

TR - A - 0054

Perceptual Units of the Infant Cry

乳幼児の泣声の知覚における単位

*Taeko TSUKAMOTO * and Yoh'ichi TOHKURA*

**Konan Women's University*

塚本妙子 (甲南女子大) 東倉洋一

1989. 6. 15

ATR 視聴覚機構研究所

〒 619-02 京都府相楽郡精華町乾谷 ☎ 07749-5-1411

ATR Auditory and Visual Perception Research Laboratories

Inuidani, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411

Facsimile: +81-7749-5-1408

Telex: 5452-516 ATR J

Perceptual units of the infant cry

TAEKO TSUKAMOTO

Department of Psychology, Faculty of Letters, Konan Women's University, Higashinada-ku, Kobe 658

YOH'ICHI TOHKURA

Hearing & Speech Perception Department, ATR Auditory and Visual Perception Research Laboratories, Seika-cho, Kyoto 619-02

Perceptual units for category identification of infant cries are studied. Original samples are classified into three categories: the hunger cry, the call cry, the anger cry and employed in perceptual experiments. In order to generate stimuli, each of the cry samples is first segmented into single segment units according to the infant's breath group vocalization. The single segment units are combined with each other in temporal order to generate two-, three-, five- and seven-segment unit stimuli. In identifying these stimuli, 52 subjects are instructed to make a forced choice among the three categories. As a result, it is shown that category identification rates are dependent upon the number of segments constructing each stimulus. However, the identification rates are saturated at two-segment units in the call cry and three- to five-segment units in the hunger and the anger cry. It indicates that these are the perceptual units which are about six to eight seconds in temporal duration across all three categories.

INTRODUCTION

The infant cry is produced in various situations where infants feel uncomfortable. In spite of powerful effects of the infant cry on the infant-mother interaction, it is not very easy for the mother to identify the cause of the infant cry only by listening to it. However, there is some experimental and observational evidence suggesting that the mother can identify the infant cry categorically only by listening to it (Wolff, 1969 ; Tsukamoto, 1983). On the other hand, several studies of the analysis of acoustic features in the infant cry have claimed that there is a positive relationship between melody types observed in breath group vocalizations and the infant's situation (Wasz-Höckert et al., 1968 ; Rosenhouse, 1977 ; D'Odorico, 1985). Moreover, it was found that there are categorical differences in time series patterns according to phonation/non-phonation of the infant cry (Kobayashi et al., 1986; Tsukamoto & Katagiri, 1988). The results indicate that there are two kinds of vocal information in the infant cry, i.e., segmental acoustic features based upon one or a small number of infant

This paper has been presented in the 117th ASA Meeting, Syracuse, May 22-26, 1989 and in the 3rd International Infant Cry Workshop, Helsinki, July 13-16, 1989.

breath groups and supra-segmental acoustic features over several or more breath groups. It is also inferred that the cry categories are identified using these two kinds of vocal information by the mother. Here, our questions are what acoustic features are important and how long must they be for cry category identification. The purpose of this study is to answer these questions by finding perceptual units which include acoustic features sufficient to represent each cry category. These perceptual units should have the two kinds of vocal information.

When we conduct perceptual experiments of the infant cry, it is important how we chose cry samples for the experiment. Taking account of some suggestions in previous research, the cry samples used in the experiment were chosen by considering the following two points: the infant's age and the situation in which the cry is produced. (1) Infant age: The samples are produced by infants aged from 1 to 5 months. The reason is that cry categories are not yet differentiated completely in the month after birth. Furthermore, rhythmic patterns of the cry sounds which are quite simple in neonates become more variable prior to one month of age to reflect the organizational development of the central nervous system (Illigworth, 1955 ; Futatsugi, 1979 ; Lester, 1985). Another reason is that the cry as a communication behavior changes its role and function around six months of age (Tsukamoto, 1985). (2) Situation in which the cry is produced: The samples are produced in situations observed in infant-mother interaction. In most previous studies, cry samples produced by infants in experimental situations have been used. The kinds of cry categories produced in such situations are very limited, and each cry category thus obtained is not always the same as the one which was intended by the experimenter.

I. METHOD

A. Recording of the infant cry

Four infants' cries were video taped (SONY SL-F1; SONY CCD-V8AF2) when observing the infant-mother interaction. The infant ages observed were from 3 weeks to 12 months. A team of observers was composed of several students of psychology and one of the authors.

B. Classification of the infant cry

Classification of the infant cry was performed by a team of observers. First, the observers identified the cause of the infant cries produced in each observational situation through watching repeated playbacks of VTR and by investigating all materials obtained during the observed situations. Second, the infant cries thus identified were evaluated by an identification coincidence among the observers. They were eventually classified into several cry categories according to the situations which caused the infant cry.

C. Samples

Cry samples were employed with three categories which included the hunger cry, the call cry^{a)}, and the anger cry produced by all four infants (YU, TE, TO, MA) used in the recording. The samples were chosen from those when infant ages were 9 to 17 weeks for the reason described in the introduction. They were digitized by an A/D converter with a sampling frequency of 20 kHz and stored in the computer (MASSCOMP 5600). These were original samples used to generate various cry stimuli in perceptual experiments of cry identification.

D. Subjects

Subjects were selected from undergraduate and graduate students and members of the ATR Auditory and Visual Perception Research Laboratories. There were 23 males and 29 females ranged in age from 18 to 40 years old.

E. Experiments

Experiments were composed of two parts: A learning session and a test session.

1. Learning session

Stimulus. The contents of stimuli used in the learning session are shown in Table 1. These three stimuli, each of which represents each of the three categories, were selected from the original samples described above. They were produced by three different infants' samples (YU, TE, TO). Stimulus durations were long enough (about 26 to 30 seconds) to include both segmental and supra-segmental features for cry category identification.

Procedure. Experiment procedures in the learning session are described below. They consist of three trials --- learning, identification and confirmation of learning. (1) In the learning trial, stimuli were presented twice in succession --- the call, the hunger, and the anger cry. Subjects were requested to learn distinctive features of each cry category by studying each stimulus category prior to listening. (2) In the identification trial, stimuli were presented twice in random order. Subjects were instructed to identify them by forced choice among the three categories. The results of this trial were regarded as an achievement index of categorical learning. (3) As confirmation of learning, the stimuli were presented again in order --- the call, the hunger, and the anger cry with each category name given to the subjects prior to listening.

2. Test session

Stimulus. The contents of stimuli used in the test session are shown in Table 2. The samples as well as the infants used to produce them for the test session were both different from those used in the learning session. The three stimuli were also used as full segment units. Their durations were as long as the stimuli used in the learning session. Full segment units were segmented into single segment units based on the fact that the breath group of the infant cry consists of the total vocalization occurring during a single expiration or inspiration. Segmentation into single segment units was carried out using a sound waveform editing program on a computer (MASSCOMP 5600). The waveform editor effi-

Table 1 Cry stimuli used in learning sessions

Category	Infant [Age in weeks]	Duration (sec)
Call	YU [17]	26.0
Hunger	TE [17]	28.6
Anger	TO [9]	26.8

Table 2 Cry stimuli used in testing sessions

Category	Infant [Age:weeks]	Duration (sec)	Number of Segments	Segmental Duration (sec)		
				Ave.	Max.	Min.
Call	MA [15]	26.4	9	2.93	8.82	0.53
Hunger	MA [12]	22.7	18	1.26	2.79	0.31
Anger	MA [13]	25.5	16	1.60	3.02	0.52

a) The call cry is defined as the cry calling for infant-mother interaction.

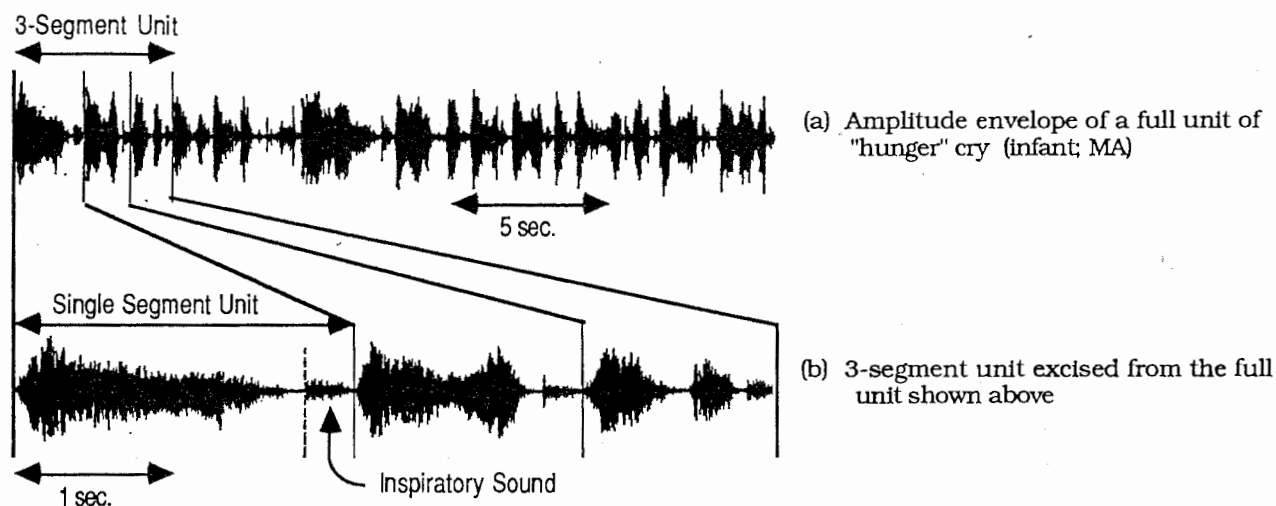
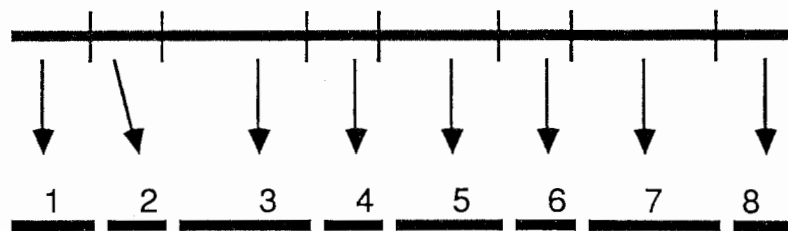


Fig.1 Segmentation of cry waveforms. A full unit cry (a) is segmented into single segment units based upon inspiratory sound portions. These single segment units are combined in temporal order to produce n-segment unit stimuli (e.g., 3-segment unit (b)).

ciently segmented cry samples with observations of both their sound spectrograms and waveforms. It was also able to confirm the segmentation results by listening to each cry sample after segmentation. As examples, amplitude envelopes of a full segment unit of the hunger cry and single segment units segmented from it are shown in Fig.1. These single segment units thus obtained were used as one segment unit stimuli in the test session. Consequently, three stimuli, i.e., full segment units of the call, the hunger and the anger cry in Table 2, were segmented into nine, eighteen, sixteen segments, respectively. They were combined in temporal order to produce two-, three-, five- and seven-segment unit stimuli for each category. As illustrated in Fig.2, initial segments to produce each were chosen by one segment shift in two- and three-segment units, two segment shifts in five-segment units, and three segment shifts in seven-segment units. Thus, there were 43, 43, 43, 22, and 15 units as one-, two-, three-, five-, and seven-segment units, respectively. There were 166 stimuli in total. In the test session experiment, 169 stimuli, including three full segment units, were prepared.

Procedure. In the test session, all the 169 stimuli were randomized and presented to the subjects after the 10 dummy stimuli. To avoid stimulus order effects in presentation, three stimulus series randomized differently were prepared. In each stimulus series, inter stimulus intervals (ISI) were set at four seconds. One of three stimulus series was presented to balanced groups of subjects. The subjects were requested to identify each stimulus by a forced choice among three categories, the hunger, the call, and the anger cry. In both the learning and the test sessions, all the stimuli were recorded once by DAT (DIGITAL AUDIO TAPE DECK, SONY DTC-1000ES) and presented to subjects with headphones in a sound booth. A learning session procedure lasted about 10 minutes and a test session procedure about 40 minutes.

Original Sample
(full segment unit)



Single segment unit
and segment number

Multi Segment Units

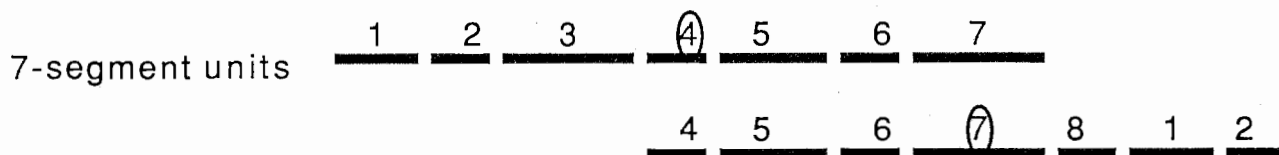
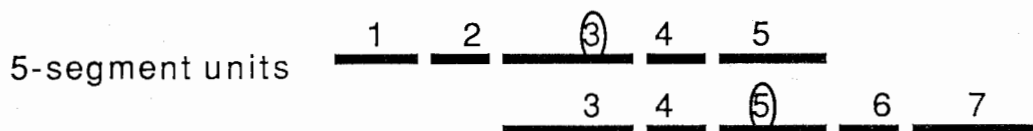
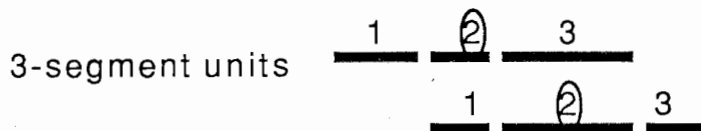
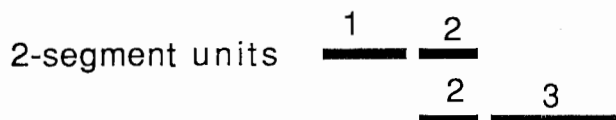


Fig.2 Illustration of cry segment units used in category identification tests. A full cry unit is divided into n cry segments. A 1-segment unit is a unit with only one segment. An i -segment unit is composed of successive i segments. In the identification test, 1-, 2-, 3-, 5-, 7-segment and full units are used as stimuli. The numbers marked by a circle represent central segment numbers of each stimulus.

II. RESULTS

According to the results of the learning session, subjects were classified into two groups: One group of 44 subjects (85%) correctly identified all six stimuli consisting of the three cry categories. Second group of 8 subjects failed to do so. Taking into account the learning session procedures and their purpose, it can be assumed that the former group of subjects (L for learned) were able to learn some distinctive features of each of the three cry categories. Also, it can be said that the latter group (N for non-learned) failed to do so. In the following discussions, the data given by the N group was omitted.

Based upon the test session results, the L group was divided again into two groups: The A group of 20 subjects who identified correctly three full segment units, one for each of the three cry categories

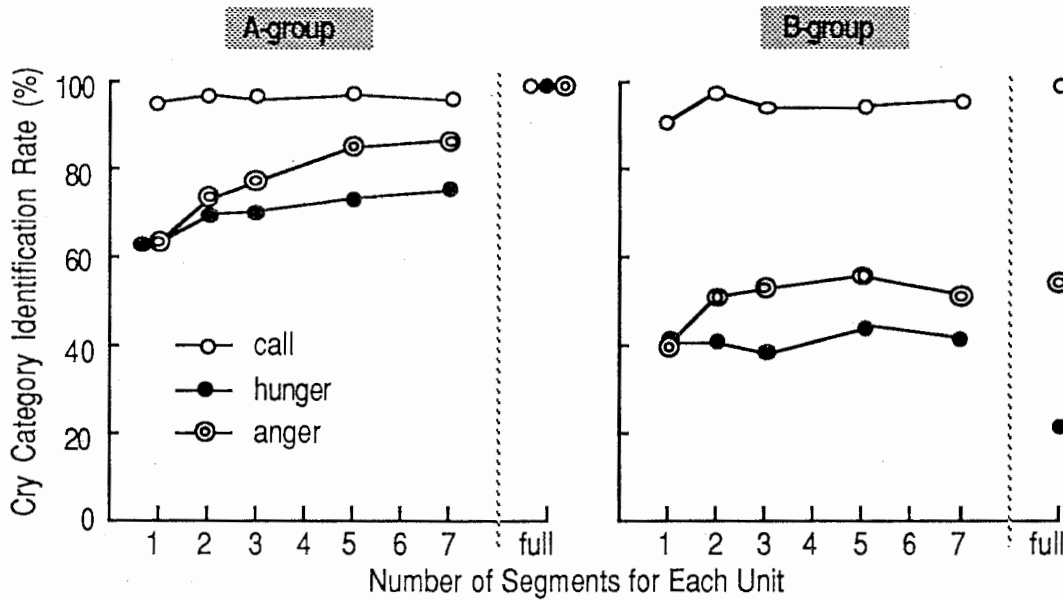


Fig.3 Cry category identification rates for A- and B-group. The A-group consists of 20 subjects who correctly identified each full unit of three cry categories by 100%. The B-group consists of 24 subjects who failed to identify full unit cries correctly.

and the B group of 24 subjects who failed to identify them. For both A and B groups, cry category identification rates were shown as a function of the number of segments consisting of a unit in Fig.3. For the call cry, identification rates are quite high across all the number of segments in both A and B subject groups. The call cry stimuli were identified correctly even when presented as one segment unit stimuli. For the hunger and the anger cry, the identification rates showed a clear dependency on the number of segments in the A group. That is, the more the number of segments increases, the higher the identification rates become. However, in the B group, it is difficult to observe the same kind of identification rate dependency. Identification characteristics for the A group are discussed based upon error properties in identifying the three cry categories. The identification error properties for the A group are shown as a function of number of segments for each cry unit in Fig.4. In two kinds of errors, confusing the hunger

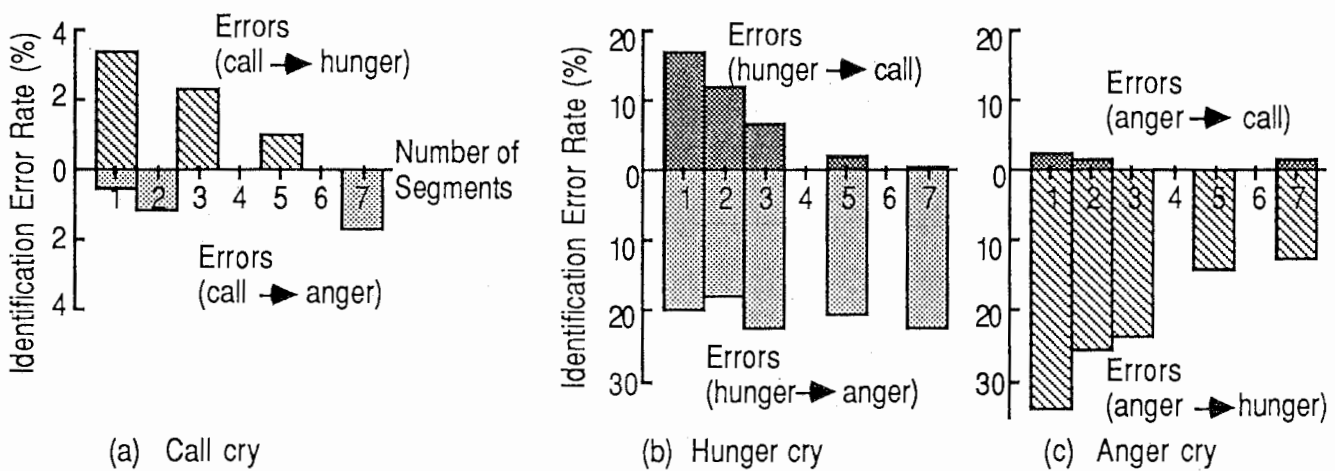


Fig.4 Error properties in identifying cry units of three categories for A-group. Error rates are shown as a function of the number of segments for each cry unit. Two kinds of errors -- misjudgement of "hunger" as "call" and misjudgement of "anger" as "hunger" -- decrease significantly as the number of cry unit segment increases.

and call cries, and the anger and hunger cries, the error rates decrease significantly dependent upon the number of segments. As a whole, the most frequent errors involved confusing the hunger and anger cries. In the B group, error rates were higher than those in the A group, but their characteristics were not so clear as those in the A group.

The analysis of variance (ANOVA) was performed for the results in the A group. The main effects were significant ($p < .001$) and indicated that the A group made significant categorical judgments in infant cry identification. Identification rates for each category were high in the following order: call, anger, and hunger cries. They showed a clear dependency on the number of segments. The identification rates increased as the number of segments in each unit became larger, but they were saturated at two segments for the call cry, and at three to five segments for the hunger and the anger cries. Concerning the hunger and the anger cries which were most frequently confused with each other, the identification rates for each segment unit stimulus were calculated. The results for the hunger cry were illustrated in Fig.5. As can be seen in Fig.5, the identification rates for each stimulus are quite different even when it has the same number of segments. Roughly speaking, identification rates become high at the same stimulus locations (i.e., stimulus numbers) across the number of segments. This tendency was more apparent for the hunger cry.

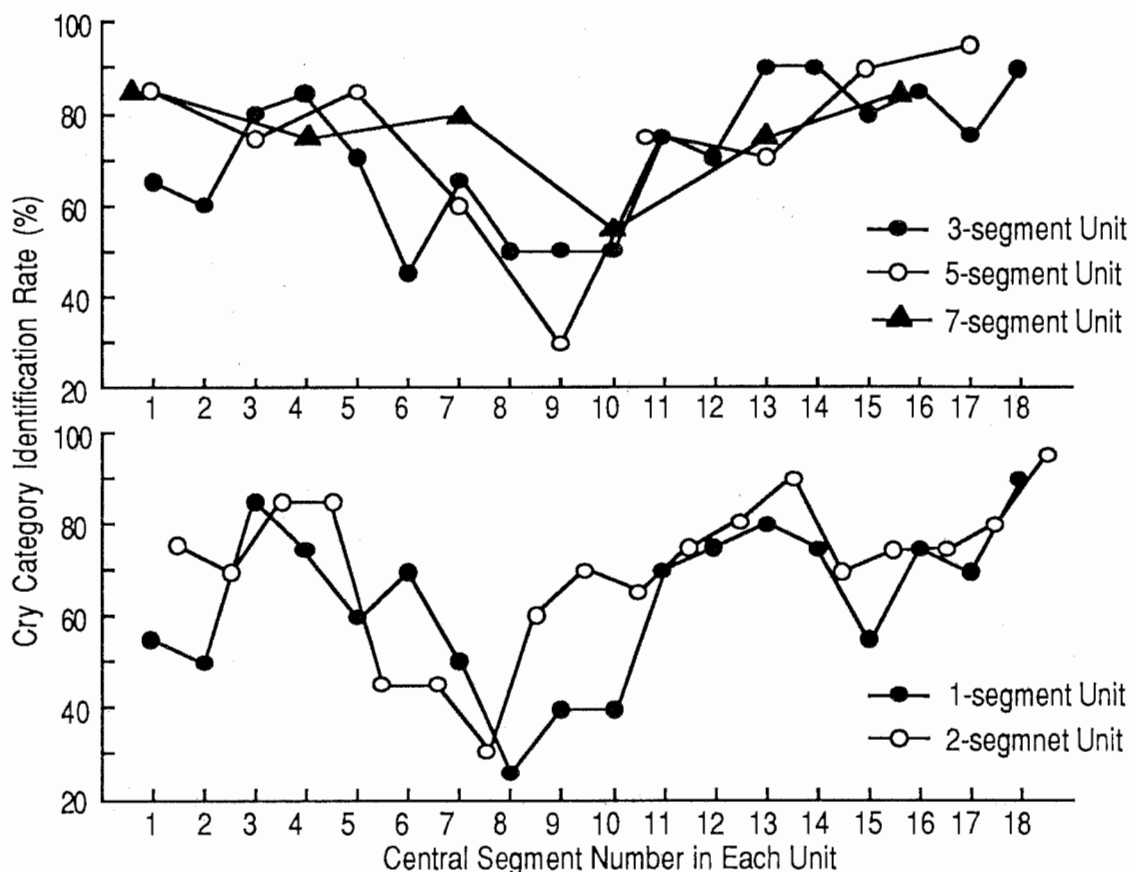


Fig.5 Category identification rates for "hunger" cries as a function of segment numbers in each unit for each of 1-, 2-, 3-, 5- and 7-segment units.

III. DISCUSSIONS

Subject properties are discussed in details. Through both the learning session and the test session, subjects were classified into groups. First, as results of the learning session, the subjects were divided into two groups: L group and N group. The L group consists of the subjects who succeeded in identifying all of the six stimuli: two repetitions of the three cry categories. The N group consists of subjects who failed to identify them. Taking the learning session purpose into account, it is concluded that the L group was able to learn some distinctive features of the three categories. Therefore, further analyses of experimental results were carried out only for the L group. Secondly, the L group was divided into two groups: A and B group based upon the identification results of three full segment units in the test session. The A group consists of subjects who were able to identify all the three full segment units correctly. On the other hand, the B group consists of subjects who failed to identify them.

As can be seen in Fig.3 different characteristics in category identification between the A and the B group are observed. It is obvious that the A group shows better identification rates on average than the B group. In addition to this, however, identification rates in the A group increase significantly as the number of segments in each stimulus unit becomes larger. On the other hand, in the B group, the identification rates show no clear dependency on the number of segments and they are not particularly high even when the number of segments is as large as seven.

These results tell us that distinctive features for the cry categories learned by the A group are somewhat different from those learned by the B group. As far as the experiment results in the learning session are concerned, both groups (i.e., they belong to the L group) were believed to have learned some of the distinctive features. However, the results in the test session revealed that the distinctive features learned by the B group are neither sufficient nor reliable enough to identify the cry categories in the test session. What was learned in the learning session might not be the distinctive features but the feature differences among the three used in the learning session experiments. According to the results for the A group, identification rate characteristics agree with our assumption described in the introduction. The rates increase as a function of the number of segments but saturate at about two segments in identifying the call cry and at about three to five segments in identifying the hunger and the anger cries. The number of segments where the identification rates saturate are thought to be the minimum "perceptual units" of the infant cry to allow us to identify the cry categories reliably. Two segments in the call cry and three to five segments in the hunger and the anger cries correspond to about six to eight seconds in length, referring to mean segment durations of each cry category in Table 2. It is very interesting that perceptual units are approximately the same across the three cry categories in their durations although the number of segments consisting of the perceptual units are different among the three categories. Detailed analyses of identification rates for each stimulus unit showed that there are quite large differences in identification rates among stimulus units even when they belong to the same cry category. As is seen for the hunger cry in Fig.5, these experiment results are different from each other. Presumably, there exists a stimulus unit in which the distinctive features are rich enough to be identified into the correct cry category even if the stimulus unit is a

short unit such as a one- or two-segment unit. Also, these discussions suggest that, if there is a segment with rich distinctive features located somewhere in the full segment unit, not only does this segment give a high recognition rate but also a multi-segment unit including this segment tends to be identified correctly. As a matter of fact, this kind of phenomena can be found in Fig.5 when the stimulus (or unit) identification rate differences among the number of segments in each stimulus unit are observed carefully. The same analyses described above were also performed for the anger cry. However, those observations discussed for the hunger cry were not as clear as for the anger cry.

The reasons that phenomena for the hunger cry were observed more clearly than for the anger cry seem to be as follows. For the cry samples used in our experiments, segments with high and low identification rates are temporally localized for the hunger cry, but not for the anger cry. In other words, a high identification rate segment adjacent to a low identification rate, and vice versa, produce a sort of compensation effect between two segments with different features.

In this paper, perceptual units of the infant cry were studied. We summarize our findings as follows. (1) In perceptual experiments to study relationships between cry category identification rates and the number of segments in each stimulus units, it was shown that the identification rates increase as the number of segments becomes larger. (2) The identification rates saturate where the number of segments is two for the call cry and three to five for the hunger and the anger cries. They are the perceptual units of the infant cry. Duration of the stimulus units consisting of these numbers of segments for each cry category is about six to eight seconds, which are the same across the category. (3) For short segments such as a one- or two-segment unit, there is a variety of distinctive features included in the unit. Due to this, differences in identification rates for each stimulus unit are quite large even when the unit belongs to the same category. (4) Identification rates for a one-segment unit consisting of rich distinctive features are high. Also, a multi-segment unit including the segment unit with a high identification rate tends to be identified correctly. (5) Further studies on relationships between the perceptual units obtained in this paper and their acoustic features are needed.

REFERENCES

- D'Odorico, L. 1984 Non-segmental features in prelinguistic communications: An analysis of some types of infant cry and non-cry vocalization. *Journal of Language*, 11, 1, 17-27.
- Futatsugi, T. 1979 Cry characteristics in newborn infants and developmental meanings of the infant cry. *Japanese Journal of Child Psychiatria*, 20, 3, 161-177. (In Japanese).
- Illingworth, R.S. 1955 Crying in infants and children. *British Medical Journal*, 8, 75-78.
- Kobayashi, K., Oda, T., & Murooka, H. 1986 Characteristic pattern in time series of crying vocal in early infant. *Japanese Society of Medical Electronics & Biological Electronics*. (In Japanese).
- Lester, B.M. 1984 Infant crying and the development of communication. In N.A.Fox & R.J. Davidson (Eds.), *The Psychobiology of Affective development*. London: Hillsdale, New Jersey. Pp.231-258.
- Rosenhouse, J. 1977 A preliminary report: An analysis of some types of a baby's cry. *Journal of*

Phonetics, 5, 299-312.

Tsukamoto, T. 1983 Mother's recognition of the infant cry. Annual Reports of Konan Women's University, 2, 31-44. (In Japanese).

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

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

Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T., & Valanne, E. 1968 The infant cry: A spectrographic and auditory analysis. Clinics in Developmental Medicine No.29. London: Spastics International Medical Publications Heinemann.

Wolff, P.H. 1969 The natural history of crying and other infant vocalization. In B.M.Foss (Ed.),