GENERAL DISCUSSION

Emergence and development of infant vocalizations

The results of this study have shown, that the nonverbal vocalizations of infants in their first year of life can be classified into 12 call types based on their acoustical structure. All call types were uttered in positive as well as in negative emotional context. Both normally hearing (NH) and hearing-impaired (HI) infants shared the same vocal repertoire. There were no differences in the time of emergence of the individual preverbal call types between NH and HI infants; the only exception was the emergence of babbling. Babbling emerged later or not at all in the vocal repertoire of the HI infants, an observation in line with results of other studies (e.g., Eilers & Oller 1994; Oller et al. 2001; Stark 1983). I found minor age-related changes in the acoustic structure of the call types cry, short cry, coo/wail and moan in both infant groups. The acoustic changes were characterized mainly by an increase in harmonic-to-noise ratio and homogeneity of the vocalizations, a decrease in frequency range and a shift of acoustic energy from higher to lower frequencies. These changes are presumed to reflect maturation processes. The downward shift of acoustic energy may be a consequence of the growth and descensus of the larynx (Lieberman 1984; Papoušek 1994; Ploog 1992). The age-related changes in call homogeneity and harmonic-to-noise ratio might be the result of the improvement of subglottal air-pressure control (Langlois et al. 1980), brought about by neuromuscular maturation processes (such as progressing myelination) as well as by training of muscular coordination (Boliek et al. 1996).

Together, these findings show that the majority of preverbal utterances are predetermined (see also Bloom et al. 1993; Oller 2000) and their emergence and production are to a great extent independent of the degree of auditory input. The only call type for which an influence of hearing impairment on its emergence could be found was babbling. It is important to note that the HI infants in this study were profoundly hearing-impaired (about 100 dB or more hearing loss on both ears), but not absolutely deaf. They had residual hearing ability that was supported by hearing aids from the second recording on. However, in my HI group, babbling did not emerge at a younger age than known for totally deaf infants (Oller & Eilers 1988; Oller et al. 1985). This indicates that even with hearing aids the hearing ability of the HI infants was far from normal.

Differences in the acoustic structure of calls related to hearing ability

Of the three call types whose acoustic structure could be compared in detail (cry, coo/wail and moan), only cries showed subtle differences between NH and HI infants. The major outcome of hearing deficiency was an increase in peak frequency and frequency range in the cries of HI infants. Furthermore, the cries of HI infants in this study showed an earlier maximum of the 1st dominant frequency band or peak frequency compared to the cries of NH infants. A study by Möller and Schönweiler (1999) also showed that cries of HI infants differed from those of NH infants. Cries of HI infants, in their study, had a longer duration and a more complex melody contour. The changes in the maximum of the 1st dominant frequency and the peak frequency found in my analysis may be interpreted as a change in melody parameters. In contrast to my results, however, Möller and Schönweiler found a decrease in the energy parameters of HI cries. This discrepancy might be due to the fact that Möller and Schönweiler analyzed only distress cries, while I, additionally, included cries uttered in positive emotional contexts.

It is possible that the increase in peak frequency and frequency range points to an increase in amplitude in HI cries. For methodical reasons, amplitude could not be measured in this study, because the infants were recorded at home in a variety of different situations and partially by their parents. A relation between frequency parameters and amplitude has been shown in a study by Hammerschmidt and Jürgens (2000). They found that there was a positive correlation between call amplitude and various frequency parameters (fundamental frequency, peak frequency, distribution of frequency amplitudes) in the vocalizations of squirrel monkeys. Additionally, Hammerschmidt and Mundry (in prep.) showed that an upward shift in frequencies under specific circumstances is interpreted by listeners as an increase in loudness.

A number of studies reported that people with no or severely reduced auditory feedback speak with an increased sound pressure level (SPL) and SPL inflection (Lane et al. 1997; Letowsky et al. 1993; Perkell et al. 2000). Similar results were found for other mammals. In experiments with cats in which deafness was induced with ototoxic substances or surgery, subjects were found to emit sounds of double intensity when compared to the sounds emitted by normal-hearing animals. With increasing age, deafened cats continued to increase the intensity of their vocalizations, suggesting that the auditory feedback is necessary to adjust the emitted volume (Hultcranz et al. 1991; Leake-Jones et al. 1982; Romand & Ehret 1984; Shipley et al. 1988). In non-human primates in which deafness was induced by electric coagulation subjects were found to increase the intensity of trill calls (Coste & Pfingst 1997; Talmage-Riggs et al. 1973). However, a direct correlation between hearing impairment and loudness of human infant cries still has to be proved.

As mentioned above, the differences in HI and NH calls were only subtle in cries and not significant in other call types. Due to the generally high variability of infant vocalizations, a high number of infants as well as a high number of utterances per infant have to be recorded and compared to detect subtle differences. Therefore, the acoustic analysis of single preverbal calls does not seem to represent a practicable diagnostic tool for the detection of hearing impairment in early infancy.

Differences in acoustic structure of calls related to emotion

In contrast to the subtle differences in acoustical structure related to hearing impairment, there was a clear emotion-related influence on call structure. I found that one and the same call type uttered in positive and negative emotions differed in acoustic structure. This was true for NH as well as for HI infants. In other words, my results indicate that the basic information, whether an infant feels good or bad, is encoded in the structure of its vocalizations. Furthermore, the transmission of this important information does not seem to be affected by hearing impairment. Vocalizations uttered in negative emotions had a longer duration than vocalizations uttered in positive emotions Additionally, depending on the call structure, one or more parameters describing frequency range and pitch showed higher values in negative than in positive emotions. An increase of pitch parameters in negative emotional context was also found in a study of Papoušek (1992) which revealed that infant vocalizations expressing discomfort differ from those expressing comfort by their higher amount of spectral energy above 1000 Hz. In adults, several studies showed a correlation between frequency parameters and negative emotional context (Banse & Scherer 1996; Hammerschmidt et al. submitted; Protopapas & Lieberman 1997; Ruiz et al. 1996). Ruiz et al. (1996), for example, analyzed speech drawn from both artificial (psychological test) and real (pilots, immediately before an air crash) stressful situations. They found stress-related increases in the mean and maximum fundamental frequency and in the formant frequencies. However, it was shown by Hammerschmidt et al. (submitted) and by Banse and Scherer (1996) that frequency parameters in adult vocalizations do not only increase in negative emotional contexts, but increase with the intensity of a given emotion in general (see also General Introduction). Both studies analyzed words produced by actors that were asked to express different emotions. Thus, it was possible to control each produced utterance for intensity. In the present study, in contrast, the focus was on natural recording conditions to get a broad variety of emotional utterances. Therefore, I had no possibility to estimate the call amplitude in order to test the influence of on call structure for a given emotion. Increased pitch has proved to be an indicator of negative emotional state also in studies on vocalizations of other mammals. Various situations have been analysed: effects of painful procedures: (piglets (*Sus scrofa*): Weary et al. 1998), isolation stress (Common marmosets (*Callithrix j. jaccus*): Schrader & Todt 1993; pigs (*Sus scrofa*): Schrader & Todt 1998), and alarm situations (Redfronted lemurs (*Eulemur fulvus rufus):* Fichtel & Hammerschmidt 2002; Barbary macaques (*Maccaca sylvanus*): Fischer et al. 1995). The analyses revealed a positive correlation between the assumed aversiveness of the elicited emotional state and frequency parameters. All these studies used indirect methods to estimate the aversiveness of a vocalization-eliciting situation. Similar results, however, have also been found in a study in which the degree of aversiveness / pleasantness of the concomitant emotional state accompanying a specific vocalization has been measured more directly. That study was carried out by Jürgens (1979) in the squirrel monkey (*Saimiri sciureus*). Different calls were elicited by activating specific brain regions responsible for controlling these calls. Activation was carried out by electrical stimulation of stereotactically implanted intracerebral electrodes. The animals were able to switch on and off the vocalizing-eliciting stimulation themselves. In this way, a quantitative measure indicating the degree to which the stimulation was avoided or sought could be obtained. With this method, the aversive or pleasant quality of the affective state underlying the production of specific calls could be determined. A following acoustic analysis of the vocalizations elicited in this study revealed that the calls expressing different degrees of aversion differ in their acoustic structure (Fichtel et al. 2001). Here again, more aversive calls showed higher pitch (characterized by peak frequency, distribution of frequency amplitudes and dominant frequency), higher frequency range and a higher amount of non-harmonic energy. Additionally, a playback study showed that these differences in call structure agreed with the squirrel monkeys assessment of the calls (Fichtel & Hammerschmidt 2003).

In summary, studies on vocalizations of different species, including humans and non-human mammals, revealed that similar acoustic parameters encode the degree of aversiveness of the emotional state underlying a specific vocalization. This supports the view of Jürgens (1993; 2003), who stated that the vocal expression of emotion in humans has deep-reaching phylogenetic roots. That this way of expressing emotions is already given in human infants, and even in hearing-impaired ones, strengthens this view. This makes clear that the basic patterns of emotional expression do not need to be learned.

While it was no problem to differentiate acoustically calls uttered in positive and negative emotional contexts, even if they were of the same call type, it was impossible to distinguish specific positive or negative emotions by call structure. I found no acoustic parameter that significantly differentiated single emotions. Neither unease and anger, nor joy, contentment and interest could be separated in any of the call types tested. As discussed in Chapter 1 (Discussion), there are several possible reasons for the low discriminability of emotions of the same valence category. It might be that in young infants, the emotional system is still developing and therefore not as differentiated as the differentiation (seven emotional labels) I asked for (see Lewis 1993; Sroufe 1979). Another explanation could be that parents differ in labelling emotions which are similar in valence. One mother, for instance, might have named a certain emotional state joy, while another mother used the term contentment. Furthermore, not every vocalization uttered in a certain emotional state must be typical for that emotion. However, a cross-check analysis, using context categories instead of emotional categories (see Chapter 1) produced the same results as the analysis based on emotional ratings. Therefore, it seems unlikely that a mismatch between the infants' emotions and the parents' ratings is the main factor for the low success in discriminating similar emotions. It is more likely that the parents used other information than the acoustic structure for their detailed judgment. During the recordings, they could see the facial expressions and gestures of their infants and had a lot of contextual information surrounding the recorded situations. Furthermore, the parents did not hear single call types but sequences of vocalizations uttered by their infants. The composition of the call sequences, therefore, could have been an additional source of information about the emotional state of the infant.

Differences in the composition of call sequences related to emotion

Until now, the composition of call sequences uttered by preverbal infants has received little attention. Besides the general knowledge that extensive crying indicates distress in children and that babbling or cooing is more probable in pleasant situations, the detailed sequence of infant utterances seems to be usually seen as a random process, happening because the infants practice vocalization. Whether the composition of call sequences are of communicative relevance and which factors influence sequence composition has hardly been studied. As far as I am aware, there is no study that has inspected the composition of call sequences in relation to different emotional contexts.

Nevertheless, as studies of vocalizations in human infants, non-human primates and birds have shown, the inspection of series of calls leads to an insight into the rules of signalprocessing and communication (Hultsch 1980; Todt 1986; Todt & Hultsch 1980). A study by Todt and Wiedenmann (in Todt 1988), for example, investigated the reactions of listeners to series of cries uttered by human infants. The results of that study showed that listeners did not react with checking and distress responses in the same way to all single cries in a series. Instead, they reacted to special temporal and structural changes within the cry sequence. Furthermore, when a distress response in listeners was elicited once through a change in the cry series, the probability of further distress responses was increased during the rest of the playback. Todt (1988) concluded that a coherent string of cries can be regarded as a super-signal, wherein both the sequential and temporal organization of constituents may be communicatively significant, and thus affect the interactional role of crying.

The results of my study support the assumption that the organization of call sequences is of communicative value. Although all preverbal call types are principally uttered in positive and in negative emotional context, the rate of certain call types within sequences changes with different emotions in NH infants. The most prominent differences were found between sequences uttered in positive and negative emotions, but also sequences with similar valence could be distinguished. Specifically, the composition of joy sequences differed from the composition of sequences expressing interest and contentment. In order to test whether changes in the composition of infant call sequences really provide meaningful information for receivers, further investigations, including playback experiments, are needed.

Differences in the composition of call sequences related to hearing ability

In contrast to the in single calls, sequence composition showed clear differences between NH and HI infants.

First, while in NH infants sequence composition changed with emotional context, this correlation was lacking in the sequences of HI infants. Since all call types that showed different frequencies in relation to emotional context in NH infants could be produced by HI infants as well, the lack of emotion-related differences in HI sequences possibly reflects difficulties of HI infants in steering or adjusting the point of time at which a certain call type should be uttered. Until now, it is largely unknown which components exactly enhance or reduce the probability of uttering a certain call type within a stream of vocalizations in preverbal infants. However, my results indicate that auditory input is one of those components. Supporting evidence comes from several studies which show that the quantity and quality of infant vocal production can be influenced by social feedback (Bloom 1988; Bloom et al. 1987; Masataka 1993; Ramey & Ourth 1971; Weisberg 1963) and especially by auditory stimulation from adults (Haugan & McIntire 1972). It might be that NH infants learn through vocal feedback by their parents to produce certain call types more or less often in certain emotional contexts and that HI infants, due to their auditory deficiency, fail to do so. Further investigation should systematically test, which call types can be increased or decreased by vocal feedback from adults.

Second, there were general differences in call sequence composition of NH and HI infants. Independent of emotional context, HI infants uttered less babbling and short cries and more coo/wail and croak. There are different possible explanations for these findings. First, it is well known that the production of babbling requires auditory input (Oller & Eilers 1988). Therefore, the emergence of babbling is delayed or absent in HI infants. The higher rate of coo/wail and croak in the sequences of HI infants is possibly due to the fact that NH infants, reaching the developmental stage of canonical babbling, replace part of the vocal types of the preceding stage by canonical babbling. HI infants, in contrast, simply continue to produce coo/wail and croak with the same rate as in younger age. According to Oller (2000), all three call types are protophones (for definition see Chapter 3, Discussion), but coo/wail and croak represent protophones of an earlier developmental stage than babbling, and therefore may be seen as precursors of the latter. Second, it was suggested that HI infants increase their vocal production to increase auditory input (Locke & Pearson 1992). It is possible that it is easier for preverbal infants to voluntarily increase the rate of the more flexible protophones coo/wail and croak than to increase the rate of other, more fixed call types. This, however, requires further investigation.

In sum, the investigation of call sequences indicates that further inspection of sequence composition under different circumstances might be might be helpful for our understanding of how preverbal infants process their vocal sequences. In terms of diagnostic value, however, the inspection of call sequences seems to be too labor-intensive to provide a practicable diagnostic tool. Due to the high intra- and intersession variability of infant vocalizations, the differences between NH and HI sequences will hardly be detectable by a small sample of vocal series. Furthermore, parents will probably not be able to report, whether the relevant call types are uttered more or less often in specific emotional contexts, or if there is an increased or decreased rate of specific call types uttered by their infant compared with a certain standard. The only exception to this general statement concerns the call type babbling. As a study of Oller et al. (2001) showed, it is not complicated to inform parents what babbling sounds like, and that they should pay attention to whether babbling is uttered or not. Thus, information on the onset of canonical babbling obtained from parents can be used as a reliable tool for risk assessment of hearing impairment in infants aged 6-12 months (Oller et al. 2001).

This study revealed that the vocal production of HI infants is in many aspects very similar to the vocal production of NH infants in the first year of life. Both infant groups shared the same vocal repertoire, most of their call types emerged at the same age, there were only minor differences in the acoustic structure of NH and HI vocalizations, and HI infants were equally able to encode their general emotional state in the structure of their calls as their NH peers. These findings may explain why parents often recognize the hearing deficiency of their infants relatively late. Additionally, they imply that probably neither acoustic analysis nor the inspection of call sequences are practicable diagnostic methods. Furthermore, the results showed that NH and HI infants encode positive and negative emotional states in call structure in the same way. Apparently, these emotion-related differences in call structure do not need be learned. Similar emotional correlates in call structure are also described for other mammalian species, indicating that this kind of emotional encoding is phylogenetically old. Emotions are also encoded in the composition of call sequences. This kind of emotional encoding, however, is influenced by hearing impairment, suggesting that sequence composition is dependent on auditory learning.