eScholarship
International Journal of Comparative Psychology

Title

Effects of Modifying Human Vocal Cues on Behavioral Responses of Steller Sea Lions (Eumetopias jubatus) Under Managed Care

Permalink

https://escholarship.org/uc/item/0f34d92r

Journal

International Journal of Comparative Psychology, 39(1)

ISSN

0889-3667

Authors

Sasaki, Masahiro
Kambara, Toshimune

Publication Date

2026-11-09

DOI

10.46867/ijcp.65100

Supplemental Material

https://escholarship.org/uc/item/0f34d92r#supplemental

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

INTERNATIONAL JOURNAL OF COMPARATIVE PSYCHOLOGY

2026, 39
Charles I. Abramson
Editor
Peer-reviewed

Effects of Modifying Human Vocal Cues on Behavioral Responses of Steller Sea Lions (*Eumetopias jubatus*) Under Managed Care

Masahiro Sasaki[1] and Toshimune Kambara[2]

[1]Kinosaki Marine World, 1090 Seto, Toyooka, Hyogo 669-6192, Japan

[2]Department of Psychology, Graduate School of Humanities and Social Sciences, Hiroshima University, 1-1-1 Kagamiyama, Higashihiroshima, Hiroshima, Japan

The present study aimed to investigate the effects of modifying human vocal discriminative stimuli (hereafter referred to as vocal cues) on the behavioral responses and discrimination performance of two adult Steller sea lions (*Eumetopias jubatus*), a female (Hama) and a male (Shunta), housed under managed care. Both subjects could already discriminate standard human vocal cues. To determine if they could spontaneously respond to modified versions of these cues without prior training, we evaluated three types of modifications: presenting only the first or second half of the vocal cue (Test 1), altering a single phoneme within the vocal cue (Test 2), and changing the accent of the vocal cue (Test 3). During testing, the modified cues were presented randomly mixed with the original vocal cues that the subjects had previously learned to associate with specific behaviors. We analyzed the subjects' accuracy rates, based on a binomial distribution, and reaction times across the three tests. Accuracy rates were significantly lower for modified cues across all test types ($p < .05$), with accent alterations having the greatest negative impact on performance. A detailed analysis of behavioral responses suggested that the subjects did not use a single strategy for all vocal cues, such as always focusing on a specific phonetic position. Instead, they developed individual strategies to find a unique discriminative point for each cue. Hama, who possessed a large repertoire of learned vocal cues, relied on the vowels produced by the trainer (vowel dominance) during her discrimination process. Frame-by-frame video analysis (30 fps) also revealed that certain cue modifications significantly influenced the reaction times of correct responses for both subjects. These findings indicate that Steller sea lions possess the cognitive flexibility to spontaneously associate modified vocal cues with their original counterparts, even when the acoustic properties deviate somewhat from familiar forms. These results also suggest that Steller sea lions, like other animals (e.g., primates, dogs, and rats), do not use the phoneme-based strategy that humans apply to large vocabularies. Instead, Steller sea lions rely on markers specific to each cue.

Keywords: cognitive flexibility, discrimination, modification, reaction time, Steller sea lion, vocal cue

Please send correspondence to Masahiro Sasaki, Kinosaki Marine World, 1090 Seto, Toyooka, Hyogo 669-6192, Japan. Email: cod.beautiful.gbr@gmail.com

飼育下のトド（*Eumetopias jubatus*）における、ヒト音声キューの改変が行動反応に与える影響

本研究は、ヒト音声を用いた弁別刺激（以下、音声キューとする）の改変が、飼育下の 2 頭のトド成獣（*Eumetopias jubatus*；メスのハマとオスのシュンタ）の行動反応および弁別成績に与える影響を調査することを目的とした。両個体はすでに標準的なヒトの音声キューを弁別することができた。そこで、彼らが事前訓練なしに、改変を加えた音声キューに対して自発的に反応できるのかを調べるために、3 種類の改変を用いて評価した。テスト 1 では音声キューの前半もしくは後半のみを提示し、テスト 2 では音声キューの中の 1 つの音素を改変して提示した。そしてテスト 3 では、アクセントを変えた音声キューを提示した。テスト期間中、改変した音声キューは、各個体がすでに特定の動作との連合を学習している通常の音声キューの中にランダムに混ぜて提示した。我々は 3 つのテストにおいて、二項分布に基づいた正解率および反応時間を分析した。正解率は全てのテストタイプにおいて、改変した音声キューが通常の音声キューに対して有意に低く（$p < .05$）、アクセントの改変が弁別成績に最も大きな負の影響を与えた。行動反応の詳細な分析から、両個体は特定の音素位置に常に着目するといった、全ての音声キューに共通する単一の弁別戦略を用いているのではないことが示唆された。その代わりに、各個体がそれぞれの音声キューに対して独自の弁別ポイントを見つけて個別の弁別戦略を発達させていた。多くの学習済み音声キューのレパートリーをもつハマは、弁別プロセスにおいてトレーナーが生成する母音に依存していた（母音優勢）。また、コマ送りの動画（30 fps）を用いた解析によって、特定のキューの改変が、両個体の正解反応時の反応時間に影響を与えることを明らかにした。これらの知見はトドが、音響特性が馴染みのある形からある程度逸脱した場合であっても、改変されたキューを、それと対応する従来のキューと自発的に関連付ける認知的柔軟性をもつことを示している。また、トドが他の動物（霊長類、イヌ、ラットなど）と同様に、ヒトが大きな語彙レパートリーに対して用いる音素ベースの弁別戦略を用いないことを示唆している。その代わりに、トドは各キューに特有のマーカーに依存している。

キーワード：認知的柔軟性、弁別、改変、反応時間、トド、音声キュー

Efectos de la modificación de las señales vocales humanas en las respuestas conductuales de los leones marinos de Steller (*Eumetopias jubatus*) bajo cuidado humano

El presente estudio tuvo como objetivo investigar los efectos de la modificación de estímulos vocales humanos discriminativos (en adelante, señales vocales) sobre las respuestas conductuales y el desempeño en tareas de discriminación de dos leones marinos de Steller adultos (*Eumetopias jubatus*), una hembra (Hama) y un macho (Shunta), mantenidos bajo cuidado humano. Ambos individuos ya eran capaces de discriminar señales vocales humanas estándar. Para determinar si podían responder espontáneamente a versiones modificadas de estas señales sin entrenamiento previo, se evaluaron tres tipos de modificaciones: la presentación únicamente de la primera o segunda mitad de la señal vocal (Prueba 1), la alteración de un solo fonema dentro de la señal vocal (Prueba 2) y el cambio del acento de la señal vocal (Prueba 3). Durante las pruebas, las señales modificadas se presentaron aleatoriamente junto con las señales vocales que los individuos habían aprendido previamente a asociar con conductas específicas.

Se analizaron las tasas de acierto de los individuos, con base en una distribución binomial, así como los tiempos de reacción en las tres pruebas. Las tasas de acierto fueron significativamente menores para las señales modificadas en todos los tipos de prueba ($p < .05$), y las modificaciones del acento tuvieron el mayor impacto negativo sobre el desempeño. Un análisis detallado de las respuestas conductuales sugirió que los individuos no empleaban una única estrategia para todas las señales vocales, como centrarse siempre en una posición fonética específica. En cambio, desarrollaron estrategias individuales para identificar un punto discriminativo único para cada señal. Hama, quien poseía un amplio repertorio de señales vocales aprendidas, se basaba en las vocales producidas por el entrenador (predominio de las vocales) durante el proceso de discriminación. El análisis de video cuadro por cuadro (30 fps) también reveló que determinadas modificaciones de las señales retrasaban significativamente los tiempos de reacción de las respuestas correctas en ambos individuos. Estos hallazgos indican que los leones marinos de Steller poseen la flexibilidad cognitiva necesaria para asociar espontáneamente señales vocales modificadas con sus correspondientes señales originales, incluso cuando las propiedades acústicas se desvían parcialmente de las formas familiares. Estos resultados también sugieren que los leones marinos de Steller, al igual que otros animales (p. ej., primates, perros y ratas), no emplean la estrategia basada en fonemas que los seres humanos utilizan para manejar vocabularios amplios. En su lugar, los leones marinos de Steller se basan en marcadores específicos de cada señal.

Palabras clave: flexibilidad cognitiva, discriminación, modificación, tiempo de reacción, león marino de Steller, señal vocal

Visual and acoustic systems are two major sensory modalities used in animal communication (Rosenthal & Ryan, 2000). Accordingly, numerous studies have investigated animals' stimulus discrimination and associative learning abilities using both hand signals and human vocal cues. Research into associative learning based on human-provided visual cue discrimination has been documented extensively, particularly in primates and cetaceans. Gardner and Gardner (1969) reported that a female chimpanzee (*Pan troglodytes*), Washoe, learned hand signals and combined them to refer to objects. Patterson and Cohn (1990) also reported that a female lowland gorilla (*Gorilla gorilla*), Koko, learned to communicate using hand signs. Among cetaceans, bottlenose dolphins (*Tursiops truncatus*) have learned to associate hand signals and pointing gestures with specific actions and target objects (e.g., Herman, 2013; Herman et al., 1984; Pack & Herman, 2004; Tschudin et al., 2001), and later studies using multiple consecutive hand signals demonstrated syntactic comprehension, the ability to process complex sequences of requests, in these dolphins (Herman, 2010, 2013; Kuczaj, 2013).

In contrast, studies on associative learning based on human vocal cues have been conducted mainly with companion animals that co-evolved with humans, such as dogs (*Canis lupus familiaris*; Griebel & Oller, 2012; Pilley & Reid, 2011) and cats (*Felis catus*; Takagi et al., 2022), as well as with parrots (*Psittacus erithacus*; Pepperberg, 1981). Instead of merely reacting to sounds, these animals demonstrate referential understanding (i.e., mapping specific words to their referents): dogs can retrieve specific toys by name (Pilley & Reid, 2011), cats can recognize their own names and those of cohabiting cats (Saito et al., 2019; Takagi et al., 2022), and parrots can vocally identify objects and categories (Pepperberg, 1981). Some studies have gone beyond merely demonstrating discrimination performance, focusing instead on investigating the underlying cognitive mechanisms of vocal cue discrimination. For instance, Fukuzawa et al. (2005) reported that single phonemic changes significantly impaired word discrimination performance in dogs. Other studies have revealed a vowel bias in word discrimination in dogs (Mallikarjun et al., 2021) and rats (*Rattus norvegicus*; Bouchon & Toro, 2019). Despite these comparative studies, findings on vocal-cue communication have long been limited to a few specific taxa.

Pinnipeds comprise three families: Phocidae (seals), Otariidae (sea lions), and Odobenidae (walruses). They communicate with conspecifics using vocalization (Schusterman & Van Parijs, 2003) and behaviors (Gentry, 1974; Wilson, 2024). Sea lions, in particular, are highly social animals, as shown by complex behaviors such as precise mother-pup vocal recognition (e.g., Australian sea lions, *Neophoca cinerea*, Charrier et al., 2009, 2022), the transfer of novel foraging sites via social learning (Schakner et al., 2017), and the maintenance of social relationships and spatial territories through visual and acoustic displays (Gentry, 1974, for Steller sea lions, *Eumetopias jubatus*). The discovery of their high social and cognitive skills led to investigations of the cognitive abilities of sea lions in human care. Several studies reported that sea lions can discriminate human physical cues (Arkwright et al., 2016) or learn the association between a trainer's physical cues and their actions (Cook et al., 2021; Kastak & Schusterman, 1994; Sasaki, Tsutsumi, et al., 2025). This led to more in-depth research demonstrating syntactic comprehension (Schusterman & Krieger, 1984; Schusterman et al., 2002) and stimulus equivalence (e.g., Lindemann-Biolsi & Reichmuth, 2014; Schusterman & Kastak, 1998) in California sea lions.

Building upon the extensive research on sea lion cognition using visual cues (i.e., hand signals and object cues; e.g., Kastak & Schusterman, 1994; Schusterman & Kastak, 1998; Schusterman & Krieger, 1984; Schusterman et al., 2003), a new wave of case studies on associative learning via human vocal cues in pinnipeds emerged in the 2020s. Endo et al. (2020) reported that a male walrus (*Odobenus rosmarus*), Pou, discriminated 10 types of human vocal cues and associated them with specific actions correctly under various conditions. A female Steller sea lion, Hama, also discriminated between 10 types of human vocal cues and associated each cue with specific actions with high accuracy (Sasaki et al., 2022). In addition, Hama's son, Kanata, also discriminated human vocal cues connected to behaviors as a Steller sea lion pup (Sasaki, Tsutsumi, et al., 2025). These recent studies suggest that certain individuals within specific pinniped species possess the ability to learn the association between vocal cues and actions. However, while dog studies have advanced to

investigating the underlying cognitive processes of discrimination, comparing them to human lexical discrimination processes (Fukuzawa et al., 2005), pinniped studies have largely remained within the framework of confirming the capacity for associative learning. Most marine mammal literature has focused on whether subjects can respond to human vocal cues, rather than exploring the mechanisms of the discrimination process itself.

In this context, Sasaki and Kambara (2024) showed that the female Steller sea lion, Hama, demonstrated the spontaneous formation of behavioral categories based on previously learned cues. Specifically, when presented with a novel condition involving two consecutive vocal cues, Hama spontaneously modified her actions to produce a combined response that maintained the structural integrity of the associated behavioral categories. Although each behavioral component came from her trained repertoire, Hama spontaneously modified these elements and combined them into a familiar compound action. This process occurred without prior reinforcement, showing that she did not simply recall trained actions but adjusted them to fit the structure of a behavioral category. Thus, while progress has been made in fundamental research on vocal cue-based associative learning, the underlying mechanisms of how sea lions process vocal cues and determine their responses remain unclear. In particular, the specific factors (e.g., phonemic and prosodic properties) that affect their discrimination performance have not yet been fully investigated.

Therefore, we conducted a pilot study to explore the mechanism by which Steller sea lions process vocal cues from a behavioral response perspective. To achieve this, we used "modified vocal cues," that is, stimuli derived from original cues learned during previous training (detailed in Sasaki & Kambara, 2024) that were subjected to specific acoustic alterations. Specifically, we investigated the vocal cue discrimination performance of two well-trained Steller sea lions under three experimental conditions. Based on historical training and husbandry records, we confirmed that while these subjects had been trained with more than 20 types of human vocal cues, they had never been exposed previously to the specifically modified vocal cues used in this study. In Test 1, we presented divided cues (only the first or second halves). In Test 2, we presented cues with a single modified phoneme. In Test 3, we presented cues with a different accent. Through these experiments, we aimed to provide basic insights into the acoustic features that drive vocal cue comprehension in Steller sea lions.

Method

The Subjects

The subjects were two well-trained Steller sea lions, Hama (female) and Shunta (male), which could respond to human cues (i.e., hand signals and vocal cues). At the time of this study, Hama was 13 years old (born on July 21, 2009) and Shunta was 21 years old (born on June 26, 2001). Hama had learned 50 human vocal cues and, based on her behavioral responses to two consecutive cues, she appeared to categorize her responses to these vocal cues (Sasaki & Kambara, 2024). Detailed information about Hama's husbandry routine and learning experiences has been described in Sasaki et al. (2022).

Shunta was born at Kinosaki Marine World and participated in basic training (i.e., stationing and behavioral shaping training) immediately after weaning at one year of age (June 2002). He was housed in multiple cages according to his developmental stages. Once he was seven years old, he was kept mainly in an exhibition pool (Sasaki & Kambara, 2024). Over 20 years (2002–2022), Shunta learned approximately 30 hand signals and their associated behaviors. Vocal cue training with Shunta began in April 2017. We established stimulus control for the vocal cues by simultaneously presenting them with known hand signals. These hand signals were then gradually withdrawn using a prompt-fading technique until the behaviors were maintained by the vocal cues alone. Through this gradual process, he successfully learned to associate 20 novel human vocal cues with his previously acquired behaviors over a five-year training period, concluding in 2022.

The Experimenter

The experimenter was a male trainer with 10 years of work experience in marine mammal husbandry and training after graduating from a university with a degree in marine biology. He (this study's first author) has previously conducted research on Steller sea lions as an experimenter.

Study Location

We conducted the experiments at two locations, depending on each subject's familiar feeding location in the habitat. We conducted tests with Hama in a cage (Sasaki et al., 2022). We conducted tests with Shunta on the pool deck (1.5 × 2.5 m; see Figure 1).

Figure 1

The Study Location for Shunta



Note. Arrow 1 indicates Shunta (the subject), and Arrow 2 indicates the experimenter. The experimenter faced Shunta across the gate during the experiment.

Human Vocal Cues

We used seven types of human vocal cues (*keirei, onaka, omawari, touritsu, huse, goron, and oshiri*) that both subjects had previously learned through training. These cues correspond to specific actions: salute, show belly, spin, handstand, lie down, roll over, and show hindquarters, respectively. The behaviors associated with each vocal cue are further described in Table 1 and the supplemental Video: https://osf.io/mx5uh/overview?view_only=5259d2a4b35e49d4855e8ef60465f0e3.

Table 1

Target Vocal Cues and Associated Actions

Vocal Cue	Action
keirei	Touching the back side of the right foreflipper to the nose.
onaka	Pressing the abdomen against the cage.
omawari	Rotating on an axis perpendicular to the ground.
touritsu	Standing upside down.
huse	Lying face down.
goron	Lying on the back.
oshiri	Turning around and lowering the hindquarters.

Note. Video of each vocal cue and action can be seen from this link:
https://osf.io/mx5uh/overview?view_only=5259d2a4b35e49d4855e8ef60465f0e3

Vocal Cue Modifications

We modified each cue according to the test type:

- Test 1: Only the first or second half of the original vocal cues was presented to the subjects.
- Test 2: One phoneme in the cues was modified to another phoneme.
- Test 3: The pitch accent of the cues was modified.

In Test 1, we divided the cues into two parts (*kei/rei, ona/naka, oma/wari, tou/ritsu, hu/se, goro/ron, and oshi/shiri*) and presented them to the subjects. Because *onaka*, *goron*, and *oshiri* consist of three moras (syllables), we divided them into two parts by overlapping the middle mora, as described above. Following this procedure, we created 14 modified vocal cues (Table 2).

In Test 2, we modified one of the phonemes of the cues and presented them to the subjects. The modified cues consisted of two patterns: cues that changed the vowel for each phoneme and cues that changed the consonant for each phoneme. We created four to eight types of modified cues for each cue type, depending on the number of moras constituting the cues (e.g., eight types for *keirei* and *omawari*; six types for *onaka* and *oshiri*; and four types for *huse* and *goron*). The modified cues were created by a novel trainer who had never interacted with these subjects before. Following this procedure, we created 44 modified vocal cues (detailed in Table 2 for a complete list).

In Test 3, we altered the pitch accent of the cues and presented them to the subjects. Japanese is a pitch-accent language, meaning that emphasis is conveyed via relative changes in pitch (frequency) across successive moras, rather than through stress or loudness as in English. Originally, the pitch accent was placed on the first mora for *keirei, touritsu*, and *huse*; the second mora for *omawari* and *goron*; and the third mora for *onaka* and *oshiri*. We shifted the accent to a different position and created one to three types of modified cues for each cue type, depending on the number of moras constituting the cues (i.e., one for *huse*, two for *goron, onaka*, and *oshiri*, and three for *keirei, touritsu*, and *omawari*). Following this procedure, we created 16 modified vocal cues (Table 2).

Table 2

Modified Cues Presented in This Study

Test 1	Test 2			Test 3
kei	*eiei*	*kairei*	*keireichi*	*keírei*
rei	*keireu*	*keiroi*	*keitei*	*keiréi*
	kekirei	*keorei*		*keireí*
ona	*enaka*	*gonaka*	*onaba*	*ónaka*
naka	*onaku*	*onika*	*osaka*	*onáka*
oma	*amawari*	*omewari*	*omawati*	*omawarí*
wari	*omawori*	*somawari*	*omawaro*	*ómawari*
	omapari	*okawari*		*omawári*
tou	*tokuritsu*	*houritsu*	*tourita*	*toúritsu*
ritsu	*toupitsu*	*tauritsu*	*tourutsu*	*tourítsu*
	tourinu	*toeritsu*		*touritsú*
hu	*hushi*	*hise*	*tsuse*	*húse*
se	*huze*			
goro	*garon*	*gorin*	*horon*	*góron*
ron	*godon*			*goroń*
oshi	*oshishi*	*oshira*	*ishiri*	*óshiri*
shiri	*osuri*	*omiri*	*woshiri*	*oshíri*

Note. In Test 3, the acute accent marks (/á/, /í/, /ú/, /é/, /ó/, and /ń/) denote the respective phonemes (/a/, /i/, /u/, /e/, /o/, and /n/) where vocal emphasis (pitch accent) was placed during pronunciation.

Test Procedure

We conducted a cue discrimination test to investigate the subjects' responses to the modified cues. When we conducted tests, each subject was kept alone at each study location, and the experimenter gave cues behind protected contact.

For all tests, we presented 10 cues to both subjects during each session. We provided them with modified cues as probe stimuli mixed with original cues (each session typically included seven original cues and three modified probe cues). Therefore, the trial counts reported in the subsequent sections refer strictly to the number of probe trials. We randomly determined the cue order using the RAND function in Microsoft Excel. We presented each cue only once and reinforced all subjects' responses using food rewards, including unexpected and passive responses (i.e., a non-differential reinforcement schedule, which is a standard procedure in probe testing to prevent the extinction of responses to novel stimuli and to eliminate potential learning effects during the test period; e.g., Herman et al., 1984). The types and amounts of food were randomized to rule out the influence of reinforcement values on performance, following the same procedure described in Sasaki, Kinoshita, et al. (2025). The experimenter wore sunglasses and a mask (which did not significantly muffle the acoustic properties of the vocal cues) and standardized his spatial positioning (stationing at a fixed distance of 50 cm for Hama and 1 m for Shunta directly in front of the subjects), posture (maintaining a crouching stance with hands held stationary in front of the body), and gaze (maintaining neutral eye contact with the subject's face) to eliminate the influence of unintentional cues.

Test 1

We conducted 18 sessions (yielding 54 probe trials) for Hama from December 25, 2022 to January 3, 2023, and 19 sessions (yielding 57 probe trials) for Shunta from December 26, 2022 to January 23, 2023. In the tests for Hama, each modified cue was presented four times during the test period, except for "*kei*," "*rei*," and "*shiri*" (three times due to procedural errors) and "*se*" (five times due to procedural errors). For Shunta, each modified cue was presented four times during the test period, except for "*naka*" (five times due to procedural errors). We conducted one to three sessions per day for each subject (Hama or Shunta).

Test 2

We conducted 59 sessions (yielding 179 probe trials) each for both sea lions: for Hama from January 18 to February 26, 2023, and for Shunta from January 25 to March 11, 2023. In Hama's tests, each modified cue was tested four times, except for "*omawaro*," "*tourita*," and "*onika*" (five times due to procedural errors). In Shunta's tests, each modified cue was tested four times, except for "*omawaro*," "*onika*," and "*hise*" (five times due to procedural errors). We conducted one to three sessions per day for each subject.

Test 3

We conducted 22 sessions (yielding 64 probe trials) for Hama from March 3 to March 19, 2023, and 21 sessions (yielding 64 probe trials) for Shunta from March 15 to March 31, 2023. In Hama's tests, each modified cue was tested four times. In Shunta's tests, each modified cue was tested four times. We conducted one to three sessions per day for each subject. Two additional probe trials for Hama were excluded because the cues were accidentally produced with the standard accent, resulting in 64 trials for analysis.

Data Analysis

All test sessions (Tests 1–3) were video-recorded using an iPad for subsequent behavioral analysis. Videos were recorded from the front or at a diagonal angle to capture both the trainer and the subject within the same frame at a frame rate of 30 fps. The resulting video data were used for both accuracy-based and reaction time-based analyses according to the following procedures.

In the analysis based on the accuracy rate, we used chi-square tests to evaluate the difference in the number of correct responses between original cues and modified vocal cues. We also used binomial tests to compare the accuracy rate with chance level and the baseline (i.e., the accuracy rate for "original" cues during the test period).

In the analysis based on the subjects' reaction time, we evaluated the difference in the reaction time between original and modified vocal cues by using two analyses, Welch's *t* test and Cohen's *d*, to confirm the results.

Analysis Based on Accuracy Rate

We coded the subjects' responses for both original and modified cues as a binomial value ("correct" or "incorrect") based on the recorded videos. (Note: "Original cues" refers to the fully intact, unmodified vocal cues the subjects had previously learned). When coding the modified cues, we categorized the response as "correct" if the subjects performed the conditioned behavior associated with the original cue (i.e., when the trainer gave the subjects "*kei*," the correct response was the behavior conditioned with the original "*keirei*" cue).

We evaluated the differences in the number of correct responses to the modified cues among the three test types using the chi-square test. Because there were three multiple comparisons of the number of correct responses among the tests, the significance level was adjusted using the Bonferroni method (i.e., .05/3 = .0167). Therefore, only cases in which the p value was less than .0167 were considered statistically significant for performance comparisons among the three tests. We also evaluated the difference in number of correct responses to modified cues between each subject for each test type using the chi-square test. These comparisons were conducted between the two conditions. A p value of less than .05 was considered statistically significant in the comparisons.

We conducted the following three comparisons using the chi-square test: In Test 1, we compared the difference in number of correct responses for modified cues between cases where only the first half of the cues were given and those where only the second half were given. In Test 2, we compared the difference in number of correct responses for modified cues between cases in which vowels were modified and those in which consonants were modified. We also compared the difference in the number of correct responses for modified cues between cases in which a phoneme in the first half was modified and those in which a phoneme in the second half was modified (we rejected 25 trials for each subject in which a phoneme in the middle mora was modified in the cues consisting of three moras: *onaka*, *oshiri*, and *goron*, from the data analysis).

The accuracy rate (%) for modified cues was calculated based on the number of trials for each subject in each condition. A binomial test was used to evaluate the number of correct and incorrect responses to the modified cues for each test type. We established two types of success probabilities, chance level and baseline accuracy (accuracy rate for original cues), to evaluate the subjects' performance.

To compare the accuracy rate for modified cues to the chance level, the probability of success was set at 14.3% (1/7) reflecting the seven types of human vocal cues (i.e., original cues: *keirei*, *onaka*, *omawari*, *touritsu*, *huse*, *goron*, and *oshiri*) that were used in each test. The significance threshold was determined based on the number of trials. Subjects' performance was considered significantly higher than chance if the accuracy rate reached or exceeded this threshold.

To compare the accuracy rate for modified cues with that of the baseline (the accuracy rate for original cues during each test), the probability of success was calculated based on the number of correct responses to the original cues in each test session. When the accuracy rate for modified cues was significantly lower than this baseline threshold, the performance was considered to have decreased significantly.

The specific significant thresholds both for chance level and for baseline are described in the Results section. All statistical tests were performed at a significance level of $\alpha = .05$. The binomial test was upper-tailed for comparison to the chance level and lower-tailed for comparison to the baseline.

Analysis Based on Reaction Time

We analyzed reaction times for correct responses to investigate whether there were qualitative differences between the "original cues" and the "modified cues." We hypothesized that if the modified cues violated the subjects' expectations of the acoustic structure, it would induce cognitive conflict or dissonance, leading to increased processing times and delayed reactions (e.g., Rossmeissl & Theios, 1982; see also Stroop effect paradigms where unexpected stimuli delay response). To confirm that these differences were not due to individual factors (e.g., age or sex), we analyzed the data for each individual independently.

We calculated the reaction time using the recorded video data (as described at the beginning of this section, using an iPad positioned to capture both the trainer and subject). We imported these videos into PowerDirector, a video editing software. Specifically, we counted the number of frames and converted the frame counts into milliseconds (ms) based on the frames per second (fps). Because the subjects' responses included many small initial movements, it was difficult to identify the exact onset of the subjects' responses. We then determined the onset of the response as the point at which the experimenter started to emit the vocal cues and determined the offset of the response as the point at which the subjects' responses were completed. Although this definition limits our ability to evaluate delays in response onset, we believe that comparing the time to completion allows us to detect differences in the quality of the subjects' reactions. We used the audio waveforms in PowerDirector to precisely identify the onset timing and identified the offset timing by analyzing the video data, frame-by-frame.

We excluded some action types and trial data from analysis because of procedural limitations described below.

- Visual obstruction: Trials in which the subject's actions overlapped with the cage frame, making it impossible to identify the offset.

- Acoustic noise: Trials in which background sound noise (i.e., other individuals' voices) was significant, making it impossible to identify the onset.

We compared the difference in the subject's reaction time between "original cues" and "modified cues" by action type because the time to completion differed among action types. Considering the limited sample size of available data for "modified cues" in each action type and the potential variance, we used two analyses to confirm the results. This ensured a more robust comparison between the original and modified conditions of the study. We used Welch's *t* test for action types in which the sample size of the modified cues was five or more. If the *p* value was less than .05, we judged that the reaction times of the original and modified cues were different. We also used Cohen's *d* to measure the effect size. For action types with fewer than five modified cues, we used only Cohen's *d*. When the *d* value was 0.5 (i.e., medium effect size) or more, we judged that the cue modification affected the subjects' responses. Cohen's *d* was calculated as the mean difference divided by the standard deviation of the original-cue condition.

Ethical Notes

The research protocols followed the ethical standards set by Kinosaki Marine World, an institution affiliated with the Japanese Association of Zoos and Aquariums (JAZA). All procedures complied with the JAZA's welfare regulations, as well as the 2020 guidelines for animal treatment in behavioral research established by the Animal Behavior Society and the Association for the Study of Animal Behaviour (ASAB/ABS, 2020). The facility's ethics committee approved this study on December 1, 2022. To guarantee optimal health, the Steller sea lions underwent daily monitoring and monthly veterinary inspections.

Results

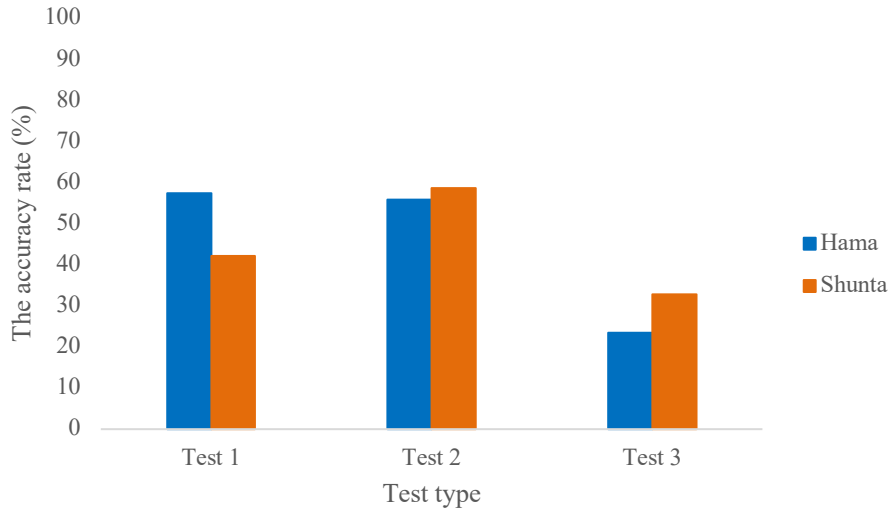
Comparison of the Accuracy Rate

Comparison of Both Subjects' Performance Between Test Types

Regarding comparisons between tests for Hama, her accuracy rate in Test 3 was significantly lower than those in both Test 1 ($\chi^2(1, N = 118) = 14.2$, $p < .001$; Figure 2) and Test 2 ($\chi^2(1, N = 243) = 19.89$, $p < .001$; Figure 2). In contrast, no significant difference was observed between Tests 1 and 2 ($\chi^2(1, N = 233) = 0.04$, $p = .84$; Figure 2). Similarly, for Shunta, his accuracy rate in Test 3 was significantly lower than that in Test 2 ($\chi^2(1, N = 243) = 12.6$, $p < .001$; Figure 2). However, no significant differences were observed in the accuracy rates between Tests 1 and 2 ($\chi^2(1, N = 236) = 4.78$, $p = .03$; Figure 2) and between Tests 1 and 3 ($\chi^2(1, N = 121) = 1.11$, $p = .29$; Figure 2), as neither reached the Bonferroni-adjusted significance level of .0167.

Figure 2

The Accuracy Rates of Each Subject for Modified Cues in Each Test Type



Comparisons of the Performance for Each Test Type in Each Subject

Chance level and baseline were calculated based on the actual number of trials in each test and the accuracy rate for original vocal cues during each test, respectively (Table 3). A binomial test showed that the accuracy rate for modified cues was significantly higher than chance ($p < .05$) but significantly lower than baseline ($p < .05$) for all test types for both subjects (Figures 3, 4, and 5).

Table 3

The Significance Thresholds Based on Both Chance Level and Baseline Accuracy for Each Subject in Each Test

Test	Subject	Chance Level (n)	Baseline Accuracy (*n*)
Test 1	Hama	24.1% (13/54)	90.7% (49/54)
	Shunta	24.6% (14/57)	89.5% (51/57)
Test 2	Hama	19.0% (34/179)	88.8% (159/179)
	Shunta	19.0% (34/179)	87.7% (157/179)
Test 3	Hama	23.4% (15/64)	78.1% (50/64)
	Shunta	23.4% (15/64)	82.8% (53/64)

Note. Values in parentheses represent the number of correct responses / total number of trials.

Figure 3

The Accuracy Rates of Each Subject for Original and Modified Cues in Test 1

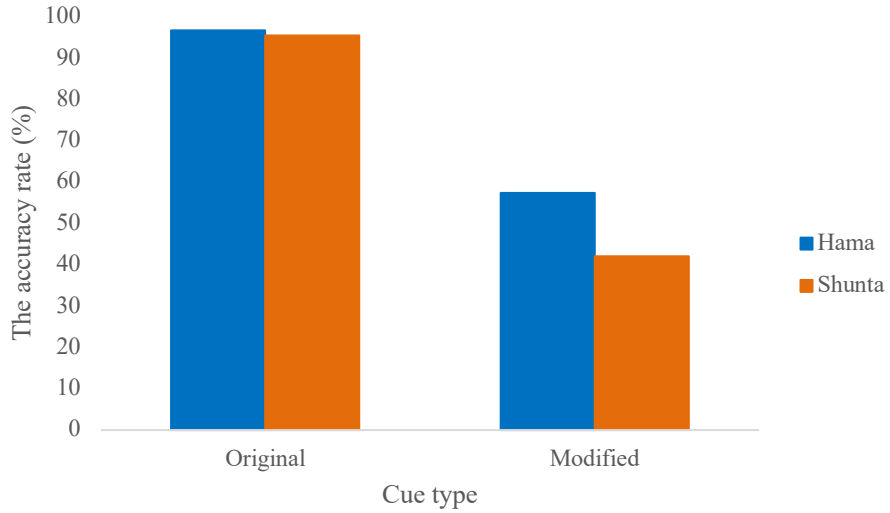


Figure 4

The Accuracy Rates of Each Subject for Original and Modified Cues in Test 2

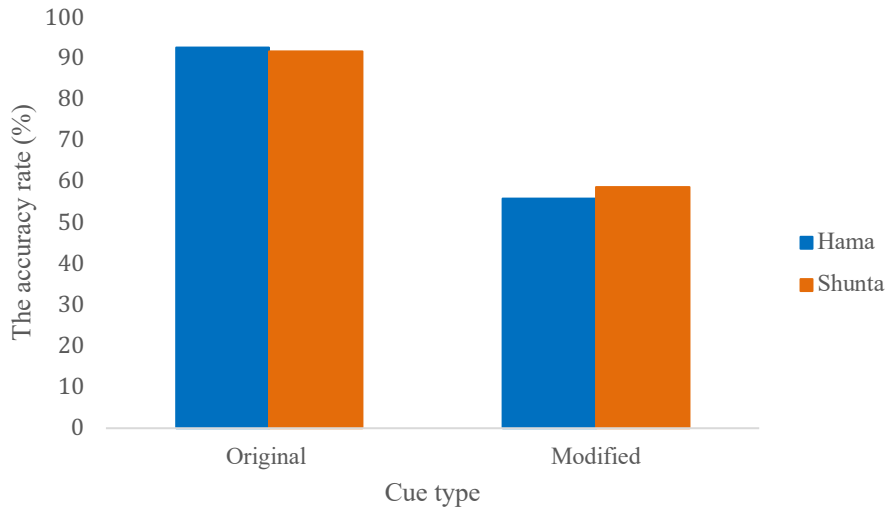
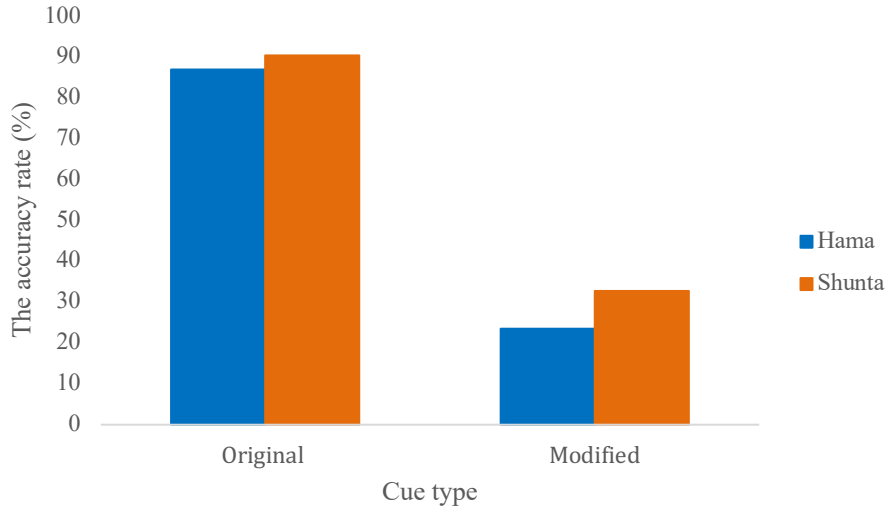


Figure 5

The Accuracy Rates of Each Subject for Original and Modified Cues in Test 3



Detailed Comparisons of the Subjects' Performance in Each Condition

Regarding the comparison between the presentation of only the first half of the vocal cue versus only the second half of the vocal cue, no significant difference in the accuracy rate was observed for either subject ($\chi^2(1, N = 54) = 1.9$, $p = .17$ for Hama; $\chi^2(1, N = 57) = 0.8$, $p = .36$ for Shunta). Similarly, when comparing the modification of a single phoneme in the first half and the second half of the vocal cue, there was no significant difference in performance for either subject ($\chi^2(1, N = 154) = 0.35$, $p = .55$ for Hama; $\chi^2(1, N = 154) = 3.4$, $p = .06$ for Shunta). Finally, we evaluated the effects of vowel versus consonant modification. For Hama, the accuracy rate was significantly higher when a vowel was modified than when a consonant was modified ($\chi^2(1, N = 179) = 5.4$, $p < .05$), whereas no such significant difference was found for Shunta ($\chi^2(1, N = 179) = 1.1$, $p = .30$).

Comparison of Performance Between Subjects

To account for different experimental conditions, we did not pool the overall data. Rather, the difference in the accuracy rates between the two subjects was compared independently within each of the three tests. The results showed no significant differences between the subjects in any of the three independent tests (Test 1: $\chi^2(1, N = 111) = 2.60$, $p = .11$; Test 2: $\chi^2(1, N = 358) = 0.29$, $p = .59$; Test 3: $\chi^2(1, N = 128) = 1.39$, $p = .24$).

When comparing performance in Test 1, Hama's accuracy rate was significantly higher than Shunta's accuracy rate where only the first half of the cue was given ($\chi^2(1, N = 56) = 4.6$, $p = .03$). Conversely, there was no significant difference between the subjects when only the second half of the cue was given ($\chi^2(1, N = 55) = 0.02$, $p = .89$).

When comparing the performance in Test 2, there was no significant difference in the accuracy rate between the subjects for any of the modification types: modified first half of the cue ($\chi^2(1, N = 162) = 1.65$, $p = .20$), modified second half of the cue ($\chi^2(1, N = 146) = 1.36$, $p = .24$), modified vowel ($\chi^2(1, N = 180) = 1.08$, $p = .30$), and modified consonant ($\chi^2(1, N = 178) = 0.07$, $p = .79$). For Test 3, we did not make a condition-level comparison, because accent applies to the whole cue and cannot be divided into first/second-half or vowel/consonant sub-conditions. The overall comparison for Test 3, reported above, showed no significant difference between subjects ($\chi^2(1, N = 128) = 1.39$, $p = .24$).

Comparison of the Reaction Times

Overall Comparison

Regarding the overall results for Hama, we analyzed six of seven action types (excluding "*touritsu*" due to data limitations, i.e., visual obstruction and acoustic noise). The reaction time for "*onaka*" was significantly longer with modified cues compared to original cues ($p < .05$; Table 4), whereas the reaction time for "*keirei*" was significantly shorter with modified cues ($p < .05$). No significant differences were observed for the other four action types (i.e., "*omawari,*" "*huse,*" "*goron,*" and "*oshiri*"; Table 4). Substantial effects (Cohen's $d \geq 0.5$) were observed for "*keirei,*" "*onaka,*" "*huse,*" and "*oshiri*" (Table 4). Specifically, reaction times were prolonged in response to modified vocal cues for "*onaka*," "*huse*," and "*oshiri*," but shortened for "*keirei*" (Table 4).

For Shunta, all seven action types were analyzed. Reaction times for "*keirei*" were significantly longer with modified vocal cues compared to original vocal cues ($p < .05$; Table 4). No significant differences were found for the other six action types. Substantial effects (Cohen's $d \geq 0.5$) were observed for "*keirei*" and "*onaka*," both showing prolonged reaction times to modified cues (Table 4).

Table 4

Overall Results of Reaction Time Analysis for the Subjects

Hama							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	96	846.0	78.4	2.79	< .05*	0.80*
	modified	15	783.2	78.5			
onaka	original	91	790.2	102.0	-4.04	< .01*	1.76*
	modified	27	969.2	219.3			
omawari	original	86	2374.8	141.5	-1.25	.21	0.25
	modified	33	2410.0	132.6			
touritsu	original	10	2286.9	372.8			
	modified	0					
huse	original	92	610.9	77.9	-1.59	.14	0.67*
	modified	12	662.8	104.8			
goron	original	88	2302.5	258.6	-0.40	.70	0.14
	modified	13	2337.9	292.8			
oshiri	original	80	2241.1	168.5	-1.47	.15	0.52*
	modified	21	2328.9	253.1			
Shunta							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	96	1322.8	211.5	-3.46	< .01*	1.38*
	modified	25	1614.4	398.7			
onaka	original	98	1002.1	271.4	-1.98	.05	0.50*
	modified	31	1138.0	344.3			
omawari	original	81	4376.0	639.0	-1.56	.14	0.49
	modified	17	4686.0	746.5			
touritsu	original	85	2049.5	290.3	-0.55	.59	0.11
	modified	22	2082.0	228.4			
huse	original	85	854.5	128.1	-1.36	.19	0.40
	modified	13	906.2	122.5			
goron	original	85	2464.5	259.3	1.22	.24	0.47
	modified	13	2343.0	332.0			
oshiri	original	93	2666.3	277.5	-0.11	.91	0.02
	modified	24	2671.6	194.9			

Note. * indicates that the *p* value was less than .05, or the *d* value was equal to or greater than 0.5.

Comparison of Reaction Time: Test 1

For Hama, six action types were analyzed (excluding "*touritsu*"). Three action types were subjected to Welch's *t* test (i.e., "*onaka,*" "*omawari,*" and "*goron*"), and the reaction time for "*onaka*" was significantly longer with modified cues ($p < .05$; Table 5). While no significant differences were observed for the other two actions, substantial effects (Cohen's $d \geq 0.5$) were observed for "*onaka,*" "*oshiri,*" and "*keirei*" (Table 5). Specifically, reaction time was prolonged for "*onaka*," but shortened for "*keirei*" and "*oshiri*" with modified vocal cues (Table 5).

For Shunta, five of the seven action types were analyzed (excluding "*huse*" and "*touritsu*"). Three action types were subjected to Welch's *t* test (i.e., "*keirei,*" "*onaka,*" and "*oshiri*"), showing no significant difference for any of them (Table 5). By contrast, substantial effects (Cohen's $d \geq 0.5$) were observed for "*omawari,*" showing shortened reaction times under modified vocal cues (Table 5).

Table 5

The Results of Reaction Times for the Subjects in Test 1

Hama							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	17	889.1	78.8			2.07*
	modified	1	726.0	0.0			
onaka	original	16	835.3	61.4	-2.37	< .05*	2.45*
	modified	8	985.9	162.9			
omawari	original	18	2288.0	136.1	0.10	.93	0.03
	modified	5	2283.6	64.0			
touritsu	original						
	modified						
huse	original	17	663.9	113.7			0.13
	modified	3	649.0	31.1			
goron	original	18	2273.3	250.6	0.64	.54	0.30
	modified	5	2197.8	203.0			
oshiri	original	13	2211.0	143.6			0.92*
	modified	2	2079.0	0.0			
Shunta							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	18	1191.7	95.8	0.41	.70	0.24
	modified	5	1168.2	105.6			
onaka	original	18	995.5	238.5	-0.42	.68	0.24
	modified	8	1051.9	321.8			
omawari	original	19	4097.2	482.9			1.65*
	modified	1	3300.0	0.0			
touritsu	original	18	2264.2	362.9			
	modified	0					
huse	original	15	891.0	92.6			
	modified	0					
goron	original	14	2590.5	156.7			0.18
	modified	3	2618.0	276.5			
oshiri	original	19	2728.6	210.9	1.03	.32	0.37
	modified	6	2651.0	125.9			

Note. * indicates that the *p* value was less than .05, or the *d* value was equal to or greater than 0.5.

Comparison of Reaction Time: Test 2

For Hama, six action types (excluding "*touritsu*") were analyzed, and all six action types were subjected to Welch's *t* test. The reaction time for "*onaka*" was significantly longer with modified cues ($p < .05$; Table 6), whereas no significant differences were observed for the other five action types (Table 6). Substantial effects (Cohen's $d \geq 0.5$) were observed for "*onaka*," "*huse*," and "*goron,*" all showing prolonged reaction times to modified vocal cues (Table 6).

For Shunta, all seven action types were subjected to Welch's *t* test, and reaction times for "*keirei*" and "*onaka*" were significantly longer with modified cues ($p < .05$; Table 6). Conversely, no significant differences were observed in the other five action types (Table 6). Substantial effects (Cohen's $d \geq 0.5$) were observed for "*keirei,*" "*onaka,*" "*omawari*," "*touritsu,*" and "*goron*," showing prolonged reaction times for "*keirei*," "*onaka*," "*omawari*," and "*touritsu*," but shortened for "*goron*" to modified vocal cues (Table 6).

Table 6

The Results of Reaction Times for the Subjects in Test 2

Hama							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	57	829.1	71.8	0.88	.39	0.21
	modified	12	814.0	47.3			
onaka	original	54	763.9	97.9	-3.22	< .01*	1.87*
	modified	17	947.3	220.9			
omawari	original	51	2376.7	132.4	-1.63	.11	0.42
	modified	24	2432.4	137.5			
touritsu	original						
	modified						
huse	original	55	594.6	63.4	-1.55	.17	1.33*
	modified	7	678.9	131.8			
goron	original	51	2281.5	218.0	-1.20	.26	0.66*
	modified	8	2425.5	306.0			
oshiri	original	48	2251.6	177.0	-0.93	.37	0.32
	modified	14	2307.6	197.8			
Shunta							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	57	1384.8	231.9	-3.45	< .01*	1.43*
	modified	18	1716.0	374.0			
onaka	original	58	1024.7	277.9	-2.39	< .05*	0.79*
	modified	18	1243.0	345.4			
omawari	original	46	4484.4	542.5	-1.80	.09	0.57*
	modified	12	4793.3	501.6			
touritsu	original	51	2014.3	250.4	-2.03	.05	0.50*
	modified	18	2139.5	208.2			
huse	original	49	831.7	120.6	-1.47	.16	0.47
	modified	11	888.0	107.6			
goron	original	52	2454.1	298.1	1.66	.13	0.72*
	modified	8	2239.9	323.5			
oshiri	original	54	2600.9	205.7	-1.12	.27	0.33
	modified	17	2669.1	215.0			

Note. * indicates that the *p* value was less than .05, or the *d* value was equal to or greater than 0.5.

Comparison of Reaction Time: Test 3

For Hama, five action types were analyzed (excluding "*touritsu*" and "*goron*"). Welch's *t* test was used for "*oshiri*," but no significant difference in reaction times was observed (Table 7). Substantial effects (Cohen's $d \geq 0.5$) were observed for "*keirei,"* "*onaka,*" and "*oshiri,*" showing prolonged reaction times for "*onaka*" and "*oshiri*," but shortened for "*keirei*" to modified vocal cues (Table 7).

For Shunta, while all seven action types were analyzed, only "*onaka*" was subjected to Welch's *t* test, with no significant difference (Table 7). Substantial effects (Cohen's $d \geq 0.5$) were observed for "*keirei*," "*touritsu*," and "*huse*," showing prolonged reaction times for "*keirei*" and "*huse*," but shortened for "*touritsu*" to modified vocal cues (Table 7).

Table 7

The Results of Reaction Times for the Subjects in Test 3

Hama							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	22	856.5	79.9			2.87*
	modified	2	627.0	33.0			
onaka	original	21	823.4	114.5			2.32*
	modified	2	1089.0	330.0			
omawari	original	17	2461.4	116.6			0.24
	modified	4	2433.8	54.1			
touritsu	original	0					
	modified	0					
huse	original	20	610.5	52.7			0.31
	modified	2	627.0	33.0			
goron	original	19	2386.4	337.1			
	modified	0					
oshiri	original	19	2235.3	159.5	-1.50	.21	1.59*
	modified	5	2488.2	328.4			
Shunta							
Action type	Cue type	Number of data	Average (ms)	Standard Deviation	Welch's *t* value	Welch's *p* value	Cohen's *d* value
keirei	original	21	1266.6	149.5			3.67*
	modified	2	1815.0	264.0			
onaka	original	22	948.0	271.2	0.48	.64	0.19
	modified	5	897.6	173.9			
omawari	original	16	4395.2	904.4			0.35
	modified	4	4710.8	1048.1			
touritsu	original	16	1920.2	165.1			0.59*
	modified	4	1823.3	100.0			
huse	original	21	881.6	153.1			0.82*
	modified	2	1006.5	148.5			
goron	original	19	2400.3	153.2			0.37
	modified	2	2343.0	165.0			
oshiri	original	20	2783.6	415.4			0.13
	modified	1	2838.0	0.0			

Note. * indicates that the *p* value was less than .05, or the *d* value was equal to or greater than 0.5.

Discussion

We investigated how two Steller sea lions discriminate vocal cues by examining their behavioral responses to modified acoustic cues. In both subjects, although the accuracy rates for the modified vocal cues were significantly higher than chance, they were significantly lower than the baseline (i.e., the accuracy rates for the original vocal cues during the test period) for all test types. The lowest accuracy rates were observed in Test 3 compared with those in the other tests. A detailed analysis suggests that the subjects did not possess a consistent discrimination strategy regarding specific positions within the vocal cues. Instead, they may have relied on individual markers tailored to each vocal cue. With respect to acoustic elements (vowels and consonants), however, while no common strategy was observed in Shunta, Hama's performance was significantly affected by vowel modifications. This suggests that Hama may use a vowel-dominant discrimination process. An analysis of the reaction times revealed that cue modification affected the subjects' responses to modified vocal cues even when they responded correctly. This suggests that Steller sea lions possess the cognitive flexibility to associate modified vocal cues with original vocal cues, even when the acoustic properties of the vocal cues deviate from familiar ones.

The accuracy rates for modified vocal cues were significantly lower than the baselines, suggesting that vocal cue modification affected the subjects' accuracy across all test types ($p < .05$). This indicates that vocal cue discrimination was influenced by all three factors: phonological information content, constituent phonemes, and accent. This result is similar to a previous study on dogs. Fukuzawa et al. (2005) reported that phonemic changes in vocal cues affected subjects' discrimination performance. Beyond this similarity, our results revealed that acoustic changes other than phonemic modifications significantly affect cue recognition. Among the modification types, changing the accent had the greatest impact on discrimination performance. In particular, for Hama, the accuracy rates for modified cues in Test 3 were significantly lower than those for the other two test types ($p < .05$), suggesting that accent plays an important role in vocal cue discrimination in this species.

Although vocal cue modification had a significant influence on the subjects' performance, their accuracy rates remained significantly higher than the chance level ($p < .05$). This suggests that the vocal cue discrimination performance was partly maintained, even when the vocal cues were modified. The subjects responded correctly (i.e., performed the actions associated with the original vocal cues) most frequently in Test 2. This suggests that vocal cue discrimination is relatively robust against a single phoneme modification in vocal cues.

In a detailed comparison, there were no significant differences between changing a phoneme in the first and second halves of the vocal cues. This suggests that there was no consistent strategy for focusing on a specific position in all vocal cues. A similar tendency was observed in the detailed comparison of Test 1. The subjects' performance did not differ between the two types of divided vocal cues (i.e., first halves vs. second halves). This supports the interpretation that the subjects' performance did not rely on universal phonological features, such as favoring a specific region of the vocal cue. In other words, they developed a strategy to obtain unique discriminative points for each vocal cue. Fukuzawa et al. (2005) reported that a dog's response to a vocal cue is not dependent on a specific word phoneme and that each dog may possess a unique criterion with which it may interpret the presented vocal cue. This comparison suggests that common strategies are observed in cue discrimination, even between species with distinct evolutionary backgrounds, such as domestic dogs and Steller sea lions.

Conversely, a comparison of the performance of vowel and consonant modification in Test 2 revealed different tendencies for each subject. Although Shunta's performance did not differ between the two modification types, Hama's performance was significantly lower when a vowel within the vocal cues was modified. This result suggests that the response decision process was not strictly dependent on a specific phonological category in Shunta, whereas vowels significantly impacted vocal cue discrimination in Hama's performance.

In human studies of word discrimination using rhyme priming, word onset has a special status in lexical access to spoken words (Marslen-Wilson & Zwitserlood, 1989). Nazzi and Cutler (2019) reported that lexical processing in human adulthood is more strongly associated with consonant processing than with vowel processing.

The importance of consonants in word recognition has also been documented in several other human studies (Kambara & Umemura, 2021; Lee et al., 2001; Nazzi et al., 2016; Soares et al., 2014). Comparing these findings with our results suggests that vocal cue discrimination in Steller sea lions may be supported by establishing a unique discrimination point (or specific acoustic markers) tailored to each of the vocal cues. In other words, an individual vocal cue discrimination process, in contrast to the systematic nature of human word discrimination, may reduce the cognitive load required to learn vocal cues.

These differences may be related to the number of vocal cues (or words) learned by each species. The lexical hypothesis proposes that an increase in vocabulary size leads to the emergence of consonant bias. This suggests that to discriminate among a larger number of lexical items, listeners must rely on consonants as specific and informative markers (Keidel et al., 2007). Conversely, for Steller sea lions, who learned far fewer vocal cues than the typical human lexicon, it may be sufficient to discriminate cues without depending on specific phonemic units (i.e., vowels and consonants). Instead, they may use specific phoneme sets or the overall acoustic profile of cues as primary markers. The significant decline in discrimination performance in Test 3 (accent modification) supports this suggestion.

Hama showed vowel-dominant discrimination in Test 2, suggesting that Hama, who has acquired a larger number of vocal cues than Shunta, may need to discriminate between vocal cues using more constrained phonological markers, similar to the consonant bias observed in humans. Benavides-Varela et al. (2025) raised the possibility that animals tend to focus more on vowel segments over consonant segments while recognizing acoustic sequences, based on previous data from various species (e.g., cotton-top tamarin, *Saguinus oedipus*, Newport et al., 2004; rat, Bouchon & Toro, 2019; dog, Mallikarjun et al., 2021). A similar vowel bias has been reported in human infants, who possess a smaller lexicon than adult humans (Poltrock & Nazzi, 2015). Benavides-Varela et al. (2025) suggested that vowel bias might be linked to the acoustic saliency of vowels. They further proposed that comparative studies could explore whether there are any conditions under which nonhuman animals might also switch from a vowel to a consonant bias, and that studies with animals exposed to language throughout their lives might be relevant to study the role of experience in such a possible change of focus from vowels to consonants. We will investigate whether Hama's vocal cue discrimination shifts toward a consonant bias as her learned repertoire increases.

Focusing on the correct response to modified vocal cues, we examined whether the subjects recognized the differences in the modified vocal cues by conducting a retrospective analysis of reaction times to examine the influence of vocal cue modifications on subjects' responses. Although there were several limitations, as mentioned in the "Data Analysis" section, due to the post-hoc and exploratory nature of the reaction time analysis, significant effects were observed in the response of each subject.

Specifically, either Welch's t value or Cohen's d value indicated that cue modification affected the reaction time for at least one vocal cue type across all test types. This result suggests that the subjects responded correctly, even though they recognized that the vocal cues differed from the original ones. Vocal cue modification had the largest influence in Test 2 for both subjects, suggesting that their high performance in Test 2 was supported by flexible cognitive processing.

Similar effects on reaction time have been reported in human studies using pseudowords. Rossmeissl and Theios (1982) revealed that reaction times in discrimination and pronunciation tasks were longer for pseudowords than for real words. Forster and Chambers (1973) also reported that both discrimination and naming times were longer for pseudowords than for words. In addition, because pseudowords are conditioned to the sensory and affective meanings of real words (Ando & Kambara, 2023; Wang et al., 2025), the reaction times to recognize pseudowords (i.e., new words in a native or second language) can be decreased by conditioning or associative learning of words and their referents (Liu et al., 2021; Obinna et al., 2025). Although these previous studies, excluding Obinna et al. (2025), used written real words or pseudowords, their results support the possibility that recognizing vocal cue modifications affected the subjects' reaction times in the current study.

The effects were not uniform across all vocal cue types for each test type. For Hama, no significant influence was observed for "*omawari*" in any of the tests. In contrast, significant differences were observed for "*onaka*" in all three test types, for "*keirei*" and "*oshiri*" in two test types, and for "*huse*" and "*goron*" in one test type. A similar variation in susceptibility to vocal cue modifications was observed in Shunta. While significant influences were observed for all vocal cue types except for "*oshiri*" at least once, "*keirei*," "*omawari*," and "*touritsu*" showed significant differences across multiple test types. These tendencies suggest that the susceptibility to vocal cue modification varied depending on the vocal cue types.

Although vocal cue modification generally resulted in increased reaction times, there were cases in which reaction times were shortened for certain vocal cue types in both subjects (i.e., "*keirei*" in Test 1 and Test 3, and "*oshiri*" in Test 1 for Hama; "*omawari*" in Test 1, "*goron*" in Test 2, and "*touritsu*" in Test 3 for Shunta). This shortening effect contradicts the delay effect reported in similar discrimination tasks of human studies (Forster & Chambers, 1973; Rossmeissl & Theios, 1982).

One possible explanation for this shortening effect is that the subjects attempted to offset the delay in initiation by accelerating their movement speed. Because we defined the onset of the response as the point at which the experimenter began to emit the vocal cue and the offset as the completion of the subjects' response, any delay in initiating the movement might be masked by an increase in execution speed. If the response delay is not large enough to exceed this offsetting effect, the total reaction time will show no numerical difference. When the delay is minimal, the total reaction time may appear to be shorter because of subsequent acceleration. Therefore, the current results may underestimate the actual impact of these vocal cue modifications. Although an increase in movement speed was observed as a visual trend during the analysis, it could not be quantified as objective numerical data because of technical limitations. Identifying the subjects' initial movements in future studies would provide stronger evidence for the delay and acceleration effects.

The cognitive flexibility observed in this study may reflect the characteristics of social communication in this species. Gentry (1974) reported that Steller sea lion pups develop their behavioral repertoire through interactions with other individuals, which include many types of "uncodable" actions. This suggests that they communicate under conditions involving ambiguous stimuli. This biological characteristic may enable them to associate modified vocal cues with the original cue, even when minor modifications are present.

While the current study provides the first preliminary findings on how vocal cue modification affects behavioral responses in pinnipeds, two large limitations remain for future studies to investigate their responses in detail. First, future studies should use target vocal cues with the same number of phonemes to allow for more effective comparisons under consistent conditions. Second, more precise methods for analyzing reaction times are required in future studies. The camera angle and method of identifying the subjects' initial movements are crucial for a deeper discussion. By addressing these limitations, we hope to further develop research on vocal cue discrimination and associative learning in Steller sea lions.

Acknowledgements

We are grateful to Dr. Akane Nagano of the RIKEN Center for Brain Science Laboratory and Dr. Hyangsun Chin of Kobe Women's University for their valuable suggestions for our research. We would like to thank the staff of Kinosaki Marine World for their support of this study. T.K. was supported by a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science (JSPS). We would like to thank our Steller sea lions (Shunta, Koharu, Konoha, Akari, and Hama), who provided us with great energy and motivation for our daily work.

References

Ando, M., & Kambara, T. (2023). Japanese written pseudowords can be conditioned to Japanese spoken words with positive, negative, and active emotions. *Cognitive Processing, 24*(3), 387–413. https://doi.org/10.1007/s10339-023-01138-0

Arkwright, T., Malassis, R., Carter, T., & Delfour, F. (2016). California sea lions (*Zalophus californianus*) can follow human finger points and glances. *International Journal of Comparative Psychology, 29,* Article 25366. https://doi.org/10.46867/ijcp.2016.29.00.02

ASAB/ABS. (2020). Guidelines for the treatment of animals in behavioural research and teaching (2020). *Animal Behaviour*, *159*, i–xi. https://doi.org/10.1016/j.anbehav.2019.11.002

Benavides-Varela, S., Bouchon, C., Hochmann, J., & Toro, J. (2025). A focus on vowels in young human infants and animals: Early sensitivities in language processing. *Journal of Language Evolution*, *10*(1), 1–7. https://doi.org/10.1093/jole/lzaf005

Bouchon, C., & Toro, J. M. (2019). Is the consonant bias specifically human? Long-Evans rats encode vowels better than consonants in words. *Animal Cognition, 22*(5), 839–850. https://doi.org/10.1007/s10071-019-01280-3

Charrier, I., Pitcher, B. J., & Harcourt, R. G. (2009). Vocal recognition of mothers by Australian sea lion pups: Individual signature and environmental constraints. *Animal Behaviour, 78*(5), 1127–1134. https://doi.org/10.1016/j.anbehav.2009.07.032

Charrier, I., Pitcher, B. J., & Harcourt, R. G. (2022). Mother–pup recognition mechanisms in Australian sea lion (*Neophoca cinerea*) using uni- and multi-modal approaches. *Animal Cognition, 25*, 1019–1028. https://doi.org/10.1007/s10071-022-01641-5

Cook, P., Reichmuth, C., & Hanke, F. D. (2021). The mind of a sea lion. In C. Campagna & R. Harcourt (Eds.), Ethology and behavioral ecology of otariids and the odobenid (pp. 323–345). Springer. https://doi.org/10.1007/978-3-030-59184-7_17

Endo, S., Kawaguchi, N., Shimizu, Y., Imagawa, A., Suzuki, T., Ashikari, H., Wakai, Y., & Murayama, T. (2020). Preliminary study of discrimination of human vocal cues in walrus (*Odobenus rosmarus divergens*). *International Journal of Comparative Psychology*, *33*(1), 1–12. https://doi.org/10.46867/ijcp.2020.33.00.03

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Fukuzawa, M., Mills, D. S., & Cooper, J. J. (2005). The effect of human cue phonetic characteristics on auditory cognition in dogs *(Canis familiaris)*. *Journal of Comparative Psychology, 119*(1), 117–120. https://doi.org/10.1037/0735-7036.119.1.117

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee: A standardized system of gestures provides a means of two-way communication with a chimpanzee. *Science*, *165*(3894), 664–672. https://doi.org/10.1126/science.165.3894.664

Gentry, R. L. (1974). The development of social behavior through play in the Steller sea lion. *American Zoologist*, *14*(1), 391–403. https://doi.org/10.1093/icb/14.1.391

Griebel, U., & Oller, D. K. (2012). Vocabulary learning in a Yorkshire terrier: Slow mapping of spoken words. *PLOS ONE*, *7*, e30182. https://doi.org/10.1371/journal.pone.0030182

Herman, L. M. (2010). What laboratory research has told us about dolphin cognition. *International Journal of Comparative Psychology*, *23*(3), 310–330. https://doi.org/10.46867/ijcp.2010.23.03.07

Herman, L. M. (2013). Cognition and language competencies of bottlenosed dolphins. In R. J. Schusterman, J. A. Thomas, & F. G. Wood (Eds.), Dolphin cognition and behavior: A comparative approach (pp. 221–252). Psychology Press. https://doi.org/10.4324/9780203767689

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, *16*(2), 129–219. https://doi.org/10.1016/0010-0277(84)90003-9

Kambara, T., & Umemura, T. (2021). The relationships between initial consonants in Japanese sound symbolic words and familiarity, multi-sensory imageability, emotional valence, and arousal. *Journal of Psycholinguistic Research, 50*(4), 831–842. https://doi.org/10.1007/s10936-020-09749-w

Kastak, D., & Schusterman, R. J. (1994). Transfer of visual identity matching-to-sample in two California sea lions (*Zalophus californianus*). *Animal Learning & Behavior, 22*, 427–435. https://doi.org/10.3758/BF03209162

Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Does grammar constrain statistical learning?: Commentary on Bonatti, Peña, Nespor, and Mehler (2005). *Psychological Science, 18*(10), 922–923. https://doi.org/10.1111/j.1467-9280.2007.02001.x

Kuczaj, S. A., II. (2013). Are conversations between dolphins and humans possible? *International Journal of Comparative Psychology, 26*(2), 114–123. https://doi.org/10.46867/ijcp.2013.26.02.05
Lee, H.-W., Rayner, K., & Pollatsek, A. (2001). The relative contribution of consonants and vowels to word identification during reading. *Journal of Memory and Language, 44*(2), 189–205. https://doi.org/10.1006/jmla.2000.2725
Lindemann-Biolsi, K. L., & Reichmuth, C. (2014). Cross-modal transitivity in a California sea lion (*Zalophus californianus*). *Animal Cognition, 17*(4), 879–890. https://doi.org/10.1007/s10071-013-0721-0
Liu, X., Horinouchi, H., Yang, Y., Yan, Y., Ando, M., Obinna, U. J., Namba, S., & Kambara, T. (2021). Pictorial referents facilitate recognition and retrieval speeds of associations between novel words in a second language (L2) and referents. *Frontiers in Communication, 6*, Article 605009. https://doi.org/10.3389/fcomm.2021.605009
Mallikarjun, A., Shroads, E., & Newman, R. S. (2021). The role of linguistic experience in the development of the consonant bias. *Animal Cognition*, *24*, 419–431. https://doi.org/10.1007/s10071-020-01436-6
Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576–585. https://doi.org/10.1037/0096-1523.15.3.576
Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics, 5*, 25–47. https://doi.org/10.1146/annurev-linguistics-011718-011919
Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science*, *25*(4), 291–296. https://doi.org/10.1177/0963721416655786
Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. https://doi.org/10.1016/j.cogpsych.2003.12.002
Obinna, U., Kabir, R. S., Yoshio, M., Yanamoto, D., & Kambara, T. (2025). Effects of pictorial referents on the associative learning of spoken word forms in a Nigerian language: An experimental investigation with L1 Japanese users. *Language, Cognition and Neuroscience, 40*(9), 1173–1185. https://doi.org/10.1080/23273798.2025.2509548
Pack, A. A., & Herman, L. M. (2004). Bottlenosed dolphins (*Tursiops truncatus*) comprehend the referent of both static and dynamic human gazing and pointing in an object-choice task. *Journal of Comparative Psychology, 118*(2), 160–171. https://doi.org/10.1037/0735-7036.118.2.160
Patterson, F. G. P., & Cohn, R. H. (1990). Language acquisition by a lowland gorilla: Koko's first ten years of vocabulary development. *Word*, *41*(2), 97–143. https://doi.org/10.1080/00437956.1990.11435816
Pepperberg, I. M. (1981). Functional vocalizations by an African grey parrot (*Psittacus erithacus*). *Zeitschrift für Tierpsychologie*, *55*(2), 139–160. https://doi.org/10.1111/j.1439-0310.1981.tb01265.x
Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*(2), 184–195. https://doi.org/10.1016/j.beproc.2010.11.007
Poltrock, S., & Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, *131*, 135–148. https://doi.org/10.1016/j.jecp.2014.11.011
Rosenthal, G. G., & Ryan, M. J. (2000). Visual and acoustic communication in non-human animals: A comparison. *Journal of Biosciences*, 25, 285–290. https://doi.org/10.1007/BF02703937
Rossmeissl, P. G., & Theios, J. (1982). Identification and pronunciation effects in a verbal reaction time task for words, pseudowords, and letters. *Memory & Cognition, 10*, 443–450. https://doi.org/10.3758/BF03197646
Saito, A., Shinozuka, K., Ito, Y., & Hasegawa, T. (2019). Domestic cats (*Felis catus*) discriminate their names from other words. *Scientific Reports, 9*, Article 5394. https://doi.org/10.1038/s41598-019-40616-4
Sasaki, M., Chiba, M., Ito, E., Tsutsumi, K., Ito, K., & Sunobe, T. (2022). Human vocal cues verify audio discrimination ability in the Steller sea lion *Eumetopias jubatus*. *International Journal of Comparative Psychology*, *35*(1), 1–11. https://doi.org/10.46867/ijcp.2022.35.5615
Sasaki, M., & Kambara, T. (2024). A case study of spontaneous category formation and behavioral expression in a language-trained Steller sea lion *Eumetopias jubatus*. *International Journal of Comparative Psychology*, *37*, 1–14. https://doi.org/10.46867/ijcp.6592
Sasaki, M., Kinoshita, H., & Konno, A. (2025). Do as I do imitation in a Steller sea lion *Eumetopias jubatus*. *Animal Cognition, 28*, Article 49. https://doi.org/10.1007/s10071-025-01971-0
Sasaki, M., Tsutsumi, K., Kinoshita, H., Nishijima, M., Matsumura, C., Toyota, A., & Kambara, T. (2025). A case study of associations between human visual-vocal cues and behaviors in a lactating Steller sea lion pup (*Eumetopias jubatus*). *International Journal of Comparative Psychology, 38*, Article 41525. https://doi.org/10.46867/IJCP.41525
Schakner, Z. A., Götz, T., Janik, V. M., & Blumstein, D. T. (2017). Can fear conditioning repel California sea lions from fishing activities? *Animal Conservation*, *20*(5), 425–432. https://doi.org/10.1111/acv.12329

Schusterman, R. J., & Kastak, D. (1998). Functional equivalence in a California sea lion: Relevance to animal social and communicative interactions. *Animal Behaviour, 55*(5), 1087–1095. https://doi.org/10.1006/anbe.1997.0654

Schusterman, R. J., & Krieger, K. (1984). California sea lions are capable of semantic comprehension. *Psychological Record*, *34*(1), 3–23. https://doi.org/10.1007/BF03394849

Schusterman, R. J., Kastak, C. R., & Kastak, D. (2002). The cognitive sea lion: Meaning and memory in the laboratory and in nature. In M. Bekoff, C. Allen, & G. M. Burghardt (Eds.), *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 217–228). MIT Press.

Schusterman, R. J., Reichmuth Kastak, C., & Kastak, D. (2003). Equivalence classification as an approach to social knowledge: From sea lions to simians. In F. B. M. de Waal & P. L. Tyack (Eds.), *Animal social complexity: Intelligence, culture, and individualized societies* (pp. 179–206). Harvard University Press.

Schusterman, R. J., & Van Parijs, S. M. (2003). Pinniped vocal communication: An introduction. *Aquatic Mammals*, *29*, 177–180. https://www.aquaticmammalsjournal.org/index.php

Soares, A. P., Perea, M., & Comesaña, M. (2014). Tracking the emergence of the consonant bias in visual-word recognition: Evidence with developing readers. *PLOS ONE, 9*(2), Article e88580. https://doi.org/10.1371/journal.pone.0088580

Takagi, S., Saito, A., Arahori, M., Chijiiwa, H., Koyasu, H., Nagasawa, M., Kikusui, T., Fujita, K., & Kuroshima, H. (2022). Cats learn the names of their friend cats in their daily lives. *Scientific Reports, 12*, Article 6155. https://doi.org/10.1038/s41598-022-10261-5

Tschudin, A., Call, J., Dunbar, R. I. M., Harris, G., & van der Elst, C. (2001). Comprehension of signs by dolphins (*Tursiops truncatus*). *Journal of Comparative Psychology, 115*(1), 100–105. https://doi.org/10.1037/0735-7036.115.1.100

Wang, N., Yoshio, M., & Kambara, T. (2025). Evaluative conditioning of sweet meaning using sweet and non-sweet food names as unconditioned stimuli and new words as conditioned stimuli. *Cognition and Emotion*. Advance online publication. https://doi.org/10.1080/02699931.2025.2566306

Wilson, S. C. (2024). The role of play in the social development of grey seal (*Halichoerus grypus*) pups with comparative notes on the harbour seal (*Phoca vitulina*). *Animals*, *14*(14), Article 2086. https://doi.org/10.3390/ani14142086

Financial conflict of interest: No stated conflicts.
Conflict of interest: No stated conflicts.

Submitted: *May 6th, 2026*
Resubmitted: *July 2nd, 2026*
Accepted: *July 14th, 2026*