# CHAPTER 1

## ACOUSTIC ANALYSES OF DEVELOPMENTAL CHANGES AND EMOTIONAL EXPRESSION IN THE PREVERBAL VOCALIZATIONS OF INFANTS


### Abstract

The nonverbal vocal utterances of seven normally hearing infants were studied within their first year of life with respect to age- and emotion-related changes. Supported by a multi-parametric acoustic analysis, it was possible to distinguish one inspiratory and eleven expiratory call types. Most of the call types appeared within the first two months; some emerged in the majority of infants not until the 5th (laugh) or 7th month (babble). Age-related changes in acoustic structure were found in only 4 call types (discomfort cry, short discomfort cry, wail, moan). The acoustic changes were characterized mainly by an increase in harmonic-to-noise ratio and homogeneity of the call, a decrease in frequency range and a downward shift of acoustic energy from higher to lower frequencies. Emotion-related differences were found in the acoustic structure of single call types as well as in the frequency of occurrence of different call types. A change from positive to negative emotional state was accompanied by an increase in call duration, frequency range and peak frequency (frequency with the highest amplitude within the power spectrum). Negative emotions, in addition, were characterized by a significantly higher rate of cry, hic and ingressive vocalizations than positive emotions, while positive emotions showed a significantly higher rate of babble, laugh and raspberry.

## Introduction

In the past, research on infant vocal development has focused almost completely on verbal development, that is, the question of how meaningful referential speech arises out of preverbal utterances (Locke 1993; Murai 1963; Nakazima 1962; Oller 1980; Papoušek 1994; Stark 1980; Tonkova-Yampol'skaya 1969). Only very little work has been done on the development of nonverbal emotional vocal utterances on their own. The present study, therefore, will focus on this question.

Most authors agree that there are essentially five preverbal stages in the infant’s vocal development (Oller 1978; Oller 1980; Papoušek 1994; Stark 1980; Tonkova-Yampol'skaya 1969). In the first stage (0-1 month), infants produce sounds that do not show any systematic contrast between opening and closing of the vocal tract, nor do they make use of the full potential of the vocal cavity as a resonating tube (Oller 1980). In the second stage (2-3 months), infants start to produce an increasing number of vocalizations with differing modulations. Also, at this age, the first appearance of laughing occurs (Stark 1980). Typical for the third stage (4-6 months) is the exploration of the vocal tract capacities, called ‘vocal play’ by Stark (1980), in the form of squealing, growling or yelling sounds, production of noises by blowing air, food or saliva through oropharyngeal constrictions, and nasal murmurs (Stark 1980). In the fourth stage (7-10 months), infants begin to produce regular syllables (‘reduplicated babbling’, according to Stark (1980); ‘canonical babbling’, according to Oller (1978; 1980)). Typical for this period are sequences, like ‘bababa’ or ‘dadada’. During the fifth stage (11-12 months), infants produce alternating syllables with differing consonantal and vocalic elements (‘variegated babbling’, according to Stark (1980); ‘nonreduplicated babbling’, according to Oller (1980); ‘play with complex vocal patterns’, according to Papoušek and Papoušek (1981).

Vocal development studies usually characterize the vocalizations of infants phonetically. The small number of studies that describe the physical structure of preverbal vocalizations focused mainly on crying behavior. According to Wasz-Höckert et al. (1968), changes in the acoustic structure of cries can be seen only in the first 3-4 days of life. No changes could be detected by the authors in the following seven months. Zlatin-Laufer and Horii (1977) report that infants display an increasing range and standard deviation of fundamental frequencies as well as an increasing duration of ‘nonreflexive’ vocalizations during the second to fifth month of life.

Very few studies also dealt with the question as to which extent differences in the acoustic structure of infant vocalizations code differences in the underlying emotional state. Only crying was investigated more extensively. The results are controversial, however. According to Wasz-Höckert et al. (1968), four types of cries can be discriminated: birth, hunger, pain and pleasure cries. Muller et al. (1974) and Murry et al. (1975), in contrast, report that in their studies, mothers were not able to identify the stimulus that elicited crying in their infant. Other authors suggest that infant crying is a graded signal, mirroring a continuum of emotional states reaching from arousal to urgency (Brennan & Kirkland 1982; Porter et al. 1986; Protopapas & Eimas 1997; Zeskind et al. 1985). Porter et al. (1986), for instance, investigated infant cries during painful circumcision procedures. They found correlations between the acoustic structure of the elicited cries and the degree of intrusiveness of the particular procedure. In the perceptual test, the cries elicited by the most intrusive procedures were judged by adult listeners also to be the most urgent ones.

According to Keller and Schölmerich (1987), parents interpret the vocalizations of even two-weeks old infants as expressions of emotional states and respond in a differentiated way to different vocalizations. There is no general agreement, however, whether infants during their first months of life are able to express specific emotions in their behavior at all (for an overview, see Strongman 1996). Some authors suggest that, in the first months, the expressive behavior is to a large extent random. This assumption is based on the observation that specific vocal or facial patterns often occur without a specific stimulus preceding them. One reason for this low predictability of specific behavior patterns could be a low degree of emotional differentiation in early infanthood. Most authors agree that from the very beginning, there is a differentiation into at least two emotional states: aversive and nonaversive (Giblin 1981; Izard & Malatesta 1987; Lewis 1993; Malatesta-Magai et al. 1991; Sroufe 1979). Some authors, however, argue that more than two emotional states can be distinguished (Giblin 1981; Izard & Malatesta 1987; Malatesta-Magai et al. 1991).

Ratings of emotional speech intonations in adults revealed that, to a certain extent, emotions do influence the acoustic structure of vocal utterances (Banse & Scherer 1996; Murray & Arnott 1993). If rating errors occur, they are not randomly distributed. Confusion happens on three dimensions of similarity: quality, intensity, and valence. For example, anxiety and panic fear are often confused because of similar quality; hot anger (the intense form of anger – in contrast to cold anger) and panic fear (the intense form of fear – in contrast to anxiety), which have a different quality, are confused because of similar intensity; with respect to valence, positive emotions are more likely confused with each other than with negative emotions (Banse & Scherer 1996).

There is some disagreement as to which are the specific acoustic variables encoding different emotions (Banse & Scherer 1996; Murray & Arnott 1993; Pittam et al. 1990). Anger, hot anger, and fear seem to be characterized by an increase in level and variability of the fundamental frequency ($F_0$), increase in high-frequency energy, and $F_0$ range. However, joy is characterized by nearly the same acoustic features. Sadness, in its quiet, subdued form, usually shows a decrease in mean $F_0$, $F_0$ range and high-frequency energy as well as a downward-directed $F_0$ contour. The findings for disgust are inconsistent. In some studies, disgust was associated with an increase in mean $F_0$, while others reported a decrease in mean $F_0$ (for a review, see Pittam et al. 1990). It is also unclear which of the acoustic parameters code valence and quality of emotional states and which code unspecific arousal.

In the present study, a multiparametric acoustic analysis has been carried out with the aim to characterize the vocal expression of emotion and its development in preverbal infants. Apart from such common parameters as fundamental frequency, frequency range, intensity, and duration, a number of additional parameters describing the energy distribution in the frequency spectra, were used. These parameters had turned out in recent studies to be useful in the description of subtle differences in nonhuman primate vocalizations (Fichtel et al. 2001; Fischer et al. 2001). In the following, we will first characterize the developmental changes during the first year of life, focusing on those call types, for which we had a sufficient number of samples for all infants. In a second step, we will look for emotion-related differences in acoustic structure and call type usage.

## Material and Methods

### Subjects

The 7 infants selected to participate in the investigation consisted of 5 boys and 2 girls, all members of middle-class families. All parents of the infants were native German speakers. All infants were born at term and healthy. Contact was made through two cooperating pediatricians, who asked the parents in the third medical check-up of the infants at week four to six whether they were interested in participating in the study. After the parents gave their consent, the infants were examined in the Department of Phoniatrics and Pedaudiology of the

Georg-August University of Göttingen to make sure that they had normal hearing. The examinations included a complete otorhinolaryngological status, sound-field audiometry, tympanometry, acoustic reflex threshold, and measurement of transient evoked otoacoustic emissions.

**Vocal recordings**

The vocalizations of the infants were recorded in intervals of 4-6 weeks during the course of one year. Recordings started at the age of 7-10 weeks and ended at the age of 53-58 weeks (see Table 1). One recording from infant 4 and two recordings from infant 5 are missing. The recordings were made with Sony WM TCD-100 DAT recorders and Sennheiser directional microphones (K6 power module and ME64 recording head).

***TABLE 1.*** *Age in weeks of the recorded infants at the end of each recording session. Each recording session lasted one week.*

| Recording No. | Infant 1 | Infant 2 | Infant 3 | Infant 4 | Infant 5 | Infant 6 | Infant 7 |
|---|---|---|---|---|---|---|---|
| 1 | 7 | 10 | 7 | 9 | 7 | 9 | 8 |
| 2 | 14 | 14 | 14 | 15 | 16 | 15 | 14 |
| 3 | 20 | 20 | 20 | 21 | | 22 | 20 |
| 4 | 26 | 26 | 25 | | 24 | 26 | 25 |
| 5 | 31 | 33 | 31 | 29 | 32 | 32 | 31 |
| 6 | 38 | 37 | 38 | 36 | | 39 | 37 |
| 7 | 45 | 45 | 47 | 46 | 47 | 49 | 44 |
| 8 | 54 | 53 | 54 | 56 | 56 | 58 | 53 |

To obtain a comprehensive vocal repertoire for the infants, the parents themselves recorded their children in familiar surroundings, after an introduction to the recording method. Each session lasted one week and contained recordings of vocalizations from 11 defined situations (Appendix 1) of normal infant life. The parents were instructed orally and in writing how to record the situations. Each of the situations had to be recorded twice during the course of one week. For each recorded situation the parents had to name the emotion they assumed their infant expressed, choosing between joy, contentment, interest, surprise, unease, anger, and pain.

**Acoustic analysis**

In order to extract acoustic parameters correlating with the emotional state, we carried out a multiparametric analysis. First, the vocalizations were inspected for quality and digitized, using RTS 2.0 (Engineering Design, Belmont, MA). Only calls of good quality and low background noise were taken. Depending on the quality of the recordings, we took 20-30 calls from each recording (more specifically, we took this amount of calls from each recording of the 11 defined situations; since each situation had to be recorded twice, we digitized 20-30 calls out of each of 22 recordings per infant and month). This resulted in a total sample size of 16,322 digitized vocalizations. If the recording of a situation contained more than 20 calls of good quality and low background noise, half of the calls digitized were chosen from the beginning of the recording and the other half from the end. Sampling frequency was 30 kHz. For each call type, two fast Fourier transformations (1024 points) were conducted with SIGNAL 3.0 (Engineering Design, Belmont, MA): the first with a frequency range of 4 kHz, the second with a frequency range of 12 kHz, resulting in a frequency resolution of 10 and 29 Hz, respectively. Time resolution was 10 ms in both cases. The resulting frequency-time spectra were analyzed with LMA 9.2 (developed by K. Hammerschmidt). LMA is a software tool to extract different sets of call parameters from acoustic signals (Hammerschmidt 1990; Hammerschmidt & Fischer 1998). We used the spectra with the better frequency resolution (frequency range: 4 kHz, frequency resolution: 10 Hz) to calculate the fundamental frequency ($F_0$) and parameters related to $F_0$ and its variations. For the calculation of parameters describing the energy distribution, we used the spectra with the higher frequency range (frequency range: 12 kHz, frequency resolution: 29 Hz).

Parameter calculations were carried out in the following way. First, we calculated an autocorrelation function for each time segment in a given vocalization (segment length: 10 ms). Depending on the number of peaks and the periodicity of the autocorrelation function, each time segment was classified as either noisy (no peaks could be detected), complex (some peaks could be detected, but they were not periodic), or tonal (periodic peaks). We then determined the percentage of time segments that were tonal (tonality). If a time segment was classified as tonal we determined the fundamental frequency ($F_0$). Additionally, we calculated from the tonal time segments call parameters, such as the mean and maximum harmonic-to-noise ratio (HNR), the mean of the frequency with the highest amplitude (peak frequency = PF tonal), and the coefficient of peak to fundamental frequency ($PF/F_0$ coeff.; see also Appendix 2).

Second, we measured the statistical distribution of the frequency amplitudes across the spectrum (hereafter distribution of frequency amplitudes = dfa). From these values, the frequency was calculated at which the dfa reached the first, second and third quartile of the total distribution, respectively (Figure 1). Third, we calculated a set of parameters describing the first three dominant frequency bands (dfb). The dominant frequency bands are characterized by amplitudes that exceed a given threshold as calculated from the adjacent frequency bins. Note that the numbers of the dominant frequency bands count from the lowest frequency upwards. The first dfb (dfb1) is not necessarily the dfb with the highest amplitude. In tonal calls or call segments, like in Figure 1, the first dfb corresponds to the fundamental frequency ($F_0$).

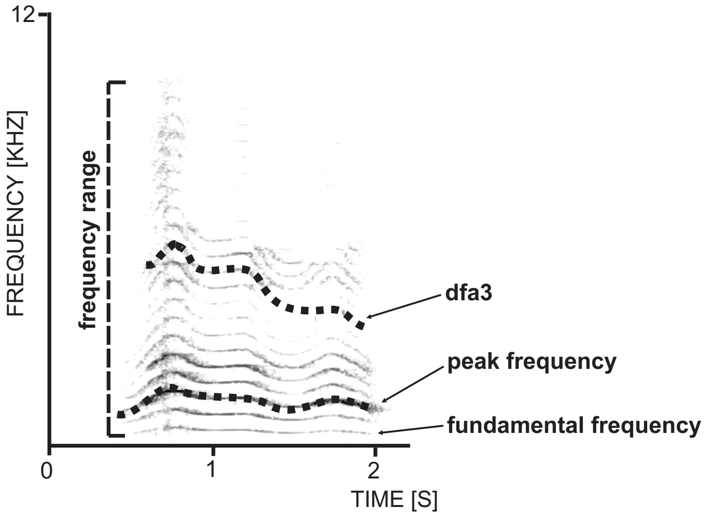


**FIGURE 1.** Examples of the kind of parameters extracted from the digitized spectrograms. The dashed lines mark the different values of peak frequency and dfa3 across the call. In this case the fundamental frequency corresponds to the dfb1. For further details, see Appendix 2.

Fourth, we determined the global energy distribution in the calls (formant like structures). Fifth, we specified the location and the modulation of the peak frequency (= frequency of highest amplitude) over the entire call including the noisy time segments. Sixth, we determined the mean and maximum frequency range. A description of the algorithms is given in Schrader and Hammerschmidt (1997).

Additionally, we conducted a linear prediction coefficient (LPC) analysis on a subset (n=807) of sampled vocalizations using SoundScope (GW Instruments, Somerville, MA). In this approach, we calculated 12 LPC coefficients for tonal parts of 100 ms duration from a 4096 pt FFT spectrum at a frequency range of 4000 Hz. To obtain a better characterization of the higher amplitude parts, we used a pre-emphasis of 6 dB/octave. The following parameters were calculated: fundamental frequency ($F_0$), jitter, shimmer, breathiness and harmonic-to-noise ratio. A comparison with the parameters calculated by LMA showed that there was no advantage of LPC over LMA in describing age- or emotion-related differences.

**Categorization of call types**

The infants produced a great variety of different nonverbal vocalizations. Their repertoire contains harmonic as well as noisy calls, calls with numerous overtones, as well as calls with only two or three overtones. Some calls had a rhythmic frequency or amplitude modulation; other calls lacked frequency or amplitude modulation almost completely. Most of these calls occur in numerous variants, sometimes forming a vocal continuum between different call types. Therefore, an exclusively mathematical procedure, such as cluster analysis, was not helpful to establish different call types (for detailed discussion see Hammerschmidt and Fischer (1998). Based on call type descriptions in the literature and the call variation of the own sample, we defined eleven call types. We established multiparametric limits for every call type. In this way, the calls could be assigned to one of the eleven call types, according to their frequency-time characteristics. Calls that did not fit in this classification were put into a ‘rest’ group. The following call types were distinguished: (1) cry, (2) short cry, (3) coo/wail, (4) moan, (5) babble, (6) whoop/squeal, (7) laugh, (8) hic, (9) groan, (10) croak and (11) raspberry. Additionally, we described a special type of vocalization, the ingressive vocalization (=IV), which differs from all other call types by its inspiratory rather than expiratory nature. For detailed description of the call types, see Results.

**Definition of emotional category and subdivision of call types**

As mentioned before, the parents had the choice between seven different emotions for describing the emotional state of their infants during recordings**.** These seven emotions were further classified into positive and negative emotions. The emotions pain, unease and anger were classified as negative emotions, the emotions surprise, interest contentment and joy as positive emotions. If useful for the analysis, the call types mentioned before were subdivided according to their occurrence in positive or negative emotional context. The call type coo/wail was subdivided in coo (positive) and wail (negative); cry was subdivided in pleasure cry and discomfort cry; short cry was divided into short pleasure cry and short discomfort cry; and whoop/squeal was divided into whoop (positive) and squeal (negative). For the other call types, we used the signs + (positive emotion) and - (negative emotion) behind the name of the call type to indicate the emotional context in which the call was uttered (e.g.: moan$^{+}$ /moan$^{-}$).

**Statistical analysis**

For the statistical analysis of the age-related and emotion-related changes, we first reduced the number of call parameters by eliminating strongly correlating parameters (Spearman rho rank correlation, two-tailed). Whenever there were two or more highly correlating parameters, we aimed at retaining the most basic ones. For instance, we preferred parameters describing the fundamental frequency over parameters describing the second or third harmonic.

Second, we used a stepwise discriminant function analysis to identify the most important parameters describing the age-related and emotion-related differences in vocalization. Discriminant function analysis identifies a linear combination of quantitative predictor variables that best characterize the differences among groups (Bortz 1993). Variables are combined into one or more discriminant functions. Variables that fail a tolerance test, i.e., variables that represent an almost linear combination of other variables, do not enter the analysis.

**Changes related to age**

To find out whether the acoustic structure of the various call types changed with age, calls of the first recording session (week 7-10) were tested against the last recording session (week 53-58). For two call types, there were not enough recordings in the last session for all infants. In these cases, the first session was tested against the penultimate session. The following call types occurred in sufficient number to be compared between the first and seventh/eighth session: (coo ($n1 = 220$; $n2 = 226$), wail ($n1 = 197$; $n2 = 145$), pleasure cry ($n1 = 30$; $n2 = 38$), discomfort cry ($n1 = 259$; $n2 = 180$), short discomfort cry ($n1 = 84$; $n2 = 17$) and moan$^{+}$($n1 = 31$; $n2 = 36$)). A stepwise discriminant function analysis was carried out for each call type. The three most important parameters found by the analysis were used for the multivariate General Linear Model (GLM) test. These three call parameters and the other parameters found to be important in describing age-related differences, were also tested with a univariate GLM test. For the test, the vocalizations were taken, as far as possible, from the same emotional state, in order to avoid mixing age-related changes with possible emotion-related effects on the acoustic structure. Wail, discomfort cry and short discomfort cry were taken from the emotional state unease; coo, pleasure cry and moan$^{+}$ were taken from situations classified as contentment. In order to obtain a sufficient number of moan$^{+}$, we had to include for one infant calls uttered under the emotion interest. For pleasure cries, we had to include for two infants calls uttered under the emotion joy. In cases in which the global hypothesis could not be rejected with the multivariate tests, we applied a Bonferroni correction in subsequent univariate tests (Bonferroni correction: $\alpha = \alpha / (k - n + 1)$, where k = number of tests, n = number of significant tests (Bortz et al. 1990).

Since the selection of specific emotions described above might have biased the results, we tested, in those cases in which sufficient data were available, the call type variants also with other emotions as well: Wail, discomfort cry and short discomfort cry were additionally tested with the emotion anger; for coo, a second test was conducted with the emotion interest.

**Differences related to emotion**

In principle, the same procedure was used as in the tests on age-related differences. We tested the four most frequent call types with a basically tonal structure: cry (n = 2545), short cry (n = 779), coo/wail (n = 4630) and moan (n = 816) (Figure 2A). Again, for each of these call types a stepwise discriminant function was used to find the three most important parameters for the separation of positive and negative emotions. For each of these parameters we calculated the mean per infant and emotional category (positive or negative) for each age class. These means were used to carry out a multivariate repeated GLM to test the global hypothesis. The repeated measure design was chosen to get a higher test power and a control for age-related interactions. Only the short cries were tested in a multivariate form without repeated measurement because of missing values. After the test of the global hypothesis, the univariate repeated GLM was undertaken with characteristic parameters for emotion-related differences.

For further differentiation within the positive and negative emotion categories, again stepwise discriminant analyses were carried out. For the call type variants coo, moan$^{+}$, pleasure cry and short pleasure cry, the discriminant analysis was used to distinguish between the emotions joy, contentment and interest. For the call type variants wail, moan$^{-}$, discomfort cry and short discomfort cry, discriminant analysis was used to distinguish between unease and anger. Surprise and pain were not included in the analysis, because both emotions were only seldom recorded. The means per individual and emotion of the three most important parameters were tested for significant differences, using the multivariate GLM test.

**Differences related to context**

In order to find out to what degree the emotion-related differences were influenced by rating biases of the parents, we did a cross-check with the situations in which the recordings were made. Out of the eleven situations listed in Appendix 1, we chose those six that were classified by the parents as mainly positive or negative, respectively. The positive situations were situations 2, 4, and 6, the negative ones were situations 1, 9, and 10. From the positive situations, we selected only those recordings that were also attributed by the parents to positive emotions; corresponding holds for the negative situations. Testing was carried out in the same way as in the tests for emotional differences, that is, first, a discriminant analysis was made, followed by multivariate and univariate GLM tests. We compared the three positive with the three negative situations and then pairwise the positive situations with each other and the negative situations with each other.

### Usage of call types

In order to find out whether different emotions can be differentiated by the relative frequencies of specific call types, we took three recordings of each infant and emotion made at the age of 29-39 weeks. Again surprise and pain were not used due to their low occurrence. We counted the number of calls for each call type uttered in the first 60 seconds of a recording. Vocalizations that could not be attributed to a specific call type were put into the rest class. We then calculated the mean number of calls for each call type per emotion for each individual and tested with the Wilcoxon signed ranks test (exact, two-tailed) for significant differences. All tests were made with SPSS 10 (Statistical Program for the Social Sciences).

## Results

### Call types

The preverbal utterances of the first year can be classified into one inspiratory and eleven expiratory call types. Figure 2 shows typical examples of the expiratory call types. Five call types, wail/coo, moan, cry, short cry, and babble show a predominantly tonal structure, but they can include atonal parts, such as the middle segment of the coo/wail call (Figure 2). Cry and short cry can be distinguished by call duration (cry: 1191 ± 688 ms, short cry: 380 ± 268 ms). Moan is the only other short tonal call type (319 ± 182 ms). Other tonal calls have mean call durations between 703 ± 544 ms (coo/wail) and 646 ± 600 ms (babble). The fundamental frequency distinguishes both cry call types ($F_0$ *mean*: cry = 416 ± 121 Hz, short cry = 393 ± 144 Hz) from the other call types ($F_0$ *mean*: coo/wail = 345 ± 107 Hz, moan

= 333 ± 138 Hz, babble = 344 ± 118 Hz). The same is true for the frequency range. Cry has a mean range of 3069 ± 1502 Hz, short cry of 3551 ± 1753 Hz. The other tonal call types had a mean range of 2197 ± 987 Hz (babble), 1592 ± 852 Hz (coo/wail) and 1524 ± 1135 Hz (moan), respectively. Babble can be easily identified by its characteristic rhythmic structure and the fact that it consists of at least one consonant and one vowel. Most of these utterances are speech-like, but meaningless. In contrast to Oller (1975), we include some meaningful utterances, such as mama, in this category. By these criteria, all tonal call types are well separable.

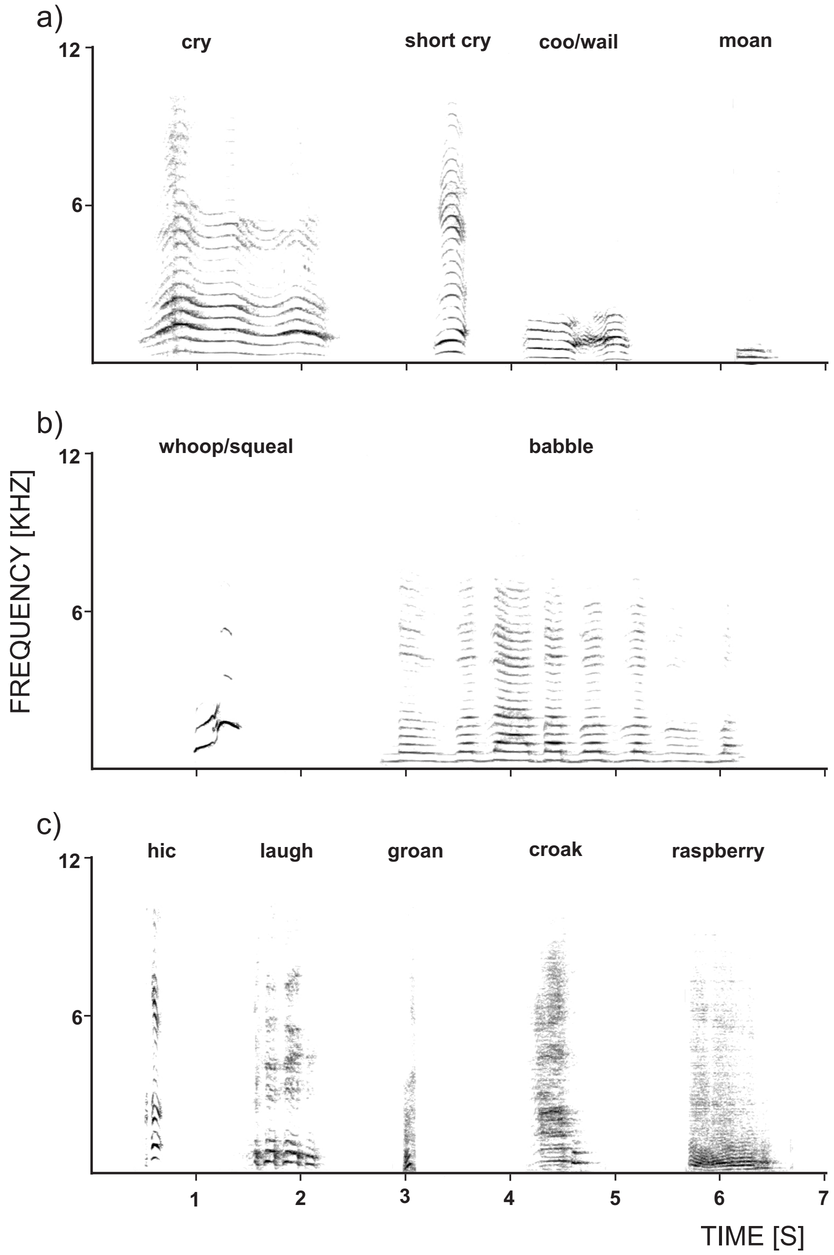


**FIGURE 2.** Frequency-time spectrograms of the call types used in this study.

Figure 2C shows the call types with predominantly noisy structure. Groan and hic can be distinguished from the other three noisy call types by call duration (mean duration: groan = 212 ± 131 ms, hic = 248 ± 255 ms, laugh = 391 ± 293 ms, croak = 379 ± 381 ms and whoop/squeal = 585 ± 579 ms). Hic has the highest mean frequency range (4405 ± 2186 Hz), groan the next highest (3410 ± 2213 Hz). Hic can also be differentiated from other calls by its typical rhythmic structure. It consists of two parts separated by a short amplitude gap (see Figure 2). Laugh is another rhythmic vocalization, made up of alternating noisy and tonal parts, separated by short pauses (see also Kipper & Todt 2001). Laughs, however, are longer than hics. If consisting of only two parts, the first part of laugh, in contrast to hic, is longer and includes tonal structures. Often laugh consists of more than two vocal elements. Laugh usually appears in positive emotional context; but a structurally similar call type also occurs in negative emotional context. We, therefore, use the term laugh independent of the underlying emotional state. Whoop/squeal (Figure 2) can be separated from all other call types by the mean frequency of the first dominant frequency band (dfb1 mean). The dfb1 mean of whoop/squeal (1151 ± 353 Hz) is substantially higher than the dfb1 mean of the other call types, which ranged between 636 ± 187 Hz (coo/wail) and 759 ± 300 Hz (croak). Additionally, whoop/squeal are characterized by their prominent frequency modulation (Figure 2). Raspberries and ingressive vocalizations were not included in the acoustic analysis. They were evaluated only with respect to their frequency of occurrence. Raspberry is defined as labial or labiolingual trill and/or vibrant. Ingressive vocalization is characterized by its inspiratory nature.

In sum, all acoustically analyzed call types could be well differentiated by a small number of parameters.

**Emergence of call types**

During negative emotions (pain, anger and unease), the call types coo/wail, moan, cry, short cry, hic, groan and croak were regularly uttered from the first recording on (Table 2). Regularly, in this case, means that more than 3 infants per recording session uttered these call types. Whoop/squeal was rare during the entire first year. Babble did not start before the 29th week.

During positive emotions (surprise, interest, contentment and joy) coo/wail, moan and groan were regularly uttered from the first recording on. Cries and short cries also appeared in the first recording session (Table 3), but were found less often than in older infants. Hic was common at the age of 20 weeks and laugh at the age of 24 weeks. As in negative emotions, babbling, with one exception, did not appear before the age of 29 weeks. Whoop/squeal again was rare during the entire first year.

***TABLE 2.*** *Occurrence of call types in negative emotions. The numbers refer to the number of individuals that produce the call type in the corresponding age class. Note that the total number of recorded individuals differ in the age classes.*

| Age (weeks) | **7-10** | **12-16** | **20-22** | **24-26** | **29-33** | **36-39** | **44-49** | **53-58** |
|---|---|---|---|---|---|---|---|---|
| Total (individuals) | 7 | 7 | 6 | 6 | 7 | 6 | 7 | 7 |
| Call type | | | | | | | | |
| Coo/wail | 7 | 7 | 6 | 6 | 7 | 6 | 7 | 6 |
| Moan | 6 | 5 | 6 | 6 | 6 | 6 | 5 | 5 |
| Cry | 7 | 7 | 6 | 6 | 7 | 6 | 7 | 6 |
| Short cry | 7 | 7 | 5 | 5 | 7 | 6 | 7 | 7 |
| Babble | 0 | 0 | 0 | 0 | 5 | 5 | 5 | 5 |
| Hic | 6 | 6 | 6 | 5 | 6 | 6 | 6 | 5 |
| Laugh | 2 | 4 | 3 | 5 | 5 | 5 | 6 | 4 |
| Groan | 6 | 6 | 4 | 5 | 7 | 6 | 6 | 5 |
| Croak | 7 | 7 | 6 | 5 | 7 | 6 | 7 | 4 |
| Whoop/squeal | 1 | 2 | 1 | 1 | 0 | 1 | 2 | 3 |

***TABLE 3.*** *Occurrence of call types in positive emotions. The numbers refer to the number of individuals that produce the call type in the corresponding age class. Note that the total number of recorded individuals differ in the age classes.*

| ***Age (weeks)*** | **7-10** | **12-16** | **20-22** | **24-26** | **29-33** | **36-39** | **44-49** | **53-58** |
|---|---|---|---|---|---|---|---|---|
| *Total (individuals)* | 7 | 7 | 6 | 6 | 7 | 6 | 7 | 7 |
| **Call type** | | | | | | | | |
| Coo/wail | 7 | 7 | 6 | 6 | 7 | 6 | 7 | 7 |
| Moan | 7 | 7 | 6 | 6 | 7 | 6 | 7 | 7 |
| Cry | 4 | 6 | 6 | 5 | 7 | 6 | 7 | 7 |
| Short cry | 5 | 5 | 5 | 3 | 6 | 6 | 7 | 7 |
| Babble | 0 | 0 | 0 | 1 | 7 | 6 | 7 | 7 |
| Hic | 2 | 4 | 6 | 5 | 6 | 3 | 2 | 4 |
| Laugh | 1 | 4 | 5 | 6 | 7 | 6 | 7 | 6 |
| Groan | 6 | 5 | 5 | 5 | 7 | 4 | 6 | 6 |
| Croak | 7 | 7 | 6 | 5 | 7 | 5 | 7 | 7 |
| Whoop/squeal | 2 | 1 | 1 | 2 | 2 | 3 | 1 | 3 |

**Structural changes related to age**

In the tests on age-related changes in acoustic structure, only tonal call types were analyzed, as only these calls occurred in satisfactory number over all recordings. For a better description of the changes over time, the call types cry, short cry, coo/wail and moan were subdivided into two variants each (see Methods). Of these, six variants were submitted to an acoustic analysis. These were coo, wail, moan$^{+}$, pleasure cry, discomfort cry and short discomfort cry. Four out of the six call type variants yielded significant results in the first examination, for which we selected vocalizations of contentment to test age-related changes of the positive call type variants and vocalizations of unease to test the negative call type variants (wail: HT = 4.5, F = 15.00, p = 0.000; moan$^{+}$: HT = 2.7, F = 9.00, p = 0.003; discomfort cry: HT = 1.26, F = 4.19, p = 0.036; short discomfort cry: HT = 1.99, F = 6.63, p = 0.01). The changes in single call parameters for these four call variants are presented in Table 4.

***TABLE 4.*** *Structural changes in call type variants with age. Arrow upward (↑) marks an increase of the corresponding parameter with age. Arrow downward (↓) marks a decrease of the corresponding parameter. Simple arrows indicate a p-value below 0.1; p-values below 0.05 are indicated by an additional asterisk (*).*

| **Call type variant** | **Wail** | **Moan$^{+}$** | **Discomfort Cry** | **Short Discomfort Cry** |
|---|---|---|---|---|
| ***Parameter*** | | | | |
| *Duration* | | | | |
| *Tonality* | | | | |
| *HNR mean* | ↑ | ↑* | ↑* | ↑* |
| *HNR max* | | ↑* | ↑* | ↑* |
| *$F_0$ median* | | | | |
| *dfa2 start* | ↓* | | ↓* | ↓* |
| *dfa2 mean* | | ↓* | | |
| *dfa3 mean* | ↓* | ↓* | ↓* | ↓* |
| *Range mean* | | ↓* | ↓* | |
| *Range max* | | ↓ | ↓* | ↓ |
| *FLS2 mean* | ↓* | ↓* | ↓* | |
| *Pf mean* | | | | |
| *PF max* | | | | |
| *dfb ratio 1* | | ↑* | ↑ | |
| *CS mean* | ↑ | | ↑* | ↑* |

The call variants moan$^+$, discomfort cry and short discomfort cry showed a significant increase in the harmonic-to-noise ratio (HNR mean and HNR max) during the first year of life. Wail showed a trend in the same direction (Table 4). The changes in mean and standard deviation for HNR mean was for moan$^+$: 0.58 ± 0.029 (age: 7 – 16 weeks) to 0.63 ± 0.016 (age: 45 – 58 weeks); for discomfort cry: 0.56 ± 0.023 to 0.61 ± 0.029; for short discomfort cry: 0.54 ± 0.03 to 0.58 ± 0.023; and for wail: 0.59 ± 0.028 to 0.61 ± 0.021. An increase in the harmonic-to-noise ratio means that these vocalizations became less noisy; that is, they sounded clearer. Additionally, all four call variants showed a decrease in call parameters describing timbre (dfa2 start, dfa2 mean, dfa3 mean). Here, the changes in mean and standard deviation for dfa3 mean was for moan$^+$: 4410 Hz ± 1384 Hz to 2510 Hz ± 956 Hz; for discomfort cry: 5146 Hz ± 369 Hz to 3971 Hz ± 833 Hz; for short discomfort cry: 5445 Hz ± 704 Hz to 4395 Hz ± 512 Hz; and for wail: 3842 Hz ± 668 Hz to 2667 Hz ± 587 Hz. Three of the call variants also showed a decrease in frequency range (Table 4). Furthermore, discomfort cries became more homogeneous in their structure (CS mean). The call type variant coo failed to reach significance in the multivariate test, but showed the tendency ($p < 0.1$) to increase in duration and to decrease in range. No significant age-related changes in call duration, tonality, fundamental frequency or peak frequency were found for wail, moan$^+$, discomfort cry and short discomfort cry.

As mentioned in the Method section, we carried out the same tests also with vocalizations of other emotional contexts (anger, interest). In general, in these cases, the results are in line with the findings described above. For example, discomfort cry uttered under the emotion anger showed a significant increase in HNR mean (mean: 0.56 ± 0.022 to 0.61 ± 0.025; $p< 0.05$; univariate GLM test), a significant decrease in dfa3 mean (mean: 5146 Hz ± 387 Hz to 4398 Hz ± 740 Hz) and a significant decrease in range max (mean: 8675 Hz ± 1025 Hz to 6864 Hz ± 912 Hz). In some cases, the parameters did not reach significance level, but just showed a trend. In these cases, the trend was always in harmony with the results shown in Table 4. The small number of calls, when broken down according to emotional state, led to an uneven distribution. This made it impossible to decide whether, there were no age-related changes or the data base was simply too small in the cases in which the parameters did not reach significance.

**Structural differences between emotions**

Again, only tonal call types occurring in satisfactory number during the recording sessions were analyzed (cry, short cry, coo/wail and moan). First, a discriminant function analysis was carried out to find the acoustic parameters that best differentiated the calls given during negative and positive emotions. The three most important call parameters found by the discriminant function analysis were tested with a multivariate test (multivariate repeated GLM, SPSS 10). Short cry was tested with the normal multivariate GLM procedure, because of missing values. For all four call types, namely coo/wail, moan, cry and short cry, significant differences in the acoustic structure were found (cry: HT = 2.08 F = 6.93, p = 0.008; coo/wail: HT = 2.55, F = 8.5, p = 0.004; moan: HT = 2.49, F = 6.64, p = 0.015; short cry: HT = 3.65, F = 12.16, p = 0.001). The results of the subsequent univariate tests are given in Table 5.

All call parameters listed in Table 5 showed significant differences in at least one call type. All call types showed a significant increase in call duration when changing from positive to negative emotion (see also Figure 3). Depending on call structure, one or more parameters describing pitch and frequency range increased as well when calls changed from positive to negative emotion (Figure 3). Specifically, in three out of four call types, the parameters dfa3 mean, FLS2 mean and PF tonal mean increased significantly with the change from positive to negative emotion. A similar trend was found for the parameters range mean and range max. Other parameters, such as $F_0$ median or dfa1 mean, increased in only one or two call types. Still others, such as dfb ratio 1, showed an antagonistic tendency in different call types. In sum, the results indicate that changes from positive to negative emotional states are accompanied by increases in duration, frequency range, and a number of pitch parameters. Some of these parameters are quite call type specific, others are of a more general nature.

For the differentiation of specific emotions, such as contentment and joy or unease and anger, the same analytical procedures were used as before. We carried out a stepwise discriminant function analysis to find out those call parameters best suited for the differentiation of emotional contexts. The three most important call parameters were submitted to a multivariate GLM test. The multivariate test revealed overall significant differences between unease and anger for wail (HT = 2.11, F = 3.7 p = 0.005) and discomfort cry (HT = 1.71, F = 3.41 p = 0.014), but no single call parameter could be found which was able to differentiate significantly between pairs of negative emotions (unease, anger) or positive emotions (interest, contentment, joy).

***TABLE 5.*** *Structural differences between calls uttered during negative and positive emotions.*

| Call Types | Cries | Short cries | Coo/Wail | Moan |
|---|---|---|---|---|
| ***Parameter*** | | | | |
| *Duration* | ↑* | ↑* | ↑* | ↑* |
| *$F_0$ median* | | ↑* | | |
| *Pf/$F_0$ coeff.* | ↑* | | | ↑* |
| *dfa1 mean* | ↑* | ↑* | | |
| *dfa2 max* | ↑* | ↑* | | |
| *dfa3 start* | | | | |
| *dfa3 mean* | ↑* | ↑* | ↑* | |
| *Range mean* | ↑ | ↑* | | ↑* |
| *Range max* | ↑ | ↑* | ↑* | |
| *FLS2 start* | | ↑* | | |
| *FLS2 mean* | ↑* | ↑* | ↑* | |
| *FLS1 amp mean* | ↓ | ↓* | ↓* | |
| *FLS2 %* | ↑ | ↑* | | |
| *PF tonal mean* | ↑* | | ↑* | ↑* |
| *PF tonal max* | | | | ↑* |
| *PF max location* | ↑* | ↑* | ↑ | |
| *dfb1 %* | | | ↑* | ↑* |
| *dfb2 %* | ↑* | | | |
| *dfb ratio 1* | ↓* | ↑ | | |
| *dfb ratio 2* | ↓* | | | |
| *CS mean* | | | ↑* | |
| *CS std dev* | | | ↓* | |
| *CS max location* | ↑* | ↑* | | |

*Arrow upward (↑) indicates an increase of the corresponding call parameter in negative emotions relative to positive emotions. Arrow downward (↓) indicates a decrease of the corresponding parameter. Parameters marked with an arrow have a p-value below 0.1; p-values below 0.05 are marked in addition with an asterisk.*

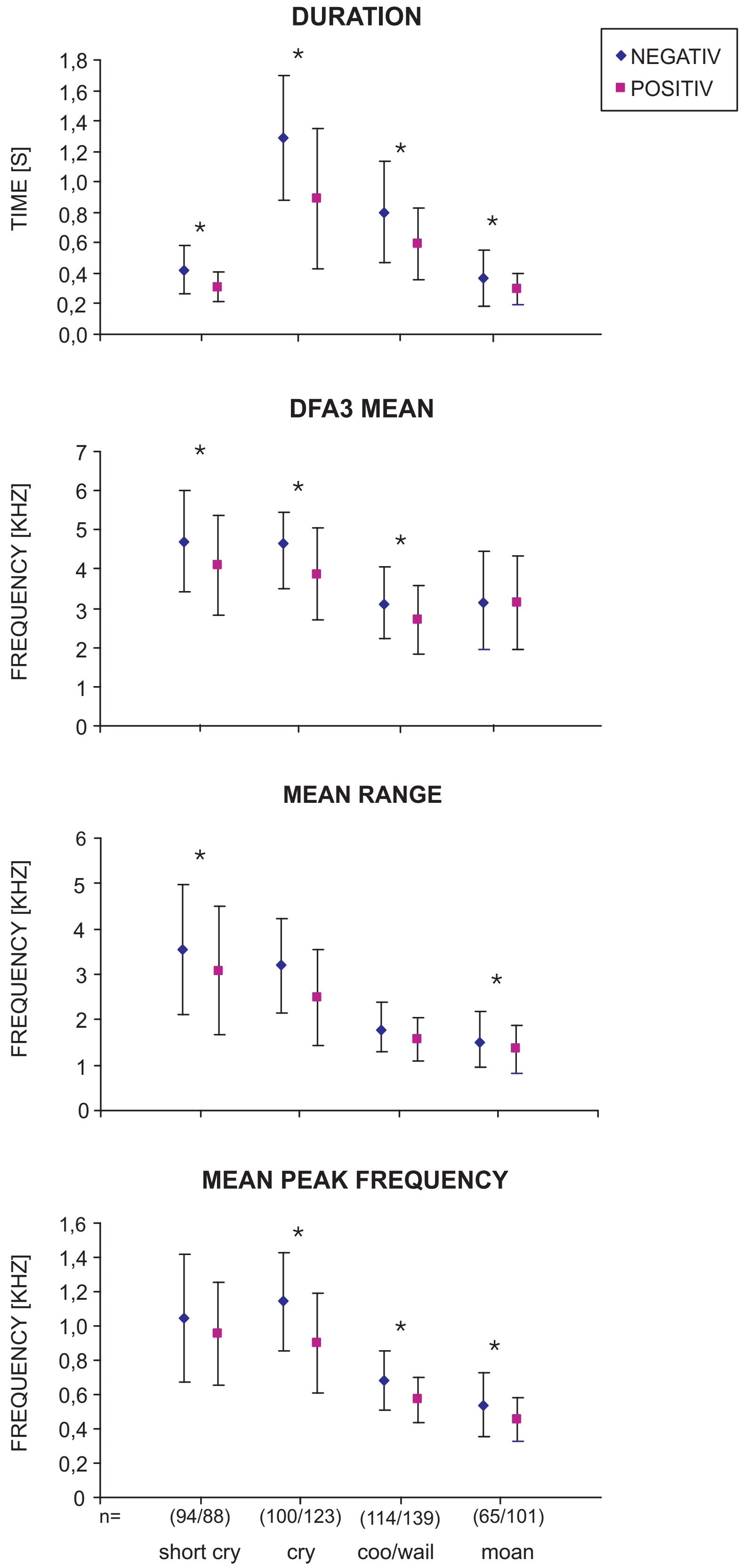


**FIGURE 3.** Differences between positive and negative emotions in the call types short cry, cry, coo/wail and moan. The diagrams show the mean values ± standard deviation for the parameters duration, dfa3 mean, mean range and mean peak frequency. Significant differences ($p < 0.05$; univariate GLM Test, SPSS 10) are marked by an asterisk.

**Structural differences between situations**

The call types cry, short cry, coo/wail and moan were analyzed with respect to situation-specific acoustic differences. We found significant differences for all four call types (cry: HT = 1.45, F = 18.44, p = 0.000; coo/wail: HT = 0.84, F = 10.74, p = 0.000; moan: HT = 0.389, F = 4.93, p = 0.005; short cry: HT = 0.83, F = 10.48, p = 0.000). Due to the close relationship between situation and emotional rating, these findings were not unexpected. The results of the subsequent univariate test confirmed the results of the test for emotional differences. We used the same procedure to differentiate between single positive (after feeding, being tickled, playing) or negative situations (hungry, tired, wet diapers). Only in five out of 24 cases the multivariate test reached significance. Significant differences were found between being tickled and playing for the call type variants pleasure cry (HT = 1.29, F = 4.19, p = 0.033) and short pleasure cry (HT = 1.50, F = 5.01, p = 0.022); between after feeding and being tickled for the call type variant short pleasure cry (HT = 1.29, F = 1.95, p = 0.040); between hungry and tired for the call type variant moan$^{-}$ (HT = 1.20, F = 4.01, p = 0.041) and between tired and wet diapers for the call type variant discomfort cry (HT = 1.61, F = 5.37, p = 0.018). In the subsequent univariate tests only few call parameters reached significance. These parameters are listed in Table 6.

***TABLE 6.*** *Structural differences between calls uttered in different situations. All call parameters listed have a p-value below 0.05. For detailed description of the situations, see Appendix 1.*

| **Call Type Variant** | **Parameter** | **Mean Value ± Standard Deviation** | |
|---|---|---|---|
| **Pleasure cry** | | **Situation 4: being tickled** | **Situation 6: playing alone** |
| | *duration [ms]* | 650.7 ± 171.9 | 982.9 ± 168.5 |
| | *CS std dev* | 0.041 ± 0.008 | 0.032 ± 0.005 |
| | *CS max location* | 330.7 ± 121.6 | 504.6 ± 91.5 |
| | *tonality [%]* | 51.2 ± 14.9 | 65.9 ± 8.2 |
| **Short pleasure cry** | | **Situation 4: being tickled** | **Situation 2: after feeding** |
| | *dfb1 mean modulation* | 54.5 ± 20.0 | 86.5 ± 29.4 |
| **Moan$^{-}$** | | **Situation 1: hungry** | **Situation 9: tired** |
| | *dfb1 min location* | 0.5 ± 0.122 | 0.72 ± 0.21 |
| | *dfb1 trend* | 0.14 ± 0.14 | -0.06 ± 0.1 |
| | *range max [Hz]* | 6074.3 ± 2028.6 | 3618.6 ± 1189.7 |
| | *PF max location* | 0.56 ± 0.25 | 0.83 ± 0.13 |

**Usage of call types**

Since it was not possible to find acoustic parameters characterizing specific emotions, the question arose of whether emotions are encoded by call type. Five emotions, joy, contentment, interest, unease, and anger, appeared often enough in the recording sessions to allow an analysis. Figure 4 shows the mean number of different call types per 60 seconds.

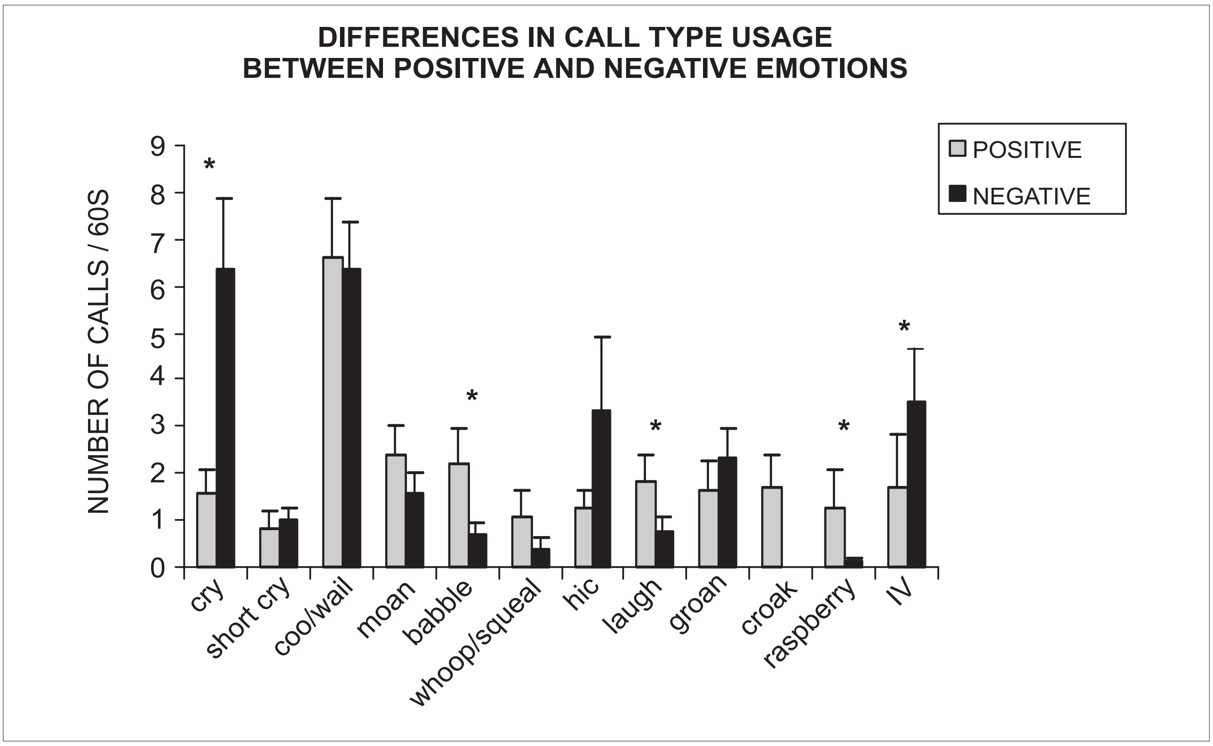


**FIGURE 4.** Differences in call type usage between positive and negative emotions. The diagram shows the mean number of each call type, including the ingressive vocalization (IV), per 60 seconds. Significant differences ($p < 0.05$; Wilcoxon signed ranks test, exact, two-tailed) are marked by an asterisk.

A comparison of the number of specific call types in positive and negative emotional contexts reveals that there are significant differences for cry, babble, laugh, raspberry and ingressive vocalization (cry: $n = 7$, $T = 1$, $p = 0.031$; babble: $n = 7$, $T = 1$, $p = 0.031$; laugh: $n = 7$, $T = 0$, $p = 0.16$; raspberry: $n = 7$, $T = 0$, $p = 0.031$; ingressive vocalization: $n = 7$, $T = 2$, $p = 0.47$; Wilcoxon Signed Ranks Test, exact, two-tailed) (Figure 4). Specifically, in negative emotions, infants showed more cries and ingressive vocalizations; in positive emotions, they uttered more babbles, laughs and raspberries. Croak was exclusively uttered in positive emotions. The other call types were not uttered in significantly different numbers in positive and negative emotion.

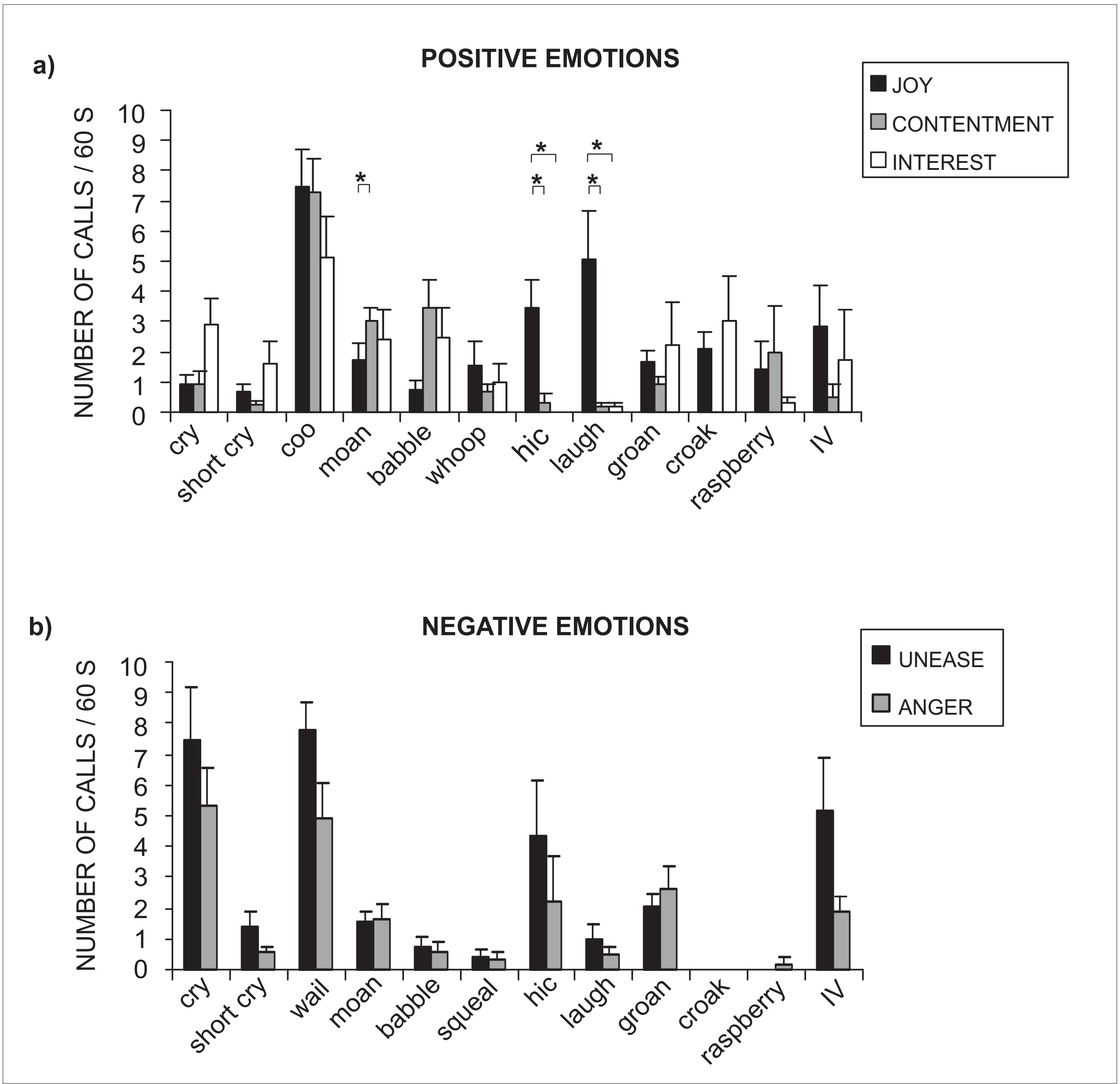


**FIGURE 5.** Call type usage during specific emotions. The diagram shows the mean of calls for each call type, including ingressive vocalization (IV), for the emotions joy, contentment and interest (A) as well as for unease and anger (B). Significant differences are marked by an asterisk ($p < 0.05$; Wilcoxon signed ranks test, exact, two-tailed).

The following differences in call usage between joy, contentment and interest were found (Figure 5A). Joy had a high proportion of hic, laugh and ingressive vocalization and a low proportion of babbling. Contentment showed no or almost no croak, laugh and hic, but had a high amount of babbling and moan. Interest, like contentment, had no or almost no hic and laugh. In contrast to contentment, however, interest showed a higher number of ingressive vocalizations. Interest also was characterized by a relatively high number of cry and short cry calls. All three emotions showed a high proportion of coo. The following differences proved to be significant (Wilcoxon Signed Ranks Test, exact, two-tailed): Hic: joy-contentment (n =

7, T = 0, p = 0.016); joy-interest (n = 6, T = 0, p = 0.031); laugh: joy-contentment (n = 7, T = 0, p = 0.016); joy-interest (n = 6, T = 0, p = 0.031); moan: joy-contentment (n = 7, T = 2, p = 0.047).

Comparing unease with anger, we found more cries, short cries, wails, hics and ingressive vocalizations in unease than in anger (Figure 5B), however, none of these differences reached significance at the 0.05 level (Wilcoxon signed ranks test, exact, two-tailed). Raspberry was exclusively uttered in anger.

Finally, we tested the hypothesis that the total number of uttered calls differs between emotions. We found that in joy more calls were uttered than in contentment (n = 7, T = 1, p = 0.031; Wilcoxon signed ranks test, exact, two-tailed). No significant differences, however, were found for the other pairs of emotions.

## Discussion

The present study has shown that the nonverbal vocalizations of infants in their first year of life may be classified into 12 call types, all of which can be differentiated on the basis of a small number of acoustic parameters. These parameters are fundamental frequency, frequency range, frequency of the lowest dominant frequency band (dfb1), amount of nonharmonic energy in the call (tonality), rhythmicity, and duration of the call. Not all vocalizations are present from the very beginning. Babbling normally does not appear before the 29th week. Laughing is another vocalization that is very rare before the fourth month. The majority of vocalizations, however, appear within the first two months. As we have only one recording session from the first two months, we are not able to further specify the date of emergence of these vocalizations, nor can we say something about the changes in acoustic structure going on in this period. Data on this period, however, have been presented by Lind and Wermke (1997) as well as Wasz-Höckert et al (1968) with respect to crying. According to Lind and Wermke, crying shows an increase in duration, stability of the fundamental frequency contour, and extent of frequency modulations during the first 8 weeks. Wasz-Höckert et al. agree with Lind and Wermke that there are changes in the acoustic structure of crying in the first days. They did not find changes beyond the fifth day, however. This discrepancy is partly due to the fact that Wasz-Höckert and colleagues measured only fundamental frequency and duration, but not the other parameters used by Lind and Wermke.

Fundamental frequency, in fact, also did not show a significant change in the Lind and Wermke study beyond the first eight weeks. The results of the present study show that neither crying nor any of the other vocalizations undergo a change in fundamental frequency or duration between the 2nd and 12th month. There are, however, other parameters that show a significant change beyond the 2nd month. Such parameters are the harmonic-to-noise ratio (HNR mean, HNR max), frequency range (range mean, range max), distribution of frequency amplitudes (dfa2 start, dfa2 mean, dfa3 mean), frequency of 2nd formant-like structure (FLS2 mean), amplitude ratio between first and second dominant frequency band (dfb ratio 1) and structure homogeneity (CS mean). Not all these parameters show an age-related change in all call types. Whenever there is a change, however, the direction of change is the same in all call types affected. This suggests that the factors responsible for those changes are of a more general nature. The downward shift of the main energy from higher towards lower frequencies (decrease of dfa 2 start, dfa 2 mean, dfa3 mean, FLS2 mean), for instance, may be a consequence of the growth and descensus of the larynx (Lieberman 1984; Papoušek 1994; Ploog 1992). In other words, the change in resonance frequencies induced by the lengthening and widening of the vocal tract during ontogeny might be one reason for the observed decrease in dfa and FLS parameters. Another factor might be improvement of subglottal air-pressure control. According to Langlois et al. (1980), such a control, which is a necessary prerequisite for stable phonation, does not evolve before the 3rd month. This is the time at which infants start to imitate the intonation of their mothers (Lieberman et al. 1982). Improvement of subglottal air-pressure control probably is brought about by neuromuscular maturation processes (such as progressing myelination) as well as by training of muscular coordination (Boliek et al. 1996). The age-related increase in call homogeneity (CS mean) and harmonic-to-noise ratio (HNR mean, HNR max) found in the present study might be the result of such an improvement in subglottal air pressure control.

The present study also showed that one and the same call type uttered in positive and negative emotional context differs in acoustic structure. In other words, the important information about whether an infant feels good or bad, is encoded in the acoustic structure of single call types. This finding is in line with a study of Papoušek (1989) who showed that listeners are able to distinguish utterances expressing comfort and joy from utterances expressing discomfort on the basis of single vocalizations. In an attempt to characterize the vocal expression of positive and negative emotional states, Papoušek carried out an acoustical analysis in a second study (Papoušek 1992). In this study, it was found that vocalizations expressing discomfort differ from those expressing comfort by their higher amount of spectral energy above 1000 Hz. Fundamental frequency, intensity and duration increase with increasing positive as well as negative arousal. The present study agrees with Papoušek in that there is an upward shift of energy in the power spectrum from lower to higher frequencies when positive emotion changes into negative emotion. This upward shift manifests itself in an increase of the parameters PF/$F_0$ coeff, dfa1 mean, dfa2 max, dfa3 mean, FLS2 mean and PF tonal mean. In contrast to Papoušek (1992), however, the parameter duration was found in the present study to increase significantly when emotion changed from positive to negative. Furthermore, frequency range turned out to be another parameter differentiating positive and negative emotions.

The finding that the fundamental frequency does not differentiate positive and negative emotions seems to contradict several studies, according to which negative emotions are characterized by an increase in fundamental frequency (Banse & Scherer 1996; Protopapas & Lieberman 1997; Ruiz et al. 1996; Williams & Stevens 1972). There are two explanations for this discrepancy. In our study, in contrast to the aforementioned ones, we compared the negative emotion directly with the positive emotion – and not a neutral emotional state. As we may assume that an increase in negative as well as positive emotion is accompanied by an increase in arousal, and an increase in arousal is accompanied by an increase in fundamental frequency (Banse & Scherer 1996), the increase in fundamental frequency reported by some authors during negative emotions might simply reflect the higher degree of arousal during negative emotions than during neutral emotional states. Another explanation could be that increases in fundamental frequency are only found in extremely aversive situations, not in moderately aversive situations. The increase in fundamental frequency reported by Protopapas and Lieberman (1997) as well as Ruiz et al. (1996) was found in the speech of pilots immediately before an air crash. In our study, the infants were recorded at home in a protective environment with only mildly aversive stimuli. In those few cases in which, in fact, an accident occurred, the parents normally first supported their infants, before they started the tape recorder. Severely aversive situations thus escaped our analysis.

While it is no problem to acoustically differentiate calls uttered in positive and negative emotional context, even if they are of the same call type, severe problems arise when specific positive or negative emotions had to be distinguished. We found no single acoustic parameter that differentiated significantly unease and anger in any of the call types tested. The same holds for the differentiation of joy, contentment and interest. Regarding the call usage in sequences of vocalizations uttered in different emotions, there was again no difference between anger and unease and between contentment and interest. Joy and interest could be distinguished in only 2 out of 12 call types, joy and contentment in 3 call types. There are several possible reasons for this low discriminability. One reason could be that in young infants the emotional system is not as differentiated as in older children or adults. As mentioned in the Introduction, this point is still controversial in the literature (Giblin 1981; Lewis 1993; Malatesta-Magai et al. 1991; Sroufe 1979). Another reason could be that infants are able to express various emotions very early, but parents have difficulty in identifying these emotions correctly. Such difficulties might arise, for instance, from interferences between the parents own mood or expectations and the infants' behavior, or from incoherence in emotional labeling. One mother, for instance, might name a certain emotional expression joy, while another mother uses the term contentment. A third reason might be that not every vocalization uttered in a certain emotional context is typical for this context. In order to evaluate the influence of such weakening effects, we did a cross-check analysis with the same calls, but instead of using the emotional categories, we used the context categories in which the calls were uttered (see description Appendix 1). This analysis produced the same results as the analysis based on the emotional ratings. It, therefore, seems unlikely that a mismatch between the infants emotions and the parents ratings is the main factor for the low discriminability of emotions in single call types. The results of the present study suggest that a differentiation of emotions beyond positive and negative is made mainly on the basis of call type selection. That is, different call types occur with different probability in different emotional contexts. Joy, for instance, is characterized by a high relative number of coos and laughs and a low number of cries, short cries and babbles. Contentment is also characterized by a high number of coos, but in contrast to joy, by a low number of laughs and a high number of babbles. Interest differs from contentment by its high number of croaks, which are completely lacking during contentment. Furthermore, there is evidence from nonhuman primate vocal communication that also the temporal structure of vocal sequences may contain information on the emotional state of the vocalizer (Todt 1988). Finally, during normal life, infant vocalizations are accompanied by facial expressions and body gestures. These visual signals represent additional information used by the parents to specify the emotional states of their infants (Young & Decarie 1977).

## Appendix 1: Situations in which the infants were recorded by their parents.

| Situation | Description for the recording parents |
|---|---|
| 1. | You are preparing food for the infant (respectively: its time for nursing). Your infant is hungry and complains about it. |
| 2. | The infant lies in your arms, immediately after feeding. |
| 3. | You are approaching your quiet infant, who becomes aware of you at this moment. |
| 4. | Your infant is lying in front of you and is tickled by you. |
| 5. | You are speaking with your infant. |
| 6. | Your infant is playing alone with toys, its fingers, or something else. |
| 7. | Your infant tries impatiently and in vain to reach a toy (or something else) out of reach. |
| 8. | Your infant is lying in bed, content and tired, shortly before falling asleep. |
| 9. | Your infant is lying in bed, overtired and fretful. |
| 10. | Your infants diaper is wet. |
| 11. | Your infant has hurt itself. |

*Each of the situations should be recorded twice during the course of one week. For each recorded situation the parents had to name the emotion they assumed their infant expressed, choosing between joy, contentment, interest, surprise, unease, anger, and pain.*

## Appendix 2: Acoustic parameters used in the analysis.

*Parameters calculated on the basis of frequency-time spectra with a range of 4 kHz and a frequency resolution of 10 Hz are marked by an asterisk. The other parameters are calculated on the basis of spectra with a range of 12 kHz and a frequency resolution of 29 Hz. For a detailed description of the parameters, see also Schrader and Hammerschmidt (1997).*

| **Parameter:** | **Description:** |
|---|---|
| duration [ms] | Time from onset to end of vocalization |
| tonality [%] * | Percent of time segments in which a clear harmonic structure can be detected |
| HNR mean * | Difference between the amplitude peaks of the dominant frequency bands and the noise level of the same frequency point, mean across all time segments |
| HNR max. * | Maximal difference between the amplitude peaks of the dominant frequency bands and the noise level of the same frequency point |
| $F_0$ mean [Hz] * | Fundamental frequency (see Figure1), mean across all time segments |
| $PF/F_0$ coeff. * | Ratio between peak frequency (see below) and fundamental frequency $(PF-F_0)/F_0$ |
| dfa1 mean [Hz] | Frequency at which the distribution of frequency amplitudes reaches the 1st quartile, mean across all time segments |
| dfa2 start [Hz] | Frequency at which the amplitude distribution reaches the 2nd quartile at start of vocalization |
| dfa2 mean [Hz] | Frequency at which the amplitude distribution reaches the 2nd quartile, mean across all time segments. |
| dfa2 max [Hz] | Maximal frequency at which the amplitude distribution reaches the 2nd quartile |
| dfa3 start [Hz] | Frequency at which the amplitude distribution reaches the 3rd quartile at start |
| dfa3 mean [Hz] | Frequency at which the distribution of frequency amplitudes reaches the 3rd quartile (see Figure1), mean across all time segments |
| Range max [Hz] | Maximum difference between highest and lowest frequency (see Figure1) across all time segments |
| Range mean [Hz] | Mean difference across all time segments |
| PF mean [Hz] | Mean peak frequency (frequency with the highest amplitude, see Figure1) |
| PF max [Hz] | Maximum peak frequency across all time segments |
| PF tonal mean [Hz] * | Mean of the frequencies with the highest amplitude across all tonal time segments |
| PF tonal max [Hz] * | Maximum of the frequencies with the highest amplitude in the tonal time segments |

| | |
|---|---|
| PF max location | Relative position of maximum peak frequency (1/duration [ms] * location PFmax [ms]; ranges between 0 and 1 |
| dfb1 mean [Hz] | Mean of the lowest dominant frequency band across all time segments; corresponds in tonal segments to $F_0$ |
| dfb1 % | Percentage of time segments with a dominant frequency band |
| dfb2 % | Percentage of time segments with a second dominant frequency band |
| dfb1 mean modulation [Hz] | Mean difference between original dfb 1 course and average |
| dfb1 min location | Relative position of dfb1 minimum in the call, ranges between 0 and 1 |
| dfb1 trend | Linear trend of the dfb1 course, calculated with the method of least squares |
| dfb ratio 1 | Amplitude ratio between dfb1 and dfb2 (dfb1/dfb2) |
| dfb ratio 2 | Amplitude ratio between dfb1 and dfb3 (dfb1/dfb3) |
| FLS1amp mean | Mean amplitude of the $1^{st}$ formant-like structure |
| FLS2 % | Percentage of time segments in which a second formant-like structure could be detected |
| FLS2 start [Hz] | Frequency of the 2nd formant-like structure at the start of a vocalization |
| FLS2 mean [Hz] | Mean frequency of the second formant-like structure |
| CS mean | Mean consistency of vocalization, measured as the correlation coefficient of successive time segments |
| CS std dev | Standard deviation of the mean consistency |
| CS max location | Location of the maximal difference in the consistency values |