Title
Preliminary Study of Object Labeling Using Sound Production in a Beluga

Journal
International Journal of Comparative Psychology, 25(3)

ISSN
0889-3675

Authors
Tsukasa, Murayama
Fujii, Yuki
Hashimoto, Takayuki
et al.

Publication Date
2012

DOI
10.46867/ijcp.2012.25.03.04

International Journal of Comparative Psychology, 2012, 25, 195-207.

Preliminary Study of Object Labeling Using Sound Production in a Beluga

Tsukasa Murayama, Yuki Fujii, Takayuki Hashimoto, Aya Shimoda, So Iijima, Kohei Hayasaka, Narumi Shiroma, and Mana Koshikawa
Tokai University, Japan

Hiroshi Katsumata, Makoto Soichi, and Kazutoshi Arai
Kamogawa Sea World, Japan

A beluga was tested to label objects using vocal symbols in order to form a bidirectional relationship between the visual symbols and sounds. In the Training session, the subject was first trained to distinguish four objects by four separate calls that he made. He learned to emit different calls corresponding to the sample stimuli. Next, these three recorded calls were played back to the subject, and he was required to select the objects (comparative stimuli) corresponding to the presented sound. He succeeded in correctly choosing the objects corresponding to the sounds played back. In the Test session, when a completely new recorded sound was presented to the subject, he could choose the correct object by matching the sample sound with the object. It is suggested that a beluga have realized bidirectional relationships or symmetrical relationship between visual symbols and sounds. This is a preliminary study that shows both production and comprehension of symbols in marine mammals.

Study of artificial language is one of the approaches used to investigate animal cognition, and teaching artificial language to animals plays an important role in the burgeoning of research in comparative animal cognition. Research on teaching language has been conducted on chimpanzees (Asano, Kojima, Matsuzawa, Kubota, & Murofushi, 1982; Gardner & Gardner, 1969; Gardner, Gardner, & Drumm, 1989; Matsuzawa, 1985; Premack & Premack, 1972; Rumbaugh, 1977; Savage-Rumbaugh, 1986; Savage-Rumbaugh, Rumbaugh, Smith, & Lawson, 1980), bonobos (Savage-Rumbaugh, 1993), orangutans (Miles, 1990; Shumaker & Beck, 2003), gorillas (Patterson, 1978), parrots (Pepperberg, 1990, 1999), California sea lions (Gisner & Schusterman, 1992; Schusterman & Krieger, 1984), and bottlenose dolphins (reviewed in Herman, 1980, 1986).

Labeling matter is a fundamental function of language. In general, in an artificial language study, language functions as a label for various types of matter. Consequently, the practice of labeling specific matter using language implies the formation of a bidirectional relationship between the matter and the medium such as sounds, gestures, and lexigrams.

Numerous behavioral studies on the cognitive abilities of dolphins suggest that they are advanced cognitive animals (e.g., reviewed in Herman, 1980; Kuczaj & Walker, 2006; Pack & Herman, 2006). As one of the methods employed in

We sincerely thank the staff members of Kamogawa Sea World for their cooperation. The authors would like to thank two anonymous reviewers for their useful comments on this manuscript. This study was supported by a grant of International Marine Biological Research Institute in the Kamogawa Sea World, Japan. Correspondence concerning this article should be addressed to Tsukasa Murayama, Department of Marine Biology, School of Marine Science and Technology, Tokai University, 3-20-1 Orido, Shimizu, Shizuoka, Japan 424-8610. (beluga@scc.u-tokai.ac.jp)

dolphin cognition research, language comprehension was studied in bottlenose dolphins using artificial languages such as gestural language (hand signs) and auditory language (computer-generated sounds) (Herman, 1980, 1986; Herman, Richards, & Wolz, 1984). However, since language studies performed on dolphins have focused on symbol comprehension, a bidirectional relationship between the symbols (gestures or sounds) and objects/actions could not be formed.

The present study aims to form a bidirectional relationship between some specific objects and sounds (calls), that is, production (sound to object) and comprehension (object to sound), and to label the objects using vocal symbols made by the subject himself.

Method

Subject

The subject was a male beluga (nicknamed Nack, 384 cm in body length, 794 kg in body weight, 24 years of age) that was kept in the Kamogawa Sea World in Chiba prefecture, Japan. The subject was maintained in a pool (3.5 m in depth, 17.0 C in water temperature) and underwent several kinds of behavioral and cognitive experiments (e.g., Murayama, Fujii, Katsumata, Arai, & Soichi, 2008; Murayama, Iochi, & Tobayama, 2001; Murayama, Kobayashi, & Ito, 2002; Murayama & Tobayama, 1995, 1997) unrelated to the present study. Another beluga, not involved in the experiment, was kept in the same pool; however, this beluga was isolated during the experiment to avoid any possible influence on the experiment.

Objects for Stimulus

The stimuli employed in the experiment were a swimming fin (hereafter "fin"), a bucket, a swimming mask (hereafter "mask"), a boot, a scrubbing brush**,** and a calling rod (Figure 1). Because these objects were used in the aquarium every day, the subject was familiar with them. In the present study, these objects (Figure 1) were presented to the subject as samples or as comparative stimuli by being held out by an experimenter or using an apparatus.

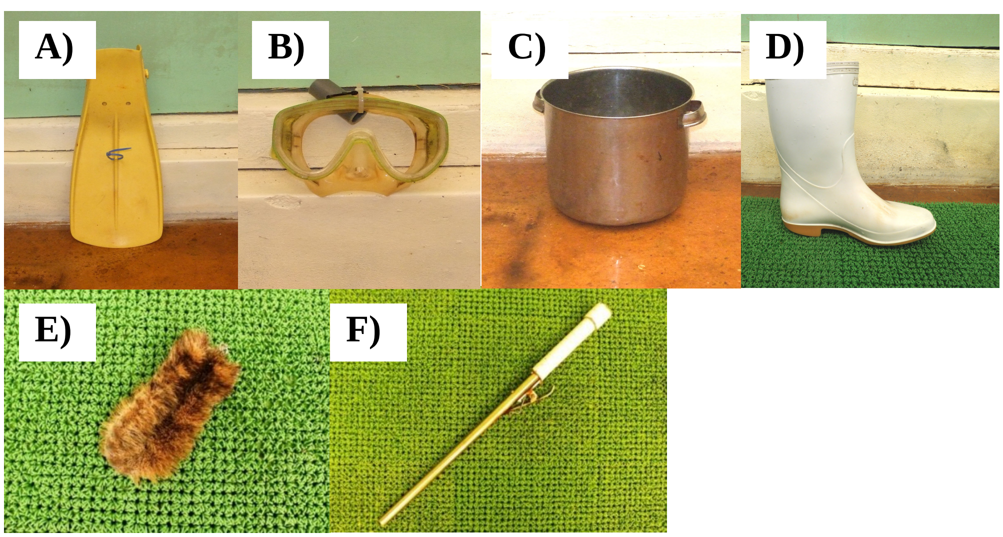

Figure 1. Objects used in the present study. **A)** fin, B) mask, **C)** bucket, **D)** boot, **E)** scrubbing brush, **F)** calling rod.

Procedure

The experiment consisted of two sessions: the Training session, which included two tasks, and the Test session, which also included two tasks. The experiment was conducted using a conditional discrimination task.

Several measures were taken to avoid the "Clever Hans effect" in the experiment. Two experimenters participated in the experiment and wore brown-tinted goggles at all times in order to not influence the subject's behavior through their eyes. Furthermore, one experimenter observed the other's actions to ensure that he did not give any cues to the subject while performing the experiment.

Training session. *Training 1 (Different call task).* The subject was trained to distinguish four objects by emitting different calls. In the experiment, a small red light was used as a cue light and the experimenter had this light with his hand during experiment. One of the four objects was presented to the subject as a sample and the cue light was turned on. Next, the subject was required to identify four objects and emit a different call corresponding to the object presented. He had to emit a short, high-pitched sound when the fin was presented; a long high-pitched sound when he saw the mask; a short, low-pitched sound when he saw the bucket; and a short, medium-pitched sound when the boot was presented (Figure 2). These call types were determined by the experimenters arbitrarily. The sounds were easily audible, and the experimenter could correctly distinguish them with his normal hearing ability. During the experiment, each emitted call was recorded with a digital audio recorder and the frequency, duration**,** and wave form of the sound were checked through the audio software "Audition"(Adobe Systems) to verify that the subject had emitted the correct call in response to the object.

If the subject emitted a call correctly, he was rewarded with a piece of fish; however, a 10-s time-out was introduced for the incorrect responses. A session comprised 10-15 trials, and the interval between each trial was approximately five seconds. Each object was displayed in a semi-random order on the basis of a random number so that the number of presentations of each object was uniform in a session. The percentages of the correct responses ((the number of correct responses / total number of trials in a session) x 100 %) were measured in each session.

Training 2 (Sound discrimination task). When the subject learned to emit distinct calls corresponding to the presented objects in Training 1, each call was recorded with a digital audio recorder. In Training 2, the playback tasks were performed, and the experiment was conducted with an alternative method. At the beginning of each trial, two objects were set as comparative stimuli with the apparatus, in front of the subject. Then, one of the three recorded sounds that were recorded in Training 1, except the sound for the boot, was presented as a sample stimulus through an audio speaker. When the cue light was switched on, the subject was required to touch one of the objects (comparative stimuli) with his snout (Figure 3), corresponding to the presented sound. That is, he had to select a fin when a short, high-pitched sound was presented; the bucket when he heard a short, low-pitched sound; and the mask on hearing a long high-pitched sound. One session comprised 10-15 trials, and the interval between each trial was approximately 5 s. Each recoded sound was presented in a semi-random order so that the number of presentations of each sound was uniform in a session. The subject was rewarded with a piece of fish if he chose the correct object when the sound was played back to him; however, a 10-s time-out was introduced for the incorrect responses. The percentages of correct responses were measured in each session.

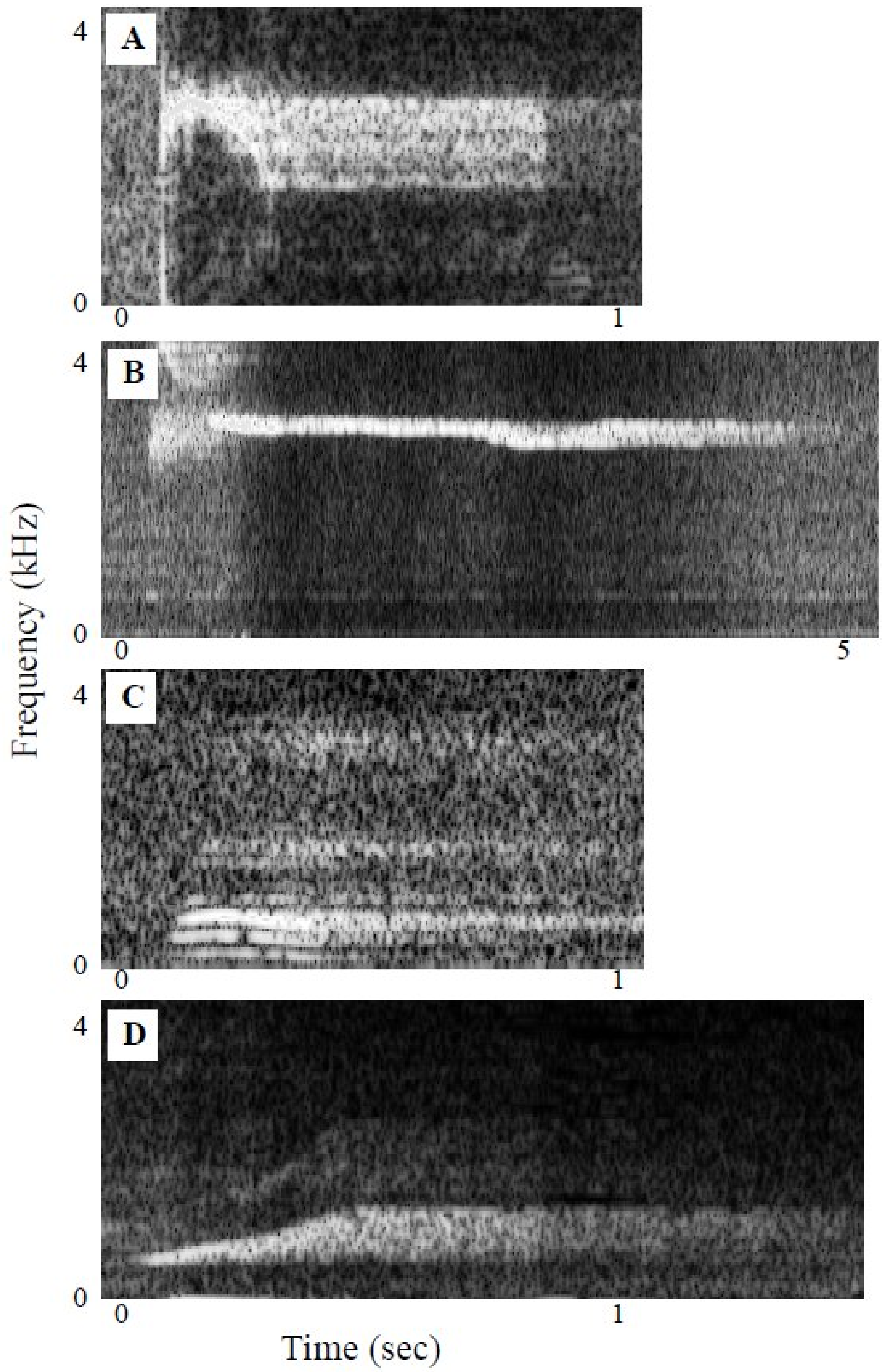

Figure 2. Spectrograms of calls emitted by the subject corresponding to the presented objects. **A)** fin, **B)** mask, **C)** bucket, **D)** boot.

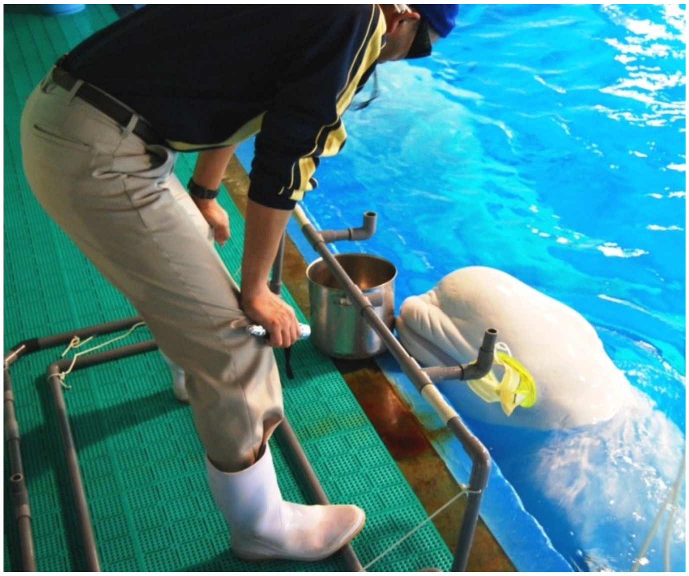

Figure 3. Selection behavior of the subject with object presented by apparatus.

Test session. In the Test session, the subject was not rewarded irrespective of whether he responded correctly in both the baseline and probe trials because the subject would learn the correct choice itself if rewarded when he responded correctly. In addition, the subject was not given any clue by the experimenter not by rewarding.

Test 1 (Sound discrimination for each object). Test 1 session consists of two phases, a baseline trial and a probe trial. At the beginning of each trial in both phases, two objects were set as comparative stimuli with the apparatus, in front of the subject. As a baseline trial, one of the three recorded sounds, fin, mask, and bucket, was presented as a sample stimulus, and the subject was required to select one of two objects (comparative stimuli) that corresponded to the presented (sample) sound. In the probe trial, one of the sample sounds for the fin, mask, bucket, and boot was presented, and then one of these trained objects and a novel object, a scrubbing brush or a calling rod (Figure 1E, F), were presented to the subject as comparative stimuli. The recorded sound for the boot was novel to the subject because he had not heard it previously and the scrubbing brush and the calling rod were novel to the subject because they had never been presented before in the present study. Then, the subject was required to select one of these comparative stimuli in response to the sample sound.

The probe trial was performed after every 3-4 baseline trials. A session comprised of 10-15 trials, and the interval between each trial was approximately 5 s. Each recorded sound was presented in a semi-random order so that the number of presentations of each sound was uniform in a session.

Test 2 (Mixture of tasks). Each trial performed in the Training sessions and the boot sound discrimination task were administered to the subject in a random order. However, in the boot sound discrimination task, the comparative stimulus that was paired with the boot was not the scrubbing brush or calling rod, but the fin, mask and bucket. (An example of the presentation is given in Table 1.) Experimentation of Test 2 session continued until 12-21 trials per one object were conducted for each task, and the percentages of the correct responses were calculated.

Statistics

The subject was required to select one of two (in Training 2 and Test 1) or one of four (in Training 1 and Test 2) stimuli. So, the criterion level for the threshold was defined on the basis of the binomial test results. That is, in the Training session, each session usually consisted of 15-20 trials. So, when the percentage of correct responses was more than 75% in two consecutive sessions, the subject was judged to have mastered the task. In the Test session, whether the percentages of correct responses were significant or not was checked by a binomial test.

Table 1
An example of presentation menu in a session in the Test 2 session. Each trial was imposed at random order.

Sample Stimulus	Comparisons	Task
fin (object)		Different call task
mask (recorded sound)	mask vs bucket	Sound discrimination task
mask (object)		Different call task
bucket (object)		Different call task
fin (recorded sound)	bucket vs fin	Sound discrimination task
boot (recorded sound)	boot vs fin	Sound discrimination task
fin (object)		Different call task
boot (recorded sound)	bucket vs boot	Sound discrimination task

Results

Training Session

Training1 (Different call task). The changes in the percentages of correct responses in Training 1 are shown in Figure 4. From sessions 1 to 14, two objects – the fin and bucket - were used as samples and each object was displayed to the subject. The subject initially confused the calls corresponding to the objects presented and was unable to emit correct calls; therefore, the percentages of correct responses were low.

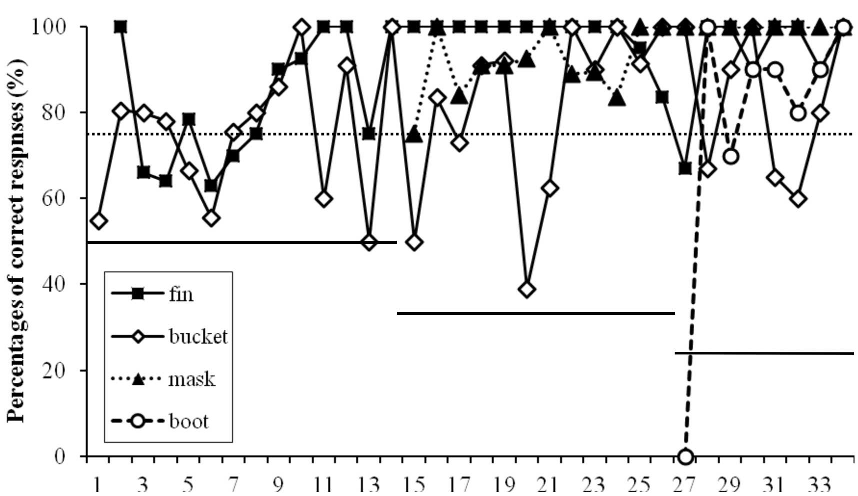

Figure 4. Changes in the percentages of correct responses in the Different call task (Training 1). Solid lines indicate a chance level, and a dashed line indicates a criterion (75%).

However, he gradually learned to emit correct calls, and the percentages of correct responses increased, indicating that the subject could identify the two objects (sample stimuli) and emit the appropriate calls corresponding to them.

After 15 sessions, the mask was added to the samples and one of these three objects was shown to the subject as a sample in a semi-random order so that the number of presentation of each object was uniform. While the percentages of correct responses for the fin remained high, those for the mask and bucket fluctuated. However, as the session proceeded, there was a gradual increase in the percentages for each object. After 22 sessions, the percentages became high, and finally the subject reached the criterion set in the method section.

In the 27th session, the boot was added to the samples, and one of these four objects was shown to the subject as a sample. The subject responded incorrectly to the presentation of the boot in the 27th session. However, the percentages of correct responses were immediately high in the next session. And it was more than 75% or the significant level in two consecutive sessions after 29 sessions.

Training 2 (Sound discrimination task). First, discrimination of the recorded sounds for the fin and mask was tested (Figure 5A). When the recorded sounds for the fin and mask were played to the subject, he was unable to choose the correct object, and the rate of correct responses changed considerably. However, as the session progressed, the subject learned to select correctly, and the percentages of correct responses reached a high value after 37 sessions.

Second, the fin or bucket sound discrimination was performed (Figure 5B). The correct response rate for the fin was low in earlier sessions, but it gradually increased. Then, the percentages of correct responses for both objects reached a high and significant level ($p < 0.05$), which was achieved in only 10 sessions.

Third, the recorded sounds for the mask and bucket were presented (Figure 5C). Although the subject could not distinguish between each recorded sound in the earlier sessions, he was able to choose the correct objects corresponding to the sounds played back to him. Eventually, it took only five sessions to reach the criterion set to succeed in this task.

During those two training sessions, the subject was rewarded with a piece of fish when it responded correctly.

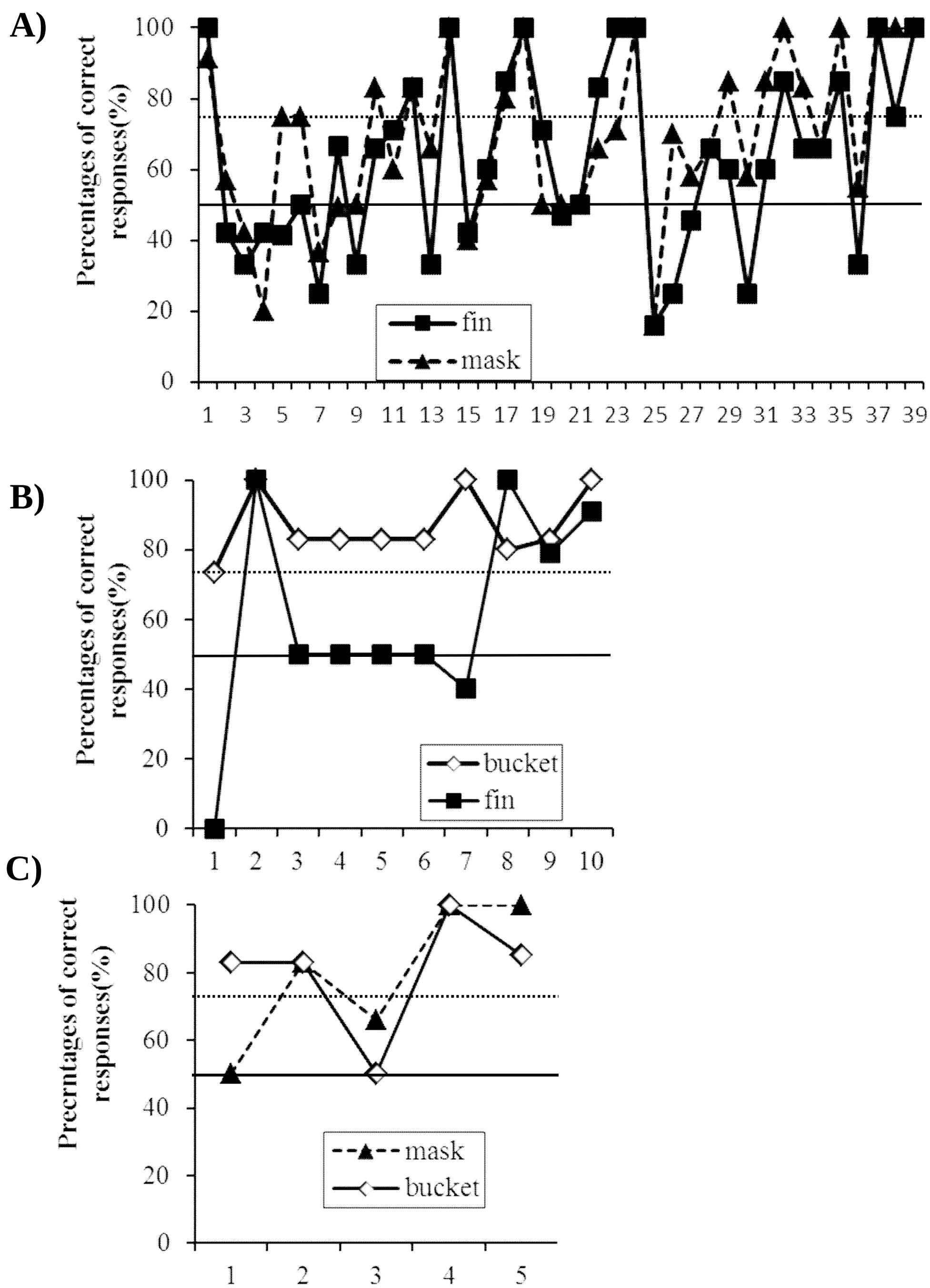

Figure 5. Changes in the percentages of correct responses in the Sound discrimination task (Training 2). **A)** fin vs. mask, **B)** bucket vs. fin, **C)** mask vs. bucket. Solid lines indicate a chance level, and dashed lines indicate criterion (75%).

Test Ssession

Test 1 (Sound discrimination for each object). The percentages of the Test session are shown in Figure 6. In the baseline trial, sound discrimination with familiar objects was correct at significant level ($p < 0.05$). In the probe trial, when the sounds for the fin, mask, and bucket were presented, the subject chose the fin, mask, and bucket, respectively. That is, the subject chose the correct object (comparative stimuli) selectively, not the novel objects (a scrubbing brush, and a calling rod), corresponding to the presented sound. The percentages of correct responses were at significant level ($p < 0.01$). Then, when the sound for the boot was presented, the subject responded correctly ($p < 0.01$). That is, he chose the boot not the novel object, without confusion. Since the sound for the boot was presented only in the probe trial, this result suggests that the subject made a discrimination choice without being trained.

Test 2 (Mixture of tasks). The percentages of correct responses for every task are shown in Figure 7. The percentages values were much higher than the chance level. Moreover, each value exceeded a significant level of $p < 0.05$. Therefore, these results demonstrated that the subject could respond correctly when any sample was presented or any task administered in a random order.

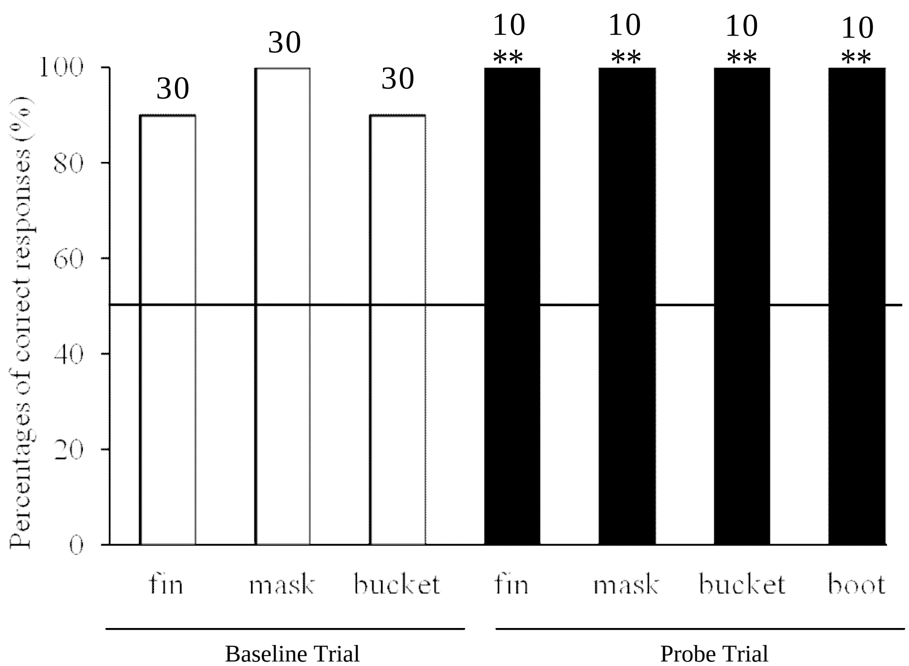

Figure 6. The percentages of correct responses of the Sound discrimination for each object in the Test 1 session. Each numeral means the number of trials. A solid line indicates a chance level. Asterisks above the bars indicate the significant level, 1% (**) of the binomial test.

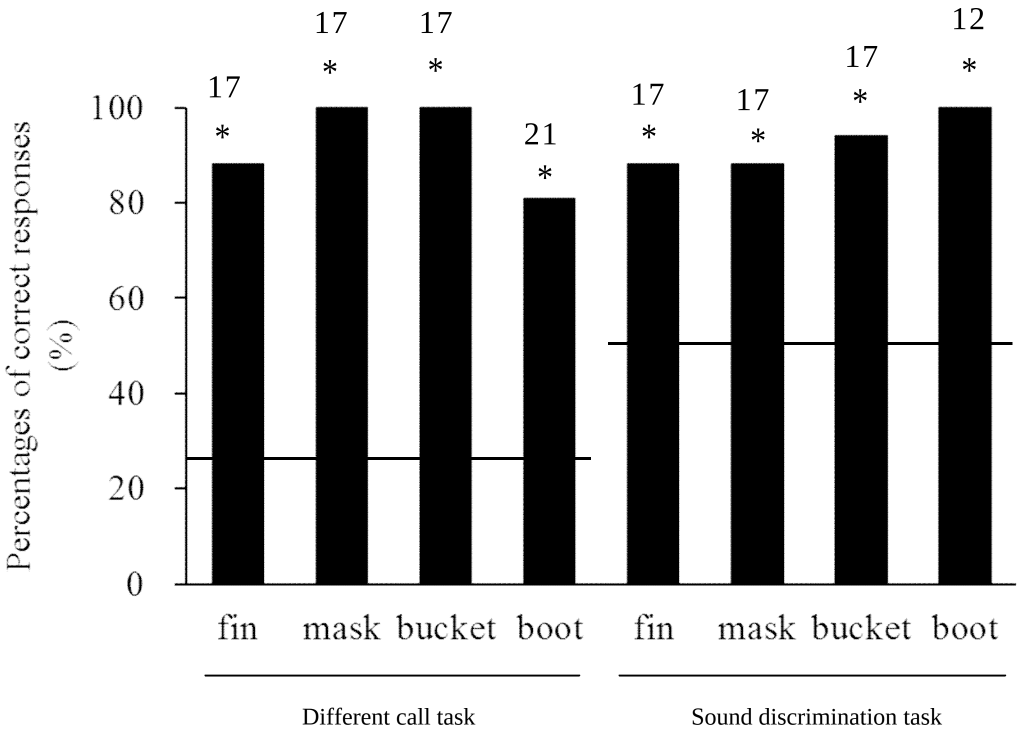

Figure 7. The percentages of correct responses of every task. Numerals mean the number of trials. Solid lines indicate a chance level. Asterisks above the bars indicate the significant level, 5% (*) of the binomial test.

Discussion

Studies performed with several terrestrial animals demonstrated they can comprehend and produce symbols such as lexigrams and sounds (Savage-Rumbaugh, Rumbaugh, & Fields, 2006). The bidirectional relationship between symbols and specific matters has been established among these species.

In marine mammals, on the other hand, the object labeling studies that employ an auditory stimulus and a visual stimulus have been conducted by Richards, Wolz, and Herman (1984) and Herman (reviewed in 1980, 1986) with dolphins and Schusterman and Krieger (1984) with a California sea lion. Richards et al. (1984) performed a vocal labeling examination and demonstrated that the dolphins gave unique vocal labels to some objects. Herman (1980, 1986) succeeded in connecting a gesture with a specific object or action. However, since he focused on symbol comprehension, the dolphin had no means of producing symbols; therefore, a bidirectional relationship, that is, production and comprehension, between gesture and object was not constructed in that study.

In the present study, the beluga succeeded in emitting a different call corresponding to the presented sample object in the Different call task, indicating that the subject could learn to produce particular sounds in response to displayed objects. In reverse, the subject was able to learn to choose the objects selectively, corresponding to the sample sounds. For the familiar recorded sounds for the fin, mask, and bucket, the subject chose the fin, mask, and bucket, respectively, by matching the sample sound with the object, not by matching by excluding the novel objects. These results indicate that he understood the relationship between the sounds and objects. In addition, when the sound for the boot was presented and the boot and the novel objects were set as comparative stimuli, the subject chose the boot selectively, that is, he was able to respond correctly to this untrained sample sound. These results suggest that the subject generalized a concept and transferred it to a novel stimulus.

The trial of the Different call task had a symmetrical relationship with that of the Sound discrimination task. Since the subject had succeeded in the Different call task, he had already learned that certain calls correspond to certain objects. Then, during the Sound discrimination task, he may have gradually grasped the symmetrical relationships in which a certain object corresponds to a certain call during the Sound discrimination task. Therefore, in the probe trial of the Test 1 session, he could choose the correct object when he was presented with a novel recorded sound - the sound for the boot, suggesting that a symmetrical relationship was formed without being trained. These results indicate that the beluga learned to label four objects and comprehended the labels for these objects.

The ecology of an animal contributes to the formation of flexible relationships between stimuli (Schusterman & Kastak, 1998). The beluga is a highly socialized animal, hence, complicated social relationships between the individuals are related to the development of symmetrical relationships between the stimuli. Schusterman and Kastak (1993) demonstrated, by performing the symmetry test using various stimuli, that the symmetry was formed voluntarily in a California sea lion, and suggested that those experiences contributed to the voluntary formation of symmetry. The subject in the present study had much experience with cognitive tasks (e.g., Murayama et al., 2001, 2002; Murayama & Tobayama, 1995, 1997). In particular, Murayama et al. (2008) reported that the subject was observed to exhibit symmetry in a symmetry test using the visual stimuli. While experiencing such tasks, there may also have been an opportunity to develop symmetrical relationships between auditory stimuli, and it might have occurred also in the present study. In the Mixture of tasks of the Test 2, the subject responded correctly, even when he was tested with all the tasks randomly mixed. Therefore, it was thought that the subject had acquired a bidirectional relationship between object and sound.

Although it is still in a preliminary stage, this is the first examination to show both production and comprehension of symbols in a marine mammal. Hence, we can conclude that the subject was able to label four objects with his own calls. However, this current study was a single subject design, that is, the sample size was small. In addition, each experiment was limited as for the number of trials. Therefore, further examination will be needed after improving these points.

References

Asano, T., Kojima, T., Matsuzawa, T., Kubota, K., & Murofushi, K. (1982). Object and color naming in chimpanzees (*Pan troglodytes*). *Proceedings of Japan Academy, 58(B),* 118-122.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science, 165,* 664-672.

Gardner, R. A., Gardner, B. T., & Drumm, P. (1989). *Teaching sign language to chimpanzees*. New York: University of New York Press.

Gisner, R., & Schusterman, R. J. (1992). Sequence, syntax and semantics: Responses of a language trained sea lion (*Zalophus californianus*) to novel sign combinations. *Journal of Comparative Psychology, 106,* 78-91.

Herman, L. M. (1980). Cognitive characteristics of dolphins. In L. M. Herman (Ed.), *Cetacean behavior: Mechanisms and functions* (pp. 363-427). New York: Wiley Interscience.

Herman, L. M. (1986). Cognition and language competencies of bottlenosed dolphins. In R. J. Schusterman, J. Thomas, & F. J. Wood (Eds.), *Dolphin cognition and behavior: A comparative approach* (pp. 221-251). Hillsdale, NJ: Lawrence Erlbaum.

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition, 16,* 129-219.

Kuczaj, S. A., II, & Walker, R. T. (2006). How dolphins solve problems? In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 580-601). New York: Oxford University Press.

Matsuzawa, T. (1985). Color naming and classification in a chimpanzee (*Pan troglodytes*). *Journal of Human Evolution, 14,* 283-291.

Miles, H. L. (1990). The cognitive foundations for reference in a signing orangutan. In S. T. Parker & K. R. Gibson (Eds.), *Language and intelligence in monkeys and apes: Comparative developmental perspectives* (pp. 511-539). Cambridge, MA: Cambridge University Press.

Murayama, T., Fujii, Y., Katsumata, H., Arai, K., & Soichi, M. (2008). Formation of symmetry in beluga. *Cognitive Studies 15,* 358-365 (in Japanese with English abstract).

Murayama, T., Iochi, A., & Tobayama, T. (2001). Discrimination of ellipse from circle in white whale. *Nippon Suisan Gakkaishi, 64,* 745-746 (in Japanese).

Murayama, T., Kobayashi, H., & Ito, M. (2002). Preliminary study on the cognition by vision. Can the dolphin count? *Fisheries Science 68 Supplement I. (Proceeding of International Commemorative Symposium of 70th Anniversary of the Japanese Society of Fisheries Science),* 302-305.

Murayama, T., & Tobayama, T. (1995). *Preliminary study of mental rotation in beluga*. The XXIV International Ethological Conference, Honolulu, USA.

Murayama, T., & Tobayama, T. (1997). Preliminary study on stimulus equivalence in beluga (*Delphinapterus leucas*). *The Japanese Journal of Animal Psychology, 47,* 79－89 (in Japanese with English abstract).

Pack, A. A., & Herman, L. M. (2006). Dolphin social cognition and joint attention: Our current understanding. *Aquatic Mammals, 32,* 443-460.

Patterson, F. (1978). The gesture of a gorilla: Language acquisition in another Pongid. *Brain and Language, 5,* 72-97.

Pepperberg, I. M. (1990). Cognition in an African gray parrot (*Psittacus erithacus*): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology, 104,* 41-52.

Pepperberg, I. M. (1999). *The Alex studies*. Cambridge, MA : Harvard University Press.

Premack, A. J., & Premack, D. (1972). Teaching language to an ape. *Scientific American, 227,* 92-99.

Richards, D. G., Wolz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology, 98,* 10-28.

Rumbaugh, D. M. (1977). *Language learning by a chimpanzee.* New York: Academic Press.

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. New York: Columbia University Press.

Savage-Rumbaugh, E. S. (1993). *Kanzi: A most improbable ape*. Tokyo: NHK Publishing.

Savage-Rumbaugh, E. S, Rumbaugh, D. M., & Fields, W. M. (2006). Language as a window on rationality. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 513-552). New York: Oxford University Press.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., Smith, S. T., & Lawson, J. (1980). Reference: The linguistic essential. *Science, 210,* 922-925.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record*, *43*, 823-839.

Schusterman, R. J., & Kastak, D. (1998). Functional equivalence in a California sea lion: Relevance to animal social and communicative interactions. *Animal Behaviour, 55,* 1087-1195.

Schusterman, R. J., & Krieger, K. (1984). California sea lion are capable of semantic comprehension. *The Psychological Record, 34,* 3-23.

Shumaker, R., & Beck, B. (2003). *Primates in question: The Smithsonian answer book*. Washington: The Smithsonian Books.