3 and 4 show the effects of pitch on each cry in each of the four instruction conditions. Note the general tendency toward higher ratings as the mean F_0 gets higher, in agreement with previous findings.

The cry is also significant in all four conditions ($p < 0.00005$), as can be seen by the differences in the curves for the different cries. This means that the cries were different, and is not important in itself, but there is a significant interaction between pitch and cry in the urgency ($f(28,252) = 2.34$, $p = 0.0003$) and the distress conditions ($f(28,252) = 1.92$, $p < 0.005$), but not in the health ($f(28,252) = 0.94, p = 0.55$) and feeling ($f(28,252) = 1.02$, $p = 0.44$) conditions. This means that the effects of pitch on the urgency and distress ratings are not the same in all cries.

The analysis for jitter shows that it has a significant effect in the urgency condition ($f(3,27) = 5.04$, $p < 0.007$), but not in in the distress, health, or feeling conditions. This means that differences in jitter did not influence subjects' ratings, with the exception of the urgency condition. Figure 2 (bottom) shows the effects of jitter in all instructions conditions. Note that the effects are small, and there is no general trend in any of the curves. Figures 5 and 6 show the effects (or lack thereof) of the jitter manipulation on the ratings of each cry in each condition. The cry is again significant ($p < 0.00005$) in all four conditions, but there is no significant interaction between jitter and cry in any of the conditions.

Three-way analyses of variance, one for Instruction×Cry×Pitch and one for Instruction×Cry×Jitter, show no significant effect of instruction condition, indicating a high interrelation between the different perceptual features that were examined. This does not mean that the four conditions are equivalent, because the results of the two-way analyses differ in the different conditions, and because there is a significant interaction between pitch and instruction ($F(12,108) = 2.39$, $p < 0.01$), showing that pitch has a different effect in the four conditions. There is no significant interaction between jitter and instruction.

Discussion

The effects of jitter are neither consistent nor significant, as expected after Experiment 3. In fact, everything said in the discussion of Experiment 3 applies here as well, and I defer further discussion until after Experiment 6. The results on the effects of pitch were not surprising. Pitch has been shown to relate to the perceived intensity of crying in studies involving acoustic cry analyses (Frodi and Senchak, 1990; Zeskind and Marshall, 1988) or acoustic cry manipulation (Bisping et al., 1990). This experiment confirms, in a precisely controlled way, that pitch does play a significant role in infant cry perception. Examination of the graphs and the sizes of the effects shows that the perceptual consequences of altering the pitch are relatively minor, and not the same for all the cries, thereby suggesting that pitch might not be the acoustic cue of primary importance, again in agreement with previous findings (Gustafson and Green, 1989). Please refer to the final discussion and conclusion section

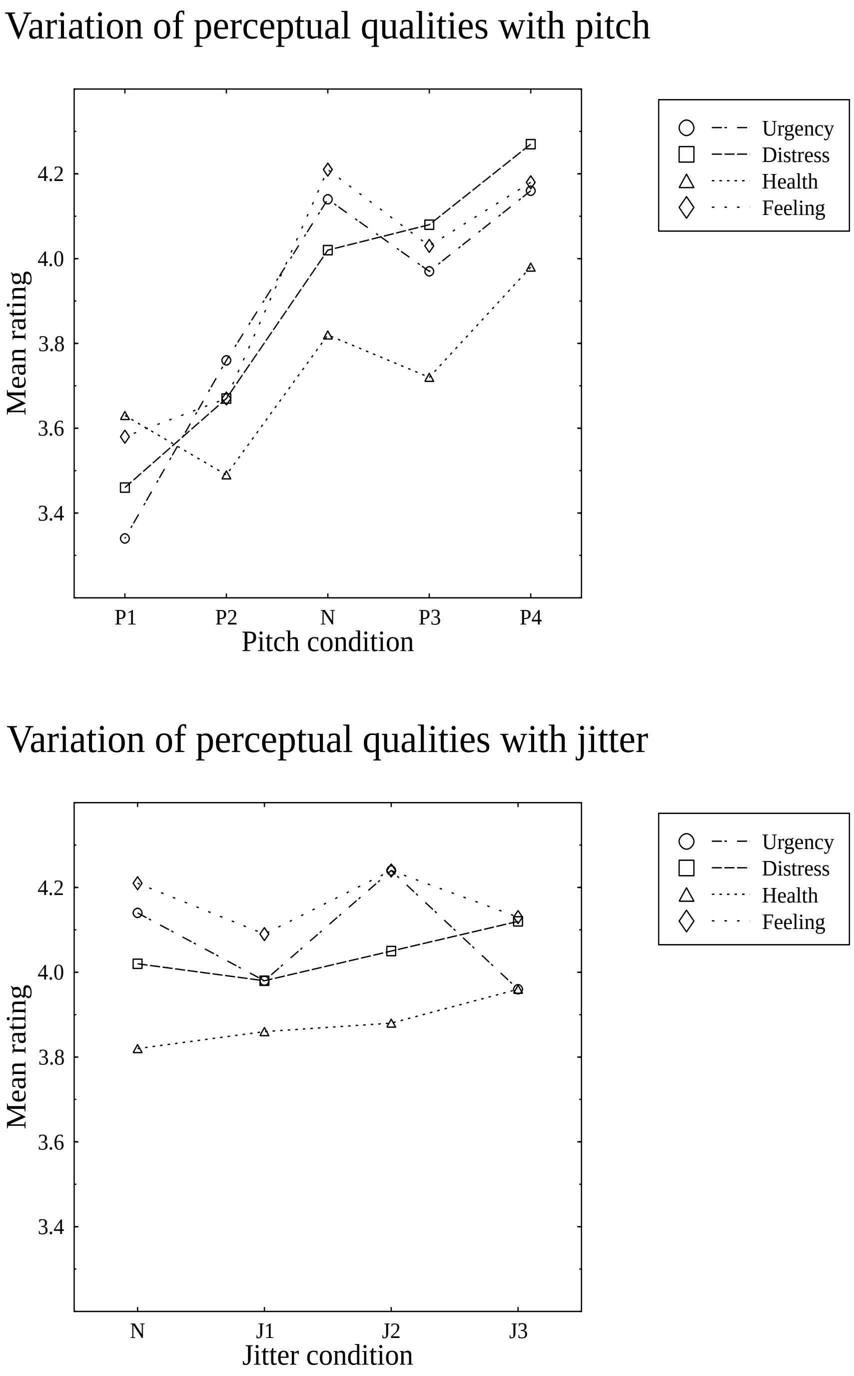
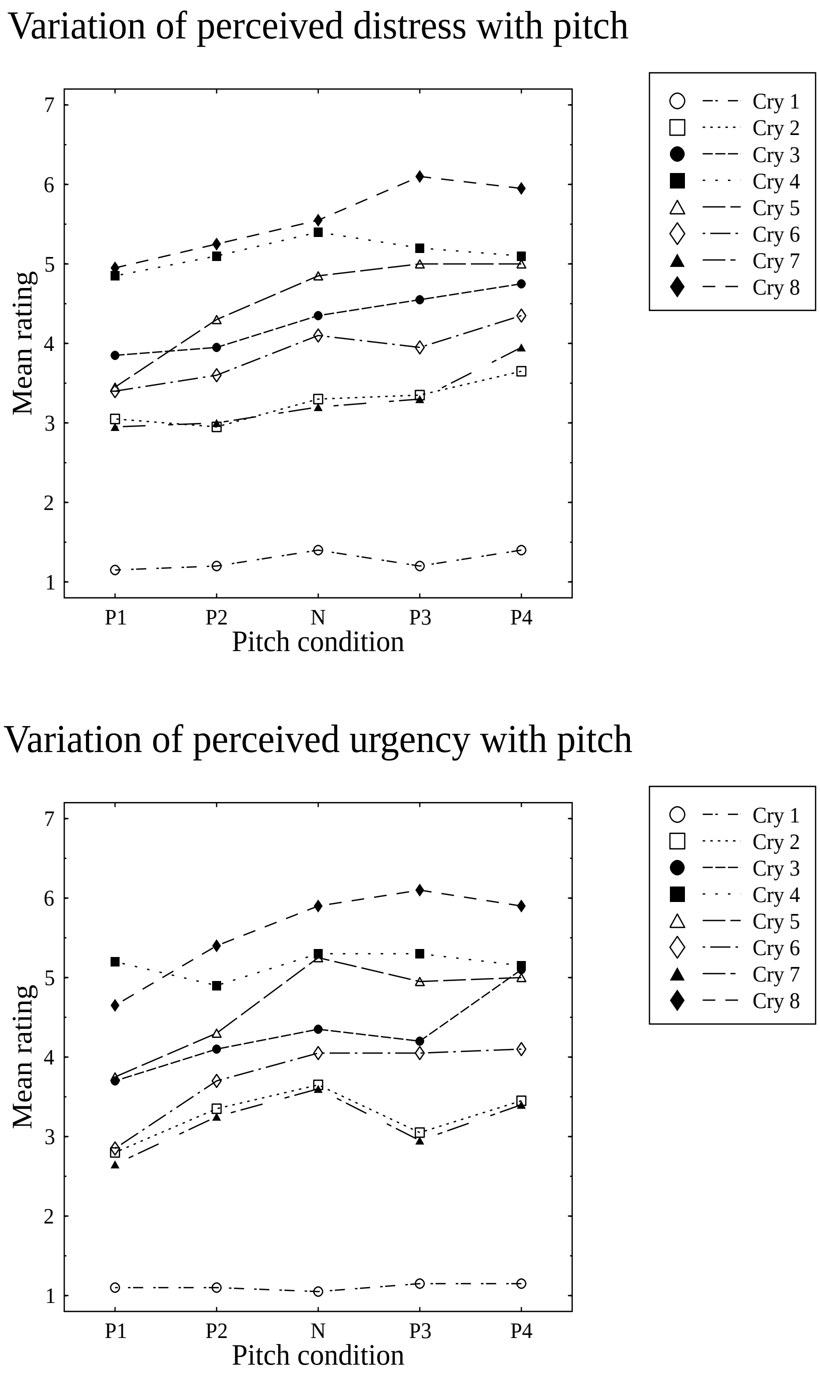
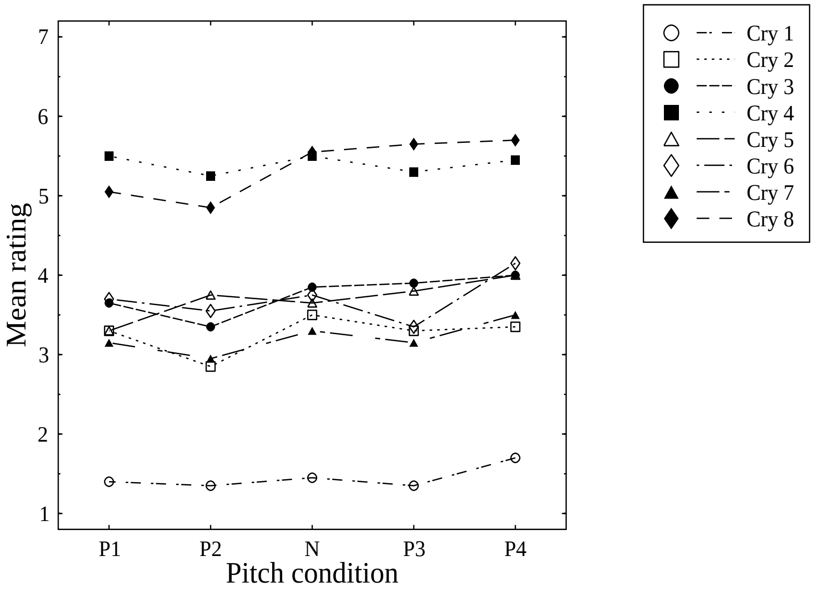


Figure 2: Effects of pitch and jitter in the different rating conditions, averaged over all cries

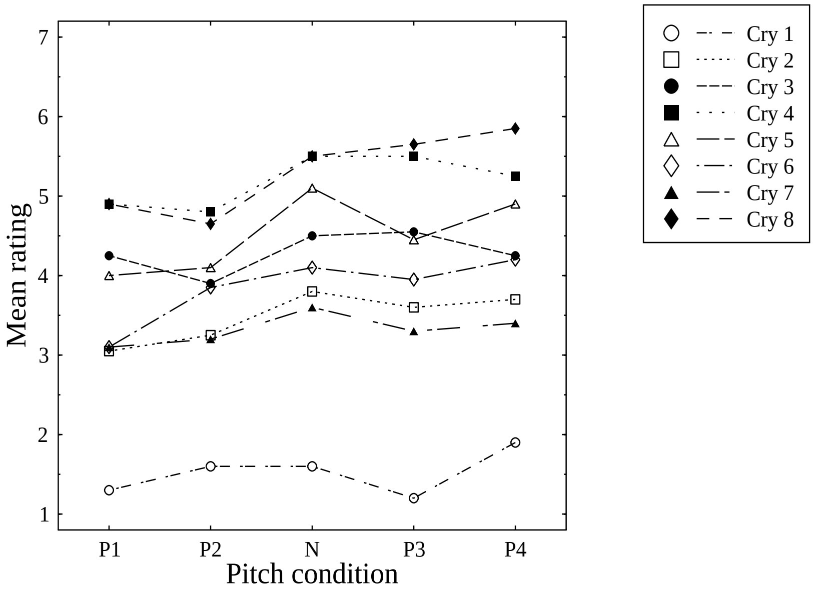

Variation of perceived distress with pitch

Variation of perceived urgency with pitch

Figure 3: Effects of pitch on individual cries in two conditions

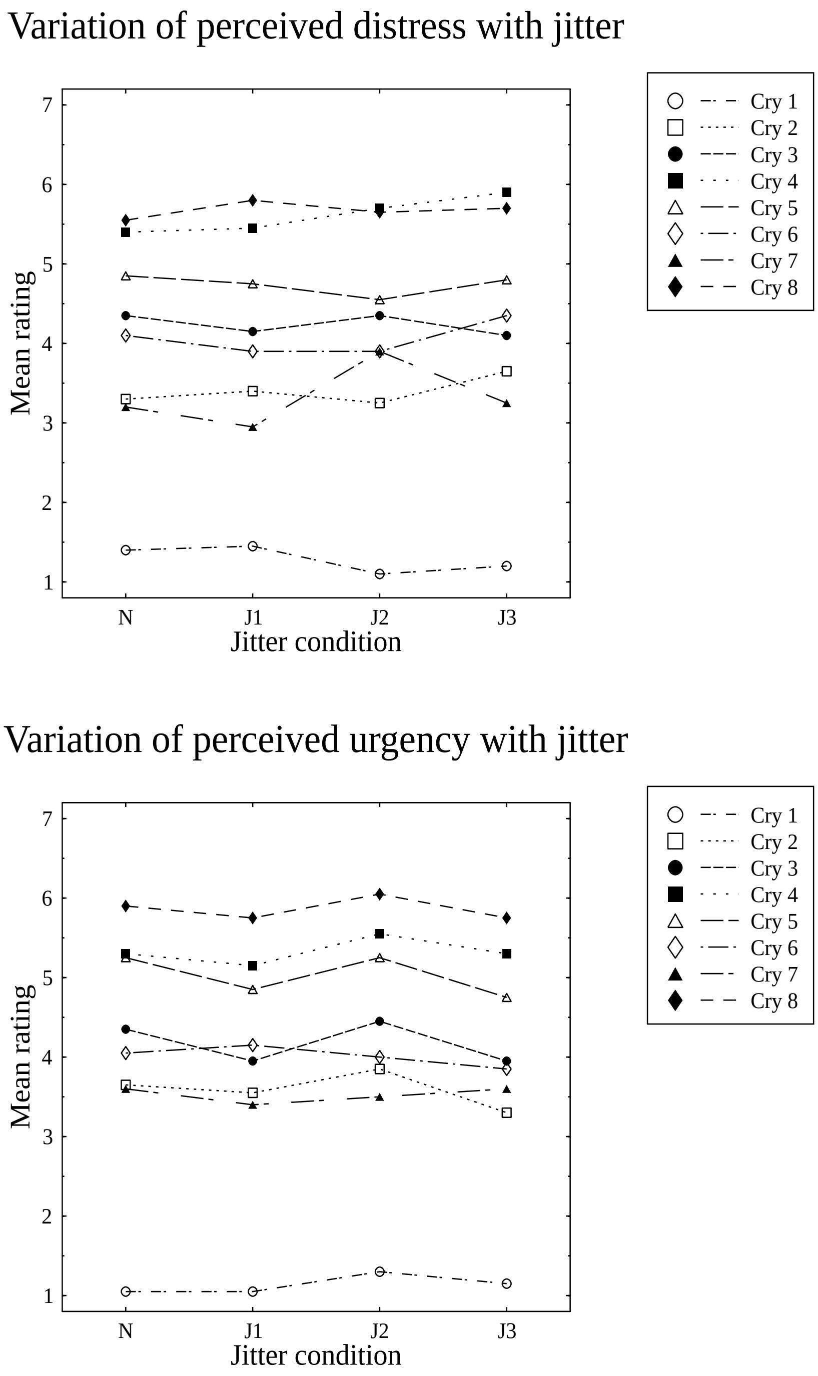


Variation of perceived health status with pitch



Variation of negative emotions with pitch

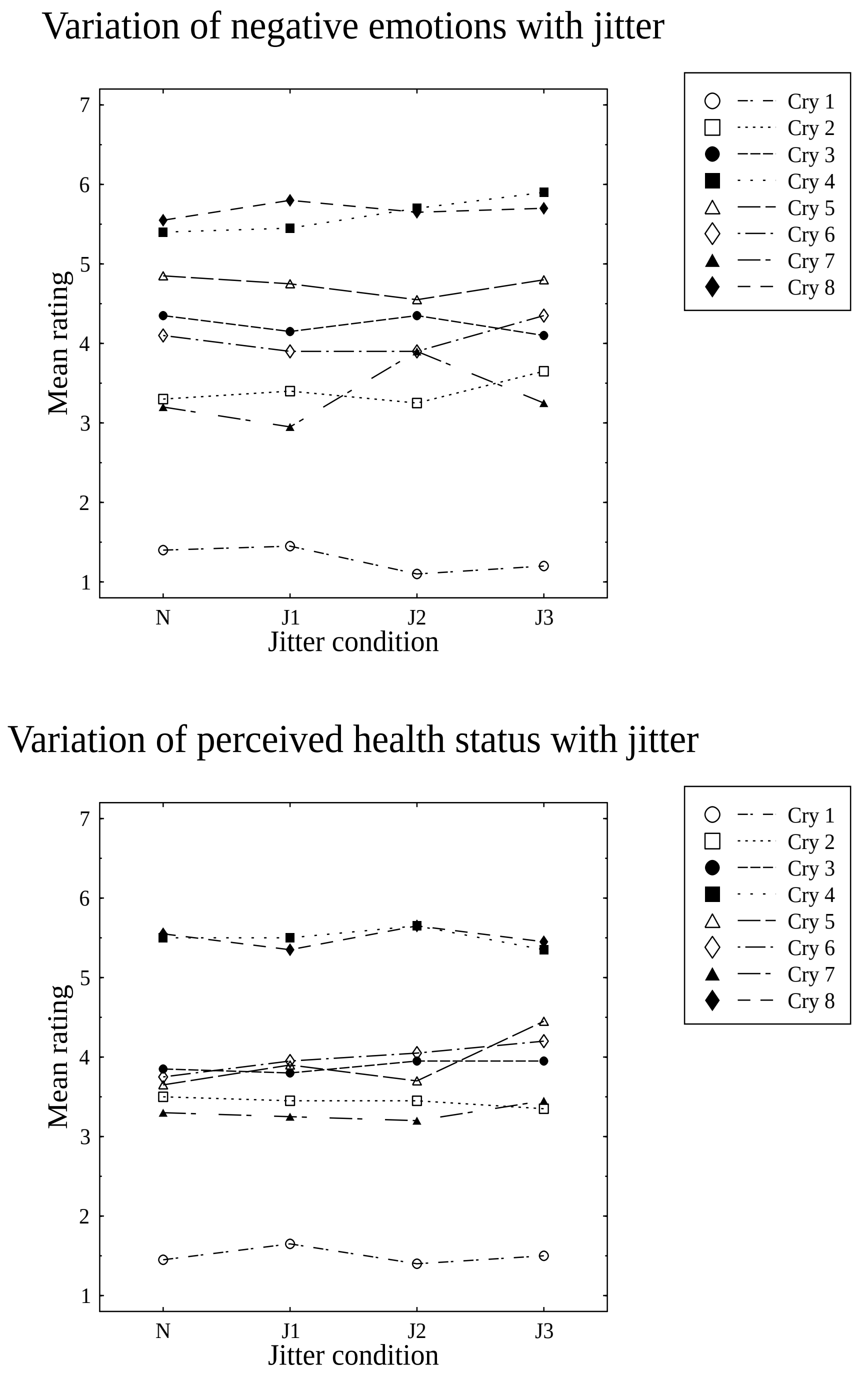
Figure 4: Effects of pitch on individual cries in two conditions



Variation of perceived distress with jitter

Variation of perceived urgency with jitter

Figure 5: Effects of jitter on individual cries in two conditions


Variation of negative emotions with jitter

Variation of perceived health status with jitter

Figure 6: Effects of jitter on individual cries in two conditions

Pitch condition	Instruction Condition			
	Urgency	Distress	Health	Feeling
P1	3.34	3.46	3.63	3.58
P2	3.76	3.67	3.49	3.67
N	4.14	4.02	3.82	4.21
P4	3.97	4.08	3.72	4.03
P5	4.16	4.27	3.98	4.18

Table 8: Mean ratings for different pitch levels (averaged across cries)

Jitter condition	Instruction Condition			
	Urgency	Distress	Health	Feeling
N	4.14	4.02	3.82	4.21
J1	3.98	3.98	3.86	4.09
J2	4.24	4.05	3.88	4.24
J3	3.96	4.12	3.96	4.13

Table 9: Mean ratings for different jitter levels (averaged across cries)

for more details and general remarks.

As the findings about the effects of free-field vs. headphone presentation of synthetic audio stimuli on jitter perception (Wilde et al., 1986) came to my attention, I performed a pilot experiment on jitter discrimination (identical to Experiment 3) using a loudspeaker to present the cries. The results ($d' \approx 1.5$) indicated that subjects could now discriminate between cries with different amounts of jitter, as initially anticipated, and that the method of presentation was the most likely factor that influencing the results of the previous experiments. In order to verify this hypothesis I conducted another experiment on cry perception, using a loudspeaker to present the stimuli to subjects. Unfortunately, subsequent pilot experiments on the naturalness of the cries used in Experiments 2–4 showed that several of them sounded very artificial over the loudspeaker, thus necessitating synthesis of a new series of stimuli. The two reasons for this result were, a) inadequate order of the LPC filter, and b) too little jitter in the N cries. It is long known in the speech synthesis field that synthesized speech sounds unnatural unless it contains some minimum amount of irregularity, including jitter. Apparently the same is true for infant cries. As far as the LPC order is concerned, it appears that, in agreement with Wilde et al.'s findings, the inadequate approximation of the spectral envelope resulted in overly smoothed characteristics in the synthetic cries, which were perceptually amplified in the reverberant room, but were imperceptible when the cries were presented through headphones.

Experiment 5: Synthetic vs. natural cries

This experiment was done in order to justify the use of the computer generated cries in Experiment 6. Subjects were asked to decide whether each stimulus was natural or synthetic. I expected to find a lack of discriminability between the natural and the synthetic cries.

Method

Subjects The subjects were six female and five male graduate students at Brown University, ranging in age from 23 to 38 years, who volunteered their participation. Five of the females reported having experience in childcare. One male was the father of a two-year-old.

Stimuli Eight cries were chosen for processing, seven of which were also used in Experiments 2–4. Cry 6 of the previous series of experiments was not used, because of problems in synthesis, and another cry was used instead, which had received similar ratings in Experiment 1. The stimuli were synthesized with the method described in Chapter 2, using upsampling to convert them to effective sampling rate of 80KHz prior to the analysis. There were four manipulation conditions during synthesis, Pitch, Jitter, Risetime-P, and Risetime-T. The Pitch condition was identical to the Pitch condition of Experiments 2–4. Five levels of F_0 change were created, P10, P05, N00, P95, and P90, which again corresponded to F_0 approximately equal[27] to 90%, 95%, 100%, 105%, and 110% of the original corresponding cry respectively. There was a little jitter (as little as possible) in all these cries in order to make them sound natural. Table 10 shows the minimum jitter values that were used to synthesize all but the Jxx variants of the eight cries.

Cry	Range of fluctuations	
	samples	µsec
1	1.5	18.75
2	2.5	31.25
3	1.5	18.75
4	2.5	31.25
5	2.1	26.25
6	2.5	31.25
7	1.5	18.75
8	3.1	38.75

Table 10: Baseline period fluctuations for the eight cries

In the Jitter condition, there were again four levels (namely N00, J45, J70, and J90) but there were three differences from the Jitter cries used in Experiments 2, 3, and 4. First, the maximum amount of jitter in each series was higher, so as to cover a wider range comparable to that found in natural cries. Secondly, because of the finer temporal resolution, it was possible to synthesize periods with more accurately specified lengths, thus reducing the quantization effects. Finally, the N00 cry (which was the same in the Pitch condition) had some jitter, therefore the baseline was farther from zero jitter. The range of fluctuations allowed in each Jitter level were minimal (depending on th cry, see Table 10), ±56.25µsec, ±87.5µsec, and ±112.5µsec respectively. Given that the sampling period was 12.5µsec, the range of fluctuations for levels J45, J70, and J90 were 4.5 samples, 7.0 samples, and 9.0 samples respectively. The actual amount of jitter contained in these cries was measured and found to be comparable to that found in naturally occurring cries, covering the entire range of values measured in infant crying.

There were two manipulations involving the melodic contour of the fundamental frequency, in particular the time between the onset of phonation and the peak of F_0 within the cry, usually referred to as rise time of the cry. One manipulation was done with times relative to the duration of the cry and the other was done with absolute time (P and T for percent and time respectively). In the Risetime-P condition, the rise time of each cry was manipulated in three levels (MP10, MP25, and MP50) in relation to the total time of phonation. In level MP10 the rise time was fixed at 10% of the cry time, in level MP25 at 25%, and in level MP50 at 50%. Note that all of the parameters were moved in parallel, so that the local characteristics of the cry would remain unchanged, and only the overall pattern was warped. This way there was no issue of F_0 interaction with formants, or any of the problems one might run into when attempting to change the fundamental frequency only. See Chapter 2 for more details about the method of synthesis.

In the Risetime-T condition, there were again three levels (MT100, MT250, and MT500), wherein the rise time was fixed at 100msec, 250msec, and 500 msec respectively, regardless of the length of the vocalization. The same method was used as in the Risetime-P condition, only with absolute time parameters instead of percentages.

Of the 15 cries of each series (1 natural and 14 synthetic as described above), six synthetic and the natural one were used in this experiment. The same six variants were chosen from all conditions. I used N00, J90, P10, P90, MP50, and MT100 from each set, so as to cover the entire range of manipulations by testing the most extreme levels, which were judged by informal listening and from detailed analysis of the results of Experiment 2 to be those most likely to be considered artificially-sounding. The natural cries were used three times each, so as to bring the number of the "Natural" trials closer to that of the "Synthetic" trials. No natural cries other than those used to create the synthetic ones were used in this experiment. In all there were eight natural cries three times each plus six synthetic variants from each of eight cry series, adding up to 72 stimuli per session.

Procedure The procedure was identical to that of Part 1 of Experiment 2. Subjects listened to one cry per trial and were required to decide whether it was natural or synthetic and indicate their choice by pressing the appropriate key on the response box. There were 72 stimuli, each of which was judged twice (but not in consecutive trials). Because of this repetition each natural cry was in fact used six times, so I was able to obtain a reliable base level measure of naturalness judgements. The trials were presented in random order, and the subjects were instructed to respond within three

---

[27] The numbers in the level names refer to percentages of pitch period, as used in the synthesis program, and thus the percentages in terms of frequency are not accurate

seconds.

Results and discussion

The total number of responses in each case is shown in Table 11. (Subjects failed to respond within three seconds in 20 trials, which were excluded from the analysis). Clearly, subjects were overall 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. The individual d' scores for each cry ranged from -0.47 to 1.20, and the individual d' scores for subjects ranged from -0.22 to 1.02. The highest discriminability scores were those of the two males who had no experience with infants (d' equal to 1.02 and 0.94). There were 4 cries with d' greater than 1.0, and 7 with d' between 0.75 and 1.0. Of these 11 cries, 5 were in one cry group, that of Cry 6. Consequently, all stimuli of the Cry 6 series were excluded from the next experiment. Although there remained a cry with $d' = 1.03$, I chose to include it in the perception experiment, because the amount by which it exceeded the commonly accepted cutoff point of 1.0 was not large enough to justify rejection of a whole series of cries[28]. Most of the remaining cries were not discriminable from the natural ones at more than 0.5 level (d' score).

Experiment 6: Varying pitch, jitter, and rise time

The purpose of this experiment was twofold: First, to verify and extend the results of Experiment 4, by repeating the Pitch and Jitter conditions and including the rise time manipulations, and secondly, to test the hypothesis that perception of aperiodicities in synthetic stimuli is better when the stimuli are presented over a loudspeaker than when they are presented through headphones. The design of this experiment was identical to that of Experiment 4. I expected to replicate the results on the effects of F_0 on infant cry perception. I also expected to find an effect of jitter, if the presentation mode hypothesis were true for infant cries. In the Risetime conditions, I expected to find that cries with faster rising times would be perceived as being more aversive. According to the findings of Wasz-Höckert et al. (1968), a falling melody form was characteristic of pain cries, and a rising-falling melody form was characteristic of hunger cries. Given that a cry rarely starts off at its maximum pitch, usually taking some time to reach it, it seems that cries with faster rising times correspond to higher levels of arousal. This is exactly what Porter et al. (1986) reported 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. I expected to see this pattern reflected in subjects' ratings of the synthetic cries in the Risetime conditions.

Method

Subjects The subjects were 20 undergraduate students at Brown University, 8 males and 12 females. Their age ranged from 19 to 23. None of the subjects were parents, but several of them reported some experience in childcare. Informed consent was obtained from all subjects prior to the experiment. Each subject was paid $12 for his or her participation.

Stimuli The total number of stimuli in this experiment was 105, 98 of which were synthetic and the remaining seven were natural. They were divided into seven groups, namely Cry 1–Cry 5, Cry 7, and Cry 8. (Cry 6 was not used in this experiment because it was not judged to sound natural in Experiment 5.) Each group contained 14 synthetic cries, divided into four manipulation conditions, synthesized as described in the "Stimuli" section of Experiment 5. There were three cries in each of the two Risetime conditions, five cries in the Pitch condition, and four cries in the Jitter condition. Actually, because N00 was in both the Pitch and the Jitter conditions, there was a total of eight cries in these two conditions combined. The seven natural cries that had been originally analyzed were used so as to provide a measure of the level of perceptual overlap between them and the corresponding resynthesized ones. The ratings of the natural cries were excluded from all the analyses for effects of the manipulation conditions.

Procedure The procedure was similar to that of Experiment 4. Subjects heard one cry per trial and responded by pressing one of seven buttons arranged along an arc on a specially designed response device that was connected to the computer.

---

[28] It was not feasible to only discard that particular stimulus, because the design of Experiment 6 required that cries and levels be completely crossed.

Actual	Response				d' score
	Natural		Synthetic		
	N	Probability	N	Probability	
Natural	282	**0.55**	232	**0.45**	0.31
Synthetic	448	**0.43**	602	**0.57**	

Table 11: Results of the natural vs. synthetic cry discrimination task

The direction (polarity) of the rating scale was randomly selected for each subject. Each stimulus was presented twice during each rating session and the average of the two responses was used for the analysis. The order of the stimuli was randomized before each rating session.

The stimuli were presented to the subjects through a single loudspeaker positioned to face the subject directly. The sound level was adjusted at a comfortable level, and remained unchanged throughout the experiment. The low-pass anti-aliasing filters of the D/A system were set at 4.9KHz in one half of the sessions and at 9.8KHz in the other half, to test whether the absence of high frequencies from the signal affects jitter perception.

The same perceptual scales were used as in Experiment 4, i.e., “Urgency”, “Distress”, “Health”, and “Feeling”. The instructions given to subjects were to rate each cry on a scale from one to seven (by pressing the appropriate button) based on how urgent the cry sounded, how distressing the cry sounded, how sick the infant sounded, and how angry or sad the cry made them feel, respectively. Each subject rated all cries on one scale before proceeding to the next scale. The experiment was divided into two parts, each lasting about half an hour. The order of the four scales was randomized across subjects using a latin square design.

Results

Table 12 shows the correlation values between the ratings of the natural cries and the corresponding N00 cries. Note that the N00 cries are not always the closest to the natural ones, either in terms of acoustics or in terms of ratings, because the natural cries often contain more jitter than that contained in the N00 synthetic ones. This is particularly true for the worst sounding cry (natural Cry 8), which was rated on average around 6.5, whereas Cry8.N00 was rated around 5.9. This is one reason that the regression coefficient is less than one (one would indicate a perfect match, given these correlation values). The other reason has to do with the fact that the spectral flattening of the most aversive sounding cries that is caused by noise excitation produced at the glottal source cannot be perfectly accommodated in the LPC model with periodic excitation, no matter how many poles one uses. Other than this minor tendency of the analysis-synthesis procedure to decrease the perceptual range, there is a very satisfactory degree of correlation between the two sets of ratings in all conditions, indicating that ordering is well preserved.

Instruction condition	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

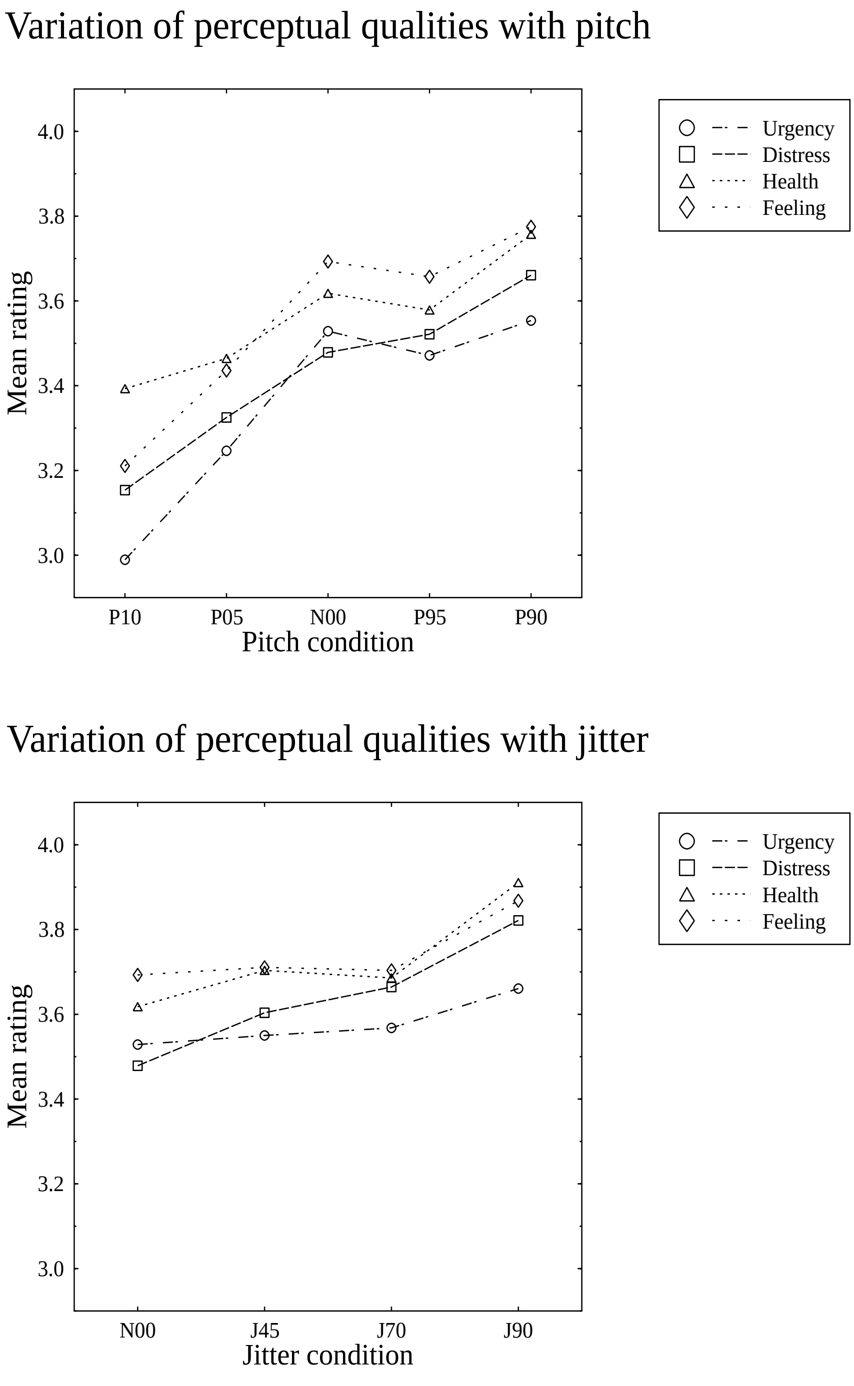
Table 12: Correlation of ratings of natural cries with ratings of corresponding N00 cries

The main effects of the four manipulation conditions in the four instruction conditions are shown graphically in Figures 7 and 8. Figure 7, top, shows the expected effect of F_0, whereby ratings increase as F_0 increases. Figure 7, bottom, shows that there is a consistent trend in the data towards higher ratings as the amount of jitter increases, in agreement with initial expectations. Figure 8 shows the effects of the rise time manipulations, which are very small, although Risetime-P does follow the expected pattern. Risetime-T only follows the pattern for small rising times, and that is excluding the Health ratings. Tables 13, 14, and 15 list the actual mean ratings in the different conditions.

Two-way univariate analyses of variance[29] (Cry×Level treated as fixed-effects within-subjects factors) for the effects of the different manipulation conditions showed that the effect of Pitch is always significant at the 0.005 level or less, and jitter is significant in all but the Urgency condition at the 0.04 level or less (see Table 16 on page 46 for the exact F and p values). Risetime-P

---

[29] The BMDP statistical package was used for all the analyses.



Variation of perceptual qualities with pitch

Variation of perceptual qualities with jitter

Figure 7: Effects of pitch and jitter in the different instruction conditions

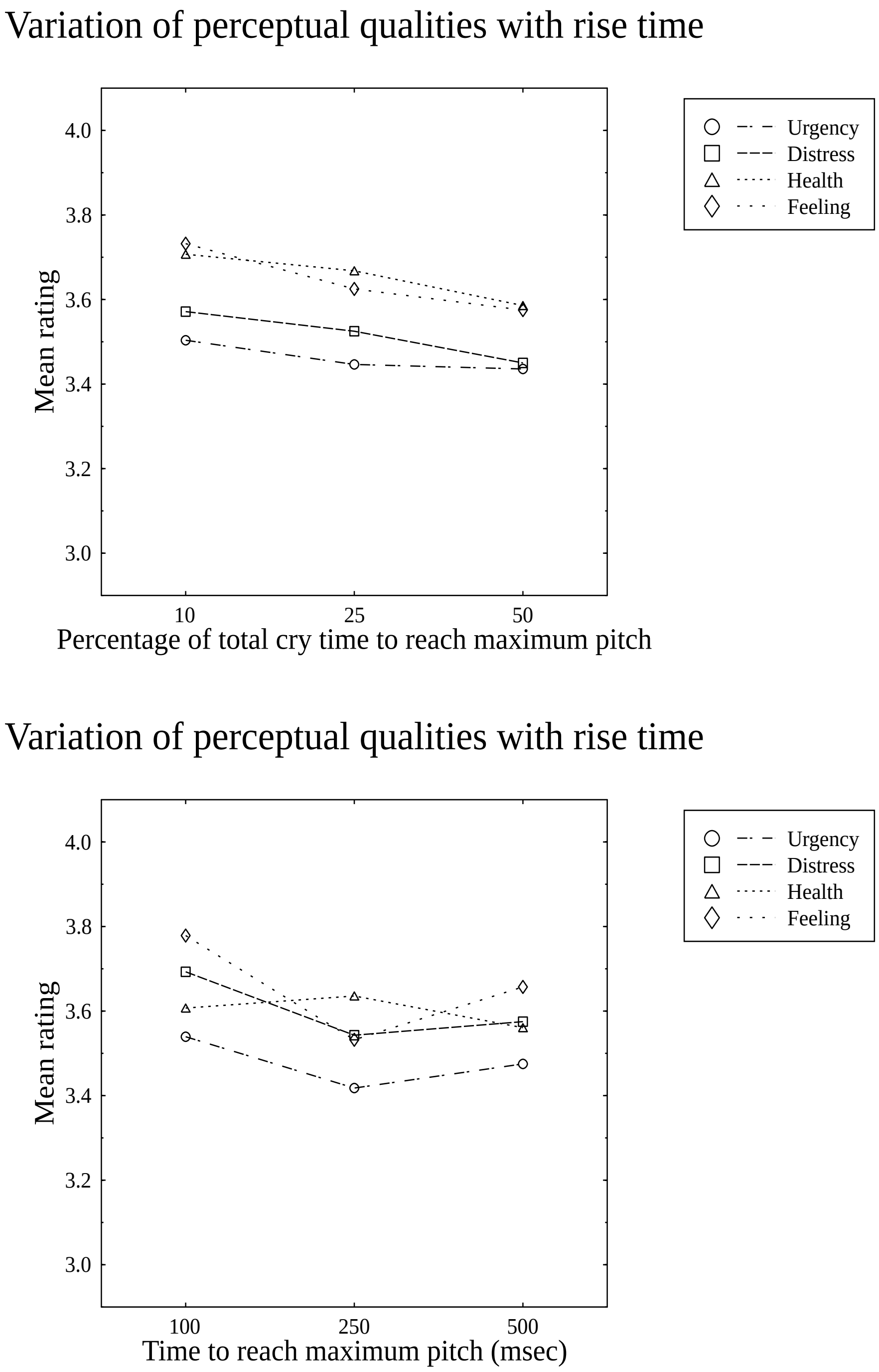


Figure 8: Effects of rise time (both manipulations) in the different instruction conditions

Pitch condition	Instruction Condition			
	Urgency	Distress	Health	Feeling
P10	2.99	3.15	3.39	3.21
P05	3.25	3.33	3.46	3.44
N00	3.53	3.48	3.62	3.69
P95	3.47	3.52	3.58	3.66
P90	3.55	3.66	3.76	3.78

Table 13: Mean ratings for different pitch levels (averaged across cries)

Jitter condition	Instruction Condition			
	Urgency	Distress	Health	Feeling
N00	3.53	3.48	3.62	3.69
J45	3.55	3.60	3.70	3.71
J70	3.57	3.66	3.69	3.70
J90	3.66	3.82	3.91	3.87

Table 14: Mean ratings for different jitter levels (averaged across cries)

is never significant, whereas Risetime-T is significant in two of the four instruction conditions at the 0.05 level or less. There is always a significant effect of Cry at the 0.00005 level, and a significant interaction of Cry with Level in some cases (see Table 17 on page 47), which means that each cry was rated differently, and that the effects of the acoustic manipulations in some cases depended on the initial characteristics of each cry and were not uniform. These values were calculated with the Huynh-Feldt method for adjusting the degrees of freedom (BMDP manual, program 4V).

Separate analyses with BMDP program 2V for the between-subjects factors (Cry×Level×Sex, Cry×Level×Session, and Cry×Level×Experience) showed no effect of Session (i.e., of cutoff frequency of the output filter) or of Experience, but showed a significant effect of Sex, usually at the 0.01 level, with males giving consistently higher ratings than females in all conditions. A test of Experience (for the Pitch condition) using data from females only (because most males had little or no experience) was performed and showed significant effect at the 0.05 level (see Table 18 on page 47).

Figures 9 and 10 show the effects of Pitch on each Cry[30] in the four instruction conditions. It is clear that the effect of changing F_0 is not the same in all cries, but tends to be consistent and more pronounced in the cries that received the highest overall ratings. Note that the Pitch manipulation appears to have similar effects in all four instruction conditions, although the p values indicate that the effect is less significant in the Health condition.

Figures 11 and 12 show the effects of Jitter on each cry in the four instruction conditions. The trend is more pronounced in the Distress and Feeling conditions, whereas the lines in the Urgency and Health conditions appear to be flatter and less consistent. Overall, although the effect is small, it is present, and the pooled data from all cries, as graphed in Figure 7, bottom, leave no doubt for that (the scale is expanded in Figure 7 so as to make the effect clearer). There is no interaction with Cry in any of the Jitter conditions, indicating that the jitter effects are uniform across cries.

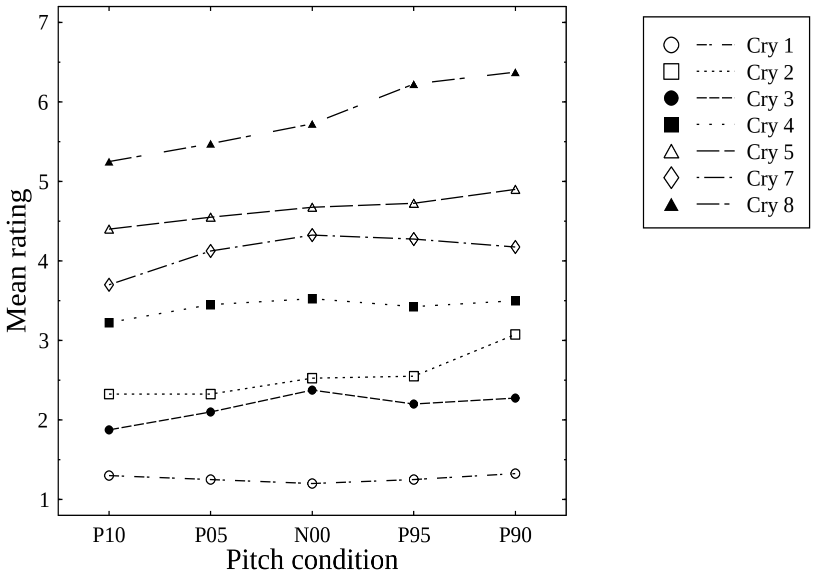
On the other hand, the effects of Risetime-P and Risetime-T are very small, and in addition, the effects of Melody-P are not significant and the effects of Risetime-T fail to show a linear trend. Figures 13 and 14 show the effects of Risetime-P on the individual cries in the different instruction conditions. Here there is no clear global effect, but several cries show a trend one way or the other. For some cries shorter rise time leads to higher ratings, whereas for others longer rise times lead to higher ratings. In all, most of the curves are quite flat, and it remains unlikely that there are any valid general conclusions.

Similarly, the effects of Risetime-T on the individual cries (Figures 15 and 16) look quite messy and fail to show a global pattern, although some curves show the originally expected trend. Nevertheless, the effects of Risetime-T levels are in two cases significant and this fact should be taken into

---

[30] Data for Cry 6 is not shown because Cry 6 was not used in this experiment due to unsatisfactory ratings in Experiment 5

Variation of perceived distress with pitch



Variation of perceived urgency with pitch

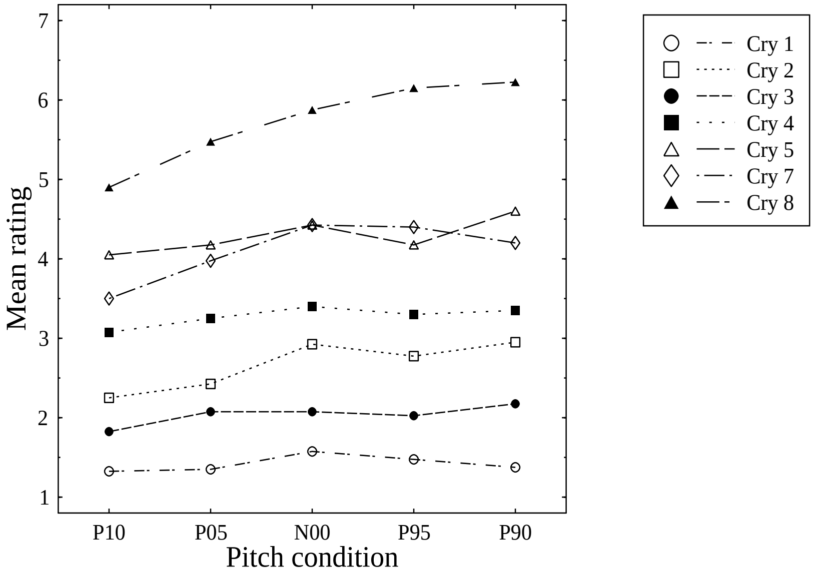
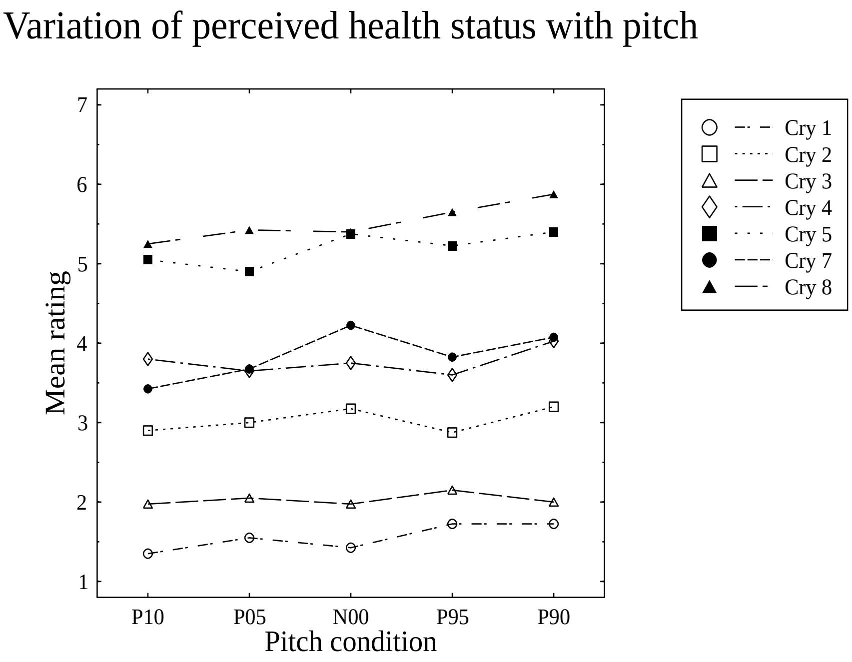


Figure 9: Effects of pitch on individual cries in two conditions



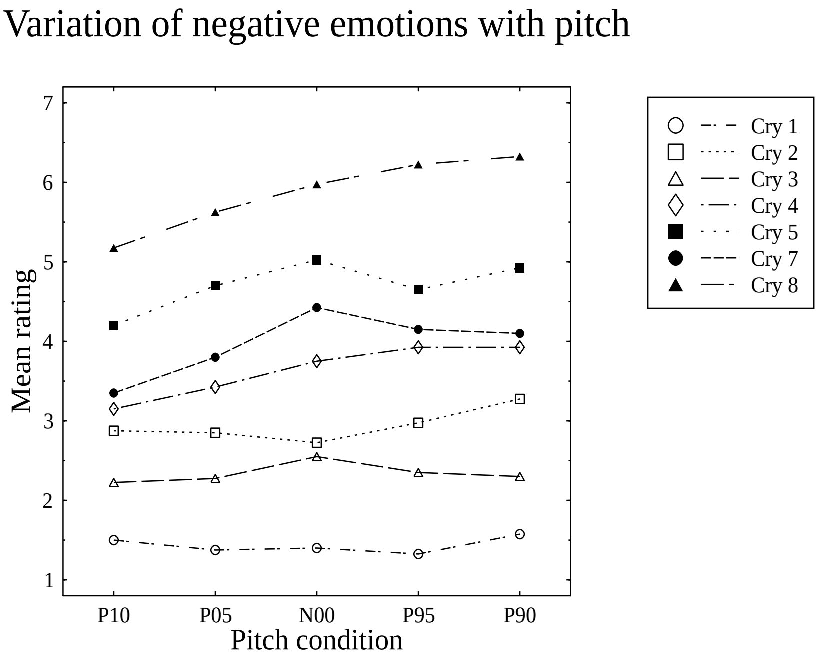


Figure 10: Effects of pitch on individual cries in two conditions

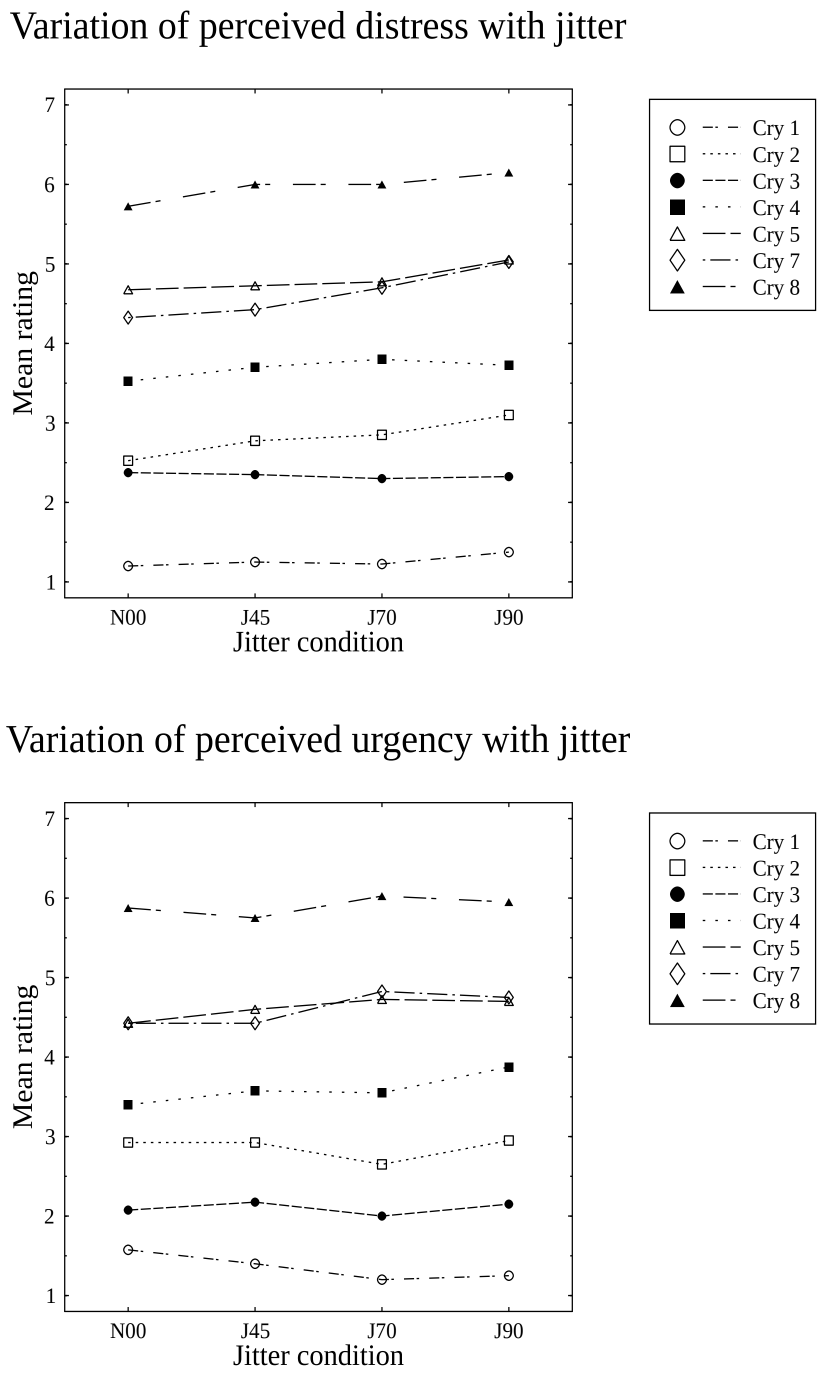
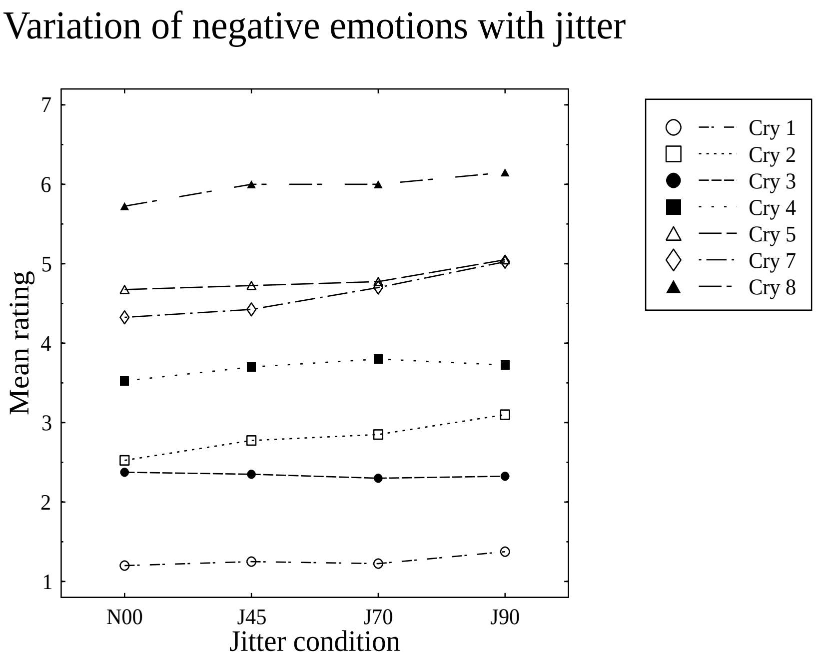


Figure 11: Effects of jitter on individual cries in two conditions


Variation of negative emotions with jitter

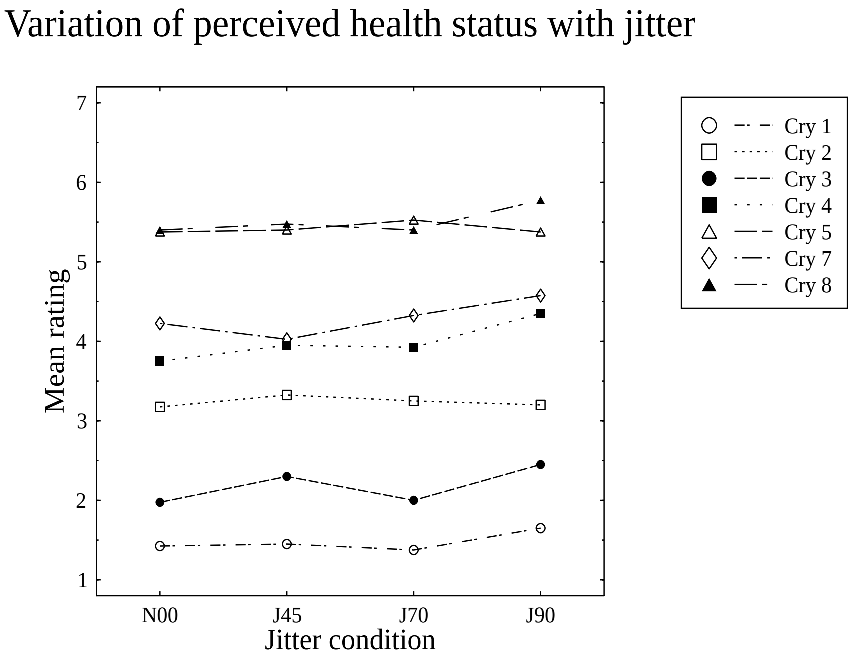
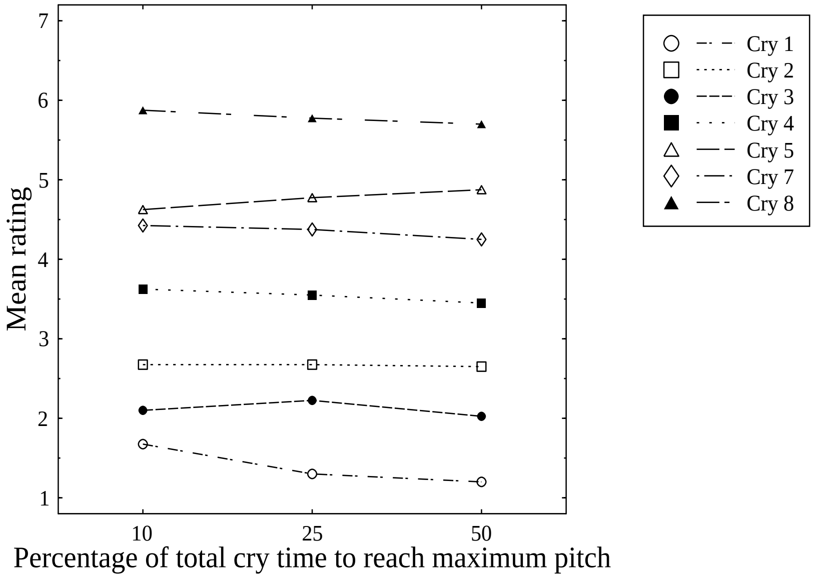

Variation of perceived health status with jitter

Figure 12: Effects of jitter on individual cries in two conditions

Variation of perceived distress with rise time



Variation of perceived urgency with rise time

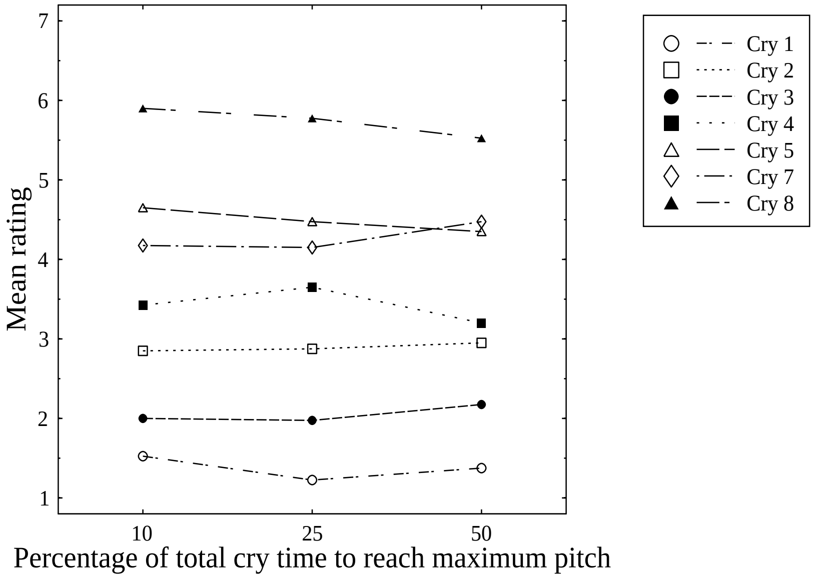
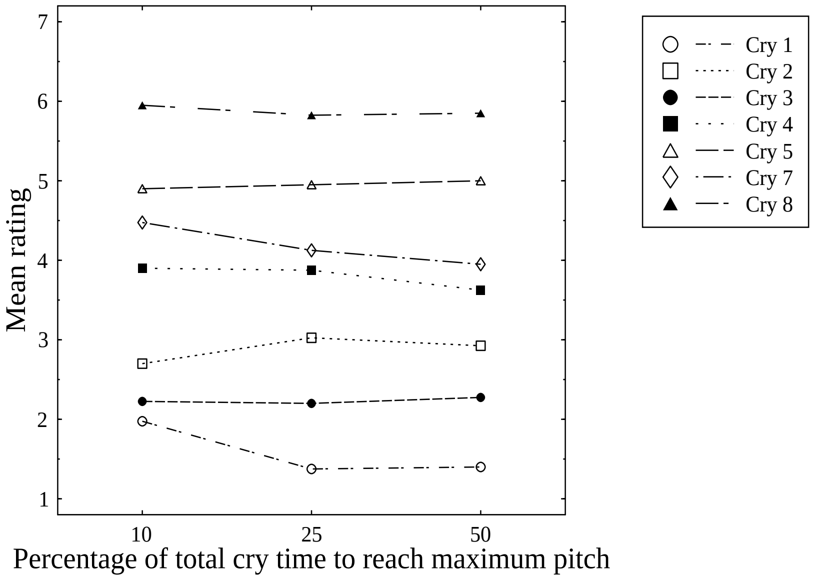
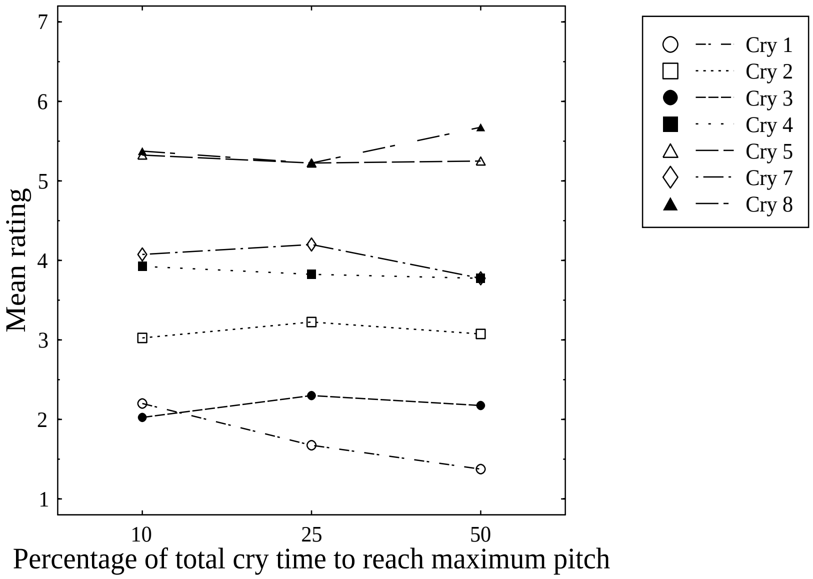


Figure 13: Effects of rise time (in percentage of total cry time) on individual cries in two conditions

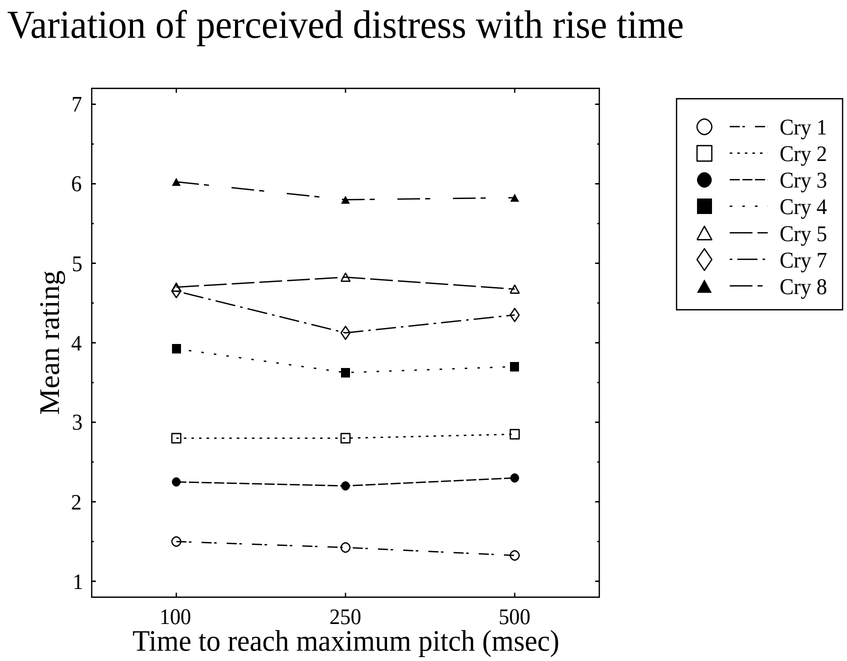


Variation of negative emotions with rise time



Variation of perceived health status with rise time

Figure 14: Effects of rise time (in percentage of total cry time) on individual cries in two conditions


Variation of perceived distress with rise time

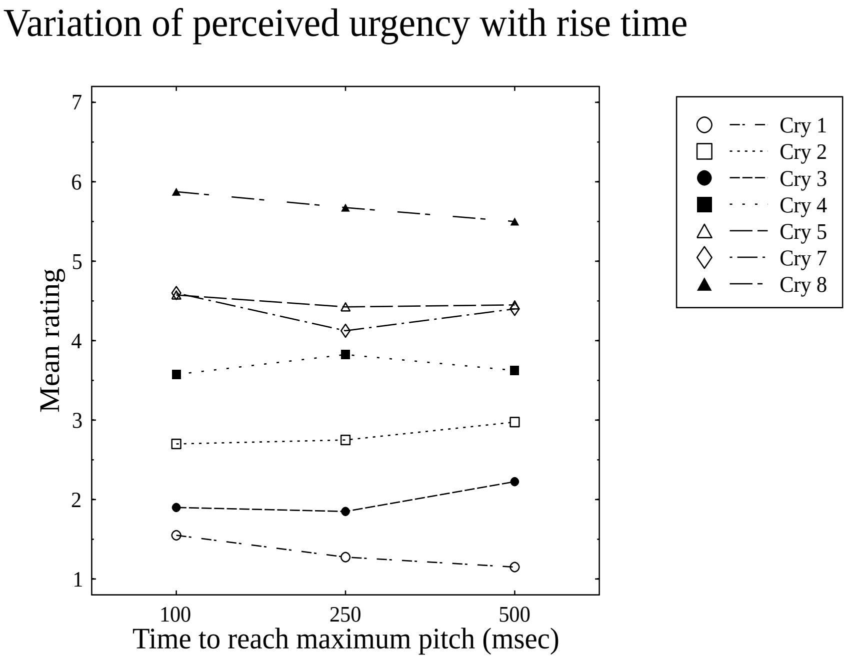

Variation of perceived urgency with rise time

Figure 15: Effects of rise time (in absolute time) on individual cries in two conditions

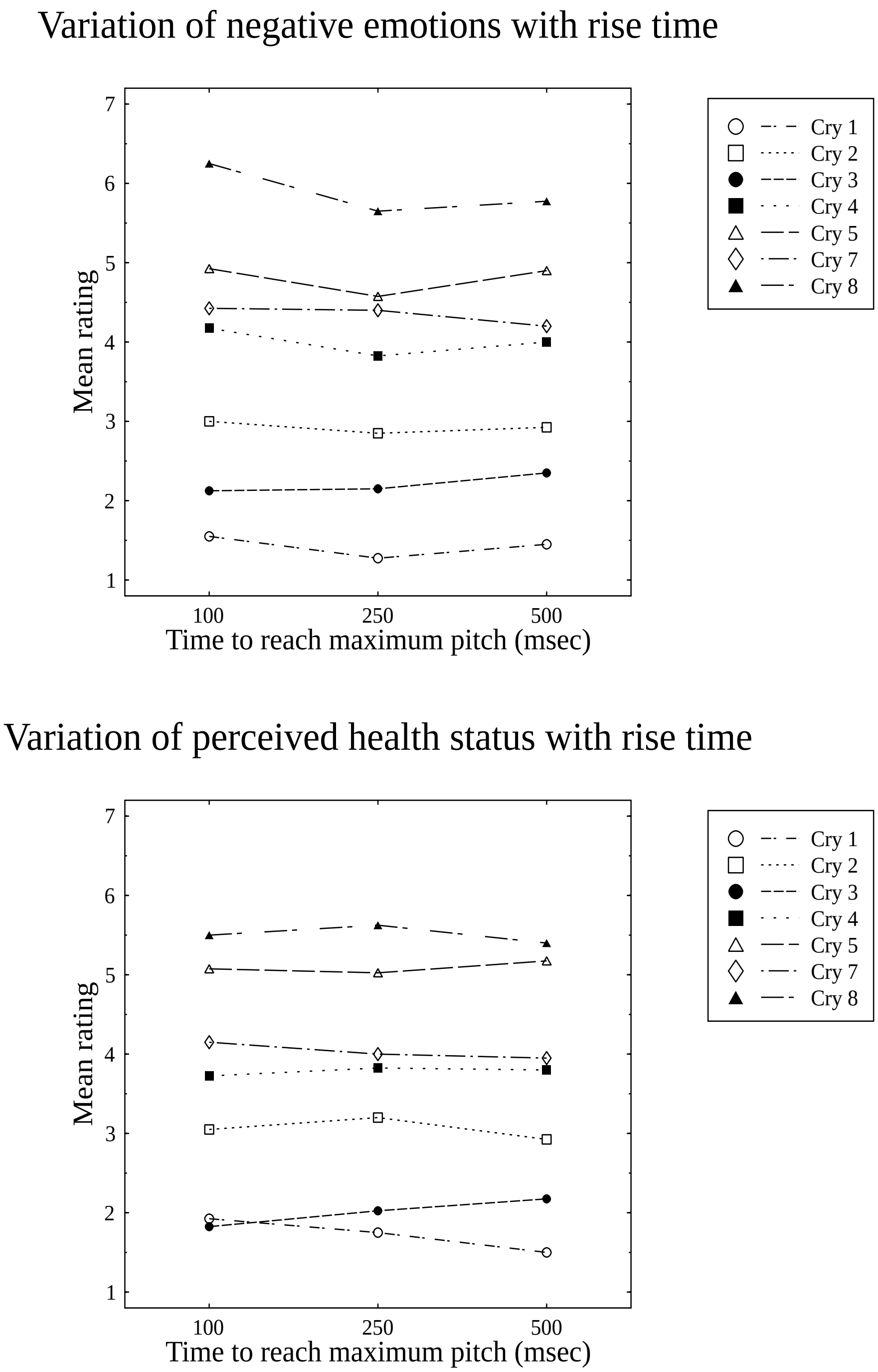


Figure 16: Effects of rise time (in absolute time) on individual cries in two conditions

Rise time condition	Instruction Condition			
	Urgency	Distress	Health	Feeling
MP10	3.50	3.57	3.71	3.73
MP25	3.45	3.53	3.67	3.63
MP50	3.44	3.45	3.59	3.58
MT100	3.54	3.69	3.61	3.78
MT250	3.42	3.54	3.64	3.53
MT500	3.48	3.58	3.56	3.66

Table 15: Mean ratings for different rise time levels of both manipulations (averaged across cries)

Acoustic manipulation condition	Instruction condition	Fixed-effects analysis			Mixed-model analysis		
		F	adj. *df*	*p*	Quasi-*F*	adj. *df*	*p*
Pitch	Urgency	19.67	3.7,69.7	**0.0000**	9.48	4,26	< **0.001**
	Distress	15.94	4.0,76.0	**0.0000**	7.59	4,22	< **0.001**
	Health	4.87	3.5,66.6	**0.0026**	4.70	4,23	< **0.01**
	Feeling	18.30	4.0,76.0	**0.0000**	7.42	4,21	< **0.001**
Jitter	Urgency	1.21	2.5,46.7	0.3134	0.93	3,16	> 0.25
	Distress	7.88	2.6,50.2	**0.0004**	7.19	3,20	< **0.005**
	Health	4.14	2.5,47.6	**0.0152**	4.15	3,21	< **0.025**
	Feeling	3.14	2.9,55.2	**0.0337**	2.53	3,9	> 0.10
Risetime-P (percent)	Urgency	0.50	2.0,38.0	0.6098	0.31	2,11	> 0.25
	Distress	1.88	1.7,32.6	0.1735	2.04	2,7	> 0.10
	Health	0.88	2.0,38.0	0.4227	0.48	2,14	> 0.25
	Feeling	1.87	1.7,31.7	0.1757	1.22	2,12	> 0.25
Risetime-T (absolute)	Urgency	1.38	2.0,38.0	0.2634	0.87	2,12	> 0.25
	Distress	3.33	2.0,38.0	**0.0465**	5.98	2,2	> 0.10
	Health	0.47	2.0,37.5	0.6276	0.68	2,5	> 0.25
	Feeling	6.73	2.0,38.0	**0.0032**	7.33	2,5	< **0.05**

Table 16: Results of the significance tests for the effects of acoustic manipulations in all conditions

account.

Given these main effects of the various factors, 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-factor in the ANOVA only indicates significance of effects in the particular items that were used. Because the cries were chosen so as to cover a wide perceptual range, and because their acoustic characteristics are typical of clear phonated cries yet diverse enough within that class, we can assume that they constitute a representative sample of all clearly phonated cries of comparable duration. We can then test the significance of the effects of the acoustic manipulations in a mixed-model design, where Cry is considered a random factor. The significant interaction between the random and the fixed factor that was found in some cases from the previous analysis precludes generalizations that are based on that analysis alone (Maxwell and Delaney, 1990). Therefore, program 8V of the BMDP package was used to calculate the appropriate mean square terms, and the method of Quasi-F (Myers, 1979) was used to calculate the appropriate denominator terms and degrees of freedom. There were no between-subject factors in this test. The results are shown in Table 16, where the significant effects are printed in bold type. Note that Pitch is significant in all conditions, and jitter is significant in two of the four conditions, namely Distress and Health. The effects of Risetime-P never approach significance, which was expected because the mixed-model analysis requires a larger effect to reach a small p value, because of the extra error terms previously included in the numerator mean squares that are now accounted for. Finally, note that Risetime-T has a significant effect in the Feeling condition, and I will

Acoustic manipulation condition	Instruction condition	Effect of Cry			Interaction with acoustic manipulation		
		F	adj. df	p	F	adj. df	p
Pitch	Urgency	101.62	4.5,86.3	0.0000	2.27	17.0,323.7	**0.0030**
	Distress	134.36	4.0,76.1	0.0000	2.10	17.0,323.8	**0.0070**
	Health	71.29	4.0,75.0	0.0000	1.04	16.8,318.2	0.4110
	Feeling	83.79	4.9,92.5	0.0000	2.37	16.8,318.2	**0.0020**
Jitter	Urgency	131.84	4.1,78.5	0.0000	1.30	13.1,249.7	0.2102
	Distress	112.07	4.1,77.6	0.0000	1.12	13.8,261.5	0.3373
	Health	62.34	4.1,77.4	0.0000	1.00	15.5,294.3	0.4590
	Feeling	104.19	4.8,90.7	0.0000	1.17	16.4,311.4	0.2878
Risetime-P (percent)	Urgency	83.04	3.5,67.0	0.0000	1.64	9.7,183.7	0.1007
	Distress	104.08	3.6,68.0	0.0000	0.93	12.0,228.0	0.5163
	Health	59.40	5.0,94.8	0.0000	2.14	11.9,225.5	**0.0162**
	Feeling	87.10	4.9,93.7	0.0000	1.62	10.1,192.2	0.1032
Risetime-T (absolute)	Urgency	98.93	4.9,92.8	0.0000	1.89	11.4,216.2	**0.0401**
	Distress	110.87	4.7,89.7	0.0000	0.70	11.6,220.3	0.7413
	Health	61.62	4.4,82.8	0.0000	0.74	8.7,164.9	0.6629
	Feeling	88.69	5.3,101.2	0.0000	0.94	10.9,206.1	0.5040

Table 17: Results of the significance tests for the effects of Cry and their interaction with the acoustic manipulations in all conditions

Acoustic manipulation condition	Instruction condition	Effect of Sex			Effect of Experience (females only)		
		F	df	p	F	df	p
Pitch	Urgency	13.13	1,18	0.0019	1.09	2,9	0.3761
	Distress	6.83	1,18	0.0176	5.94	2,9	0.0227
	Health	6.53	1,18	0.0199	1.17	2,9	0.3526
	Feeling	9.33	1,18	0.0068	5.09	2,9	0.0332
Jitter	Urgency	16.36	1,18	0.0008	—	—	—
	Distress	6.47	1,18	0.0204	—	—	—
	Health	4.15	1,18	0.0565	—	—	—
	Feeling	10.77	1,18	0.0041	—	—	—
Risetime-P (percent)	Urgency	18.98	1,18	0.0004	—	—	—
	Distress	4.65	1,18	0.0449	—	—	—
	Health	10.19	1,18	0.0051	—	—	—
	Feeling	8.83	1,18	0.0082	—	—	—
Risetime-T (absolute)	Urgency	12.17	1,18	0.0026	—	—	—
	Distress	7.35	1,18	0.0143	—	—	—
	Health	5.09	1,18	0.0367	—	—	—
	Feeling	5.23	1,18	0.0345	—	—	—

Table 18: Results of the significance tests for the effects of Sex and Experience in all conditions (only data from female subjects were used in this test for Experience effects and only the Pitch condition was tested)

return to that when I discuss the implications of my findings.

Discussion and conclusion

Several subjects reported that it was hard not to remember the previous cries during the rating sessions. They said that they had learned the seven basic patterns and had assigned a rating to each one of them. Some of them said that they were just trying to be consistent. Because of the global characteristics of each cry, it was very easy to identify the basic patterns and effectively ignore minor variations in pitch or otherwise, thus highly suppressing the effects that were expected in the ratings. This has to be kept in mind when evaluating the size of the effects of the manipulations. It should also be a lesson for the future. A greater range of cries should be used, and each subject should be used in one rating session only. This way the potential for memorizing the cries and ignoring the relatively small differences between the different versions will be minimized. Another solution would be to employ a between-subjects design, so that each subject will not hear different versions of any given cry. Unfortunately that would be more risky, because within-group designs have been found to create reliable differences not found in between-subjects comparisons (Zeskind and Huntington, 1984).

The statistical analysis of the ratings in the Pitch condition leaves no doubt that F_0 plays an important role in infant cry perception. In agreement with previous findings, I have shown that increases in F_0 lead to perception of cries as being more urgent, more distressing, coming from a sicker infant, and making the listener feel increasingly angry or sad. However, the size of the effect is relatively small, and it appears that fundamental frequency of phonation might not be the primary factor in determining the aversiveness of a cry. There are many other aspects of crying that play a role in determining its perceptual effect. Perhaps a consonance of several features is necessary to elicit major shifts in listeners' response[31]. Because of the interactions of several systems during cry production, it is likely that many acoustic parameters are simultaneously affected by changes in the infant's state. Thus a global perceptual effect is produced, which might not be directly decomposable into parts that can be safely ascribed to individual acoustic features. Nevertheless, it is very important to understand the first order effects before we can investigate interaction effects, so as to establish the necessary reference points for meaningful research. From this perspective, this study constitutes a major contribution to the field, because it shows the perceptual effects of individual acoustic features in a controlled and replicable way.

The perceptual effects of short term F_0 variability were also successfully demonstrated. Given recent findings on the potential value of jitter analysis for diagnostic purposes and the correlation of F_0 variability with the short-term affect state and the long-term disposition of infants, there are many issues related to jitter that are worth investigating. The effect size of the Jitter manipulation was found to be small yet significant, and there is a clear trend in the data towards higher ratings as the jitter content increases. This finding complements similar findings in adult speech research and contributes in filling part of the gap in the perturbations literature. This is the first time that a perceptual effect of infant cry jitter has been demonstrated. It remains to be found what kinds of jitter are present in natural cries, and how each kind correlates with production mechanisms and perceptual impact. Studies using synthetic stimuli would be ideal in the investigation of the latter part.

The results on the melodic contour manipulations (rise time conditions) do not lend themselves to clear interpretation. In the case of Risetime-P the expected trend is found, but the size of the effect is very small and statistically insignificant. In the Risetime-T manipulation there is a significant effect but the lack of linear trend in the expected direction creates reluctance to draw definite conclusions. Clearly, the initial expectations were not met in either Risetime case, and it is not clear whether this is a result of an improper methodology or some property of the cries. The effects of the infants' state on the rise time of the cry are documented enough that I believe it is more likely that there is some problem with the method than with the cries. It is also very likely that rise time is not an indicator of the infants state *per se*, but it correlates with the actual indicators because of constraints in the cry production systems.

A conclusion that one may draw given the findings I have presented is that no single acoustic parameter of crying is likely to have major effects on the listeners' perception of crying. Even F_0, which has usually been treated as a major acoustic determinant of aversiveness fails to show a big

---

[31] On the other hand, it might be the case that the 10% change in F_0 was just a small change and that in fact the perceptual difference observed is quite substantial.

effect. Moreover, rise time, also frequently considered a reliable indicator of cry aversiveness, fails to show a clear effect at all. More detailed analyses of the cry acoustics are necessary before one can answer what in the acoustics elicits all the responses discussed in Chapter 1. Furthermore, there are a number of methodological issues that are still far from being resolved.

One methodological issue which has concerned many researchers pertains to the appropriateness of using short cry segments, or even single-expiration vocalizations in perceptual studies. Some argue that removing a brief vocalization from its context greatly diminishes the perceptual effect and does not allow conclusions to be drawn for real-life effects of infant crying. Others posit that it is unreasonable to use long cry bouts, because they are so variable that no meaningful data can be extracted from them. I believe that there is truth in both sides, and that this is not an issue to be resolved easily. It might be most reasonable to utilize both approaches and attempt to tie in the results of one with those of the other. For example, this study used short vocalizations, consisting of a single expiration. The conclusions that can be safely drawn are valid only for short, out of context vocalizations, and thus indicate a lower-level effect in comparison to results from studies using whole cry episodes. In studies such as this one, the results are not generalizable to aspects of infant-caretaker communication or parenting. However, they provide well defined guidelines for higher-level research, which could not be found without examining the most elementary features of crying, i.e., the acoustic features.

Another methodological issue, which is not as much a subject of debate, concerns the use of rating scales for investigating the perceptual effects of individual cries. Although it has been generally considered a major improvement over earlier studies of cry categorization by eliciting stimuli, the use of cry rating might not have very much to offer beyond consistent results, because such ratings are far removed from situations where the listener has a chance to take action and possibly affect the factors that led to crying. There is also little account for different listener states. Moreover, it is not clear, and certainly not theoretically justified, which rating scales should be used and why, or how to interpret the ratings and put them in the appropriate perspective. Again, this seems to be another issue of levels, where ratings attempt to reveal specific elements of perceptual significance, whereas categorization and reaction monitoring are better suited to investigations of global effects. I believe that it is important to have some idea of the elements to guide research of the whole. Naturally, this need not be true, if gestalt phenomena prevail. This is a critical question in itself, and only by comparing the results of low level studies with those of high level studies will it be possible to gain some insight into the matter.

One of the goals of this study was to show that it is practically possible and theoretically appropriate to use artificially produced infant cries for this kind of research. In other words, whether it can be done with the available technology, and whether it should be done in terms of generalizeability of the results to natural cry perception. I believe that both questions have been successfully addressed, and the answer is in both cases affirmative. I have applied a widely used, relatively well behaved algorithm to infant cry processing and succeeded in producing stimuli that sounded very much like infant cries. Furthermore, I showed experimentally, for the first time, that these artificial stimuli were indeed indistinguishable from natural cries. The correlation of the ratings of the synthetic cries with those of the natural cries leaves no doubt that, at least in this kind of experiment, it is appropriate to use artificially produced stimuli.

Finally, I showed that the mode of presentation of the stimuli influenced the perceptual effect of jitter. This result is in agreement with what is known about perception of synthesized adult voices, and should certainly be taken into account in any future studies. In fact, even if F_0 variability is not under investigation, it would be advisable to use loudspeakers for presenting artificially created auditory stimuli anyway, because interactions with perceived roughness might influence the results. Also, it is not known whether any other aspects of the acoustics are perceived differently when heard through headphones.

To conclude, this study has addressed a number of methodological and theoretical issues, and has introduced a method of investigating infant cry perception by using digital signal processing techniques to alter the acoustics of cries. I have solved a number of problems associated with processing the cry signals so as to produce acceptable, natural sounding stimuli, and I have found significant effects of individual acoustic features in adults' perception of infant cries. Further research is necessary to reveal the role of the acoustics in more detail including more acoustic parameters, as well as combinations of parameters. By integrating such findings with those from studies on other levels we can greatly improve our understanding of the communicative functions of infant crying, with possible implications for developmental research and parenting.

References

Adachi, T., Murai, N., Okada, H., and Nihei, Y. (1985). Acoustic properties of infant cries and maternal perception. *Tohoku Psychologica Folia*, 44:51–58.

Aldrich, C. A., Sung, J., and Knopp, C. (1945). The crying of newly born babies. *Journal of Pediatrics*, 26:313.

Anand, K. J. S. and Hickey, P. R. (1987). Pain and its effects in the human neonate and fetus. *The New England Journal of Medicine*, 317(21):1321–1329.

Atal, B. S. and Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, 50(2):637–655.

Bennett, S. (1971). Infant-caretaker interactions. *American Academy of Child Psychiatry Journal*, 10:321–335.

Beyer, J. E. and Wells, N. (1989). The assessment of pain in children. *Pediatric Clinics of North America*, 36(4):837–854.

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

Bleichfeld, B. and Moely, B. E. (1984). Psychophysiological responses to an infant cry: Comparison of groups of women in different phases of the maternal cycle. *Developmental Psychology*, 20(6):1082–1091.

BMDP Statistical software, Inc. (1990). *BMDP statistical software manual, Vols. 1 & 2*. University of California Press, Berkeley, CA.

Boukydis, C. F. Z. (1980). An analog study of adult physiological and self-report responding to infant cries. *Cry Research Newsletter*, 2(2):2–5.

Boukydis, C. F. Z. (1985). Perception of infant crying as an interpersonal event. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 9, pages 187–215. Plenum Press, New York.

Boukydis, C. F. Z. and Burgess, R. L. (1982). Adult physiological response to infant cries: Effects of temperament of infant, parental status, and gender. *Child Development*, 53:1291–1298.

Brennan, M. and Kirkland, J. (1982). Classification of infant cries using descriptive scales. *Infant Behavior and Development*, 5:341–346.

Brennan, M. and Kirkland, J. (1983). Perceptual dimensions of infant cry signals: A semantic differential analysis. *Perceptual and Motor Skills*, 57:575–581.

Brennan, M. and Kirkland, J. (1985). Comparison of perceptual dimensions uncovered from infant cry signals using the method of pair-comparison and the semantic differential. *Scandinavian Journal of Psychology*, 26:12–20.

Brenner, M., Branskomb, H. H., and Schwartz, G. E. (1979). Psychological stress evaluator — two tests of a vocal measure. *Psychobiology*, 16(4):351–357.

Bryan, Y. E. and Newman, J. D. (1988). Influence of infant cry structure on the heart rate of the listener. In J. D. Newman (Ed.), *The physiological control of mammalian vocalizations*. Plenum Publishing Corporation.

Cismaresco, A.-S. and Montagner, H. (1990). Mothers' discrimination of their neonates' cry in relation to cry acoustics: The first week of life. *Early Child Development and Care*, 65:3–11.

Cohen, A. and Zmora, E. (1984). Automatic classification of infants' hunger and pain cry. In V. Cappellini and A. G. Constantinides (Eds.), *Digital Signal Processing — 84*, pages 667–672. Elsevier Science Publications, Amsterdam.

Colton, R. H., Steinschneider, A., Black, L., and Gleason, J. (1985). The newborn infant cry: Its potential implications for development and SIDS. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 6, pages 119–137. Plenum Press, New York.

Deem, J. F., Manning, W. H., Knack, J. V., and Matesich, J. S. (1989). The automatic extraction of pitch perturbation using microcomputers: some methodological considerations. *Journal of Speech and Hearing Research*, 32:689–697.

Dunn, J. (1977). *Distress and Comfort*, chapter 3, pages 17–25. Harvard University Press, Cambridge, Mass.

Fisichelli, V. R., Coxe, M., Rosenfeld, L., Haber, A., Davis, J., and Karelitz, S. (1966). The phonetic content of the cries of normal infants and those with brain damage. *Journal of Psychology*, 64:119–126.

Fisichelli, V. R., Karelitz, S., Fisichelli, R. M., and Cooper, J. (1974). The course of induced crying activity in the first year of life. *Pediatric Research*, 8:921–928.

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.

Flanagan, J. L. (1972). *Speech Analysis, Synthesis, and Perception.* Springer-Verlag, New York, 2nd edition.

Freeburg, T. J. and Lippman, M. Z. (1986). Factors affecting discrimination of infant cries. *Journal of Child Language*, 13:3–13.

Friedman, S. L., Zahn-Waxler, C., and Radke-Yarrow, M. (1982). Perception of cries of full-term and preterm infants. *Infant Behavior and Development*, 5:161–173.

Frodi, A. (1985). When empathy fails: Aversive infant crying and child abuse. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 12, pages 263–277. Plenum Press, New York.

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

Frodi, A. M. and Lamb, M. E. (1978). Fathers' and mothers' responses to the faces and cries of normal and premature infants. *Developmental Psychology*, 14(5):490–498.

Frodi, A. M. and Lamb, M. E. (1980). Child abusers' responses to infant smiles and cries. *Child Development*, 51:238–241.

Frodi, A. M., Lamb, M. E., Leavitt, L. A., and Donovan, W. L. (1978). Fathers' and mothers' responses to infant smiles and cries. *Infant Behavior and Development*, 1:187–198.

Fuller, B. F. (1991). Acoustic discrimination of three types of infant cries. *Nursing Research*, 40(3):156–160.

Fuller, B. F. and Horii, Y. (1986). Differences in fundamental frequency, jitter, and shimmer among four types of infant vocalizations. *Journal of Communication Disorders*, 19:441–447.

Fuller, B. F. and Horii, Y. (1988). Spectral energy distribution in four types of infant vocalizations. *Journal of Communication Disorders*, 21:251–261.

Fuller, B. F., Horii, Y., and Conner, D. (1989). Vocal measurements of infant pain. In S. G. Funk, E. M. Tornquist, M. T. Champagne, L. A. Copp, and R. A. Wiese (Eds.), *Key aspects of comfort*, pages 46–51. Springer, New York.

Gardiner, W. (1838). *The music of nature.* Wilkins and Carter, Boston.

Gladding, S. T. (1979). Effects of training versus non-training in identification of infant cry-signals: A longitudinal study. *Perceptual and Motor Skills*, 48:752–754.

Golub, H. L. and Corwin, M. J. (1982). Infant cry, a clue to diagnosis. *Pediatrics*, 69:197–201.

Golub, H. L. and Corwin, M. J. (1985). A physioacoustic model of the infant cry. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 3, pages 59–81. Plenum Press, New York.

Grauel, E. L., Höck, S., and Rothgänger, H. (1990). Jitter-index of the fundamental frequency infant cry as a possible diagnostic tool to predict future development problems, Part 2: Clinical considerations. *Early Child Development and Care*, 65:23–29.

Green, J. A. and Gustafson, G. E. (1983). Individual recognition of human infants on the basis of cries alone. *Developmental Psychobiology*, 16:485–493.

Green, J. A. and Gustafson, G. E. (1990). Interrelations among the acoustic features of cries: How many features do we need? *Early Child Development and Care*, 65:31–44.

Green, J. A., Jones, L. E., and Gustafson, G. E. (1987). Perception of cries by parents and nonparents: Relation to cry acoustics. *Developmental Psychology*, 23(3):370–382.

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

Gustafson, G. E., Green, J. A., and Tomic, T. (1984). Acoustic correlates of individuality in the cries of human infants. *Developmental Psychobiology*, 17:311–324.

Gustafson, G. E. and Harris, K. L. (1990). Women's responses to young infants' cries. *Developmental Psychology*, 26(1):144–152.

Hess, W. (1983). *Pitch determination of speech signals.* Springer series in information sciences. Springer-Verlag, Berlin Heidelberg.

Hillenbrand, J. (1988). Perception os aperiodicities in synthetically generated voices. *Journal of the Acoustical Society of America*, 83(6):2361–2371.

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

Hollien, H., Michel, J., and Doherty, E. T. (1973). A method for analyzing vocal jitter in sustained phonation. *Journal of Phonetics*, 1:85–91.

Inbar, G. F. and Eden, G. (1976). Psychological stress evaluators: EMG correlation with voice tremor. *Biological Cybernetics*, 24:165–167.

Johnston, C. C. and O'Shaughnessy, D. (1988). Acoustical attributes of infant pain cries: Discriminating features. In R. Dubner, G. F. Gebhart, and M. R. Bond (Eds.), *Proceedings of the Vth World Congress on Pain*, pages 336–340. Elsevier, Amsterdam.

Johnston, C. C. and Strada, M. E. (1986). Acute pain responce in infants: A multidimensional description. *Pain*, 24:373–382.

Kaplan, H. L., MacMillan, N. A., and Creelman, C. D. (1978). Methods and designs, tables of d' for variable-standard discrimination paradigms. *Behavior Research Methods & Instrumentation*, 10(6):796–813.

Karnell, M. P., Scherer, R. S., and Fischer, L. B. (1991). Comparison of acoustic voice perturbation measures among three independent laboratories. *Journal of Speech and Hearing Research*, 34:781–790.

Kirkland, J. (1990). Adults' perceptual shifts between infant fussing and crying: A proposition. *Early Child Development and Care*, 65:77–82.

Konner, M. (1972). Aspects of a developmental ethology of a foraging people. In N. Blurton-Jones (Ed.), *Ethological studies of child behavior.* Cambridge University Press, Cambridge, England.

Kuroda, I., Fujiwara, O., Okamura, N., and Utsuki, N. (1976). Method for determining pilot stress through analysis of voice communication. *Aviation, Space, and Environmental Medicine*, 47(5):528–533.

Lester, B. M. (1984). A biosocial model of infant crying. In L. Lipsitt and C. Rovee-Collier (Eds.), *Advances in infancy research Vol. 3*, pages 167–212. Ablex, Norwood, NJ.

Lester, B. M. (1985). There's more to crying than meets the ear. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 1, pages 1–27. Plenum Press, New York.

Lester, B. M. (1987). Developmental outcome prediction from acoustic cry analysis in term and preterm infants. *Pediatrics*, 80(4):529–534.

Lester, B. M., Boukydis, C. F. Z., Garcia-Coll, C. T., Hole, W., and Peucker, M. (1992). Infantile colic: Acoustic cry characteristics, maternal perception of cry, and temperament. *Infant Behavior and Development*, 15:15–26.

Lester, B. M., Boukydis, C. F. Z., Garcia-Coll, C. T., Peucker, M., Vohr, B. R., Brem, F., Anderson, L., and Oh, W. (1993). Developmental outcome as a function of the goodness of fit between the infant's cry characteristics and the mother's perception of her infant's cry. *Child Development.* Submitted for publication.

Lester, B. M., Corwin, M. J., Sepkoski, C., Seifer, R., Peucker, M., McLaughlin, S., and Golub, H. L. (1991). Neurobehavioral syndromes in cocaine-exposed newborn infants. *Child development*, 62:694–705.

Lester, B. M., Garcia-Coll, C. T., and Valcarcel, M. (1989). Perception of infant cries in adolescent and adult mothers. *Journal of Youth and Adolescence*, 18(3):231–243.

Lester, B. M. and Zeskind, P. S. (1982). A biobehavioral perspective on crying in early infancy. In *Theory and research in behavioral pediatrics, Vol. 1*, pages 133–180. Plenum Press, New York.

Levine, J. D. and Gordon, N. C. (1982). Pain in prelingual children and its evaluation by pain-induced vocalization. *Pain*, 14:85–93.

Lieberman, P. (1961). Perturbations in vocal pitch. *The Journal of the Acoustical Society of America*, 33(5):597–603.

Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *The Journal of the Acoustical Society of America*, 35(3):344–353.

Lieberman, P. (1980). On the development of vowel production in young children. In G. Yeni-Komshian and J. Kavanaugh (Eds.), *Child Phonology: Perception and production*, pages 113–142. Academic Press, New York.

Lieberman, P. (1985). The physiology of cry and speech in relation to linguistic behavior. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 2, pages 29–57. Plenum Press, New York.

Lieberman, P. and Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics.* Cambridge studies in speech science and communication. Cambridge University Press, Cambridge.

Lieberman, P. and Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *The Journal of the Acoustical Society of America*, 32(7):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 Development and Care*, 65:83–90.

Lundh, P. (1986). A new baby-alarm based on the tenseness of the cry signal. *Scandinavian Audiology*, 15:191–196.

Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580.

Markel, J. D. and Gray, Jr., A. H. (1976). *Linear Prediction od Speech.* Springer-Verlag, Berlin Heidelberg.

Maxwell, S. E. and Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*, chapter 10. Wadsworth publishing company, Belmont, California.

Mead, M. and Newton, N. (1967). Cultural patterning of perinatal behavior. In S. Richardson and A. Guttmacher (Eds.), *Childbearing: Its social and psychological aspects.* Williams and Wilkins, Baltimore, MD.

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 Development and Care*, 65:95–107.

Michelsson, K. (1980). Cry characteristics in sound spectrographic cry analysis. In T. Murry and J. Murry (Eds.), *Infant Communication: Cry and Early Speech*, chapter 4, pages 85–105. College-Hill Press, Houston, Texas.

Michelsson, K., Järvenpää, A. L., and Rinne, A. (1983). Sound spectrographic analysis of pain cry in preterm infants. *Early Human Development*, 8:141–149.

Michelsson, K., Paajanen, S., Rinn, A., Tervo, H., Kunnamo, I., and Kiminkin, T. (1990). Mothers' perceptions of and feelings towards their babies' crying. *Early Child Development and Care*, 65:109–116.

Michelsson, K., Raes, J., Thoden, C. J., and Wasz-Höckert, O. (1982). Sound spectrographic cry analysis in neonatal diagnostics. an evaluative study. *Journal of Phonetics*, 10:79–88.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research*, 30:529–538.

Milenkovic, P. H. (1992). Voice source model for continuous control of pitch period. *NCVS Status and Progress report*, pages 39–58.

Müller, E., Hollien, H., and Murry, T. (1974). Perceptual responces to infant crying: Identification of cry types. *Journal of Child Lnaguage*, 1:89–95.

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

Murray, A. D. (1985). Aversiveness is in the mind of the beholder: Perception of infant crying by adults. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 10, pages 217–239. Plenum Press, New York.

Murry, T., Amudson, P., and Hollien, H. (1977). Acoustical characteristics of infant cries: Fundamental frequency. *Journal of Child Language*, 4:321–328.

Muta, H., Baer, T., Wagatsuma, K., Muraoka, T., and Fukuda, H. (1988). A pitch-synchronous analysis of hoarseness in running speech. *The Journal of the Acoustical Society of America*, 84(4):1292–1301.

Myers, J. L. (1979). *Fundamentals of experimental design*, chapter 7, pages 188–192. Allyn and Bacon, Inc., Boston, MA, third edition.

Nakazima, S. (1980). The reorganization process of babbling. In T. Murry and J. Murry (Eds.), *Infant Communication: Cry and Early Speech*, chapter 12, pages 272–283. College-Hill Press, Houston, Texas.

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.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-time signal processing*, chapter 7, pages 444–457. Prentice Hall, Englewood Cliffs, New Jersey.

Owens, M. E. (1984). Pain in infancy: Conceptual and methodological issues. *Pain*, 20:213–230.

Papoušek, M. (1989). Determinants of responsiveness to infant vocal expression of emotional state. *Infant Behavior and Development*, 12:507–524.

Pinto, N. B. and Titze, I. R. (1990). Unification of perturbation measures in speech signals. *Journal of the Acoustical Society of America*, 87(3):1278–1289.

Pollack, I. (1968). Detection and relative discrimination of auditory "jitter". *The Journal of the Acoustical Society of America*, 43(2):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 Development*, 57:790–802.

Rabiner, L. R. and Schafer, R. W. (1978). *Digital processing of speech signals.* Prentice-Hall signal processing series. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632.

Rosenberg, A. E. (1966). Pitch discrimination of jittered pulse trains. *The Journal of the Acoustical Society of America*, 39(5):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 Development and Care*, 65:145–152.

Saito, S. and Nakata, K. (1985). *Fundamentals of Speech Signal Processing.* Academic Press, Orlando, Florida.

Scherer, K. R. (1981). Vocal indicators of stress. In J. Darby (Ed.), *Speech evaluation in psychiatry*, pages 171–187. Grune & Stratton, New York.

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

Sherman, M. (1927). The differentiation of emotional responses in infants. *Journal of Comparative Psychology*, 7:265–351.

Shimura, Y., Imaizumi, S., and Shutou, T. (1992). Perceptual evaluation of emotional aspects of infants' vocalizations. Annual Bulletin 26, Research Institute of Logopedics and Phoniatrics, University of Tokyo.

Smith, G. A. (1977). Voice analysis for the measurement of anxiety. *British journal of Medical Psychology*, 50:367–373.

Strobach, P. (1990). *Linear Prediction Theory.* Springer series in information sciences. Springer-Verlag, Berlin Heidelberg.

Thodén, C.-J., Järvenpää, A.-L., and Michelsson, K. (1985). Sound spectrographic analysis of pain cry in prematures. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 5, pages 105–117. Plenum Press, New York.

Thodén, C.-J. and Koivisto, M. (1980). Acoustic analyses of the normal pain cry. In T. Murry and J. Murry (Eds.), *Infant Communication: Cry and Early Speech*, chapter 5, pages 124–151. College-Hill Press, Houston, Texas.

Titze, I. R., Horii, Y., and Scherer, R. C. (1987). Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research*, 30:252–260.

Tsakalos, N. and Zigouris, E. (1991). Autocorrelation-based pitch determination algorithms for real time vocoders with the TMS320/C25. *Microprocessors and Microsystems*, 14(8):511–516.

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

Tsukamoto, T. and Tohkura, Y. (1992). Tempo as a perceptual cue for judgement of infant cries. *Perceptual and Motor Skills*, 74:258.

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

Wasz-Höckert, O., Michelsson, K., and Lind, J. (1985). Twenty-five years of scandinavial cry research. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 4, pages 83–103. Plenum Press, New York.

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

Wiggins, S. L., McCranie, M. L., and Bailey, P. (1975). Assessment of voice stress in children. *The Journal of Nervous and Mental Disease*, 160(6):402–408.

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*, pages 219–225, San Francisco. Computer Music Association.

Williams, C. E. and Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, 52(4):1238–1250.

Wolfe, V., Cornell, R., and Palmer, C. (1991). Acoustic correlates of pathologic voice types. *Journal of Speech and Hearing Research*, 34:509–516.

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

Wolff, P. H. (1969). The natural history of crying and other vocalizations in early infancy. In M. Foss (Ed.), *Determinants of Infant Behavior IV*, pages 81–109. Methuen, London.

Wong, D., Ito, M. R., Cox, N. B., and Titze, I. R. (1991). Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. *The Journal of the Acoustical Society of America*, 89(1):383–394.

Zajac, D. J. and Linville, R. N. (1989). Voice perturbations of children with perceived nasality and hoarseness. *The Cleft Palate Journal*, 26(3):226–232.

Zeskind, P. S. (1980). Adult responces to cries of low and high risk infants. *Infant Behavior and Development*, 3:167–177.

Zeskind, P. S. (1981). Behavioral dimensions and cry sounds of infants of differential fetal growth. *Infant Behavior and Development*, 4:297–306.

Zeskind, P. S. (1983a). Cross-cultural differences in maternal perception of cries of low- and high-risk infants. *Child Development*, 54:1119–1128.

Zeskind, P. S. (1983b). Production and spectral analysis of neonatal crying and its relation to other biobehavioral systems in the infant at risk. In T. Field and A. Sostek (Eds.), *Infants born at risk*, chapter 2, pages 23–43. Grune & Stratton, New York.

Zeskind, P. S. (1985). A developmental perspective on infant crying. In B. M. Lester and C. F. Z. Boukydis (Eds.), *Infant Crying*, chapter 8, pages 159–185. Plenum Press, New York.

Zeskind, P. S. (1987). Adult heart-rate responses to infant cry sounds. *British Journal of Developmental Psychology*, 5:73–79.

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 Development*, 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 Development*, 49:580–589.

Zeskind, P. S. and Lester, B. M. (1981). Cry features of newborns with differential fetal growth. *Child Development*, 52:207–212.

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

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 Development*, 56:549–554.