Title
Vocal Imitation of Human Speech, Synthetic Sounds and Beluga Sounds, by a Beluga (Delphinapterus leucas)

Journal
International Journal of Comparative Psychology, 27(3)

ISSN
0889-3675

Authors
Murayama, Tsukasa
Iijima, So
Katsumata, Hiroshi
et al.

Publication Date
2014

DOI
10.46867/ijcp.2014.27.03.10

INTERNATIONAL JOURNAL OF COMPARATIVE PSYCHOLOGY

2014, 27(3), 369-384
Dani Brunner & Lori Marino
Editors
Research Article
Peer-reviewed

Vocal Imitation of Human Speech, Synthetic Sounds and Beluga Sounds, by a Beluga (*Delphinapterus leucas*)

Tsukasa Murayama, So Iijima
Tokai University, Japan
&
Hiroshi Katsumata, Kazutoshi Arai
Kamogawa Sea World, Japan

We tested the ability of a male beluga (*Delphinapterus leucas*) to imitate 11 sounds presented to it. During the training session, we presented the subject 3 recorded sounds emitted by the subject itself, and the subject was trained to imitate them. The subject learned to correctly imitate the sounds. During the test session, 2 novel computer-generated artificial sounds were presented through an audio speaker. In addition, 9 arbitrary vocal sounds produced by the experimenter were presented to the subject, and the subject was required to imitate them. 9 persons, who were not involved in the experiment, were presented the sample sounds and imitated calls; subsequently, they judged whether both sounds were similar to each other. Further, the sound spectrums of the sample sounds and imitated calls were analyzed. As a result, we found that some components of the sound spectrums were similar, and most imitated calls possessed spectral features similar to those of the sample sounds. These results demonstrated that the beluga correctly imitated novel sounds and spontaneously displayed an aptitude for imitation.

Imitation requires higher cognitive abilities in animals. It is a form of social learning and is acquired by learning rather than by instinct. That is, learning ability is necessary for imitation. Therefore, the animals whose lives depend upon learning are skilled in imitation. Imitation is classified into vocal imitation and motor imitation. Vocal imitation involves a modification of the sound that the animal produces based on the sound that another animal (the partner) emits. Some evidence exists for vocal imitation among animals such as birds (Pepperberg, 2000; Todt, 1975; West, Stroud, & King, 1983), African elephants (Poole, Tyack, Stoeger-Horwath, & Watwood., 2006), and marine mammals. In odontocetes, vocal imitation is well known, particularly in bottlenose dolphins. The bottlenose dolphin is a highly social animal that emits a variety of sounds that are believed to have a social function (reviewed in Herman & Tavolga, 1988). Considerable information concerning the ability to imitate has been obtained from bottlenose dolphins. Several cases of spontaneous vocal imitation in bottlenose dolphins are reported (Lilly, 1961, 1962; Lilly, Miller, & Truby, 1968), and the bottlenose dolphin is able to imitate computer-generated artificial sounds (Reiss & McCowan, 1993; Richards, Wolz, & Herman, 1984). In addition, Tyack (1986), Smolker and Popper (1999), and Watwood, Tyack, and Wells (2004) suggested that the bottlenose dolphin could imitate the whistle of a social partner. However, how this capability may be used in nature remains unclear.

In the present study, we examined the imitation ability of another cetacean species, the beluga (*Delphinapterus leucas*). Belugas form schools comprising many individuals (Gregory & O'Corry, 2002). The beluga is called the "sea canary" because of its loquaciousness, and the sounds it produces may have specific social functions within the school. Further, to determine whether the beluga exhibits the same aptitude for imitation as that observed in bottlenose dolphins, we examined whether the beluga could imitate a range of sounds, including computer-generated and vocal sounds of humans.

Correspondence concerning this article should be sent to Dr. Tsukasa Murayama (Email:beluga@scc.u-tokai.ac.jp).

Method

Subject

The subject was a male beluga (nicknamed Nack; body length, 384 cm; body weight, 881 kg; age, 24 years) that was kept at the Kamogawa Sea World in the Chiba Prefecture, Japan. The subject was maintained in an indoor pool (depth, 3.5 m; width, 18 m; water temperature, 17.0°C), and it had previously undergone numerous behavioral and cognitive experiments (Murayama & Tobayama, 1995, 1997, Murayama, Iochi, & Tobayama, 2001; Murayama, Kobayashi, & Ito, 2002; Murayama, Fujii, Katsumata, Arai, & Soichi, 2008; Murayama et al., 2012) that were unrelated to the present study. Another beluga was kept in the same pool; however, it was not involved in the experiment and was isolated so as not to influence the experiment.

Procedure

All trials were performed in air of the indoor pool, i.e., the experimenter presented the stimuli in air. The subject responded with his head (face) lifted up from the water surface. The experimenter wore brown-tinted goggles at all times so as not to influence the subject's behavior with his eyes.

Training session.

Training 1 (Call recording task). As reported in Murayama et al. (2012), the subject learned to emit different calls that corresponded to the sample stimuli, such as a swimming fin (hereafter "fin"), a swimming mask (hereafter "mask"), and a bucket. That is, the subject emitted a short, high-pitched sound when a fin was presented to the subject; a long high-pitched sound when he saw a mask; a short, low-pitched sound when he saw a bucket. The emitted sounds were considerably different and readily audible, thereby allowing the experimenter to correctly distinguish them by listening. The subject was rewarded with a piece of fish when the correct call was emitted, whereas a 10-s time-out was imposed for an incorrect response. A session comprised of 15-20 trials, and the interval between each trial was approximately 2s.

We recorded each sound emitted by the subject using a digital audio recorder (ICD-UX70, SONY) when the corresponding sample object was presented. These recorded sounds were subsequently used in Training 2.

Training 2 (Imitation task). One of the three sounds recorded in Training 1 was presented in air through an audio speaker (SRS-Z510, Sony) as a sample stimulus. When the cue light was turned on, the subject was required to imitate the sound that was presented. The subject earned a reward when the call emitted matched the one that was played. The sample sounds were presented in a semi-random order. Because each emitted call was clearly different and audible to humans, the experimenter judged the success or failure of the imitation as he heard the sounds.

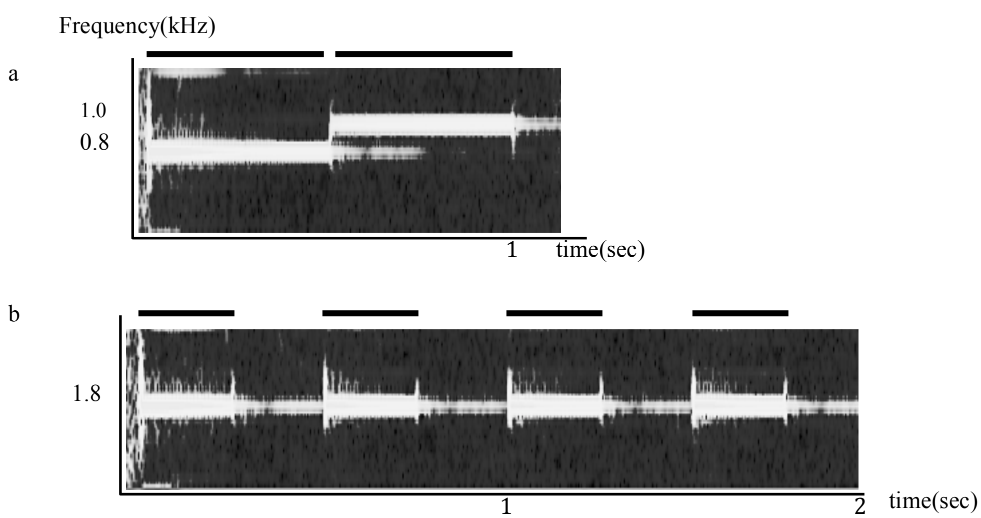


Figure 1. Sound spectrograms of "Artificial sound I" (a). and "Artificial sound II" (b). Solid lines indicate the sound presented.

Test session.

Prior to the test, two different computer-generated artificial sounds were added as extra sample sounds. One comprised 0.8- and 1.0-kHz pure tones (Figure 1a), and the other comprised four consecutive 1.8-kHz pure tones (Figure 1b).

As a baseline trial, one of the three imitated sounds for the fin, mask, and bucket recorded during Training 1 was presented to the subject through an audio speaker. Subsequently, the two different above-mentioned computer-generated artificial sounds were presented through an audio speaker as a probe trial. Since these sounds were presented only during the test session, the animal was not trained to imitate them, *i.e.*, they were novel stimuli for the subject. In addition, nine arbitrary vocal sounds (Table 1) emitted by the experimenter's voice were presented to the subject. These vocal sounds were also novel to the subject. Further, the baseline and probe trials were performed in a random order. Subsequently, the subject was required to imitate those sample sounds, and both the sample sounds and imitated calls were recorded using the digital audio recorder.

During the test session, the subject was given no rewards irrespective of whether he responded correctly or not in both the baseline and probe trials so that the experimenter did not provide any cues. However, the subject was made to perform an unrelated performance every five test trials for which it was rewarded with a piece of fish.

Whether the sample sounds and imitated calls were similar was decided according to the following two methods. One involved a "Subjective similarity decision" in which the imitated sounds were presented to a person who was not involved in the experiment and who had no information about the nature of the experiment, and the person then was asked to choose which of the eleven sample sounds was most similar to the imitated sounds. This process was repeated with nine persons. Thus, the experimenter's bias could be eliminated by this method.

The other method to estimate the similarity of sounds was an analysis of spectrograms of the sample sounds and imitated calls ("Analysis of imitated calls"). These spectrograms were analyzed using audio analysis software ("Audition", Adobe). According to Ridgway Carder, Jeffries, and Todd (2012) and Stoeger et al. (2012), both sounds were compared in terms of the total number of vocal bursts, time duration, and spectral peak frequency. To eliminate the analyzer's bias, this analysis was also performed by a person who had no information about the experiment.

Table 1
Sample sounds presented to the subject

List of sample sounds	
Computer-generated artificial sounds	
Artificial sound I	
Artificial sound II	
Vocal sounds emitted by the experimenter	
Hahaha…	(Laughter of human)
Hou?	(Voice to ask to again)
A wawawa…	(Voice for dandling a baby in Japanese)
Duke	(The nickname of another beluga in the pool)
Ohayo	(It means "good morning" in Japanese)
Tsukasa	(The name of author of the present study.)
Piyo piyo	(Call of chick in Japanese)
Hoh kekyo	(Call of bush warbler in Japanese)
Oh!	(Shout)

Results

Training Session

Since the subject had already learned to emit different calls that corresponded to the sample sounds (Murayama et al., 2012), we recorded each call for the fin, mask, and bucket in Training 1 (call recording task). In Training 2 (imitation task), those recorded sounds for the fin, mask and bucket were employed as samples. The subject was presented with each sample sound and was required to imitate it.

The spectrograms of the sample sounds and imitated calls are shown in Figure 2. Since the experiment was performed in the indoor pool, calls emitted by the subject echoed. As those echoes were included in the spectrograms, the resulting quality of the spectrograms was poor. As mentioned previously, each emitted call was audible and distinguishable by listening for the experimenter, and the spectrograms were clearly different depending on the sample sounds, as shown in Figure 2. In addition, these spectrograms showed that the subject correctly imitated the sample sounds since the spectrum of the sample sounds and imitated calls appeared similar to each other. Subsequently, when the emitted call matched the sample sound, the subject was rewarded by the experimenter.

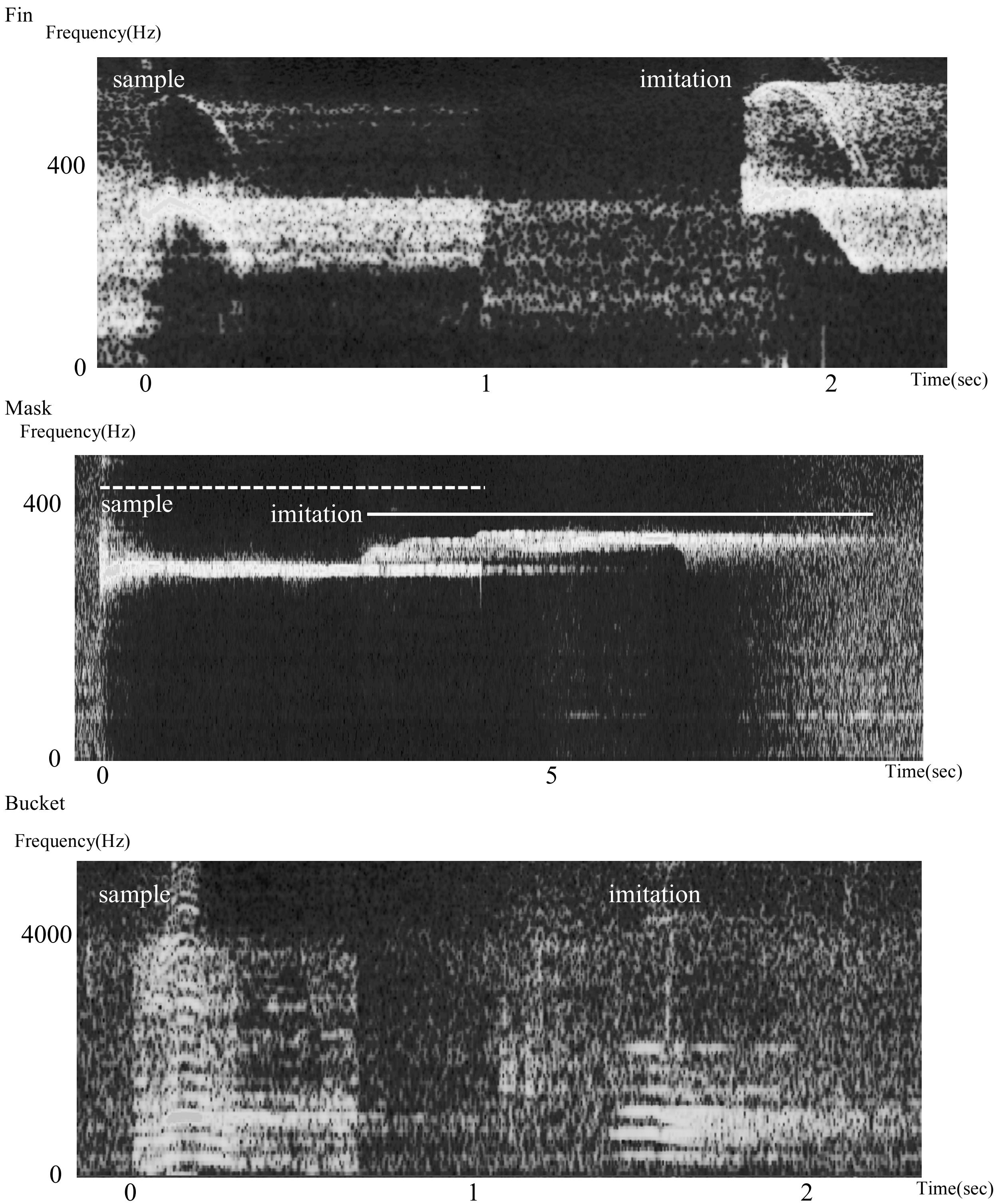


Figure 2. Spectrograms of the sample sounds and imitated calls used in the training session. The imitated calls were different depending on the sample sounds; however, each imitated sound was similar to the sample sound.

Figure 3 shows the changes in the percentages of correct imitation responses in Training 2. During the earlier sessions, the recorded sounds for the fin (○) and bucket (♦) were employed as the sample sounds. However, the subject did not understand what to do when presented with these sample sounds; therefore, he did not emit any calls or he produced arbitrarily incorrect calls when the sample sounds were played. After several sessions, he learned to distinguish the sample sounds and responded by imitating them. Although the percentages of responses for the bucket sometimes decreased, those for both objects gradually increased and reached a high level as the session progressed.

After 18 sessions, the recorded sound for the mask (▲) was added to the sample stimulus, following which the sounds for the fin, bucket, and mask were randomly presented. From sessions 18-22, the subject did not emit any calls when presented with the recorded mask sound. However, he correctly imitated the recorded sounds after 23 sessions, and there was a rapid increase in the related percentages. Finally, he was able to distinguish the sound for each object and correctly imitated each one without confusion.

Test Session

In the probe trials of the test session, the sample sounds, *i.e.,* two computer-generated artificial sounds generated, and the nine vocal sounds produced by the experimenter, were presented to the subject. Each sample sound was presented 2-9 times in a random order (The number of presentations of each sound was shown in Table 2). The subject responded quickly and without being confused by imitating the sound each time. In no case did the subject emit sounds that differ from the sounds that were played. Thus, the subject imitated the sounds without confusion, although the subject was not given a reward irrespective of whether he responded correctly or not.

Subjective similarity decision. The imitated calls were recorded using the audio digital recorder, and the recorded calls were presented to nine persons. Each person judged verbally which of the eleven model sounds was similar to the imitated sounds.

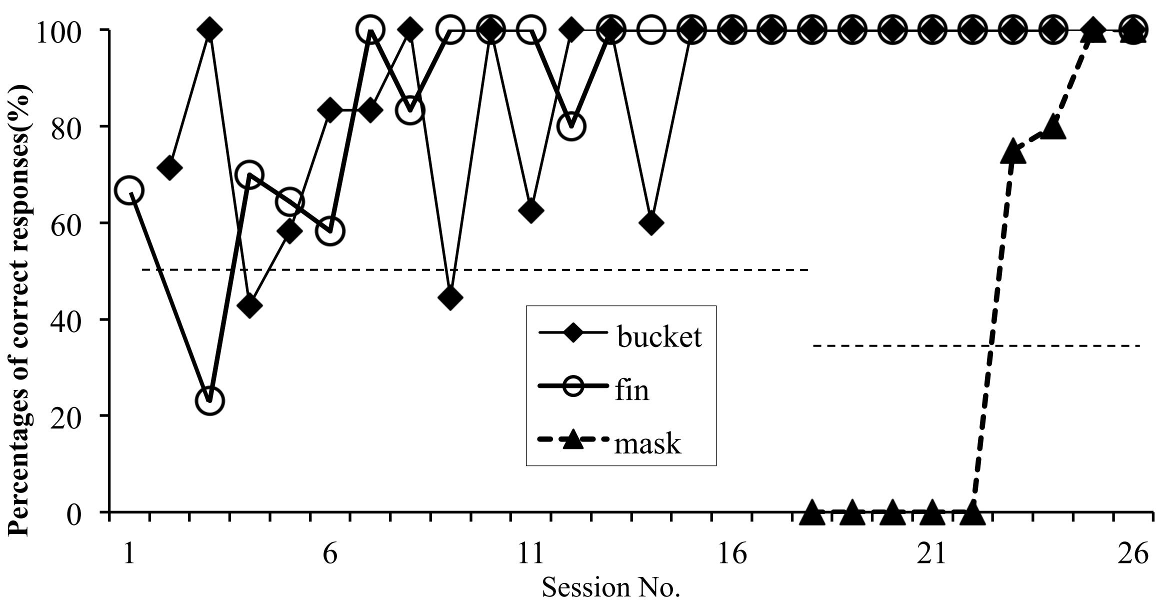


Figure 3. Changes in the percentages of correct responses in Training 2 (imitation task). Dashed lines indicate the chance level.

As described above, each sample sound was presented 2-9 times, and the subject imitated it each time. The recorded imitated calls were presented to each of the nine persons, and a ratio of the number of presentations that were judged to be similar among the total number of presentations was calculated for each person for each stimulus. The average across persons for each stimulus was then taken. The resulting average ratios are shown in Figure 4.

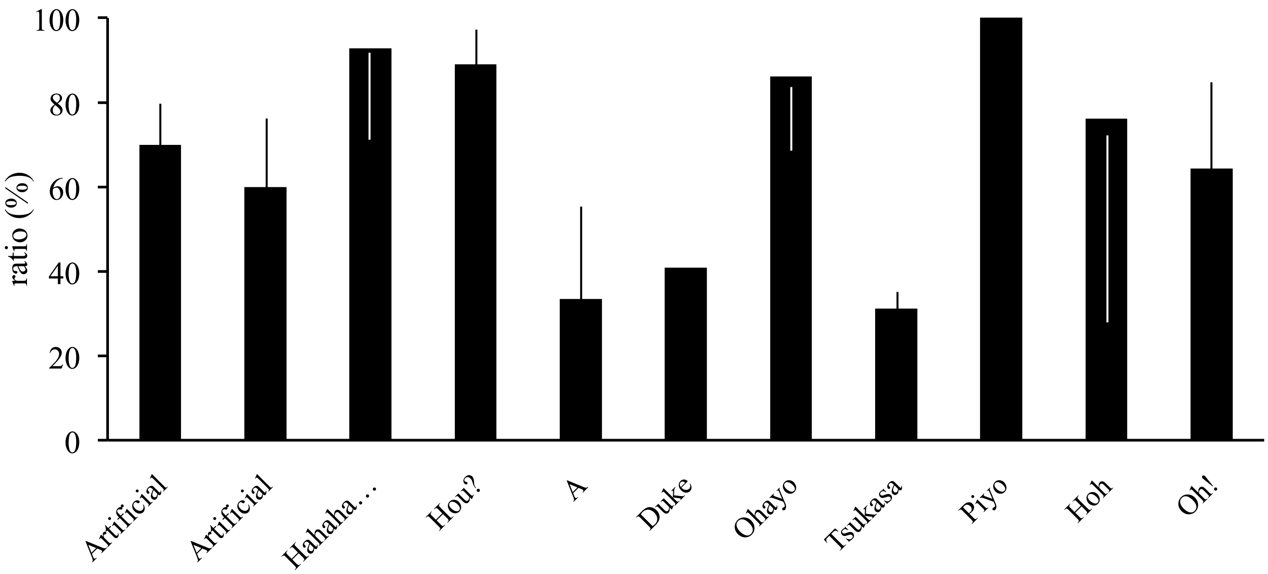


Figure 4. The average ratios that were judged to be similar across nine persons in subjective similarity decision. (mean±*SD*) The number of presentations of each stimulus was shown in Table 2.

The ratios reached more than 50%, except for "A wawawa…", "Duke" and "Tsukasa", and the ratios were higher than chance level for these sounds ($p < 0.001$, Chi-square test). That is, according to the nine persons, these imitated calls sounded similar to the sample sounds.

Analysis of imitated calls. Figure 5 shows the spectrograms of the sample sounds and imitated calls in the test trial. Since the experiments were performed in the indoor pool, the echoes of calls were reflected in the spectrograms. However, each spectrum was clearly observed, and the waveforms, number and pattern of each vocal burst in the spectrograms of the sample and the imitated sounds appeared similar to each other. However, an objective quantitative analysis was required to compare the characteristics of both sound waves. According to Ridgway et al. (2012) and Stoeger et al. (2012), some acoustic features such as the total number of vocal bursts, time duration, and spectral peak frequency of each vocal burst were measured. (Each factor is illustrated in Figure 6.)

The measured values of each component of the spectrum are shown in Table 2. No differences were observed between the sample sounds and imitated calls in the total number of vocal bursts, except for "A wawawa…". The values for "A wawawa…" were considerably different between these sounds, because the gap between each vocal burst was not unclear in the spectrum. For the time duration, few differences were observed between the sample sounds and imitated calls, except for "Artificial sounds II" and "A wawawa…". For the spectral peak frequency of each vocal burst, significant differences were observed for "Artificial sounds I", "Artificial sounds II", "A wawawa…", and "Ohayo", but few differences were observed for the other sample sounds.

Based on these values, changes in spectral peak frequency of each sound are schematically illustrated in Figure 7. This figure indicates that although some small differences were observed between the sample sounds and imitated calls in frequency and time duration, the pattern of each vocal burst, *i.e.*, rhythm or intonation of the sample sound and imitated call, was considerably similar, except for "A wawawa…". For "A wawawa…", the subject emitted a long endless call, which did not match the sample sound at all.

Table 2

The values of each sound component

		N	Total number of vocal bursts	Time duration of each vocal burst(ms) ($M \pm SD$) Number of vocal bursts ①	②	③	④	⑤	⑥	Spectral peak frequency(Hz) ($M \pm SD$) Number of vocal bursts ①	②	③	④	⑤	⑥
Artificial sound I	sample sound	5	2	500±0	500±0					732±0	991±0				
	imitated call		2	493±248	632±352					2799±61 *	3080±152 *				
Artificial sound II	sample sound	5	4	230±0	230±0	230±0	230±0			732±0	732±0	732±0	732±0		
	imitated call		4	416±1 *	460±23 *	526±98 *	589±317 *			2756±61 *	2741±100 *	2821±30 *	2821±29 *		
Hahaha···	sample call	4	6	92±3	87±10	85±7	99±2	91±28	83±6	453±30	453±30	474±0	453±30	453±30	431±20
	imitated call		6<=	76±23	73±8	54±4 *	84±3	77±11	87±20	689±61	689±61	689±61	668±30	668±30	668±20
Hou?	sample call	9	1	329±36						374±25(min)-703±50(max)					
	imitated call		1	400±76						503±90(min)-933±50(max)					
A wawawa …	sample call	6	5	121±8	164±11	157±8	151±13	156±16		237±30	539±30	517±0	517±0	517±0	
	imitated call		1	650<=						301±22					
Duke	sample call	3	2	220±43	182±44					284±87	194±30				
	imitated call		2	150±11	107±19					323±31	258±0				
Ohayo	sample call	6	3	159±12	226±37	273±52				875±5	1249±149	1077±86			
	imitated call		3	250±38	354±99	365±151				646±86 *	890±204	761±66 *			
Tsukasa	sample call	7	3	252±84	265±56	173±35				703±90	1192±100	1220±87			
	imitated call		3	236±69	217±43	273±112				660±66	689±0	718±30			
Piyo piyo	sample call	4	4	88±15	77±7	68±3	74±13			2785±90	2024±789	3000±109	1507±228		
	imitated call		4	103±14	116±32	90±17	88±9			2742±25	2527±163	3029±431	2498±187 *		
Hoh kekyo	sample call	3	3	825±89	96±45	98±19				367±30	668±30	668±30			
	imitated call		3	1995±339	100±0	122±14				668±30 *	668±30	689±0			
Oh!	sample call	2	1	174±56						280±30					
	imitated call		1	223±9						538±335					

Note. *N*: The number of presentations
* $p < 0.05$, t-test.

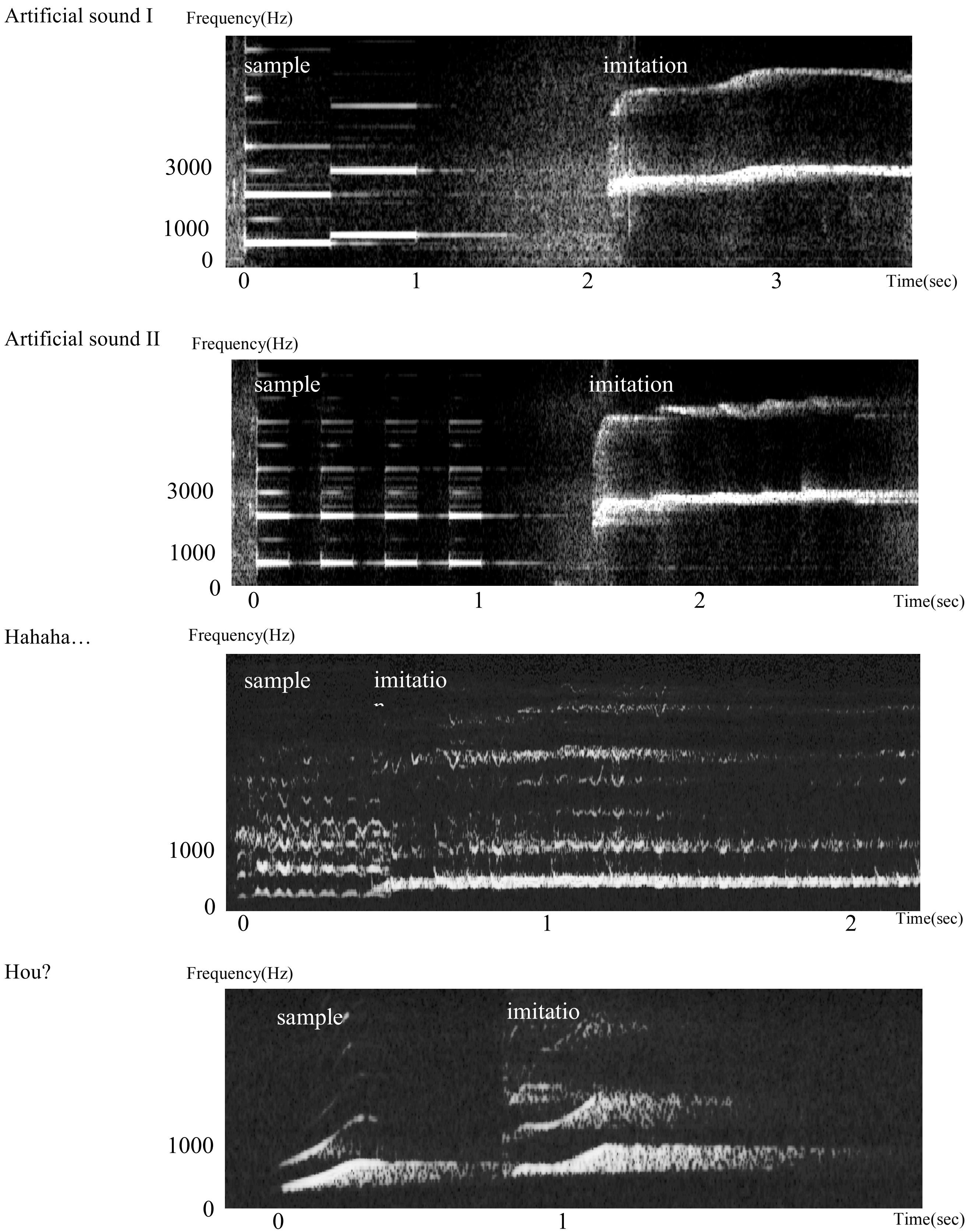
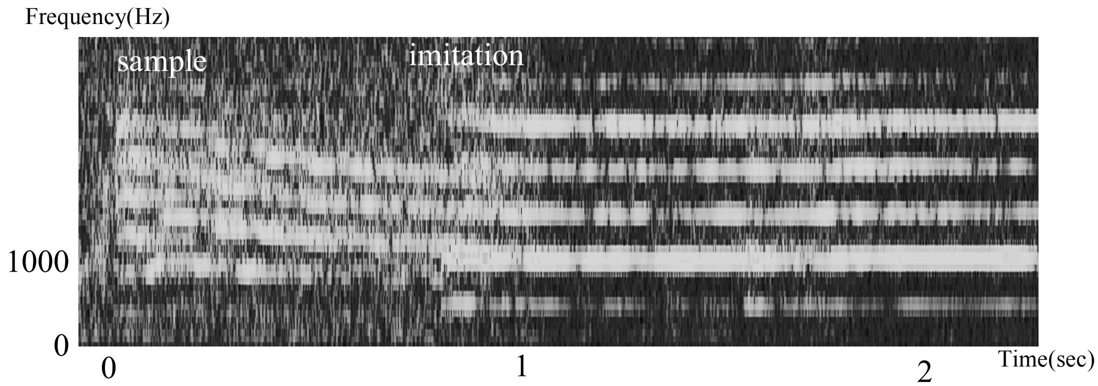
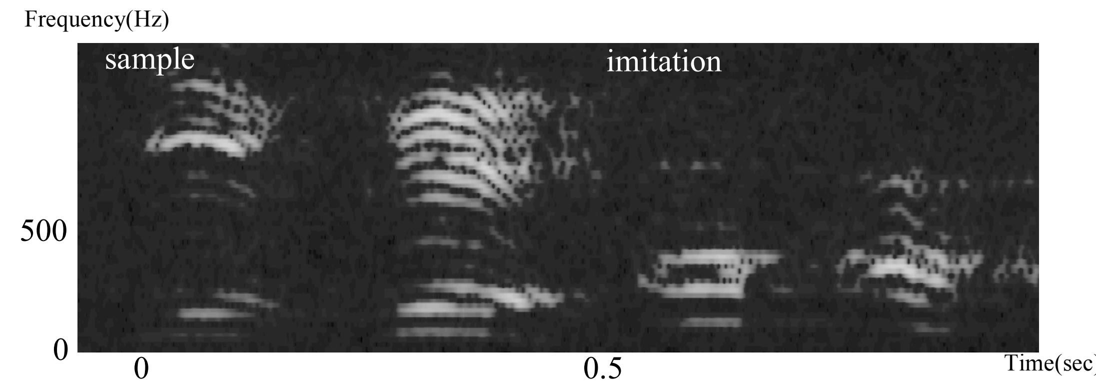


Figure 5. Spectrograms of the sample sounds and imitated calls.

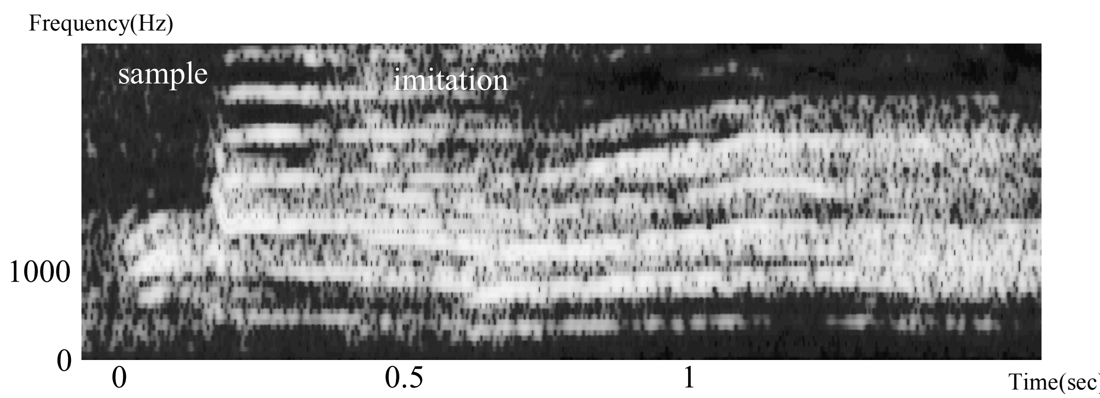
A wawawa



Duke



Ohayo



Tsukasa

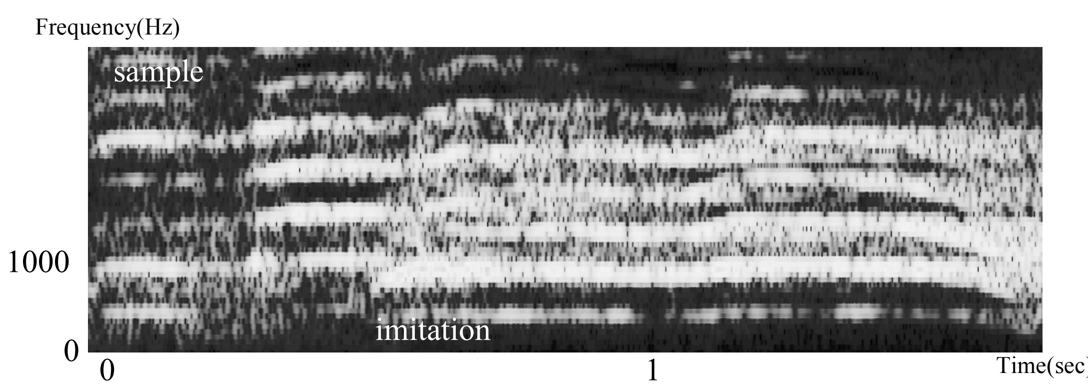


Figure 5 (continued).

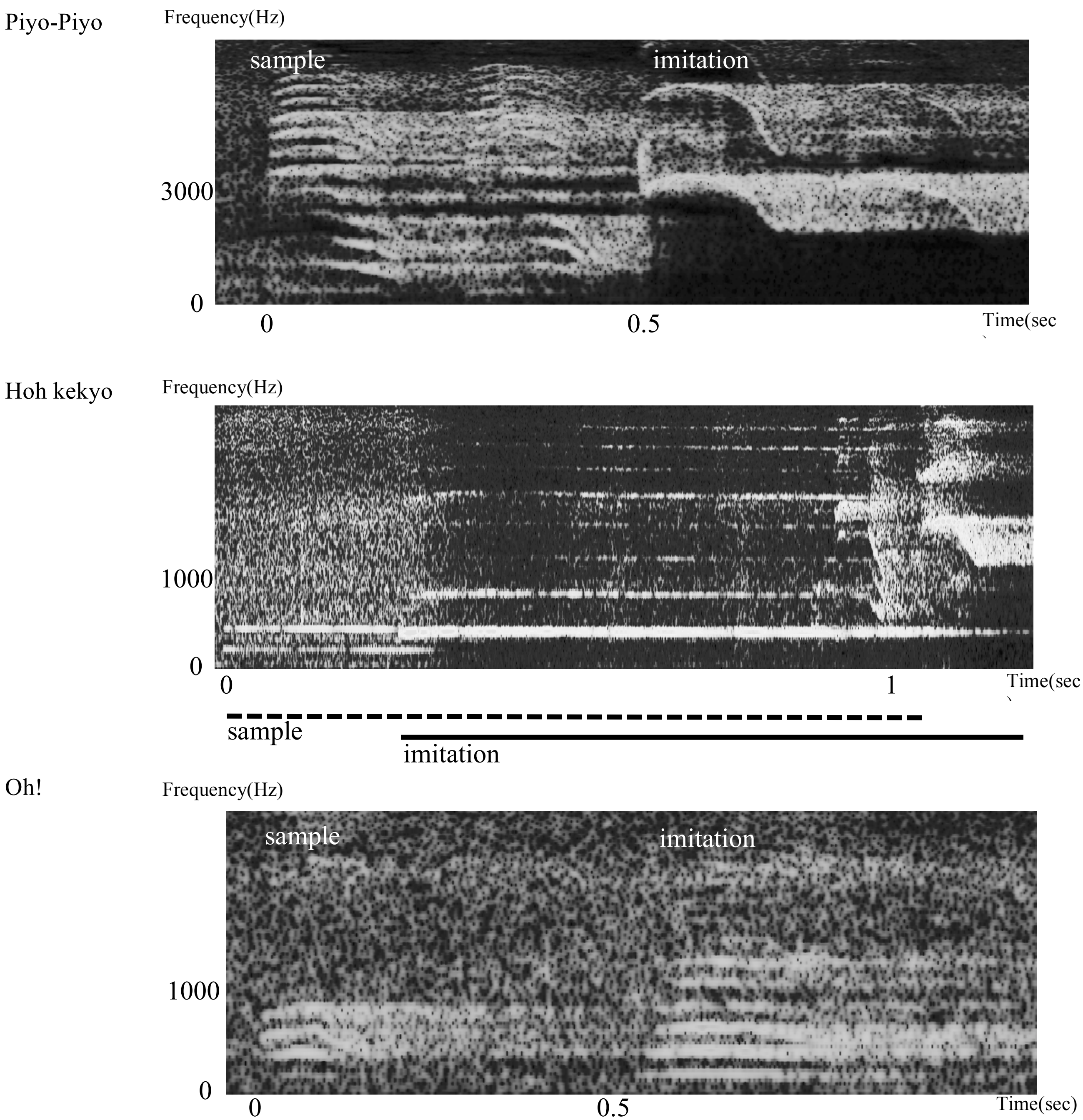


Figure 5 (continued).

a

b

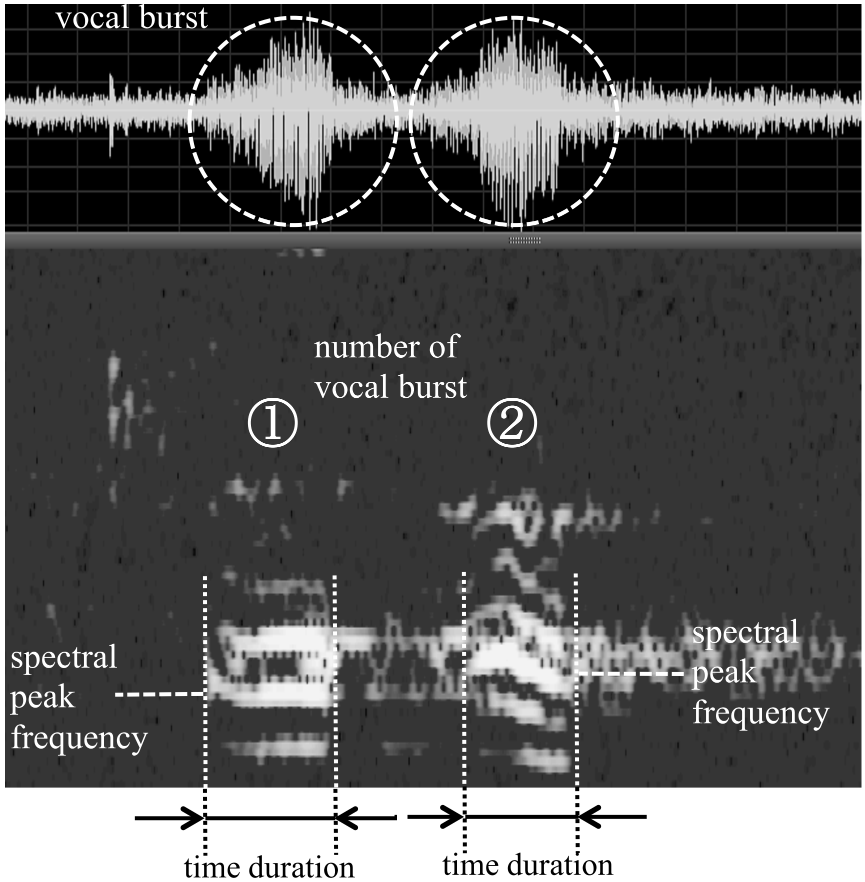


Figure 6. Example of spectrogram with legends. (a). Changes in amplitude. (b). Spectrogram of frequency.

These results demonstrated that the subject displayed an aptitude for imitation.

Discussion

Imitation is a highly cognitive ability, and belugas possess high cognitive abilities. For example, a beluga has shown the formation of transitivity and symmetrical relationships, and the ability to label objects with sound production (Murayama & Tobayama, 1997; Murayama et al., 2008, 2012). In the present study, the subject learned to imitate the sample sounds during the training session, and he could imitate not only the trained sounds (calls) but also the untrained sounds, which were two computer-generated artificial sounds and nine vocal sounds emitted by the experimenter. When the artificial sounds were tested, some of the acoustic features of the imitated sounds resembled the sample sounds, i.e., the subject imitated the artificial sounds well. These results were in agreements with those of studies that have been conducted on bottlenose dolphins (Reiss & McCowan, 1993; Richards et al., 1984). In addition, some acoustic features of the imitated human vocal sounds were considerably similar to those of the sample sounds, except for "A wawawa…", suggesting that the subject could imitate arbitrary human calls. Lilly (1962) examined whether the bottlenose dolphin could imitate human speech sounds, without success. Since the belugas, as well as bottlenose dolphins, do not possess vocal chords like humans, the subject of the present study could not exactly reproduce the human

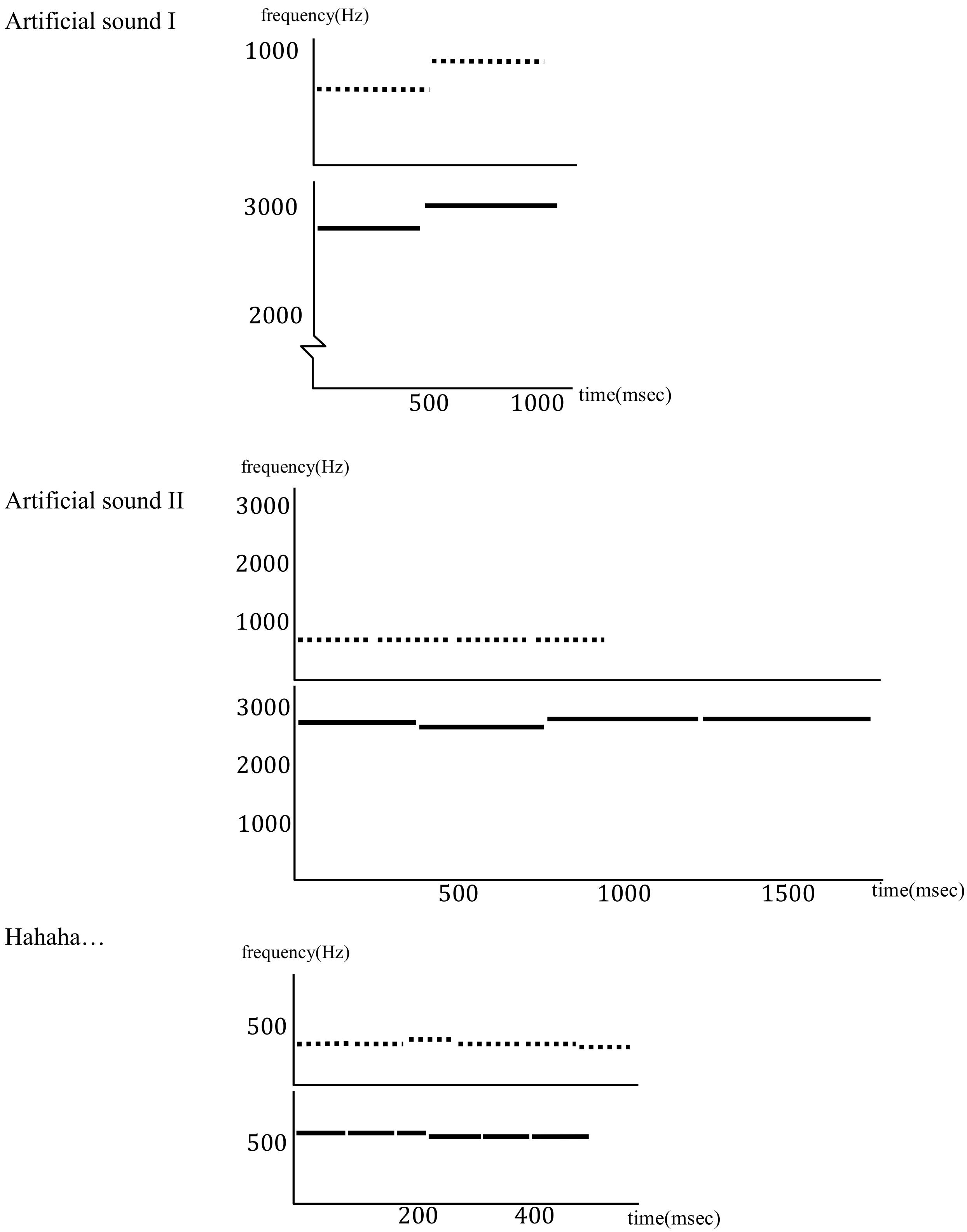
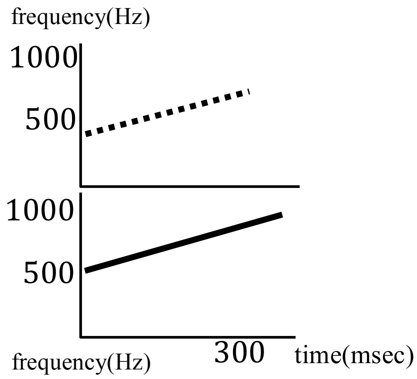
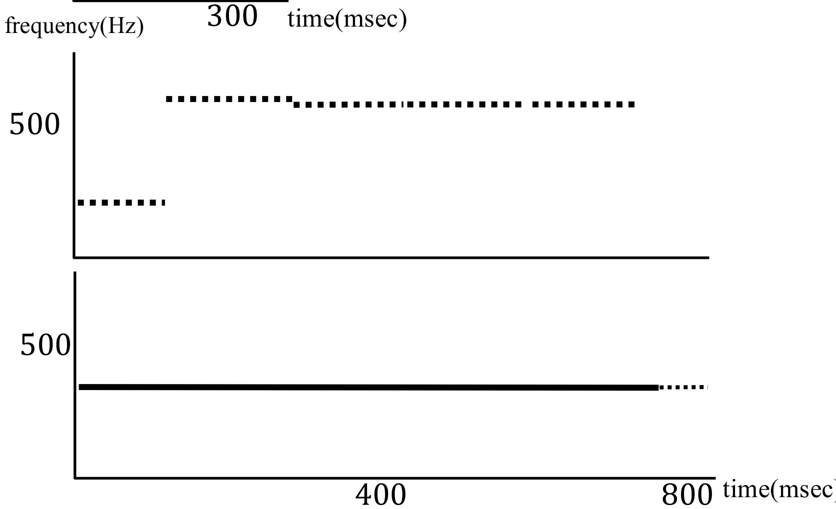


Figure 7. Schematized changes in spectral peak frequency. Dashed lines represent the changes in spectral peak frequency of the sample sounds, and solid lines indicate that of the imitated calls.

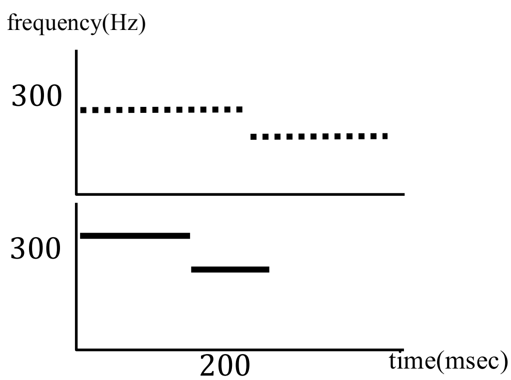
Hou?



A wawawa···



Duke



Ohayo

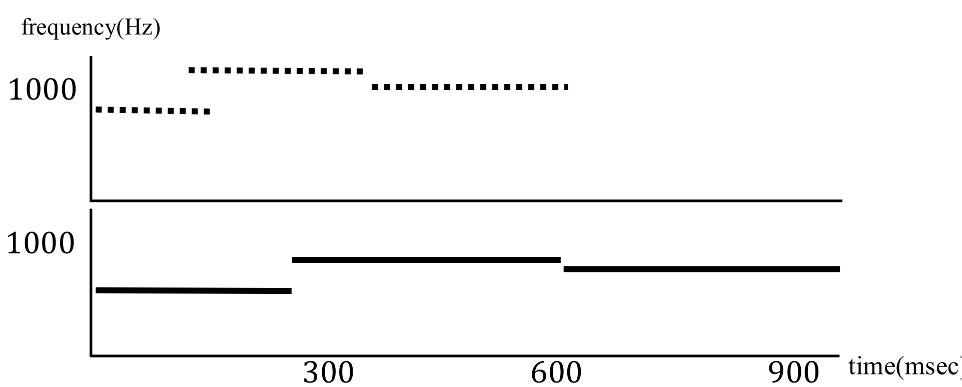


Figure 7 (continued).

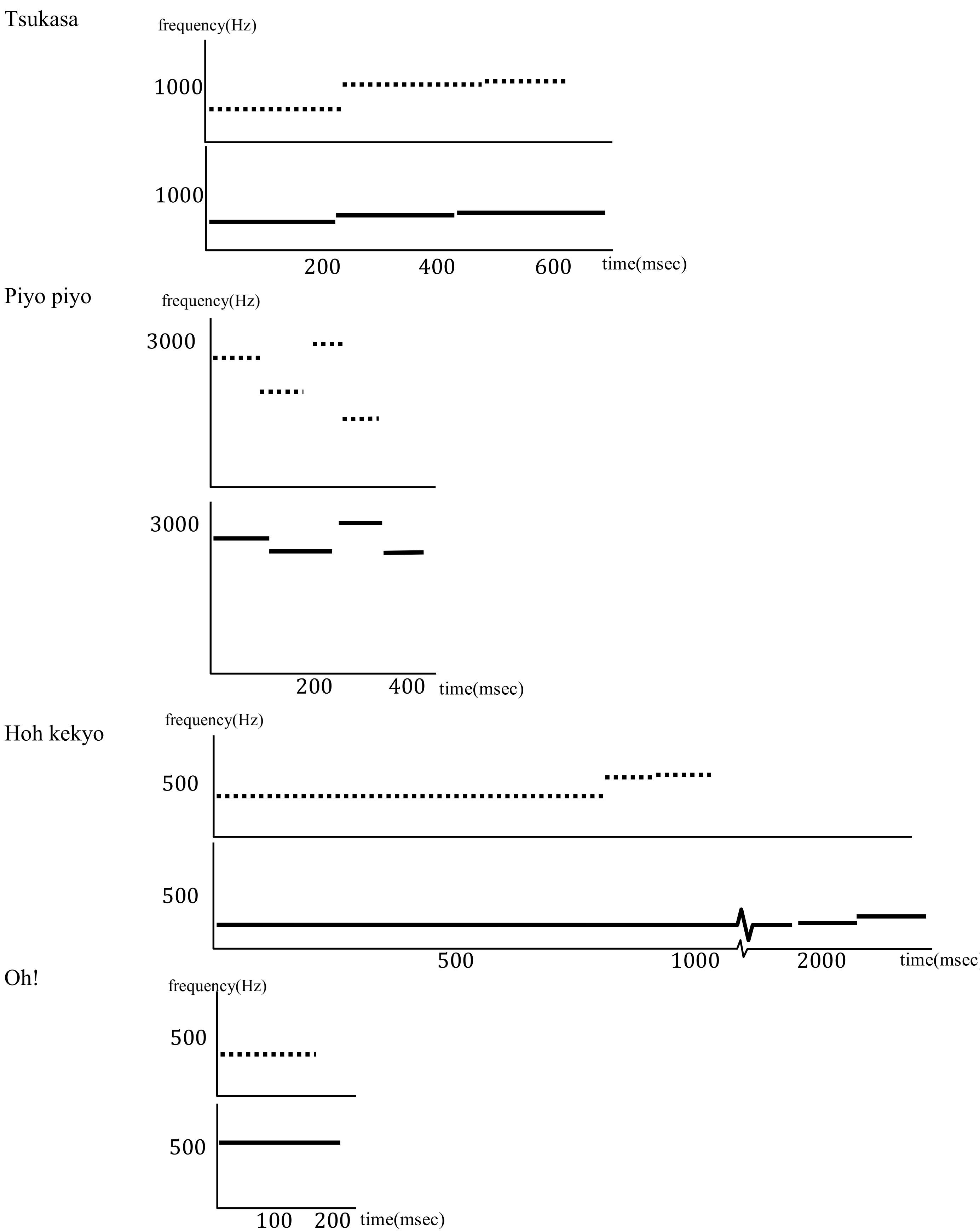


Figure 7 (continued).

vocal sounds were considerably similar to those of the sample sounds, except for "A wawawa…", suggesting that the subject could imitate arbitrary human calls. Lilly (1962) examined whether the bottlenose dolphin could imitate human speech sounds, without success. Since the belugas, as well as bottlenose dolphins, do not possess vocal chords like humans, the subject of the present study could not exactly reproduce the human voice. However, Ridgway et al. (2012) reported that a beluga spontaneously imitated human speech. In the present study, we demonstrated that the subject imitated some components of human vocal sounds and that the rhythm or intonation of imitated calls was similar to that of the sample sounds, indicating that he tried to imitate human speech. Thus, the ability of a beluga to imitate human speech was experimentally verified.

Moreover, in the test session, the subject received no rewards irrespective of correct or incorrect responses, but the subject spontaneously imitated each sound in response to the sample sound. These results suggested that the subject possessed a general aptitude for imitation.

The bottlenose dolphin has the ability to imitate sounds or calls made by another individual (Smolker & Pepper, 1999; Tyack, 1986; Tyack & Sayigh, 1997), and the bottlenose dolphin can imitate sounds that he has not heard previously and can spontaneously imitate computer-generated artificial sounds (Reiss & McCowan, 1993; Richards et al., 1984). In addition, bottlenose dolphins incorporate features of artificial sounds into their signature whistles (Miksis, Tyack, & Buck 2002). Thus, we showed that belugas, as well as bottlenose dolphins, have the ability to imitate sounds.

The role of imitation is thought to be an affiliation signal to integrate new members in a group (Mammen & Nowicki, 1981). A human, in imitation, modifies many aspects to match the partner and to establish a social relationship with the partner (Giles, 1984). Vocal imitation appears in other species that form and maintain individual specific bonds within social groupings. Although belugas produce a variety of calls in nature, the role of these calls remains unclear. Janik (2000) reported that bottlenose dolphins imitate another's whistle to contact that particular individual, suggesting that the imitation ability functions as a name for reference (Tyack, 1999) or as an affiliative signal (Smolker & Pepper, 1999) in bottlenose dolphins. In dolphins, vocal imitation allows the development of vocal communication. Therefore, the call of belugas may have a similar role. If the role of these calls is further elucidated, the function of imitation may become clear.

Acknowledgements

We sincerely thank Miss Yuri Hobo, Mr Ryohei Kawasaki and the staff members of Kamogawa Sea World aquarium for their cooperation. The authors would like to thank two anonymous reviewers and the editor for their advice on this manuscript.

Supplementary Material

Recording of vocalizations by human trainers and subsequent Beluga whale vocal responses (see Supplementary Material).

References

Giles, H. (1984). The dynamics of speech accommodation. *International Journal of the Sociology of Language*, *46*, 1-155.

Gregory, M., & O'Corry-Crowe. (2002). Beluga whale. In W. Perrin, B. Wursig, & J. G. M. Thewissen (Eds.), *Encyclopedia of Marine Mammals* (pp. 94-99). San Diego, CA: Academic Press.

Herman, L. M., & Tavolga, W. N. (1988). The communication system of cetaceans. In L. M. Herman (Ed), *Cetacean behavior:Mechanisms and functions* (pp.149-209). Florida: Krieger.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science, 289*, 1355-1357.

Lilly, J. C. (1961). *Man and dolphin: Adventures of a new scientific frontier* (1st ed.). New York, NY: Garden City.

Lilly, J.C. (1962). *Vocal behavior of the bottlenose dolphin.* Proceedings of the American Philosophical Society*, 106*, 520-529.

Lilly, J. C., Miller, A. M., & Truby, H. M. (1968). Reprogramming of the sonic output of the dolphin: Sonic burst count matching. *Journal of the Acoustical Society of America*, *43*, 1412-1424.

Mammen, D. L., & Nowicki, S. (1981). Individual differences and within-flock convergence in chickadee calls. *Behavioral Ecology and Sociobiology, 9*, 179-186.

Miksis, J. L., Tyack, P., & Buck, J. (2002). Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds. *Journal of Acoustical Society of America*, *112*, 728-739.

Murayama, T., & Tobayama, T. (1995). Preliminary study of mental rotation in beluga. *Abstract of the XXIV International Ethological Conference*, *114*, Hawaii, 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)

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., 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., Fujii, Y., Hashimoto, T., Shimoda, A., Iijima, S., Hayasaka, K., …Arai, K. (2012). Preliminary study of object labeling using sound production in a beluga. *International Journal of Comparative Psychology*, *25*, 195-207.

Pepperberg, I. M. (2000). *The Alex studies: Cognitive and communicative abilities of gray parrots*. Cambridge, MA: Harvard University Press.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Animal behavior: Elephants are capable of vocal learning. *Nature*, *434*, 455-456.

Reis, D., & McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning. *Journal of Comparative Psychology*, *107*, 301-312.

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.

Ridgway, S., Carder, D., Jeffries, M., & Todd, M. (2012). Spontaneous human speech mimicry by a cetacean. *Current Biology*, *22*, 860-861.

Smolker, R., & Pepper, J. (1999). Whistle convergence among allied male bottlenosed dolphin (*Delphinidae, Tursiops* sp.). *Ethology*, *105*, 595-617.

Stoeger, A. S., Mietchen, D., Sukhun, O., Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, *22*, 2144-2148.

Todt, D. (1975). Social learning of vocal patterns and modes of their application in gray parrots (*Psittacus erithacus)*. *Zeitschrift für Tierpsychologie*, *39*, 178 -188.

Tyack, P. L. (1986). Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: Mimicry of signature whistles? Behavioral Ecology *and* Sociobiology*, 18*, 251-257.

Tyack, P. L. (1999). Communication and Cognition. In J. E. Reynolds III, & S. A. Rommel (Eds.), Biology of Marine Mammals (pp. 287-323). Washington, D.C.: Smithsonian Institution Press.

Tyack, P. L., & Sayigh, L. S. (1997). Vocal learning in cetaceans. In C. Snowdon, & M. Hausberger (Eds.), *Social Influences on Vocal Development* (pp. 208-233). Cambridge, MA: Cambridge University Press.

Watwood, S. L., Tyack, P. L., & Wells, R. S. (2004). Whistle sharing in paired male bottlenose dolphins *Tursiops truncatus*. *Behavioral Ecology and Sociobiology*, *55*, 531-543.

West, M. J., Stroud, A. N., & King, A. P. (1983). Mimicry of the human voice by European starlings: The role of social interaction. *Wilson Bulletin*, *95*, 635-640.

Financial Support: This study was supported by a grant of International Marine Biological Research Institute in the Kamogawa Sea World, Japan.

Conflict of Interest: All authors of this paper declare no conflict of interest.

Submitted: *December 19th, 2013*
Resubmitted: *April 5th, 2014*
Accepted: *May 25th, 2014*