Vocal recognition in primate:
Comparison between Japanese macaques and humans

Takafumi Furuyama

Doshisha University
Graduate School of Life and Medical Sciences

2016

Doshisha University

Graduate School of Life and Medical Sciences

Abstract

Vocal recognition in primate: Comparison between Japanese macaques and humans

by Takafumi Furuyama

Most animals produce and perceive vocalizations to communicate with other individuals. Primates must recognize both the contents and the speakers of vocalizations from conspecific individuals accurately for maintaining social interactions and the increases their chances of survival. In addition, studies with non-human primates are required to discuss whether the voice recognition was evolutionarily maintained in primates. The purpose of this thesis was to investigate the vocal recognition in Japanese macaques and humans. This thesis is constructed of three behavioral studies. 1) This study investigated the temporal resolutions of both Japanese macaques and humans. The results of temporal resolution showed that humans were more sensitive to detecting amplitude modulation than were Japanese monkeys. 2) The acoustic characteristics used to discriminate individuals based on conspecific and heterospecific vocalizations were determined in Japanese macaques and humans. Our data about individual discrimination showed that monkeys and humans seemingly use different acoustic characteristics to distinguish conspecific and heterospecific voices. 3) The acoustic features for the discrimination of individuals were investigated in Japanese macaques. Our data suggested that formants related to vocal tract characteristics contributed to discriminating individuals based on vocalization in monkeys rather than the temporal structures of fundamental frequencies. Formants were also contributed to distinguish individuals in humans. Thus, our data may imply that the vocal processing of Japanese macaques for individual discrimination were similar to that of humans. Further studies are need to investigate the neural activities behind individual discrimination based on conspecific vocalizations.

List of Contents

List of Tables

List of Figures

Acknowledgments

First of all, I would like to thank many people who have been supporting my study. I would like to thank Professor Hiroshi Riquimaroux and Professor Kenichi Yoshikawa for giving me the opportunity to pursue a doctoral degree and for encouraging me with their grateful support throughout the undergraduate and graduate years.

I would like to thank Dr. Kohta I. Kobayasi for his advices throughout the undergraduate/graduate years and constructive comments on my study. His wide knowledge of neuroethology and his behavioral techniques have provided both the bases of behavioral protocols and fundamental knowledge. His crucial advices supported behavioral techniques and discussions in this dissertation. In addition, I wish to appreciate him for constant encouragement and helpful me.

I also wish to thank Dr. Shizuko Hiryu for both insightful discussions and providing importance of animals' auditory on my study. In addition, I would like to appreciate her for continuous encouragement.

I wish to appreciate Mr. Hiroshi Takaoka for his knowledge of fundamental programming in vocal processing. His programming skill of vocal processing have provided the techniques of vocal modification on my thesis.

I would like to appreciate my collaborators, Mr. Yuki Torigoe, Mr. Naoya Akiyama, Mr. Takeshi Morimoto, and Mr. Takashi Noguchi for supports of both animal care and training of monkeys. In addition, I thank them for improving experimental apparatus on my study.

I would like to appreciate the financial support of Japan Society for Promotion of Science.

I greatly appreciate my parents for their support through the undergraduate and graduate years. Finally, I wish to thank my wife, Aiko, for her continuous support and encouragement.

Takafumi Furuyama

Chapter 1

Introduction

Many primates communicate with other individuals by using vocalizations. It is important not only to recognize a content of vocalizations but also to identify individuals based on utterances, because most primates keep social interactions and increase their chances of survival and mating rates. In addition, comparative studies between humans and nonhuman primates is necessary to describe whether the voice recognition was evolutionarily maintained in primates. This chapter describe the brief backgrounds on vocal recognition in primates, remaining questions, and the purpose of this thesis.

1.1 Information on vocalizations of primates

Many animals produce vocalizations to maintain social interactions. Most non-human primates produce species-specific vocalizations and possess rich vocal repertoires. Several early studies classified the communication sounds of non-human primates. Conspecific sounds were recorded and analyzed in rhesus monkeys (Rowell and Hinde, 1962), squirrel monkeys (Winter *et al.*, 1966), and gorillas (Fossey, 1972). Vocalizations of Japanese macaques were recorded and classified into six classes and 37 types (Itani, 1963). Another study analyzed calls using spectro-temporal structures, classifying 10 classes and 41 subtypes (Green, 1975). These studies explored the evolution of language in humans through species-specific vocalizations of non-human primates.

Animals have to perceive the contents of utterances from conspecific individuals or predators accurately, because this increases their chances of survival and mating rates. Several studies have demonstrated that vocalizations of non-human primates include content like that of human languages. For example, Seyfarth *et al.* (1980) presented different types of alarm calls to free-ranging vervet monkeys from speakers, and differences in the behaviors of the animals resulted, depending on the type of vocalization. Additionally, animals learn the role of alarm calls during development (Seyfarth and Cheney, 1986). Non-human primates learn to use and comprehend vocalizations (Janik and Slater, 2000). Free-ranging putty-nosed monkeys are able to combine two calls to convey information on both a predator and an impending movement (Arnold and Zuberbühler, 2006; 2008). Japanese macaques emit greeting calls together with increased social interactions when they approach unrelated females (Katsu *et al.*, 2014).

In addition to context, conspecific vocalizations, including human speech, contain much more information (for review see Belin *et al.*, 2004; Taylor and Reby, 2010). Vocalizations are also used to convey information on both affective state and individualities: this is often called "paralinguistic" information. For example, humans can distinguish individuals and identify emotional state during telephone conversations. Such paralinguistic information is necessary for the construction of social and cooperative interactions among individual primates, including humans. Particularly, individual recognition based on vocalizations is important because most non-human primates live in a forest.

1.2 Vocalization mechanisms in primates

The basic mechanics and anatomy of vocalizations are broadly similar across primates, including humans. Thus, the vocalizations of primates have clear fundamental frequencies (F0s) and harmonics (Fig. 1-1). Fant (1971) developed the "source-filter theory" in the context of speech production (Fig. 1-2). Vocalizations involve a sound source (larynx) and the coupled supralaryngeal cavities (the oral and nasal cavities). The larynx is opened and closed periodically by air from the lungs during voice production. The periodic rate of the vibrations is used to determine the F0 of the vocalizations and is called the pitch. As the repetition rates of the larynx pass through the vocal tract airways above the larynx, the vocal tract characteristics (VTC) generate resonances and enhance/dampen specific frequency ranges: these are referred to as "formant".

There are important differences between speech and non-human vocalizations in some aspects (Fitch, 2000). Many primates have air sacs, outpouchings of the epithelium lining the larynx, whereas humans do not. Moreover, the position of the larynx differs between humans and non-human primates. The larynx of humans is located lower than that in non-human primates. The space encompassing the vocal tract is stretched by the descent of the larynx, and humans are able to move the tongue flexibly. Additionally, the stretched vocal tract produces low formant frequencies. However, resent study showed that vocal tract characteristics of monkeys are able to produce human speech adequately (Fitch *et al.*, 2016).

In addition to these anatomical differences, the acoustic characteristics of vocalizations differ between monkeys and humans. The F0 in adult humans is ~100–350 Hz (Bachorowski and Owren, 1999; Skuk *et al.*, 2015), whereas that of monkeys is ~300–1000 Hz (Green, 1975). In addition, the waveform of speech represents slow fluctuations in its amplitude over time. The rates of envelope changes in speech are 333–100 ms (thus, ~3–10 Hz), and the mean syllabic rates of speech correspond approximately to these fluctuations (Houtgast and Steeneken, 1985; Greenberg and Takayuki, 2004). However, such temporal fluctuations in vocalizations were not observed in field studies of Japanese macaques (Green, 1975; Sugiura, 1993).

1.3 Relationships between acoustic features and physical characteristics

Many studies in humans have investigated the relationships between acoustic characteristics and physical features such as gender, age, and body size. Previous studies demonstrated that acoustic features contribute to gender discrimination (Lass *et al.*, 1976; Childers and Wu, 1991; Wu and Childers, 1991; Bachorowski and Owren, 1999). Although listeners seem to underestimate the age of an individual, humans may try to estimate the age of a speaker from their vocalizations (Hartman and Danhauer, 1976; Hartman, 1979). Other studies showed that acoustic characteristics contributed to perceptions of physical features of the talker in humans (Smith and Patterson, 2005; Smith *et al.*, 2005).

Vocalizations of non-human primates have also been correlated with physical body features. One study measured vocal tract length using radiographs and analyzed the vocalizations of monkeys; vocal tract length in monkeys was correlated with body size (Fitch, 1997). The acoustic characteristics of vocalizations have been related to individual discrimination in baboons (Owren *et al.*, 1997) and humans (Lloyd, 2005). A study using MRI revealed a correlation between vocal tract length and body size in humans (Fitch and Giedd, 1999). Another study measured the vocal tract area while vocalizing vowels using MRI; inter-individual differences in the supralaryngeal spaces were shown to influence frequency beyond 2.5 kHz (Kitamura *et al.*, 2005).

1.4 Study on vocalizations of Japanese macaques

The species of study in this thesis was the Japanese macaque (*Macaca fuscata*). Living primates now consist of more than 70 genera and 400 species (Fleagle, 2013). *Macaca* appeared 7–8 million years ago in northern Africa (Delson and Rosenberger, 1980). Japanese macaques are a species of old world monkeys, and they inhabit the northern-most regions of the non-human primate range. The native Japanese monkeys live in Japan (30-41 N). *Macaca* are close to baboons and mangabeys phylogenetically (Fleagle and McGraw, 1999). The life span of Japanese macaques in their natural environment is about 20 years (Fukuda, 1988; Takahata *et al.*, 1998). The mean body weight of Japanese macaques is 10 kg, and their mean height is 55 cm.

Vocalizations of Japanese macaques have been investigated in field studies. Green (1975) described the contact vocalization as a “coo call” in Japanese macaques. Coo calls have both a clear fundamental frequency (F0) and rich harmonics, and these vocalizations are important for social interactions. For example, monkeys vocalize coo calls when they approach other individuals closely for grooming (Mori, 1975). Monkeys exchange coo calls to avoid separation from the group and for maintaining group organization (Mitani, 1986). Brown *et al.* (1979) showed that monkeys could distinguish sound localizations using coo calls. Japanese monkeys in the Yakushima lowland alter the acoustic features of F0 or the duration of coo calls according to context when exchanging coo calls. Japanese monkeys can match some acoustic characteristics of the F0s with playback vocalizations (Sugiura, 1993; 1998; Koda, 2004). Japanese macaques vocalize greeting calls (coo calls, grunts, and girneys) together with increased social interactions when they approach unrelated females (Katsu *et al.*, 2014). Japanese macaques, compared with other genera, have been observed frequently to discriminate individuals based on vocalizations in the natural environment.

1.5 Differences in vocal recognition between humans and non-human primates

Comparative studies between humans and non-human primates are necessary to reveal whether the voice recognition was evolutionarily maintained in primates. Several studies have compared psychoacoustic differences directly between monkeys and humans using both tone bursts and broadband noises to investigate their basic sensory abilities. Monkeys have better sensitivity than do humans in the high-frequency range (Japanese macaques: 0.028–34.5 kHz; humans: 0.031–17.6 kHz, Owren *et al.*, 1988). Additionally, the 'best' frequency for Japanese macaques differed from that for humans (Japanese macaque: 1 kHz: humans: 4 kHz, Jackson *et al.*, 1999). A previous study using pure-tone bursts revealed that the frequency discrimination limits of humans were approximately seven-fold smaller than those of monkeys (Sinnott *et al.*, 1985; Prosen *et al.*, 1990). In addition to frequency discrimination, the sensitivity of intensity limits was worse in monkeys than in humans (Sinnott *et al.*, 1985). A study using amplitude-modulated broadband noise stimuli demonstrated that the auditory system of humans was more sensitive than that of monkeys (O'Connor *et al.*, 2000). Other studies comparing macaques with humans showed that humans were better able to detect amplitude modulation (in noise bursts) when the modulation frequency was relatively low (O'Connor *et al.*, 2011). These studies showed that humans and monkeys differ in their basic hearing abilities.

Many studies using speech stimuli have also compared sensitivities between humans and monkeys in attempts to describe how non-human primates perceive human speech. Several studies used synthetic consonant-vowel sounds. One study presented a stimuli continuum between /ba/ and /da/; humans were more sensitive to formant transitions than were monkeys (Sinnott *et al.*, 1976). Another study showed that monkeys and humans had the same phoneme boundaries (Kuhl and Padden, 1983). Sinnott and Adams (1987) presented gradual voice onset times (VOTs); they demonstrated that humans were more sensitive than monkeys to VOT. Monkeys and humans had different boundaries in discriminating /ra/ and /la/ (Sinnott and Brown, 1997). These study showed that non-human primates were difficult to distinguish human speech.

Several studies have shown that humans have a specialized brain mechanism for the processing of human speech sounds. Several studies using functional magnetic resonance imaging (fMRI) revealed that vocal sounds activated neurons in the upper superior temporal sulcus (STS) (Belin *et al.*, 2000; Belin *et al.*, 2002). Another study compared the neural activity elicited by voice versus non-voice sounds (e.g., musical instruments, animal vocalizations); the left STS regions responded more strongly to speech vocal sounds than to non-vocal sounds (Fecteau *et al.*, 2004).

The auditory cortex of non-human primates is also specialized to process species-specific vocalizations, similar to humans. Neurons of the auditory cortex respond to species-specific vocalizations rather than tone bursts, click sounds, or white noise bursts (Winter and Funkenstein, 1973). Activities in the primary auditory cortex in common marmosets were elicited by natural species-specific vocalizations rather than time reversed-vocalizations and temporally changed vocalizations (Wang *et al.*, 1995). Species-specific vocalizations elicited neural activity in the left hemisphere (Poremba *et al.*, 2004). Another study using behavioral and neurophysiological techniques showed that the left temporal cortex, including the auditory cortex, was necessary to discriminate two types of vocalizations in Japanese macaques (Heffner and Heffner, 1984).

1.6 Individual recognition based on faces

Many studies on individual identification have investigated the perception and recognition of faces. Facial communication has become sophisticated in humans and non-human primates. Humans often communicate with other individuals face-to-face, and humans can readily discriminate individuals by their face. Like humans, non-human primates are able to distinguish the faces of other individuals. Chimpanzees matched the faces of mothers and sons (Parr and de Waal, 1999). Rhesus macaques and chimpanzees can distinguish unfamiliar conspecifics using facial features (Parr *et al.*, 2000). Indeed, monkeys preferred conspecific faces rather than heterospecific faces (Dufour *et al.*, 2006).

Several studies have directly compared facial recognition in non-human primates with that in humans. Eye movements contribute to facial recognition in interactions with other individuals.

In a study using eye-tracking methods, it was suggested that humans and rhesus macaques use the same strategies to scan conspecific faces (Dahl *et al.*, 2009). Another study showed that both chimpanzees and humans exhibit unique eye movement approaches for interactions with conspecifics but not heterospecifics (Kano and Tomonaga, 2010).

Many studies using functional imaging methods revealed that facial stimuli are processed in the inferior temporal cortex (IT) of primates. It was shown that the fusiform gyrus of the human brain is involved in processing facial stimuli more strongly than other complex stimuli, using fMRI (Kanwisher *et al.*, 1997; Haxby *et al.*, 2000; Kanwisher and Yovel, 2006) and positron emission tomography (PET) (Sergent *et al.*, 1992).

Previous studies also investigated how single neurons in the brain are responsible for recognizing facial features. Recorded neural activity in the superior temporal polysensory region showed that single neurons responded to both human and monkey faces, but removal of the eyes from faces reduced that neural activity (Bruce *et al.*, 1981). Neurons of the IT responded to facial stimuli rather than to other visual stimuli, such as images of hands (Desimone *et al.*, 1984). Another study demonstrated that objects might be represented by combinations of neural activities that respond to particular features in images (Tanaka *et al.*, 1991). In rhesus macaques, a study used faces that were intermediate between two individuals to show that single neurons might be linked to tuning perceptions of face identities (Leopold *et al.*, 2006). Electrophysiological studies in macaques have demonstrated neural mechanisms specialized for facial recognition (for review see Gross, 2008).

1.7 Individual recognition based on voice in primates

In addition to faces, speech sounds contain information on the identities of individuals. For example, humans can readily identify individuals from speech during telephone conversations and when listening to the radio. Voice recognition is important for both perceiving the content of a conversation and identifying individuals among primates, including humans, for maintaining social communities.

Most non-human primates are also able to identify individuals based on their vocalizations. Manly studies have demonstrated that primates can identify their own and other infants based on vocalizations alone. Mothers of adult squirrel monkeys (*Saimiri sciureus*) were able to distinguish the vocalizations of their own infants from those of other infant monkeys (Kaplan *et al.*, 1978). Snowdon and Cleveland (1980) showed that the pygmy marmoset (*Cebuella pygmaea*) could distinguish other individuals as group members. Other studies demonstrated an ability to identify infants based on calls alone in *Chlorocebus pygerythrus* (Cheney and Seyfarth, 1980), *Macaca fuscata* (Pereira, 1986), and *Macaca mulatta* (Jovanovic *et al.*, 2000).

Primates, including humans, can identify the body characteristics of individuals based on the acoustic features of vocalizations. A study based on a habituation–dishabituation paradigm showed that rhesus macaques can discriminate the species-specific communication of kin from that of non-kin (Rendall *et al.*, 1996; Rendall *et al.*, 1998). Japanese macaques also distinguish individuals using vocalizations (Ceugniet and Izumi, 2004). Untrained rhesus macaques were able to distinguish age-related body size by using voices (Ghazanfar *et al.*, 2007). Lemurs also discriminated individuals based on vocalizations (Gamba *et al.*, 2012). Previous studies using whisper speech showed that humans can readily distinguish different speakers (Tartter, 1991). Another study presented sine-wave sentences; listeners were apparently capable of determining individual vocals (Fellowes *et al.*, 1997; Remez *et al.*, 1997). A study that statistically analyzed vowels in humans suggested that various acoustic features with vocal tract resonances contribute to classification of the speaker (Bachorowski and Owren, 1999).

Many studies have demonstrated individual recognition based on vocalizations in monkeys (for review see Belin, 2006). These observations showed that monkeys were able to *discriminate* rather than *identify* individual callers. In another study, Sliwa *et al.* (2011) measured preferential looking time in rhesus macaques and evaluated whether voices are linked to faces; they showed that macaques could spontaneously match familiar faces with voices. Non-human primates may be able to memorize individuals based on their vocalizations and faces.

Several studies have investigated the brain regions involved in voice identification. A study using PET showed that the anterior temporal lobes respond bilaterally to the identification of a speaker (Imaizumi *et al.*, 1997). The left frontal pole and right temporal pole were correlated with familiar voices (Nakamura *et al.*, 2001). Another study using continuum stimuli demonstrated that the anterior temporal lobe responds to changes in voice identity (Andics *et al.*, 2010). In a study that measured neural activities in response to continuum stimuli of vocalizations among individuals using fMRI, and the right inferior frontal cortex responded to changes in perceived identity (Latinus *et al.*, 2011).

A few recent studies of non-human primates investigated brain activities of the individual recognition by using fMRI. The few neurons of prefrontal cortex in rhesus macaques responded to vocalizations from one caller and not to calls from other callers (Romanski *et al.*, 2005). A previous study presented vocalizations of conspecifics and heterospecifics to awake monkeys, and suggested that activities of anterior temporal cortex was elicited by same types of vocalizations produced different individuals (Petkov *et al.*, 2008).

1.8 Acoustic characteristics for individual discrimination of vocalizations

Humans can distinguish individuals and emotional state using F0s. In human speech, pitch can signify the emotional state of the speaker (Scherer, 1995). One study indicated that gender identification required F0 information in humans (Lass and Davis, 1976). Another study also showed that F0 information helped identify the speaker’s gender (Bachorowski and Owren, 1999).

Several monkey species have been shown to discriminate vocalizations using temporal structures of F0. Temporal structures of F0 contain information on vocalizations, because monkeys modify their vocalizations depending on different situations (Green, 1975). Trained Japanese macaques discriminated the peak positions of natural tonal vocalizations (Zoloth *et al.*, 1979). Monkeys may categorically discriminate temporal structures of F0 (May *et al.*, 1989). Additionally, monkeys distinguish synthetic vocalizations of conspecifics using the peak positions of F0 (Hopp *et al.*, 1992). Statistical analyses of the acoustic features of F0, such as the beginning frequency and maximum frequency, indicate that the F0 can be a reliable cue for identifying callers in several monkey species (Smith *et al.*, 1982; Snowdon *et al.*, 1983).

Chimpanzees could discriminate individuals based on the F0 of vocalizations (Kojima *et al.*, 2003). In related research, Japanese macaques were trained to discriminate the vocalizations of different monkeys, and the subjects responded to the F0 as a discriminant stimulus for the task, suggesting that F0 contributes to individual discrimination (Ceugniet and Izumi, 2004).

Formant frequencies are necessary for speech recognition. Different vowels match different configurations of articulators that produce different formant frequencies. Formant frequencies create an important acoustic cue for the identification of vowels (Peterson and Barney, 1952). Human listeners can obtain speech from sine-wave sounds consisting of three tone bursts following the first three formants (Remez *et al.*, 1981). In addition, humans are also able to discriminate individuals based on whisper vocalizations (Tartter, 1991).

Many non-human primate species use formants to discriminate individuals in vocal communications. Vocal tract length is necessary to distinguish individual talkers in human speech (Bachorowski and Owren, 1999). Formants related to vocal tract length were used to discriminate alarm vocalizations in a manner similar to that used in humans for the identification of speech (Owren, 1990). Similar to humans, trained Japanese macaques exhibit great sensitivity to different formant frequencies (Sommers *et al.*, 1992). Characteristics of formants contributed to individual differences (Rendall, 2003). Non-human primates were able to discriminate formant changes in species-specific vocalizations (Fitch and Fritz, 2006). Gahazanfar *et al.* (Ghazanfar *et al.*, 2007) used a preferential looking paradigm; they suggested that formants in monkeys acted as indexing cues of age-related size.

1.9 Remaining questions

Many studies have compared basic hearing abilities between humans and non-human primates in terms of psychoacoustics. However, these studies primarily assessed frequency discrimination limits. The vocalizations of primates contain spectro-temporal information (Fig. 1-1), but few studies have examined temporal resolution in monkeys.

Vocalizations of primates have complex acoustic characteristics (e.g., F0, VTC, and duration). Some studies have used sophisticated software to modify the vocalizations of primates

(McAulay and Quatieri, 1986; Narendranath *et al.*, 1995; Veldhuis and He, 1996), but there are limits to changing acoustic parameters flexibly. The acoustic characteristics of vocalizations in non-human primates have been filtered and lengthened using software (Ceugniet and Izumi, 2004; Fitch and Fritz, 2006). However, few studies have independently modified specific parameters of acoustic characteristics in the vocalizations of monkeys.

Several studies showed that the auditory and prefrontal cortex prefer species-specific vocalizations rather than heterospecific vocalizations or natural sounds (Romanski *et al.*, 2005; Petkov *et al.*, 2008). However, few behavioural studies have investigated the differences of vocal processing between vocalizations of conspecific and heterospecifics.

The integration from multiple senses (e.g., visual, auditory) is necessary for the identification of particular individuals. Individual recognition of faces has been examined in many behavioral and neurophysiological studies. Several studies have assessed the acoustic characteristics used for individual discrimination, based on conspecific vocalizations, but acoustic characteristics used to discriminate individuals have been still yet to be discussed.

Many neurophysiological studies have investigated how neurons recognize facial characteristics. However, little is known about how neurons in the brain recognize individuals based on vocalizations.

1.10 Purposes

The aim of this dissertation was to investigate the vocal recognition in both Japanese macaques and humans. Primates have to perceive both the contents and the speakers of utterances from conspecific individuals or predators accurately, because they maintain social interactions and this increases their chances of survival and mating rates. In addition, studies with non-human primates are important to describe whether the voice recognition was evolutionarily maintained in primates. In this thesis, regarding the basic abilities associated with vocal recognition, the temporal resolutions of amplitude modulations were compared between Japanese macaques and humans. Moreover, regarding the higher cognition of hearing abilities, we determined important acoustic characteristics used by Japanese macaques and humans to discriminate individual monkeys.

1.11 Organizations of thesis

This thesis is organized as follows: the temporal resolutions of amplitude modulation in both humans and monkeys were compared in Chapter 2. Chapter 3 describes continuum vocalizations modified using auditory signal processing software. This chapter explains how both monkeys and humans perceive the morphed stimuli, and the acoustic characteristics were determined to discriminate monkeys based on vocalizations alone in Japanese macaques and humans. Chapter 4 explains the important acoustic features used by Japanese macaques to discriminate individuals. Summaries of the main results and future research are presented in Chapter 5.

Chapter 2

Perception of amplitude-modulated broadband noise: Comparisons between Japanese macaques (*Macaca fuscata*) and humans

2.1 Introduction

Quite a few studies have shown how humans perceive amplitude-modulated (AM) sounds (Viemeister, 1979). A major reason for us to investigate AM sound perception is to suggest that temporal changes in amplitude envelopes of sound may play important roles in speech perception. For example, noise-vocoded speech sounds (synthesized speech sounds in which the speech signal is replaced by several bands of noise while the amplitude envelope is preserved) are able to create not only speech perception but also pitch accents (Shannon *et al.*, 1995; Riquimaroux, 2006). Previous study using AM stimuli demonstrated that the auditory system of humans was better sensitive than that of monkeys (O'Connor *et al.*, 2000). Recent studies by O’Connor and colleagues comparing rhesus macaques (*Macaca mulatta*) with humans showed that humans were better able to detect amplitude modulation (in noise bursts) when the modulation frequency was relatively low (less than 15 Hz) (O’Connor *et al.*, 2011). The results suggested the existence of differences in the temporal processing between species in primates. Several human studies measuring the durations of syllabic segments of English and Japanese showed that the temporal modulations of languages peaked at ~3–10 Hz (Houtgast and Steeneken, 1985; Greenberg and Takayuki, 2004). Taken together, humans’ superiority in sensitivity for AM sounds might be required for speech perception, which involves processing sounds with low modulation frequencies.

Many studies have compared psychoacoustic differences directly between monkeys and humans using tone bursts to investigate their basic sensory abilities. Monkeys have better sensitivity than do humans in the high-frequency range (Japanese macaques: 0.028–34.5 kHz; humans: 0.031–17.6 kHz Owren *et al.*, 1988). Additionally, the ‘best’ frequency for Japanese macaques differed from that for humans (Japanese macaque: 1 kHz: humans: 4 kHz, Jackson *et al.*, 1999). A previous study using pure-tone bursts revealed that the frequency discrimination limits of humans were approximately seven-fold smaller than those of monkeys (Sinnott *et al.*, 1985; Prosen *et al.*, 1990). In addition to frequency discrimination, the sensitivity of intensity limits was worse in monkeys than in humans (Sinnott *et al.*, 1985). These studies indicated that non-human primates and humans differ in their basic hearing abilities.

In this study, we used Japanese macaques (*Macaca fuscata*) and humans as subjects. Although Japanese macaques and rhesus macaques are in the same genus (*macaca*), the species differ in their vocalization behaviors (Owren *et al.*, 1993) and their hearing sensitivity (Heffner, 2004). To date, however, no study has compared Japanese macaques to humans in terms of sensitivity to AM sound with low modulation frequency.

The sensitivities to detect amplitude modulation (in noise bursts) in Japanese macaques and humans were examined by using standard Go/NoGo operant conditioning. Subjects were trained to discriminate continuous and repetitive white noise bursts, and their sensitivities were quantified using AM noise with various modulation depths. The results may provide a better understanding of differences in auditory temporal perception between humans and non-human primates.

2.2 Materials and Methods

2.2.1 Subjects

Two male Japanese macaques (*Macaca fuscata*) and three male humans, aged 21–22, participated in the experiment. Monkey 1 was 7 years old and Monkey 2 was 10 years old. Each animal was individually kept in a primate cage with constant light/dark cycles of 13/11 h. Their access to liquids was limited for 24h as water served as positive reinforcement in the experiments. Monkeys got total 500 ml fruits juice both during training and after training. All experiments were conducted in accordance with the guidelines approved by the Animal Experimental Committee of Doshisha University and the ethics board of Doshisha University.

2.2.2 Experimental apparatus

Figure 2-1 shows experimental settings. All trainings and tests were conducted in a sound-attenuated room (1.70 × 1.85 × 2.65 m). A loudspeaker (P-610MB, Diatone, Japan) was positioned 68 cm in front of the subject’s head. The frequency response of the speaker was flattened (± 3 dB) between 0.1 kHz and 18 kHz using a graphic equalizer (GQ2015A, Yamaha, Japan). The equalized stimuli were amplified (SRP-P2400; Sony, Tokyo, Japan). A white light-emitting diode (LED) was placed on the top of the loudspeaker and was turned off during breaks in operant conditioning. In addition to LED, a charge-coupled device (CCD) video camera was attached to monitor the experiment of monkeys.

2.2.3 Stimuli

Two types of white noise bursts were used as discriminative stimuli. The Go stimulus (S+) was a 500 ms continuous white noise burst with 10 ms linear rise/fall times presented with sound pressure level (*re*: 20 μPa) of 60 dB in Fig. 2-2 and 2-3. The NoGo stimulus (S-) consisted of three repetitive white noise bursts having the same total duration (500 ms) as S+ with two 50 ms silent gaps in Fig. 2-2 and 2-3. Detailed temporal profiles of the stimuli are shown in Fig. 2-2B.

All sound stimuli were created by using Cool Edit 2000 (Syntrillium Software) with 44.1 kHz sampling and 16 bit resolution.

Test stimuli were 500 ms AM white noise bursts in which the modulation depths of two sections, corresponding to the silent portions of S-, varied (Fig. 2-2B, test stimulus). Five different modulation depths (11, 29, 50, 75, and 87 %) were presented as test stimuli (Fig. 2-3B). Each type of test stimuli was presented for five times.

2.2.4 Procedure

Figure 2-4 shows training and test procedures. Two male monkeys and three male humans were trained to discriminate between continuous (Go stimulus: S+) and repetitive (NoGo stimulus: S-) white noise bursts with standard Go/NoGo operant conditioning. The subjects had to depress a lever for 200 ms to begin a trial. During training trials, S- was repeated 3–5 times, and then either S- or S+ was presented as a discriminative stimulus (Fig. 2-4). The inter-onset interval between adjacent stimuli was 1000 ms. When S+ was presented as the discriminative stimulus, the subjects had to release the lever within 1000 ms from the offset of S+ (Fig. 2-4A); when they did so, the reaction was scored as a “hit.” If S- continued as the discriminative stimulus (Fig. 2-4B), the subjects had to keep depressing the lever to record a “correct rejection.” Monkeys got about 1 mL of fruit juice with 80% probability at both hits and correct rejections. If the subjects failed to release the lever within 1000 ms after the offset of S+, a “miss” was scored. If the subjects released the lever during S- presentation or within 1000 ms after the offset of S-, a “false alarm” was scored. Misses and false alarms were penalized with a 3–5 s timeout, where the LED was turned off and the start of the next trial was delayed by timeout. In monkey experiment, each training session consisted of 400 trials, in which 200 Go trials and 200 NoGo trials were randomly placed. In human experiment, each training session was constituted by 20 trials, in which 10 Go trials and 10 NoGo trials were randomly ordered. Performance was measured by the correct response percentage (CRP: the total percentage of hits and correct rejections per total trials) and the reaction time (RT: latency from stimulus offset to lever-release). After verifying that the CRP exceeded 80 % for two consecutive sessions, the subjects proceeded to test sessions in which 10 % of all trials were test trials, whereas half of the remainder (45 %) were Go trials and other half (45 %) were NoGo trials; all trials were ordered randomly. A test stimulus was presented after S- was repeated 4 times. No feedback (reward or punishment) followed the response to test stimuli. If the subjects did not release the lever within 1000 ms response period, we recorded reaction time as 1000 ms. We measured Go response rates and reaction times to stimuli in all subjects. We examined the sensitivity of amplitude modulation of humans by using same procedure of monkeys. However, no juice was given to human subjects.

2.3 Results

2.3.1 Training

Monkey 1 and Monkey 2 needed 9 and 7 days of trainings respectively to learn to distinguish between the sets of cooAs and cooBs. Two days before the test day, the monkeys scored correct response rates (CPRs) of 82% and 87%. The day before the test day, the CPRs were 83% and 95%. The animals successfully learned to discriminate between continuous and repetitive white noise bursts. The CRP of Monkey 1 was 83 % and that of Monkey 2 was 92 % during test sessions. The CRPs of all humans were higher than 95 % (Human 1: 97 %, Human 2: 100 %, and Human 3: 97 %). Go response rates to training stimuli (both S+ and S-) in test sessions did not statistically differ from those in training sessions, suggesting that both monkeys and humans were maintaining the same discriminatory performance in response to training stimuli throughout the experiment.

2.3.2 Amplitude modulation depth

Go response rates to types of test stimuli for two monkeys are shown in Figure 2-6. Go response rate at 75% of modulation depth was below 50% in Monkey 1, whereas Go response rate of monkey 2 did not decrease at several types of modulation except the 99% of modulation depth (Fig. 2-5). In two humans, Go response rates to stimuli were below 50 % at 27% of modulation depth, and Go response rates of one human was below 50 % at 50% of modulation depth (Fig. 2-6).

In addition to Go response rates, Figure 2-7 shows reaction times to training and test stimuli in two monkeys. The reaction times of two monkeys were lengthen along the increase of modulation depth. In addition to monkeys, the reaction times of all humans were longer depending on the increase of modulation depth (Table 2-1). The average reaction time to the S+ stimulus differed between Monkey 1 (75 ms) and Monkey 2 (254 ms), and that of humans also varied among individuals (Human 1: 110 ms, Human 2: 362 ms, Human 3: 256 ms; Table 2-1). Thus, the reaction times to different stimuli (white noise bursts with different modulation depths) were normalized into *z*-scores based on the average and standard deviation of the reaction time to S+

(Fig. 2-8, modulation depth = 0 %) within each subject to examine interspecies and inter-subject differences in a standardized manner. The *z*-scores of reaction times to test stimuli with different modulation depths in each monkey are shown in Fig. 2-8. In all subjects (monkeys and humans), the *z*-scores increased as the modulation depth increased (Fig. 2-8, Table 2-1), suggesting that less modulated noise bursts tended to be perceived more similar to continuous noise bursts (i.e., S+). That is, monkeys and humans depressed the lever longer as the amplitude modulation deepened.

A *z*-score of 1.96, corresponding to $p = 0.05$, was used as a criterion to estimate the ability to detect the change between continuous and repetitive white noise bursts. The *z*-scores surpassed the criterion (1.96) when the modulation depth became greater than 75 % in both monkeys (at 75 %: Monkey 1, $z = 1.96$, $p = 0.05$ and Monkey 2, $z = 2.57$, $p = 0.01$), whereas the *z*-scores exceeded the criterion at an average of 27 % in humans (Human 1: 11 %, Human 2: 29 %, Human 3: 50 %). Thus, monkeys were worse than humans, by 8.9 dB (humans: 27 % vs. monkeys: 75 %), in detecting amplitude modulation.

2.4 Discussion

2.4.1 Individual differences in monkeys

The reaction time of Monkey 1 was shorter than that of Monkey 2 for all test stimuli (Table 2-1). The individual differences might have been caused by differences learning strategies. Specifically, Monkey 1 might have learned to release the lever after the stimuli that were not S-, meaning that he enacted Go responses to stimuli that differed from S-; on the other hand, Monkey 2 might have learned to release the lever after a continuous white noise burst (S+), meaning that he enacted Go responses to stimuli resembling to S+. However, the *z*-scores (normalized to the reaction time to S+) of Monkey 1 were similar to those of Monkey 2 (Fig. 2-9), suggesting that the reaction times were consistent when measured for their perception (Pfingst *et al.*, 1975a; Pfingst *et al.*, 1975b).

2.4.2 Sensitivities of AM broad-band noise in monkeys and humans

A previous study comparing rhesus macaques (*Macaca mulatta*) to humans with sinusoidal AM broadband noise showed that humans had better sensitivity in detecting amplitude modulation when the modulation frequency was low (less than 15 Hz). The difference in depth sensitivity reached about 9 dB at 5 Hz (O’Connor *et al.*, 2011). The superiority in sensitivity might be related to humans’ need to process 3–10 Hz amplitude modulation for speech perception (Houtgast and Steeneken, 1985; Greenberg and Takayuki, 2004). We used about 6 Hz AM white noise bursts and demonstrated that humans were better able, by 8.9 dB, to detect the modulation. Whereas there are many methodological differences between our experiment and previous studies (O’Connor *et al.*, 2011) our results showed the same trend regarding species differences between humans and macaques. In conclusion, this experiment strengthens the idea that humans are better able to detect amplitude modulation in low modulation frequencies than are macaques.

Chapter 3

Acoustic characteristics used for the discrimination of individuals based on vocalizations in Japanese macaques and humans

3.1 Introduction

Many primates, including humans, are able to distinguish a speaker based on vocalizations alone. Previous studies showed that mothers could distinguish the voices of their own infants from those of other juvenile individuals in *Saimiri sciureus* (Kaplan *et al.*, 1978), *Chlorocebus pygerythrus* (Cheney and Seyfarth, 1980), *Macaca fuscata* (Pereira, 1986), and *Macaca mulatta* (Jovanovic *et al.*, 2000). A previous study presented monkey contact voices from a hidden speaker, and the pygmy marmosets recognized other group members as individuals (Snowdon and Cleveland, 1980). Another study using a habituation–dishabituation paradigm showed that rhesus macaques were also able to discriminate the species-specific vocalizations of kin from the those of non-kin (Rendall *et al.*, 1996). Humans can discriminate speakers not only by natural speech but also by whispered speech (Tartter, 1991). Other studies suggested that listeners were able to determine individuals vocally by sine-wave sentences based on speech (Fellowes *et al.*, 1997; Remez *et al.*, 1997). Together, these studies indicate that the identification of individuals by their vocalizations is important in many primates.

Many studies have also examined visual recognition, especially faces, in primates using both behavioral and neurophysiological approaches. In these studies, intermediate morphs that have subtle variations have been used to examine sharp transitions in perception. Previous studies generated gradually changing continuous stimuli between the faces of two humans, and it was shown that humans could classify the morphed intermediates as one or the other face (Leopold *et al.*, 2001; Webster *et al.*, 2004; Furl *et al.*, 2007). In a neurophysiological study, such morphed stimuli were used for investigating how neurons respond to visual features. Other studies, using such stimuli as a series of faces and 3D objects, examined how neurons recognize complex visual features (Freedman *et al.*, 2001; 2003; Leopold *et al.*, 2006; Sigala *et al.*, 2011).

In addition to facial perceptions, such approaches using continuum stimuli have been used to examine transitions in perceptions in psychoacoustics. In previous studies in which formant frequencies were changed gradually, American subjects were able to distinguish ‘r’ and ‘l’ by the slight change in formant frequencies (Miyawaki *et al.*, 1975). Kuhl and Padden (1982) generated continuum phonetic voice samples from humans, and showed that non-human primates categorized the phonetic voice of humans using the second formant. Another study generated gradual temporal structures of vocalizations in monkeys, and it was shown that monkeys used temporal structures to discriminate conspecific communication calls (May *et al.*, 1989). Another study compared the transitions of perception in the discrimination of /ra-la/ in both humans and monkeys (Sinnott and Brown, 1997). Continuum stimuli of vocalizations among different individuals have also been evaluated in humans (Chakladar *et al.*, 2008).

Several studies compared the sensitivities of psychoacoustics between humans and monkeys using synthetic vocalizations. A behavioral study suggested that humans and monkeys exhibit different speech processing, even though the monkeys were able to discriminate phoneme stimuli between /ba/ and /da/ (Sinnott *et al.*, 1976). One study compared the difference in sensitivity between humans and monkeys using a continuum of voice onset times (VOTs) in English; it was suggested that the differences in sensitivity in discriminating pairs of syllables in VOT were worse in monkeys than in humans (Sinnott and Adams, 1987). These studies showed that that sensitivities of humans and monkeys differ.

Acoustic characteristics for discriminating individual primates have been investigated in many studies. Owren *et al.* (1997) analyzed the vocalizations of female chacma baboons (*Papio ursinus*) and reported that the acoustic features of vocal tract filtering may reflect individuality. Bachorowski and Owren (1999) analyzed phonemes of speech in humans and showed that vocal tract filtering may contribute to identification of individuals. The resonance of vocal tract filtering may affect individual identification in rhesus macaques (Rendall *et al.*, 1998) and lemurs (Gamba *et al.*, 2012). In addition to the formants, statistical analyses of the acoustic features of the F0, such as the beginning frequency and maximum frequency, indicate that the F0 can be a reliable cue for identifying callers in several monkey species (Smith *et al.*, 1982; Snowdon *et al.*, 1983).

Several studies have shown that auditory processing software can be used to modify the vocalizations of primates (McAulay and Quatieri, 1986; Narendranath *et al.*, 1995; Veldhuis and He, 1996). Acoustic parameters of vocalizations were modified in non-human primates using auditory processing software (Fitch and Fritz, 2006). In other research, Japanese macaques were trained to discriminate the vocalizations of different monkeys, and the frequencies of the vocalizations were filtered and lengthened (Ceugniet and Izumi, 2004). However, this software is limited in terms of changing acoustic features flexibly.

STRAIGHT (Speech Transformation and Representation based on Adaptive Interpolation of weiGHTed spectrograms) can create a F0-independent spectral envelope that represents vocal tract information, independent of its source (Kawahara *et al.*, 1999a). F0 and VTC were manipulated independently using STRAIGHT; it was shown that VTC contributed to the detection of speaker body size (Smith and Patterson, 2005). Another study modified only the formant structures of vocalizations in monkeys (Ghazanfar *et al.*, 2007). One study generated continuum vocalizations of monkeys between two individuals, and the authors presented the stimuli to humans (Chakladar *et al.*, 2008). However, few studies have investigated transitions in perceptions to discriminate vocalizations of monkeys using continuum stimuli.

The purpose of this study was to investigate how monkeys respond to the continuum stimuli between two individuals using STRAIGHT. Additionally, in the present study, acoustic features for discriminating the vocalizations of two monkeys in monkeys and humans were determined. Specifically, two monkeys and five humans were trained to discriminate the vocalizations of two monkeys. We generated two additional continuum stimuli in which only one acoustic feature, F0 or VTC, was modified from the two monkeys, while the other acoustic characteristics were maintained. In terms of behavior, the reaction times of all subjects were correlated significantly with the proportion of morphing. The reaction times in monkeys were correlated with the rates of changes in F0 and VTC, whereas the reaction times in humans were correlated only with the rate of modification of F0. These results suggest that the stimuli affected the perception of individuals systematically, and our data showed that humans and monkeys use different acoustic characteristics to discriminate the vocalizations of conspecifics and heterospecifics.

3.2 Materials and Methods

3.2.1 Subjects

Two male Japanese macaques and five humans (22–23 years old) were the subjects of these experiments. Monkeys 1 and 2 were 7 and 10 years old, respectively, at the time of testing. In addition, two monkeys were trained to discriminate continuous and repetitive white noise burst. Each monkey was housed individually in a primate cage under a constant 13/11 h light/dark cycle. Access to liquids was limited, because water served as a positive reinforcement in the experiments. All experiments were conducted in accordance with the guidelines approved by the Animal Experimental Committee of Doshisha University and the ethics board of Doshisha University.

3.2.2 Apparatus

All training and tests were conducted in a sound-attenuated room (length × width × height: 1.70 m × 1.85 m × 2.65 m). In the experiments involving the monkeys, subjects were seated in a monkey chair equipped with a drinking tube and a response lever. In the experiments involving the human subjects, the same lever was attached to a desk, and the subject was seated in a standard laboratory chair in front of the desk. A loudspeaker (SX-WD1KT; Victor, Tokyo, Japan) driven by an amplifier (SRP-P2400; Sony, Tokyo, Japan) was positioned 58 cm in front of the subject's head at the same height as the ears. The frequency response of the speaker was flattened (±3 dB) between 0.4 kHz and 16 kHz using a graphic equalizer (GQ2015A; Yamaha, Hamamatsu, Japan). A white light-emitting diode (LED) and a charge-coupled device (CCD) video camera were attached to the top of the speaker. The LED was lit during the training and test sessions for lighting, and subjects were monitored using the CCD camera.

3.2.3 Acoustic stimuli

Sound stimuli were obtained from the two adult male monkeys (Monkeys A and B). Coo calls from Monkey A (cooA) and Monkey B (cooB) were recorded using a digital audio tape recorder (TCD-D8; Sony, Tokyo, Japan) and a condenser microphone (type 2142; Aco, Tokyo, Japan) at a sampling rate of 44.1 kHz and a resolution of 16 bits. Prior to the experiment, the subjects (both monkeys and humans) did not hear the voices of the stimulus monkeys. Seven coo calls with a signal-to-noise ratio greater than 40 dB were selected randomly from the recorded sounds for use as stimuli.

Recorded coo calls (Fig. 3-1) were analyzed using a digital-signal-processing package (STRAIGHT, Kawahara *et al.*, 1999b) to measure three acoustic parameters: the F0 (Fig. 3-2), VTC (frequency structure corresponding mostly to the resonance characteristics of the vocal tract; Fig. 3-3), and the durations of the coo calls. Twelve coo calls (six per individual) were used as training stimuli (cooAs and cooBs; Fig. 3-1). One coo call from each monkey (cooA and cooB) was not played during training and was used to synthesize a test stimulus. Three continuum stimuli of coo calls were created using STRAIGHT. The program was used to break down a coo call into several acoustic parameters (F0 and VTC) and allowed us to manipulate the parameters independently of each other. For example, we could synthesize a coo call from 30% of the information from Monkey A (i.e., cooA) and 70% of the information from Monkey B (i.e., cooB) in one acoustic parameter (e.g., F0), while using no information from Monkey A in another parameter (e.g., VTC). A stimulus continuum, defined as a whole morph, consisting of cooA and cooB was created to comprise 10, 30, 50, 70, and 90% of cooB (Fig. 3-4). Each stimulus in the continuum contained equal F0 and VTC from cooB. We created two additional sets of continuum stimuli in which only one acoustic parameter, F0 or VTC, was changed from cooA to cooB, while the other acoustic feature stayed as Monkey B. Continuum stimuli, defined as the F0-morph, were created to comprise 10, 30, 50, 70, and 90% of F0 from cooB (Fig. 3-5), and another, defined as the VTC-morph, comprised 10, 30, 50, 70, and 90% of VTC from Monkey B (Fig. 3-6). Three different sound pressure level (SPL) stimuli were created for each stimulus type: 57, 60, and 63 dB SPLs (*re*: 20 μPa). All stimulus amplitudes were modified digitally and were calibrated (using a microphone: type 7016, Aco, Tokyo, Japan). The call durations were equalized to 517 ms (i.e., the average of all calls) via linear time-stretching or -compressing using STRAIGHT.

3.2.4 Procedure

Standard Go/NoGo operant conditioning was used. Figure 3-7 shows the schematized event sequence of the trials. Subjects were required to depress the lever switch on the monkey chair to begin the trial. Then, coo calls from the same subject, Monkey A or Monkey B, were presented randomly three to seven times. In the repetition, call types were selected randomly from 18 types of stimuli (1 individual × 6 types of coo calls × 3 intensities). The inter-stimulus interval between adjacent stimuli was 800 ms. While the calls from the same monkey were presented (NoGo trial), subjects were required to continue depressing the lever. When the stimulus was changed from one monkey to another (Go trial), subjects were required to release the lever within 800 ms from the offset of the stimulus. For example, a trial was started using the repeated playback of cooAs (NoGo stimulus). In the repetition, the cooA type (out of six) and the stimulus intensity (out of three: 57, 60, and 63 dB SPL) were changed randomly. The subjects were required to continue depressing the lever while cooA was repeated (correct rejection [CR]). When cooB (Go stimulus) was presented, the subjects were required to release the lever within 800 ms after the offset of cooB (Hit). Hits were reinforced by providing fruit juice (2 mL). When the subjects released the lever during the repetition period of the NoGo stimulus (false alarm) or failed to release the lever within 800 ms after the Go stimulus (miss), a 5–10 s timeout period accompanied by turning off the LED was provided as feedback. When the subjects responded successfully to the Go stimulus, the stimulus contingencies were reversed in the next trial. That is, the next trial was started using a playback of cooB instead of cooA, and the subject had to release the lever when cooA was played to receive the reward.

Performance was measured by the correct response percentage (CRP; total percentage of hits and CRs). In total, 130–180 Go trials (i.e., trials in which the stimulus changed from one monkey to the other) and 800–1000 NoGo trials were presented per day to both subjects.

After the monkey scores exceeded the CRP threshold (75%), the subjects proceeded to the test sessions. Test trials were conducted approximately every 10–20 training trials. A test stimulus was presented after cooB, repeated three to seven times, and each type of test stimulus was played six times. Neither reward nor punishment followed the test trial.

For the human subjects, no juice was given as a reward in the trials, and a CRP of 90% was used as the threshold for proceeding to the test session. Test trials were conducted every 5–10 training trials, and each type of test stimulus was presented five times.

3.2.5 Data analysis

We measured the Go response rates and reaction times of the subjects to test stimuli as the time interval between the end of each stimulus and the subjects releasing the lever switch. The coefficient of correlation (Spearman product-moment correlation coefficient) between reaction times and sets of continuum stimuli were calculated using commercial statistics software (SPSS; IBM, New York, USA).

3.3 Results

3.3.1 Training results in each subject

Monkeys 1 and 2 required 20 and 21 days of training, respectively, to distinguish between the sets of cooAs and cooBs. Two days before the test day, the monkeys scored CRPs of 85% (Monkey 1: d' = 1.81) and 91% (Monkey 2: d' = 2.48). One day before the test day, the CRPs were 85% (Monkey 1: d' = 1.89) and 86% (Monkey 2: d' = 2.09). The CRPs for all human subjects were > 90% during the training sessions. During the test period, the CRPs to training stimuli were > 75% in both monkeys (Monkey 1: 86%; Monkey 2: 87%) and > 90% in all humans (Human 1: 98%; Human 2: 98%; Human 3: 97%; Human 4: 94%; Human 5: 99%). The CRPs during the test period did not differ from those during the training sessions, indicating that the subjects maintained the same discriminatory performance as that with the training stimuli throughout the experiment.

3.3.2 Morphed stimuli between cooA and cooB: whole-morph

The Go response rates to the whole-morph stimulus continuum (whole morph) are shown in Fig. 3-8A. The Go response rates of Monkey 1 and humans decreased gradually with increasing morph proportion of test-cooB, but that of Monkey 2 did not decrease. The Go response rate of Monkey 1 decreased to < 50% when the morphing proportions increased to > 70%. In humans, average Go response rates decreased to < 50% when the morphing proportions increased to > 50%.

Figure 3-8B and Table 3-1 show the reaction times to the whole morph. Reaction times of both monkeys and humans increased gradually with the increase in morphing proportion. A significant positive correlation was observed between morphing proportions of cooB and reaction times to the stimuli in both monkeys and humans (Spearman correlation coefficients, Monkey 1: $r = 0.62$, $n = 42$, $p < 0.05$; Monkey 2: $r = 0.55$, $n = 42$, $p < 0.05$; Humans: $r = 0.84$, $n = 35$, $p < 0.05$). Both monkeys and humans depressed the lever longer as the stimuli became more similar to the test-cooB.

3.3.3 Morphed F0 continuum results

The Go response rates of Monkey 1 and humans decreased gradually with the increase in morphing proportion of F0 from test-cooB, but that of Monkey 2 did not decrease (Fig. 3-9A). The Go response rates of Monkey 1 decreased to < 50% when the morphing proportions of F0 from test-cooB increased to > 30%. In humans, the Go response rates decreased to < 50% when the morphing proportions increased to > 50%.

Figure 3-9B and Table 3-2 represent the reaction times to the F0-morph. The reaction times for each subject (in the two monkeys and two of the humans) increased as the proportions of F0 from test-cooB increased (Monkey 1: r = 0.50, n = 30, $p < 0.05$; Monkey 2: r = 0.46, n = 30, $p < 0.05$; Humans: r = 0.56, n = 25, $p < 0.05$). Both monkeys and humans depressed the lever longer as the stimuli of F0-morph became more similar to the test-cooB.

3.3.4 Morphed VTC continuum results

The Go response rate of Monkey 1 decreased with the increase in morphing proportion of VTC from test-cooB, while that of Monkey 2 did not decrease systematically and remained > 50% (Fig. 3-10A). In Monkey 1, the Go response rate decreased to < 50% when the morphing proportions of VTC of test-cooB increased to > 70%. In humans, the Go response rates remained < 50% regardless of the morphing proportion in VTC-morph.

Figure 3-10B and Table 3-3 show the reaction times to the VTC-morph. The reaction times of both monkeys increased significantly as the contribution of test-cooB to the VTC increased (Monkey 1: r = 0.71, n = 30, $p < 0.05$; Monkey 2: r = 0.40, n = 30, $p < 0.05$), whereas the reaction times of humans were not correlated significantly with the morphing rate (Humans: r = 0.31, n = 25, $p > 0.05$), and remained constant over the VTC-morph continuum.

3.3.5 Comparison between monkeys and humans

Figure 3-11 shows the distributions of correlation coefficients in F0-morph and VTC-morph. The range of correlation coefficients for F0-morph was 0.46–0.50 in monkeys and 0.27–0.80 in humans. The range of correlation coefficients for VTC-morph was 0.40–0.71 in monkeys and 0.00–0.48 in humans.

3.4 Discussion

3.4.1 Response to test stimuli

With all continuum stimuli, the Go response rates for Monkey 1 and humans decreased with the increase in morph proportion, whereas Go response rates to the test stimuli in Monkey 2 remained > 50% (Figs. 3-8A, 3-9A, 3-10A). Monkey 2 might have learned to release the lever after the stimuli that were not learned cooBs. The reaction times of the monkey, however, were correlated significantly with morph proportion (Figs. 3-8B, 3-9B, 3-10B), as in the other subjects, suggesting that Monkey 2 also exhibited stimulus generalization to the stimulus set, albeit relatively limited compared with the others. Our data suggested that all subjects (humans and monkeys) responded to the high morph proportion more similarly to cooB than cooA, whereas subjects responded to the low morph proportion more similarly to cooA than cooB.

3.4.2 Whole-morph continuum

Several studies have indicated that the vocalizations of monkeys can be modified using STRAIGHT. Previously, the vocal tract lengths of rhesus monkeys were effectively increased or decreased virtually using the software (Ghazanfar *et al.*, 2007). Chakladar *et al.* (Chakladar *et al.*, 2008) demonstrated that vocalizations of macaques could be morphed between different individuals using the software, and the quality of the morphs was evaluated by human listeners. In the present study, in both monkeys and humans, the time taken to release the lever increased gradually with an increase in the morph proportion (Fig. 3-8B). To our knowledge, a stimulus continuum synthesized by STRAIGHT was applied for the first time in monkey subjects and demonstrated that the stimuli systematically affected the perception of individuals.

The stimulus continuum has been used to investigate perception in detail and has been especially valuable in evaluating categorical perceptions (Miyawaki *et al.*, 1975; Sinnott *et al.*, 1976; Kuhl and Padden, 1983; Sinnott and Adams, 1987; Sinnott and Brown, 1997). A common feature in categorical perception is that the subject is more sensitive to a physical transition between two perceptual categories than to the same change occurring within a category. This was typically measured using a combination of both a discrimination task between adjacent stimulus pairs (e.g., 10% vs. 30% morphed) in a stimulus continuum and an identification task along the continuum. Using our stimulus scheme, the process by which monkeys categorize vocalizations of different conspecifics can be addressed quantitatively in future research.

3.4.3 Responses to F0-morph stimuli in monkeys and humans

We investigated the acoustic features for individual discrimination using continuum stimuli. The reaction times of both monkeys and humans increased gradually with the increase in morphing proportion with the F0-morph (Fig. 3-9B), suggesting that both monkeys and humans, on average, use F0 as a discriminative stimulus.

Monkeys were able to discriminate vocalizations using temporal structures of F0. In field studies, different temporal structures of F0 were observed in different situations (Green, 1975). In addition to field studies, trained Japanese macaques were able to discriminate the peak positions of natural tonal vocalizations (Zoloth *et al.*, 1979). Trained monkeys were able to classify the temporal structures of F0 categorically (May *et al.*, 1989). Another study analyzed the acoustic characteristics of F0 and indicated that F0 could be used by some monkey species to identify callers (Smith *et al.*, 1982; Snowdon *et al.*, 1983). Japanese macaques were trained to distinguish the vocalizations of different monkeys, and the subjects responded to the F0 as a discriminant stimulus for the task, suggesting that the F0 contributes to individual discrimination (Ceugniet and Izumi, 2004).

In our stimulus set, the mean frequencies of cooB were higher than those of cooA by ~350 Hz (cooA: 519±50 Hz [mean ± standard deviation]; cooB: 875±121 Hz, Fig. 4-2) or 17%; the sensitivities of difference limits for frequency in monkeys and humans have been reported to be 14–33 Hz and 2.4–4.8 Hz, respectively (Sinnott *et al.*, 1985; Prosen *et al.*, 1990), suggesting, the average F0 alone can readily serve as a discriminative stimulus in both species. Additionally, the F0 of the cooA peak was earlier than that of the cooB peak by ~60 ms (the peak position of the vocalizations: 95±22 ms for Monkey A and 134±45 ms for Monkey B). Japanese macaques and humans have shown the ability to distinguish changes in the peak position as small as 20–50 ms

(Hopp *et al.*, 1992), indicating that the temporal structure of F0 can also function as a discriminative stimulus in both species. Thus, the F0 was such that both monkeys and humans could use it as a key to distinguish the stimulus sets.

3.4.4 Responses to VTC-morph stimuli in monkeys and humans

Both monkeys took significantly longer to respond as the morphing proportion of VTC-morph increased (Fig. 3-10B). The results showed that monkeys used the formant frequencies, in addition to F0, as discriminative stimuli for the stimulus sets. Resonances of the vocal tract have physical characteristics in baboons (Owren *et al.*, 1997; Rendall, 2003). In human speech, vocal tract length was necessary to classify individual talkers (Bachorowski and Owren, 1999). It has been shown that formants are biologically significant for the vocal communication of many primate species. Owren (Owren, 1990) showed that formants were used to distinguish alarm calls in a manner similar to that used by humans for discriminating speech. Similar to humans, trained Japanese macaques showed great sensitivity to different formant frequencies (Sommers *et al.*, 1992). Non-human primates were able to discriminate formant changes in species-specific vocalizations (Fitch and Fritz, 2006). One study using a preferential looking paradigm in non-trained monkeys showed that the index characteristics of age-related size were embedded in the formants of monkeys (Ghazanfar *et al.*, 2007). Together, these results were consistent with the present results; formant information played an important role in vocal communication, and the monkeys used the information to discriminate the stimulus sets.

In contrast, the human behavioral data showed that the mean reaction times and Go response rate did not change systematically as the morphing proportion of VTC-morph increased (Fig. 3-10B). These result indicated that, unlike the monkeys, humans, on average, did not use the formant frequency as a key to discriminate the stimulus sets. This difference might stem from differences in auditory sensitivity. Japanese macaques have better high-frequency hearing (i.e., > 8 kHz) than do humans (Heffner, 2004). The power spectrum peak at 10 kHz of cooA, the most distinct feature differentiating the stimulus sets (Fig. 3-3), could be more salient to monkeys than to humans. Thus, the VTC had a greater effect on the monkeys than humans.

Another explanation, which does not necessarily contradict that of auditory sensitivity, is the difference in auditory processing. Many species are specialized to distinguish conspecific vocalizations from other sounds. Previous studies have shown that humans are more sensitive than monkeys in discriminating formant transitions, although monkeys are able to distinguish linguistic sounds (Sinnott *et al.*, 1976; Sinnott and Brown, 1997). Another study compared differences in the sensitivity of humans and monkeys using a continuum of VOT in English; it was suggested that differences in the sensitivity of discriminating pairs of syllables in VOT were worse in monkeys than in humans (Sinnott and Adams, 1987). Our behavioral data indicated that the auditory system of the monkeys is specialized to process their vocalizations, especially the biologically significant acoustic cue of VTC.

3.4.5 Comparisons between monkeys and humans

The distributions of correlation coefficients differed between monkeys and humans (Fig. 3-11). The correlation coefficients of the two monkeys showed similar distributions. Thus, the monkeys used F0 and VTC to discriminate the vocalizations of a monkey caller. However, the distributions of correlation coefficients differed among the human subjects. For example, three of the humans apparently used both F0 and VTC to distinguish the caller monkey, whereas two used only F0. Each human may have used a unique strategy to discriminate the vocalizations of the two monkeys. This result might indicate that the auditory processes for discriminating monkey vocalizations differ between humans and monkeys. A behavioral study using human speech suggested that speech processing differs between humans and monkeys, even though monkeys were able to discriminate phoneme stimuli of speech (Sinnott *et al.*, 1976). In another study that compared differences in sensitivity between humans and monkeys using a continuum of VOT in English, it was suggested that differences in the sensitivity of discriminating pairs of syllables in VOT were worse in monkeys than in humans (Sinnott and Adams, 1987). Our behavioral data demonstrated that each primate used different acoustic structures to distinguish conspecific vocalizations from other sounds (heterospecific sounds or natural sounds).

In addition to behavioral studies, neurophysiological studies have shown that primates utilize brain mechanisms specialized for processing conspecific vocalizations. A study using fMRI showed that vocal sounds activate neurons in the upper part of the STS (Belin *et al.*, 2000; Belin *et al.*, 2002). Another study compared the neural activity elicited by voice versus non-voice stimuli (i.e., musical instrument and animal vocalizations), and the left STS regions responded more strongly to speech vocal sounds than to non-vocal sounds (Fecteau *et al.*, 2004). Neurons of the auditory cortex responded to species-specific vocalizations rather than to tone bursts, click sounds, or white noise bursts (Winter and Funkenstein, 1973). The primary auditory cortex of common marmosets was activated by species-specific vocalizations (Wang *et al.*, 1995). Another study showed that the left temporal cortex, including the auditory cortex, was necessary to discriminate two types of vocalizations in Japanese macaques (Heffner and Heffner, 1984). It has also been shown that species-specific vocalizations elicit neural activity in the left hemisphere (Poremba *et al.*, 2004). Our behavioral data were consistent with those neurophysiological studies in terms of the discrimination of conspecific vocalizations from other sounds (heterospecific sounds or natural sounds).

Several studies confirmed that primates performed different behavior to perceive conspecifics and heterospecifics. Previous study showed that rhesus monkeys possessed perceptual capability for conspecific faces in similarity to that of humans (Dahl *et al.*, 2009). Other previous study demonstrated that chimpanzees and humans performed different scanning to conspecific and heterospecific faces (Kano and Tomonaga, 2010). These study imply that particular strategies of visual recognition were performed for interactions with conspecifics in primates. Our data showed that humans and Japanese macaques employed different acoustic characteristics of conspecific and heterospecific vocalizations in primates.

3.5 Conclusions

The present study confirmed a practicality of continuum stimuli between the vocalizations of two monkeys using STRAIGHT. We also investigated the acoustic features used by both monkeys and humans to discriminate two individuals. Two monkeys and five humans were trained to discriminate the vocalizations of two monkeys. The test stimuli were continuum stimuli between two monkeys, and additional test stimuli in which the specific acoustic parameter used was changed were evaluated. In our behavioral data, monkeys and humans responded to the test stimuli with high morph proportions of Monkey B. Monkeys used F0 and VTC to discriminate individuals by vocalization, whereas humans tended to prefer F0 over VTC. The present data may indicate that monkeys and humans use original methods to discriminate a caller in interactions with conspecifics.

Chapter 4
Role of vocal tract characteristics in individual discrimination by Japanese macaques (*Macaca fuscata*)

4.1 Introduction

Many studies have suggested that primates, including humans, can identify individuals by listening to their vocalizations. The pygmy marmoset (*Cebuella pygmaea*) recognizes other group members as individuals (Snowdon and Cleveland, 1980). Rendall and colleagues demonstrated that rhesus macaques (*Macaca mulatta*) can also distinguish the species-specific communication "coo calls" of kin from those of non-kin and distinguish among the coo calls of close kin using a habituation–dishabituation paradigm (Rendall *et al.*, 1996). Adult squirrel monkey (*Saimiri sciureus*) mothers are able to distinguish the voices of their own infants from those of other juvenile individuals (Kaplan *et al.*, 1978). Several other species, including vervet monkeys (*Chlorocebus pygerythrus*) (Cheney and Seyfarth, 1980), Japanese macaques (*Macaca fuscata*) (Pereira, 1986), and rhesus macaques (Jovanovic *et al.*, 2000), also exhibit the ability to identify their infants based on voice alone. These studies indicate that the identification of individuals by their vocalizations is important for many primates.

Despite the behavioural significance, there are still debates regarding how non-human primates identify individuals from their vocalizations and about the neural mechanisms underlying individual vocal identification. Most monkey vocalizations are harmonically structured such as human vowels because the vocal mechanism in monkeys are the same as those of humans (Green, 1975; Jovanovic *et al.*, 2000; Rendall, 2003; Ghazanfar and Rendall, 2008; Ackermann *et al.*, 2014; Koda *et al.*, 2015). The periodic opening and closing of the vocal folds generates pulses during vocalizations. The repetition rate of the pulses determines the fundamental frequency (F0) of the vocalization and is perceived as pitch. As pulses created by the vocal folds pass through the vocal tract, the vocal tract characteristics (VTC) produce resonances and enhance/dampen particular frequency bands; these are called the formants. It has been well documented that both pitch and formant are highly important in primate communications, whereas how each acoustic characteristic contributes to vocal identification is not fully understood.

Several lines of evidence suggest that the formants created by the filter characteristics of the VTC play significant roles in the acoustic distinctiveness of individual primates, including humans. Previous study used and presented whisper speech, and the authors showed that humans were able to distinguish speakers by phonetic cues. (Tartter, 1991). Other previous study presented sine-wave sentences based on formant frequencies, the authors suggested that listeners were able to determine vocal individuals (Fellowes *et al.*, 1997; Remez *et al.*, 1997). Bachorowski and Owren (Bachorowski and Owren, 1999) analysed phonemes of speech in humans and showed that vocal tract filtering may contribute to individual identification. Other previous study measured vocal tract area during vocalizations of vowel by using MRI, the authors demonstrated that the inter-individual differences of the supralaryngeal spaces influence frequency in range beyond about 2.5 kHz (Kitamura *et al.*, 2005). Owren et al. (Owren *et al.*, 1997) analysed the vocalizations of female chacma baboons (*Papio ursinus*) and suggested that the acoustical features of vocal tract filtering may reflect individuality. The resonance of vocal tract filtering may affect individual identification in rhesus macaques (Rendall *et al.*, 1998) and lemurs (*Eulemur rubriventer*) (Gamba *et al.*, 2012). In addition to the formants, statistical analyses of the acoustic features of the F0, such as the beginning frequency and maximum frequency, indicate that the F0 can be a reliable cue for identifying callers in several monkey species (Smith *et al.*, 1982; Snowdon *et al.*, 1983). In relatively recent research by Ceugniet and Izumi (Ceugniet and Izumi, 2004), Japanese macaques were trained to discrimination the vocalizations of different monkeys, and the subjects responded to the F0 as a discriminant stimulus for the task, which suggests that the F0 contributes to individual discrimination.

In the present study, we used the contact calls of Japanese macaques to study individual vocal recognition. Green (Green, 1975) acoustically analysed and classified the vocalizations of Japanese macaques in the field and reported that Japanese macaques have several types of call. As a result of Green’s work, many other research groups have also focused on studying vocalization behaviours, and the Japanese macaque has become one of the most valuable and well-studied non-human primate models. These macaques exchange a coo call with one another when listening to the calls of other troop members (Mitani, 1986). The function of vocal exchange has been discussed in terms locating other individuals and maintaining within-group communication (Green, 1975). This study was performed to investigate the relative importance of acoustic cues (i.e., formant and pitch) in individual vocal recognition in Japanese macaques. We used operant conditioning and speech-processing techniques to systematically compare and quantify the perceptual contribution of each acoustic parameter.

4.2 Materials and methods

4.2.1 Subjects

Two male Japanese macaques (*Macaca fuscata*) were used in this experiment. At the time of testing, subject 1 was 7 years old and subject 2 was 10 years old. Two monkeys were trained to discriminate vocalizations of different two monkeys. Each animal was kept in an individual primate cage under a constant 13-h/11-h light/dark cycle. Their access to liquids was limited because water served as the positive reinforcement in the experiments. All procedures were conducted in accordance with guidelines established by the Ethics Review Committee of Doshisha University, and the experimental protocols were approved by the Animal Experimental Committee of Doshisha University.

4.2.2 Experimental apparatus

The training and tests were conducted in a sound-attenuated room (length × width × height of 1.70 m × 1.85 m × 2.65 m). The monkey chair in which the subjects were seated during the experiment was equipped with a drinking tube and a response lever. A loudspeaker (SX-WD1KT; Victor, Tokyo, Japan) was positioned 58 cm in front of the subject's head at the same height as the ears. All acoustic stimuli were amplified (SRP-P2400; Sony, Tokyo, Japan), and the frequency response of the speaker was flattened (± 3 dB) between 0.4 kHz and 16 kHz with a graphic equalizer (GQ2015A; Yamaha, Hamamatsu, Japan). A white light-emitting diode (LED) and a charge-coupled device (CCD) video camera were attached to the top of the speaker. An LED was lit during training and test trials to provide lighting, and subjects were monitored using the CCD camera.

4.2.3 Acoustic stimuli

The sound stimuli were obtained from two adult male monkeys (Monkey A and Monkey B). The coo calls of Monkey A (cooA) and Monkey B (cooB) were recorded using a condenser microphone (type 2142; Aco, Tokyo, Japan) and digital audio tape recorder (TCD-D8; Sony, Tokyo, Japan) with a resolution of 16 bits and a sampling rate of 44.1 kHz. The monkeys (Monkey A and Monkey B) who provided the coo calls had never encountered the subject monkeys (subjects 1 and 2), and this experiment was the first time that the subjects heard the voices of the stimulus monkeys. Fourteen coo calls (seven from each monkey) with signal-to-noise ratios > 40 dB were randomly selected from the recorded sounds.

The coo calls were analysed using STRAIGHT (Kawahara *et al.*, 1999) to measure three acoustic parameters of the coo calls (Fig. 4-1): the fundamental frequencies (F0s, Fig.4-2), vocal tract characteristics (VTCs, Fig. 4-3), and durations. Twelve coo calls (six coo calls per individual) of the total of fourteen were used as training stimuli (cooAs and cooBs). One coo call from each monkey was not played during training, and these calls were used to synthesize the test stimuli. The test stimuli coo calls were synthesized by combining the F0s and VTCs of the different individuals using STRAIGHT. Two types of test stimulus were synthesized as probes. The $F0_{cooA}$-VTC_{cooB} stimulus was synthesized from the F0 of cooA and the VTC of cooB, whereas the other test stimulus, $F0_{cooB}$-VTC_{cooA}, was generated from the F0 of cooB and the VTC of cooA (Fig. 4-4). The call durations were equalized to 517 ms (i.e., the average of all of the calls) via linearly time-stretching or compressing with STRAIGHT. With this manipulation, the duration of the original call was modified by 10% in the most extreme case. The root-mean-square (RMS) envelopes were calculated with a 512-point (≈12 ms) window, and the amplitude envelopes of all calls were normalized to average shape (Fig. 4-1). The overall amplitudes of stimuli were digitally modified and calibrated (with a microphone: type 7016; Aco) at to yield three different sound pressure levels (SPL, *re*: 20 μPa), i.e., 57, 60, and 63 dB, at the position of the head. That is, three different SPL stimuli were generated for each stimulus type. The fundamental frequencies of all of the calls were also modified, and the temporal average of the F0 was normalized to 733 Hz (i.e., the average of all of the original calls, Fig. 4-2), and the vocal tract characteristics remained unmodified (Fig. 4-3). In this study, we only use the synthesized stimulus for a test. Untrained cooA and B were never presented to the subjects, and were saved for a subsequent report.

4.2.4 Procedure

We employed standard Go/NoGo operant conditioning in this study. The event sequence of the trials is schematically illustrated in Fig. 4-5. The subjects were required to depress the lever switch on the monkey chair for 200 ms to begin the trial. Then, the calls from a single subject, either Monkey A or Monkey B, were repeated 3–7 times. In each repetition, the call type was randomly selected from 18 different types of call (6 types of coo call × 3 intensities from the same monkey). The interstimulus interval between adjacent stimuli was 800 ms. While the calls from the same monkey were presented (NoGo trial), the subjects were required to continue depressing the lever (correct rejection: CR). In other words, after a CR response, the next stimulus automatically began as long as an animal continued to hold the lever. After 3 to 7 repetitions, the stimulus was changed from one monkey to the other (Go trial). The subjects were required to release the lever within 800 ms of the offset of the stimulus (Hit). After a Hit response, the next trial did not begin until an animal depressed the lever again. For example, a trial began with the repetitive playback of cooAs (NoGo stimulus). In the repetition, the individual cooA (of the total of six) and the intensity of the stimulus (57, 60, and 63 dB SPL) were changed randomly. The subjects were required to continue depressing the lever while cooA was repeated. When cooB (Go stimulus) was presented, the subjects were required to release the lever within 800 ms after the offset of the cooB. Hits were reinforced with 2 ml of fruit juice. When the subjects released the lever during the repetition period of the NoGo stimulus (false alarm: FA) or failed to release the lever within 800 ms after the Go stimulus (miss), a 15–20 s timeout period accompanied by the turning off of the LED was provided as feedback. After an FA or miss response, a trial with same stimulus contingencies was provided. When the timeout period was over, the LED was lit to inform the animal of the initiation of a new trial. If the subject responded successfully to the Go stimulus, the stimulus contingencies were reversed in the next trial. That is, the next trial began with the playback of cooB instead of cooA, and the subject had to release the lever when cooA was played to receive the reward. Performance was measured as the correct response percentage (CRP: the total percentage of the Hits and CRs). One hundred thirty to 160 Go trials (i.e., trials in which the stimulus changed from one monkey to the other) and 650 to 800 NoGo trials were presented per day to both subjects.

After the subjects' scores exceeded the CRP threshold (70%) for two consecutive days, they proceeded to the test day. A test stimulus was presented, after cooB was repeated 5 times, and each type of test stimulus was played 6 times. The test trials were interleaved with 10-20 training trials. Neither reward nor punishment followed the test trial.

4.2.5 Statistical analysis

We measured both the Go response rates and reaction times (RTs, the time period between the end of each stimulus and the release of the lever switch). If the subjects did not release the lever within the 800 ms response period, the RT was regarded as 800 ms for the analysis. The CCD camera on the speaker allowed us to monitor the behaviour of each subject, and if the subject did not look straight into the speaker during the sound playback, the data in the trial were excluded from the analysis. The RTs to the test ($F0_{cooB}$-VTC_{cooA} and $F0_{cooA}$-VTC_{cooB}) and training stimuli were analysed by Mann-Whitney U test using a commercial statistical software package (SPSS 21; IBM Armonk, NY, US).

4.3 Results

Subject 1 and 2 needed 20 and 25 days of trainings respectively to learn to distinguish between the sets of cooAs and cooBs. Two days before the test day, the monkeys scored correct response rates of 82% (subject 1: d’ = 1.85, Hit = 80%, FA = 16%) and 76% (subject 2: d’ = 1.38, Hit = 75%, FA = 24%). The day before the test day, the correct response rates were 78% (subject 1: d’ = 1.54, Hit = 75%, FA=19%) and 71% (subject 2: d’ = 1.13, Hit = 77%, FA = 65%). The Go response rates to the training stimuli in the test day did not differ from those in the training day. In the test day, the correct response rates of subject 1 and subject 2 to the training stimuli were 76% (d’ = 1.49, Hit = 72%, FA = 20%) and 73% (d’ = 1.30, Hit = 81%, FA= 34%), respectively, suggesting that the subjects maintained the same discriminatory performance with the training stimuli throughout the experiment. The Go response rates to the test stimuli for the two monkeys are illustrated in Fig. 4-6. The Go response rates to $F0_{cooA}$-VTC_{cooB} (Fig. 4-6), which had the same F0 as the Go stimulus (= cooA) and the same VTC as the NoGo stimulus (= cooB), of subjects 1 and 2 were 16.7% and 33.3%, respectively. The Go response rates of subjects 1 and 2 to $F0_{cooB}$-VTC_{cooA} (Fig. 4-6) were 83.3% and 83.3%, respectively. Our data revealed that $F0_{cooB}$-VTC_{cooA} triggered more Go responses from both monkeys than $F0_{cooA}$-VTC_{cooB}.

The RTs to the test stimuli were examined to quantify the perceptual similarity of the stimuli (Table 4-1). The median RTs of subjects 1 and 2 to $F0_{cooA}$-VTC_{cooB} were 800 (interquartile range: 753–800) ms and 800 (391–800) ms, respectively. In contrast, the median RTs of subjects 1 and 2 to $F0_{cooB}$-VTC_{cooA} were 368 (276–592) ms and 230 (161–499) ms, respectively (Fig. 4-7). The median RTs to $F0_{cooA}$-VTC_{cooB} and $F0_{cooB}$-VTC_{cooA} were compared with those to the training stimuli. Because the test stimulus was 60 dB sound pressure level (SPL), the training stimulus with same 60 dB level was treated as a comparison stimulus. The stimulus was presented 40 and 45 times to subjects 1 and 2, respectively, in the test day. Of those repetitions, 3 (in subject 1) and 4 (in subject 2) presentations were excluded from the analyses because the monkeys’ heads were not oriented towards the speaker during the presentations. The RTs to $F0_{cooB}$-VTC_{cooA} were not significantly different from those to the Go stimulus (cooA) in either subject 1 ($F0_{cooB}$-VTC_{cooA}: 368 (276–592) ms, Go stimulus: 416 (351–558) ms; $p = 0.93$) or subject 2 ($F0_{cooB}$-VTC_{cooA}: 230

(161–499) ms, Go stimulus: 226 (108–321) ms; *p* = 0.33). Additionally, the median RT of subject 1 to the NoGo stimulus was 800 (800–800) ms and that of subject 2 was 800 (581–800) ms. There were no significant differences between the RTs of either subject to F0$_{\text{cooA}}$-VTC$_{\text{cooB}}$ and the NoGo stimuli in the test day (Fig. 4-7, subject 1: *p* = 0.93; subject 2: *p* = 0.88).

4.4 Discussion

Two monkeys were able to discriminate vocalizations of two unfamiliar monkeys, while correct response rates were lower than correct response rates of previous our experiment. Previous our study used coo calls that were different fundamental frequencies. In present study, we used vocalizations of monkeys with same mean fundamental frequencies. Monkeys might be difficult to discriminate coo calls of present our study because acoustic characteristics of vocalizations to discriminate coo calls decreased. However, correct response rate of two subjects were above 75%. Thus, monkeys were able to discriminate coo calls with same fundamental frequencies.

We used acoustic synthesis and analysis software to systematically quantify the relative importance of acoustic characteristics (i.e., the VTC and the temporal structure of the F0) when the monkeys identify callers. The behavioural data suggest that the animals perceived the $F0_{cooA}$-VTC_{cooB} as the same as cooB, whereas they perceived $F0_{cooB}$-VTC_{cooA} as the same as cooA instead of recognizing them as intermediate between the two stimuli. When only the VTC was switched from one type to the other, the subjects still responded as if the call type had transitioned, whereas the animals did not respond if only the temporal pattern of F0 changed (Fig. 4-6). The subjects' behavioural responses revealed that the VTC played a critical role in distinguishing the stimulus sets, suggesting that monkeys relied more on the VTC than on the temporal pitch patterns in discriminating caller identity. The difference in the temporal pattern of the F0 may have been too small to enable the monkeys to differentiate the stimulus set, but we believe that this was not the case. Hopp et al. (Hopp *et al.*, 1992) studied the sensitivity of Japanese macaques to the peak position of F0 in synthesized coo calls and demonstrated that trained animals were able to detect changes in the peak position of as little as 20–50 ms in smooth early high coos. The F0 of the cooA peak was earlier than that of the cooB peak by approximately 60 ms (the peak position of the vocalizations of Monkey A was 195 ± 22 ms and that of Monkey B was 134 ± 45 ms [average ± standard deviation]). Thus, the subjects were able to distinguish the stimulus sets using the peak position of the vocalizations in this experiment.

Monkeys are also able to discriminate vocalizations using the end frequencies of the stimuli. A previous study using pure-tone bursts of 1000 Hz revealed that Japanese macaques are able to distinguish frequency differences limens as small as 33 Hz (i.e., a difference of approximately 3%) (Sinnott *et al.*, 1985). In our stimulus set, the mean frequencies of the stimuli were normalized, and the temporal patterns of F0 were maintained (Fig. 4-2). Therefore, the end frequencies of cooA were lower than those of cooB by approximately 120 Hz (cooA: 578 ± 57 Hz; cooB: 706 ± 26 Hz) or 15%. Thus, it is reasonable to assume that the subjects were able to distinguish the stimulus sets according to the end frequency in addition to the peak timing.

There are still several questions that remain to be answered. Whereas the past studies described above suggest that the monkeys were able to discriminate our stimulus sets by the temporal patterns of F0. It is probable that the F0 differences were sufficiently salient for use as discriminative cues compared with the VTCs. In contrast, the significance of the VTCs in the monkeys' discrimination does not necessarily mean that the VTC is only cue that used for individual discrimination. To address these questions, we would need to quantify the contribution (if any) of the F0 to the discrimination using synthesized calls without differences in VTC (i.e., vocal signals with the same VTC that differ only in the F0) and also measure the perceptual threshold of the F0 components. In addition to those studies, because our data demonstrated that the speech-processing techniques (STRAIGHT, (Kawahara *et al.*, 1999) provide reliable behavioural data, we can now create a stimulus continuum between different individuals and systematically investigate the relationships between the acoustic parameters and vocal identification.

As described in non-primate species (Reby and McComb, 2003; Reby *et al.*, 2005), the formants embedded in the acoustic structures of nonhuman primate calls provide cues about the physical characteristics of the caller (Owren *et al.*, 1997; Bachorowski and Owren, 1999; Rendall, 2003). A previous study using a preferential looking paradigm suggested that untrained rhesus monkeys use formants as indexical cues of age-related body size (Ghazanfar *et al.*, 2007). Fitch and Fritz (Fitch and Fritz, 2006) also demonstrated that nonhuman primates can perceive formant shifts in species-specific vocalizations. Owren (Owren, 1990) demonstrated that trained vervet monkeys can use formants to discriminate between their alarm calls in a manner similar to that used by humans to distinguish speech sounds. Similar to humans, with training, Japanese macaques exhibit exquisite sensitivity to different formant frequencies (Sommers *et al.*, 1992). These results indicate that formants are biologically significant in the vocal communication of many primate species.

In addition to formants, pitch has also been demonstrated to be important for communication. Japanese macaques are regarded as sensitive to the temporal patterns of the F0, particularly in coo calls, because the peak temporal position differentiates the call type; i.e., smooth early high and smooth late high (Zoloth *et al.*, 1979; May *et al.*, 1989). The F0 has also been reported to differ between individuals in several primate species, and the F0 is a statistically significant determinant of caller identity (Smith *et al.*, 1982; Snowdon *et al.*, 1983). To our knowledge, however, there have been only a few attempts to directly compare the importance of the VTC and F0 in identification. Ceugniet and Izumi (Ceugniet and Izumi, 2004) trained two Japanese macaques to discriminate the vocalizations of different individuals using operant conditioning; these authors demonstrated that macaques judge individuality via a combination of both the VTC and the frequency of the F0. In addition, chimpanzees used pitch for discriminating individuals (Kojima *et al.*, 2003). Thus, the dominant acoustic cues in the determination of individuality in non-human primates are still largely unknown. Our data indicated that the formant frequencies generated by the VTC were preferentially used over the F0 temporal structures to discriminate the stimulus sets, which strengthens the suggestion that the formant structure is significant for the perception of conspecific sounds and also possibly for individual identification.

This experiment was performed to determine the primary cues that are used for the identification of individuals. However, the monkeys may have only *discriminated* between the features of two sets of vocalizations rather than *identifying* the individual the caller. Further studies are required to determine whether monkeys perceive the stimulus sets as the vocalizations of two different monkeys.

4.5 Conclusion

Many primates, including humans, can discriminate individuality based only on listening to vocalizations. Two monkeys were trained to discriminate vocalizations of two unfamiliar individuals. Our experiments directly compared the relative importance of acoustic parameters (F0 and VTC) in Japanese macaques, and the results suggest that VTCs are more important for discriminating the caller than the temporal structure of the fundamental frequency.

Chapter 5
Conclusions

This dissertation compared the vocal recognition between Japanese macaques and humans using behavioral techniques. Primates, including humans, must recognize both the contents and the individuals of vocalizations from conspecifics or predators accurately, because they maintain social interactions and this increases their opportunities of survival and mating rates. In addition, comparative studies among primates are necessary to reveal whether the voice recognition was evolutionarily maintained in primates. This chapter provides a summary of the major results and a discussion of the future work in this thesis.

5.1 Summary of major results

5.1.1 Perception of amplitude-modulated broadband noise in primates (Chapter 2)

Temporal fluctuations in amplitude envelopes of sound are important in perceiving speech in humans (Sinnott *et al.*, 1976). As a comparison of basic hearing abilities, we compared the sensitivities of amplitude modulation in monkeys and humans using broadband noise. Two monkeys and three humans were trained to discriminate continuous and repeated white noise bursts using standard Go/NoGo operant conditioning. The sensitivities of subjects were quantified using amplitude modulation noise with various modulation depths. In our data, monkeys were 2.5-fold (8 dB) less sensitive to detecting modulation depths of broadband noise than were humans. These results showed that the temporal processing among primates differed, which may provide a greater understanding of the differences in temporal perception between humans and non-human primates.

5.1.2 Acoustic characteristics for discriminating conspecific and heterospecific vocalizations (Chapter 3)

Chapter 3 describes continuum vocalizations modified using auditory signal processing software (STRAIGHT). We confirmed that the signal-processing algorithm was appropriate for morphed stimuli between the vocalizations of two monkeys. Both monkeys and humans were trained using standard Go/NoGo operant conditioning to distinguish the vocalizations of the two unfamiliar monkeys. The continuum stimuli between the voices of the two individuals were generated using STRAIGHT. The reaction times to the stimuli were measured in the subjects. In our data, the responses of the subjects were correlated with changes in stimuli from one individual to another. Our results demonstrated that continuum stimuli among the vocalizations of different monkeys were generated successfully using STRAIGHT.

We also investigated the acoustic features used to discriminate individuals based only on voices in monkeys and humans, because several studies have debated how acoustic characteristics (F0 and VTC) contribute to individual discrimination in primates. We trained monkeys and humans to distinguish vocalizations from individuals using standard Go/NoGo operant conditioning. We created two sets of continuum stimuli in which only one acoustic feature, F0 or VTC, was changed from the two monkeys, while another acoustic characteristic was maintained from one individual. The reaction times to these stimuli were measured in monkeys and humans. The reaction times of the monkey subjects were correlated with the changes in F0 and VTC. However, the reaction times of the human subjects were correlated with the changes in F0, but humans did not respond to changes in VTC. These results suggested that the specific auditory characteristics could be modified flexibly using STRAIGHT. In addition, our data may indicate that primates use different acoustic features to discriminate conspecific and heterospecific vocalizations.

5.1.3 Acoustic features for individual discrimination in monkeys (Chapter 4)

Chapter 4 explains the acoustic features used by Japanese macaques to discriminate individuals. Monkeys have been found to discriminate individuals based only on their voices, but there is still debate regarding how the F0s and filter properties of the VTC contribute to individual discrimination in non-human primates. This study was performed to investigate the acoustic keys used by Japanese macaques in individual discrimination. Two animals were trained using standard Go/NoGo operant conditioning to distinguish the coo calls of two unfamiliar monkeys. The subjects were required to continue depressing a lever until the stimulus changed from one monkey to the other. The test stimuli were synthesized by combining the F0 and VTC from the two individuals. Both subjects released the lever when the VTC changed, whereas they did not when the F0 changed. The reaction times to the test stimuli were not significantly different among the training stimuli that shared the same VTC. Our data suggest that VTC are important for the identification of individuals by Japanese macaques.

5.2 Future works

In this dissertation project, subjects were trained to discriminate only two vocalizations of monkeys. This thesis investigated only acoustic features to *discriminate* vocalizations of individuals rather than *identify* them. Thus, this thesis does not provide substantiation for the identification of individuals by primates. Further studies are required using other behavioral protocols to investigate identifying individuals based on vocalizations alone. For example, subjects could be trained to detect specific individuals among various individuals.

We examined the discrimination of individuals based on only vocalizations in Japanese macaques and humans. However, individual identification requires memorizing and linking multiple information sources, such as faces and vocalizations. Thus, in addition to the procedure described above, a face-to-voice matching task could be used as another example behavioral protocol.

This study compared psychoacoustics between monkeys and humans using only behavioral protocols. Individual recognition requires combining multiple sources of information in the brain, such as faces and voices. However, little is known about how single neurons in the brain identify individuals based on vocalizations. Further studies are required to determine the neural activity behind individual discrimination based on voices.

We investigated acoustic characteristics to discriminate conspecific and heterospecific individuals based on vocalizations. However, this thesis presented the vocalizations of only monkeys to Japanese macaques and humans. Further studies are needed to determine which acoustic features are used by monkeys to discriminate individual humans.

We used only one type of vocalization (coo calls) from Japanese monkeys, but Japanese macaques possess a larger repertoire of vocalizations. Japanese macaques may also be able to distinguish individuals using other types of calls. Further study is needed to investigate the acoustic characteristics of other types of calls used to distinguish individuals.

We examined the vocal recognition of Japanese macaques, because the vocal communication of this species has been evaluated previously in field studies. Comparisons between humans and non-human primates that are closer to humans evolutionarily (e.g., chimpanzee, bonobo) than are Japanese macaques are required to discuss the evolution of cognitive functions involved in hearing. Further studies are required to investigate the auditory recognition of the non-human primates that are closest evolutionarily to humans.

5.3 Final remarks

The purpose of this study was to investigate the vocal recognition in Japanese macaques and humans. This thesis examined the temporal resolutions of both Japanese macaques and humans. In addition, the acoustic characteristics used to discriminate individuals based on conspecific and heterospecific vocalizations were investigated. The temporal resolution results demonstrated that humans were more sensitive to detecting amplitude modulation than were Japanese monkeys. Moreover, our data about individual discrimination showed that monkeys and humans seemingly use different acoustic characteristics to distinguish conspecific and heterospecific vocalizations, and formants contributed to discriminating individuals based on vocalization in monkeys rather than the temporal structures of F0s. We demonstrated that Japanese macaques performed the individual discrimination by using same acoustic features that humans discriminated speakers by vocalizations alone. Our results may imply that common ancestor of humans and Japanese monkeys used vocal tract characteristics to discriminate individuals. In addition, these results showed that Japanese macaques might be established as the model animal for individual recognition based on vocalizations. Further studies are need to investigate the neural activity behind individual discrimination based on vocalizations.

References

Ackermann, H., Hage, S. R., and Ziegler, W. (**2014**). "Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective," Behav. Brain Sci. **37**, 529-546.

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., and Vidnyánszky, Z. (**2010**). "Neural mechanisms for voice recognition," Neuroimage **52**, 1528-1540.

Arnold, K., and Zuberbühler, K. (**2006**). "Language evolution: semantic combinations in primate calls," Nature **441**, 303-303.

Arnold, K., and Zuberbühler, K. (**2008**). "Meaningful call combinations in a non-human primate," Curr. Biol. **18**, R202-R203.

Bachorowski, J.-A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054-1063.

Belin, P. (**2006**). "Voice processing in human and non-human primates," Philosophical Transactions of the Royal Society B: Biological Sciences **361**, 2091-2107.

Belin, P., Fecteau, S., and Bedard, C. (**2004**). "Thinking the voice: neural correlates of voice perception," Trends Cogn. Sci. **8**, 129-135.

Belin, P., Zatorre, R. J., and Ahad, P. (**2002**). "Human temporal-lobe response to vocal sounds," Cognitive Brain Research **13**, 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (**2000**). "Voice-selective areas in human auditory cortex," Nature **403**, 309-312.

Brown, C. H., Beecher, M. D., Moody, D. B., and Stebbins, W. C. (**1979**). "Locatability of vocal signals in Old World monkeys: Design features for the communication of position," J. Comp. Physiol. Psychol. **93**, 806.

Bruce, C., Desimone, R., and Gross, C. G. (**1981**). "Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque," J. Neurophysiol. **46**, 369-384.

Ceugniet, M., and Izumi, A. (**2004**). "Vocal individual discrimination in Japanese monkeys," Primates **45**, 119-128.

Chakladar, S., Logothetis, N. K., and Petkov, C. I. (**2008**). "Morphing rhesus monkey vocalizations," J. Neurosci. Methods **170**, 45-55.

Cheney, D. L., and Seyfarth, R. M. (**1980**). "Vocal recognition in free-ranging vervet monkeys," Anim. Behav. **28**, 362-367.

Childers, D. G., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841-1856.

Dahl, C. D., Wallraven, C., Bülthoff, H. H., and Logothetis, N. K. (**2009**). "Humans and macaques employ similar face-processing strategies," Curr. Biol. **19**, 509-513.

Delson, E., and Rosenberger, A. (**1980**). "Phyletic perspectives on platyrrhine origins and anthropoid relationships," in *Evolutionary biology of the new world monkeys and continental drift* (Springer), pp. 445-458.

Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (**1984**). "Stimulus-selective properties of inferior temporal neurons in the macaque," The Journal of Neuroscience **4**, 2051-2062.

Dufour, V., Pascalis, O., and Petit, O. (**2006**). "Face processing limitation to own species in primates: a comparative study in brown capuchins, Tonkean macaques and humans," Behav. Processes **73**, 107-113.

Fant, G. (**1971**). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations* (Walter de Gruyter).

Fecteau, S., Armony, J. L., Joanette, Y., and Belin, P. (**2004**). "Is voice processing species-specific in human auditory cortex? An fMRI study," Neuroimage **23**, 840-848.

Fellowes, J. M., Remez, R. E., and Rubin, P. E. (**1997**). "Perceiving the sex and identity of a talker without natural vocal timbre," Percept. Psychophys. **59**, 839-849.

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213-1222.

Fitch, W. T. (**2000**). "The evolution of speech: a comparative review," Trends Cogn. Sci. **4**, 258-267.

Fitch, W. T., de Boer, B., Mathur, N., and Ghazanfar, A. A. (**2016**). "Monkey vocal tracts are speech-ready," Sci. Adv. **2**, e1600723.

Fitch, W. T., and Fritz, J. B. (**2006**). "Rhesus macaques spontaneously perceive formants in conspecific vocalizations," J. Acoust. Soc. Am. **120**, 2132-2141.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511-1522.

Fleagle, J. G. (**2013**). *Primate adaptation and evolution* (Academic Press).

Fleagle, J. G., and McGraw, W. S. (**1999**). "Skeletal and dental morphology supports diphyletic origin of baboons and mandrills," Proc. Natl. Acad. Sci. **96**, 1157-1161.

Fossey, D. (**1972**). "Vocalizations of the mountain gorilla (Gorilla gorilla beringei)," Anim. Behav. **20**, 36-53.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (**2001**). "Categorical representation of visual stimuli in the primate prefrontal cortex," Science **291**, 312-316.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (**2003**). "A comparison of primate prefrontal and inferior temporal cortices during visual categorization," J. Neurosci. **23**, 5235-5246.

Fukuda, F. (**1988**). "Influence of artificial food supply on population parameters and dispersal in the Hakone T troop of Japanese macaques," Primates **29**, 477-492.

Furl, N., van Rijsbergen, N. J., Treves, A., and Dolan, R. J. (**2007**). "Face adaptation aftereffects reveal anterior medial temporal cortex role in high level category representation," Neuroimage **37**, 300-310.

Gamba, M., Colombo, C., and Giacoma, C. (**2012**). "Acoustic cues to caller identity in lemurs: a case study," J. ethol. **30**, 191-196.

Ghazanfar, A. A., and Rendall, D. (**2008**). "Evolution of human vocal production," Curr. Biol. **18**, R457-R460.

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., and Logothetis, N. K. (**2007**). "Vocal-tract resonances as indexical cues in rhesus monkeys," Curr. Biol. **17**, 425-430.

Green, S. (**1975**). "Variation of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): a field study," Primate behavior **4**, 1-102.

Greenberg, S., and Takayuki, A. (**2004**). "What are the essential cues for understanding spoken language?," IEICE transactions on information and systems **87**, 1059-1070.

Gross, C. G. (**2008**). "Single neuron studies of inferior temporal cortex," Neuropsychologia **46**, 841-852.

Hartman, D. E. (**1979**). "The perceptual identity and characteristics of aging in normal male adult speakers," J. Commun. Disord. **12**, 53-61.

Hartman, D. E., and Danhauer, J. L. (**1976**). "Perceptual features of speech for males in four perceived age decades," J. Acoust. Soc. Am. **59**, 713-715.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (**2000**). "The distributed human neural system for face perception," Trends Cogn. Sci. **4**, 223-233.

Heffner, H. E., and Heffner, R. S. (**1984**). "Temporal lobe lesions and perception of species-specific vocalizations by macaques," Science **226**, 75-76.

Heffner, R. S. (**2004**). "Primate hearing from a mammalian perspective," Anat. Rec. A Discov. Mol. Cell Evol. Biol. **281**, 1111-1122.

Hopp, S. L., Sinnott, J. M., Owren, M. J., and Petersen, M. R. (**1992**). "Differential sensitivity of Japanese macaques (*Macaca fuscata*) and humans (*Homo sapiens*) to peak position along a synthetic coo call continuum," J. Comp. Psychol. **106**, 128.

Houtgast, T., and Steeneken, H. J. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069-1077.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., Itoh, K., Kato, T., Nakamura, A., and Hatano, K. (**1997**). "Vocal identification of speaker and emotion activates differerent brain regions," Neuroreport **8**, 2809-2812.

Itani, J. (**1963**). "Vocal communication of the wild Japanese monkey," Primates **4**, 11-66.

Jackson, L. L., Heffner, R. S., and Heffner, H. E. (**1999**). "Free-field audiogram of the Japanese macaque (*Macaca fuscata*)," J. Acoust. Soc. Am. **106**, 3017-3023.

Janik, V. M., and Slater, P. J. (**2000**). "The different roles of social learning in vocal communication," Anim. Behav. **60**, 1-11.

Jovanovic, T., Megna, N. L., and Maestripieri, D. (**2000**). "Early maternal recognition of offspring vocalizations in rhesus macaques (*Macaca mulatta*)," Primates **41**, 421-428.

Kano, F., and Tomonaga, M. (**2010**). "Face scanning in chimpanzees and humans: Continuity and discontinuity," Anim. Behav. **79**, 227-235.

Kanwisher, N., McDermott, J., and Chun, M. M. (**1997**). "The fusiform face area: a module in human extrastriate cortex specialized for face perception," J. Neurosci. **17**, 4302-4311.

Kanwisher, N., and Yovel, G. (**2006**). "The fusiform face area: a cortical region specialized for the perception of faces," Philos. Trans. R. Soc. Lond. B Biol. Sci. **361**, 2109-2128.

Kaplan, J. N., Winship-Ball, A., and Sim, L. (**1978**). "Maternal discrimination of infant vocalizations in squirrel monkeys," Primates **19**, 187-193.

Katsu, N., Yamada, K., and Nakamichi, M. (**2014**). "Development in the Usage and Comprehension of Greeting Calls in a Free ‐ Ranging Group of Japanese Macaques (Macaca fuscata)," Ethology **120**, 1024-1034.

Kawahara, H., Masuda-Katsuse, I., and De Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech communication **27**, 187-207.

Kitamura, T., Honda, K., and Takemoto, H. (**2005**). "Individual variation of the hypopharyngeal cavities and its acoustic effects," Acoust. Sci. Technol. **26**, 16-26.

Koda, H. (**2004**). "Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques (Macaca fuscata yakui)," Behaviour **141**, 1279-1296.

Koda, H., Tokuda, I. T., Wakita, M., Ito, T., and Nishimura, T. (**2015**). "The source-filter theory of whistle-like calls in marmosets: Acoustic analysis and simulation of helium-modulated voices," J. Acoust. Soc. Am. **137**, 3068-3076.

Kojima, S., Izumi, A., and Ceugniet, M. (**2003**). "Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee," Primates **44**, 225-230.

Kuhl, P. K., and Padden, D. M. (**1983**). "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," J. Acoust. Soc. Am. **73**, 1003-1010.

Lass, N. J., and Davis, M. (**1976**). "An investigation of speaker height and weight identification," J. Acoust. Soc. Am. **60**, 700-703.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (**1976**). "Speaker sex identification from voiced, whispered, and filtered isolated vowels," J. Acoust. Soc. Am. **59**, 675-678.

Latinus, M., Crabbe, F., and Belin, P. (**2011**). "Learning-induced changes in the cerebral processing of voice identity," Cereb. Cortex **21**, 2820-2828.

Leopold, D. A., Bondar, I. V., and Giese, M. A. (**2006**). "Norm-based face encoding by single neurons in the monkey inferotemporal cortex," Nature **442**, 572-575.

Leopold, D. A., O'Toole, A. J., Vetter, T., and Blanz, V. (**2001**). "Prototype-referenced shape encoding revealed by high-level aftereffects," Nat. Neurosci. **4**, 89-94.

Lloyd, P. (**2005**). "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry," Acoust. Soc. Amer **117**, 994-1005.

May, B., Moody, D. B., and Stebbins, W. C. (**1989**). "Categorical perception of conspecific communication sounds by Japanese macaques, Macacafuscata," J. Acoust. Soc. Am. **85**, 837-847.

McAulay, R., and Quatieri, T. (**1986**). "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust. **34**, 744-754.

Mitani, M. (**1986**). "Voiceprint identification and its application to sociological studies of wild Japanese monkeys (Macaca fuscata yakui)," Primates **27**, 397-412.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**, 331-340.

Mori, A. (**1975**). "Signals found in the grooming interactions of wild Japanese monkeys of the Koshima troop," Primates **16**, 107-140.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., and Ito, K. (**2001**). "Neural substrates for recognition of familiar voices: a PET study," Neuropsychologia **39**, 1047-1054.

Narendranath, M., Murthy, H. A., Rajendran, S., and Yegnanarayana, B. (**1995**). "Transformation of formants for voice conversion using artificial neural networks," Speech Commun. **16**, 207-216.

O'Connor, K. N., Barruel, P., and Sutter, M. L. (**2000**). "Global processing of spectrally complex sounds in macaques (*Macaca mullata*) and humans," J. Comp. Physiol. A **186**, 903-912.

O’Connor, K. N., Johnson, J. S., Niwa, M., Noriega, N. C., Marshall, E. A., and Sutter, M. L. (**2011**). "Amplitude modulation detection as a function of modulation frequency and stimulus duration: comparisons between macaques and humans," Hear. Res. **277**, 37-43.

Owren, M. J. (**1990**). "Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*): II. Synthetic calls," J. Comp. Psychol. **104**, 29.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1993**). "Vocalizations of rhesus (Macaca mulatta) and Japanese (M. Fuscata) macaques cross‐fostered between species show evidence of only limited modification," Dev. Psychobiol. **26**, 389-406.

Owren, M. J., Hopp, S. L., Sinnott, J. M., and Petersen, M. R. (**1988**). "Absolute auditory thresholds in three Old World monkey species (*Cercopithecus aethiops, C. neglectus, Macaca fuscata*) and humans (*Homo sapiens*)," J. Comp. Psychol. **102**, 99.

Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like grunt calls in chacma baboons (Papio cyncephalus ursinus): Implications for production processes and functions," J. Acoust. Soc. Am. **101**, 2951-2963.

Parr, L. A., and de Waal, F. B. (**1999**). "Visual kin recognition in chimpanzees," Nature **399**, 647-648.

Parr, L. A., Winslow, J. T., Hopkins, W. D., and de Waal, F. (**2000**). "Recognizing facial cues: individual discrimination by chimpanzees (*Pan troglodytes*) and rhesus monkeys (Macaca mulatta)," J. Comp. Psychol. **114**, 47.

Pereira, M. E. (**1986**). "Maternal recognition of juvenile offspring coo vocalizations in Japanese macaques," Anim. Behav. **34**, 935-937.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175-184.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (**2008**). "A voice region in the monkey brain," Nat. Neurosci. **11**, 367-374.

Pfingst, B. E., Hienz, R., Kimm, J., and Miller, J. (**1975a**). "Reaction− time procedure for measurement of hearing. I. Suprathreshold functions," J. Acoust. Soc. Am. **57**, 421-430.

Pfingst, B. E., Hienz, R., and Miller, J. (**1975b**). "Reaction time procedure for measurement of hearing. II. Threshold functions," J. Acoust. Soc. Am. **57**, 431-436.

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (**2004**). "Species-specific calls evoke asymmetric activity in the monkey's temporal poles," Nature **427**, 448-451.

Prosen, C., Moody, D., Sommers, M., and Stebbins, W. (**1990**). "Frequency discrimination in the monkey," J. Acoust. Soc. Am. **88**, 2152-2158.

Reby, D., and McComb, K. (**2003**). "Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags," Anim. Behav. **65**, 519-530.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., and Clutton-Brock, T. (**2005**). "Red deer stags use formants as assessment cues during intrasexual agonistic interactions," Proc. R. Soc. Lond. B Biol. Sci. **272**, 941-947.

Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identification based on phonetic information," J. Exp. Psychol. Hum. Percept. Perform. **23**, 651.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947-949.

Rendall, D. (**2003**). "Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons," J. Acoust. Soc. Am. **113**, 3390-3402.

Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations," J. Acoust. Soc. Am. **103**, 602-614.

Rendall, D., Rodman, P. S., and Emond, R. E. (**1996**). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," Anim. Behav. **51**, 1007-1015.

Riquimaroux, H. (**2006**). "Perception of noise-vocoded speech sounds: Sentences, words, accents and melodies," Acoust. Sci. Technol. **27**, 325-331.

Romanski, L. M., Averbeck, B. B., and Diltz, M. (**2005**). "Neural representation of vocalizations in the primate ventrolateral prefrontal cortex," J. Neurophysiol. **93**, 734-747.

Rowell, T. E., and Hinde, R. (**1962**). "Vocal communication by the rhesus monkey (*macaca mulatta*)," in *Proceedings of the Zoological Society of London*, pp. 279-294.

Scherer, K. R. (**1995**). "Expression of emotion in voice and music," J. Voice **9**, 235-248.

Sergent, J., Signoret, J.-L., Bruce, V., and Rolls, E. (**1992**). "Functional and anatomical decomposition of face processing: Evidence from prosopagnosia and PET study of normal subjects [and discussion]," Philosophical Transactions of the Royal Society of London B: Biological Sciences **335**, 55-62.

Seyfarth, R. M., and Cheney, D. L. (**1986**). "Vocal development in vervet monkeys," Anim. Behav. **34**, 1640-1658.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (**1980**). "Vervet monkey alarm calls: semantic communication in a free-ranging primate," Anim. Behav. **28**, 1070-1094.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303.

Sigala, R., Logothetis, N. K., and Rainer, G. (**2011**). "Own-species bias in the representations of monkey and human face categories in the primate temporal lobe," J. Neurophysiol. **105**, 2740-2752.

Sinnott, J., Beecher, M. D., Moody, D., and Stebbins, W. (**1976**). "Speech sound discrimination by monkeys and humans," J. Acoust. Soc. Am. **60**, 687-695.

Sinnott, J. M., and Adams, F. S. (**1987**). "Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts," J. Acoust. Soc. Am. **82**, 1539-1547.

Sinnott, J. M., and Brown, C. H. (**1997**). "Perception of the American English liquid/ra–la/contrast by humans and monkeys," J. Acoust. Soc. Am. **102**, 588-602.

Sinnott, J. M., Petersen, M. R., and Hopp, S. L. (**1985**). "Frequency and intensity discrimination in humans and monkeys," J. Acoust. Soc. Am. **78**, 1977-1985.

Skuk, V. G., Dammann, L. M., and Schweinberger, S. R. (**2015**). "Role of timbre and fundamental frequency in voice gender adaptation," J. Acoust. Soc. Am. **138**, 1180-1193.

Smith, D. R., and Patterson, R. D. (**2005**). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and agea)," J. Acoust. Soc. Am. **118**, 3177-3186.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305-318.

Smith, H. J., Newman, J. D., Hoffman, H. J., and Fetterly, K. (**1982**). "Statistical discrimination among vocalizations of individual squirrel monkeys (*Saimiri sciureus*)," Folia Primatol. (Basel) **37**, 267-279.

Snowdon, C. T., and Cleveland, J. (**1980**). "Individual recognition of contact calls by pygmy marmosets," Anim. Behav. **28**, 717-727.

Snowdon, C. T., Cleveland, J., and French, J. A. (**1983**). "Responses to context-and individual-specific cues in cotton-top tamarin long calls," Anim. Behav. **31**, 92-101.

Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (**1992**). "Formant frequency discrimination by Japanese macaques (Macacafuscata)," J. Acoust. Soc. Am. **91**, 3499-3510.

Sugiura, H. (**1993**). "Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques," Behaviour **124**, 207-225.

Sugiura, H. (**1998**). "Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques," Anim. Behav. **55**, 673-687.

Takahata, Y., Suzuki, S., Agetsuma, N., Okayasu, N., Sugiura, H., Takahashi, H., Yamagiwa, J., Izawa, K., Furuichi, T., and Hill, D. A. (**1998**). "Reproduction of wild Japanese macaque females of Yakushima and Kinkazan Islands: a preliminary report," Primates **39**, 339-349.

Tanaka, K., Saito, H.-a., Fukada, Y., and Moriya, M. (**1991**). "Coding visual images of objects in the inferotemporal cortex of the macaque monkey," J. Neurophysiol. **66**, 170-189.

Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Percept. Psychophys. **49**, 365-372.

Taylor, A. M., and Reby, D. (**2010**). "The contribution of source–filter theory to mammal vocal communication research," J. Zool. **280**, 221-236.

Veldhuis, R., and He, H. (**1996**). "Time-scale and pitch modifications of speech signals and resynthesis from the discrete short-time Fourier transform," Speech Communication **18**, 257-279.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364-1380.

Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (**1995**). "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics," J. Neurophysiol. **74**, 2685-2706.

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (**2004**). "Adaptation to natural facial categories," Nature **428**, 557-561.

Winter, P., and Funkenstein, H. H. (**1973**). "The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey (*Saimiri sciureus*)," Exp. Brain Res. **18**, 489-504.

Winter, P., Ploog, D., and Latta, J. (**1966**). "Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance," Exp. Brain Res. **1**, 359-384.

Wu, K., and Childers, D. G. (**1991**). "Gender recognition from speech. Part I: Coarse analysis," J. Acoust. Soc. Am. **90**, 1828-1840.

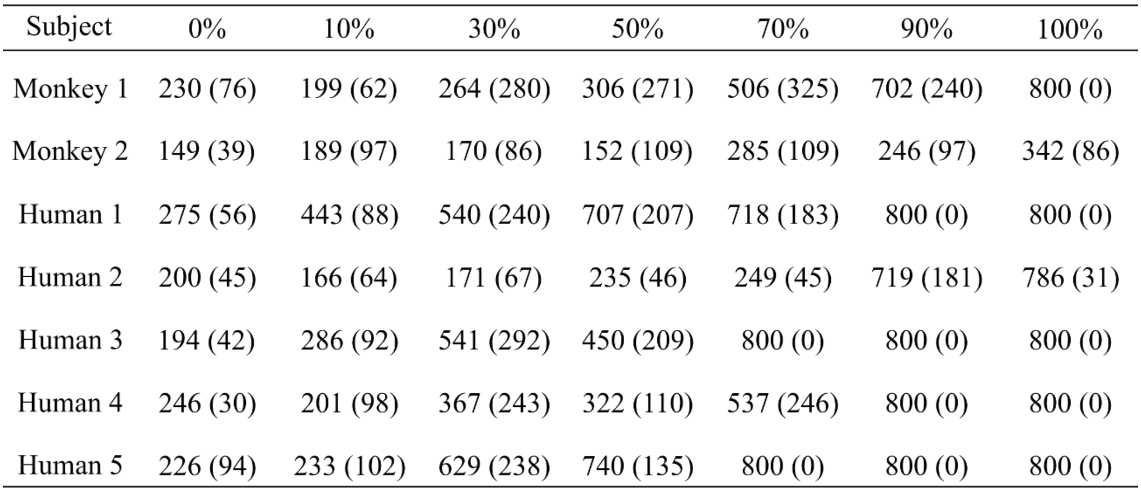
Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B., and Stebbins, W. (**1979**). "Species-specific perceptual processing of vocal sounds by monkeys," Science **204**, 870-873.

Table 2-1. Mean reaction times (RT, mean ± standard deviation [SD]) and z-scores to both of training and test stimuli. The z-scores were based on the average reaction time to the continuous white noise bursts (modulation depth = 0%) from each subject. A z-score exceeding 1.96 was considered statistically significant (*p < 0:05). Maximum modulation depth was denoted as ''> 97%'' in the leftmost column because background noise had sound pressure level of about 30 dB (re: 20μPa) or less.

Modulation depth (%)		Monkeys				Humans					
		Monkey 1		Monkey 2		Human 1		Human 2		Human 3	
		RT (ms) Mean ± *SD*	*z*-score	RT (ms) Mean ± *SD*	*z*-score	RT (ms) Mean ± *SD*	*z*-score	RT (ms) Mean ± *SD*	*z*-score	RT (ms) Mean ± *SD*	*z*-score
0	S+	75.0 ± 65	0.00	254 ± 188	0.00	110 ± 57	0.00	362 ± 149	0.00	256 ± 110	0.00
11		105 ± 68	0.47	199 ± 68	-0.30	277 ± 187	2.95 *	352 ± 120	-0.07	227 ± 60	-0.27
29		167 ± 152	1.42	242 ± 56	-0.06	707 ± 413	10.52*	856 ± 250	3.33*	452 ± 164	1.78
50		108 ± 60	0.52	424 ± 382	0.90	656 ± 473	9.64 *	934 ± 147	3.85*	1000 ± 0	6.76 *
75		202 ± 138	1.96 *	737 ± 362	2.57*	1000 ± 0	15.72*	878 ± 273	3.47*	1000 ± 0	6.76 *
87		205 ± 111	2.00 *	1000 ± 0	3.96*	1000 ± 0	15.72*	878 ± 273	3.47*	1000 ± 0	6.76 *
> 97	S-	998 ± 10	14.22*	913 ± 516	3.50*	1000 ± 0	15.72*	1000 ± 0	4.30*	1000 ± 0	6.76 *

* $p < 0.05$

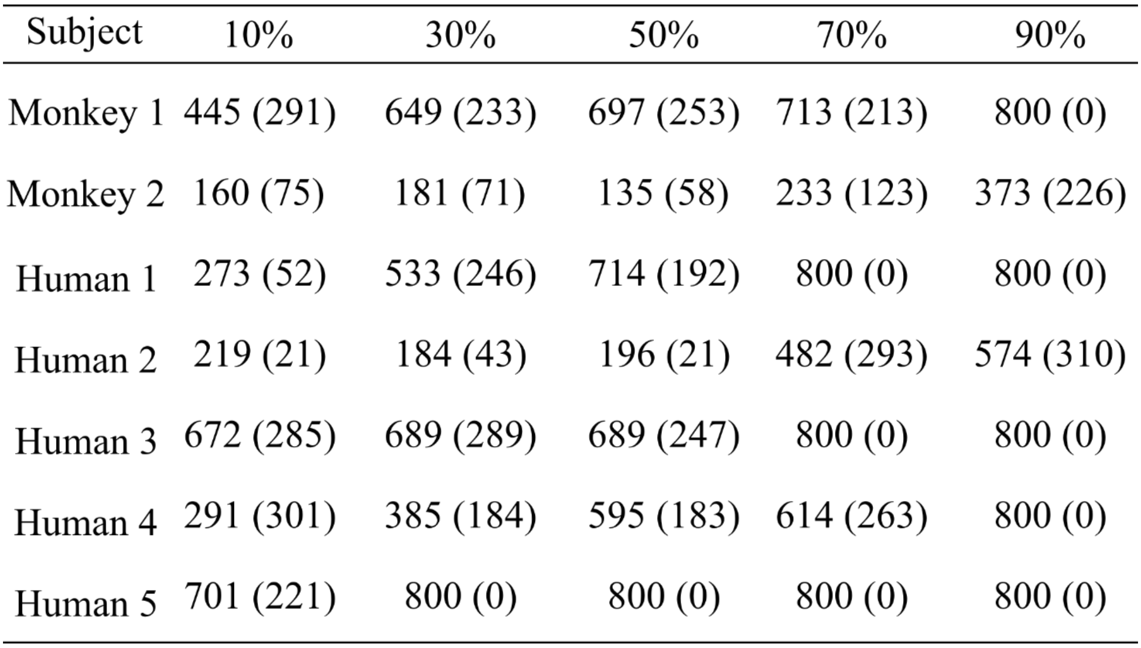
Table 3-1. Mean (SD) reaction times to whole-morph stimuli in each subject. Percentages represent the morphing proportions of information from Monkey B.



Subject	0%	10%	30%	50%	70%	90%	100%
Monkey 1	230 (76)	199 (62)	264 (280)	306 (271)	506 (325)	702 (240)	800 (0)
Monkey 2	149 (39)	189 (97)	170 (86)	152 (109)	285 (109)	246 (97)	342 (86)
Human 1	275 (56)	443 (88)	540 (240)	707 (207)	718 (183)	800 (0)	800 (0)
Human 2	200 (45)	166 (64)	171 (67)	235 (46)	249 (45)	719 (181)	786 (31)
Human 3	194 (42)	286 (92)	541 (292)	450 (209)	800 (0)	800 (0)	800 (0)
Human 4	246 (30)	201 (98)	367 (243)	322 (110)	537 (246)	800 (0)	800 (0)
Human 5	226 (94)	233 (102)	629 (238)	740 (135)	800 (0)	800 (0)	800 (0)

Table 3-2. Mean (SD) reaction times to F0-morph stimuli in each subject.

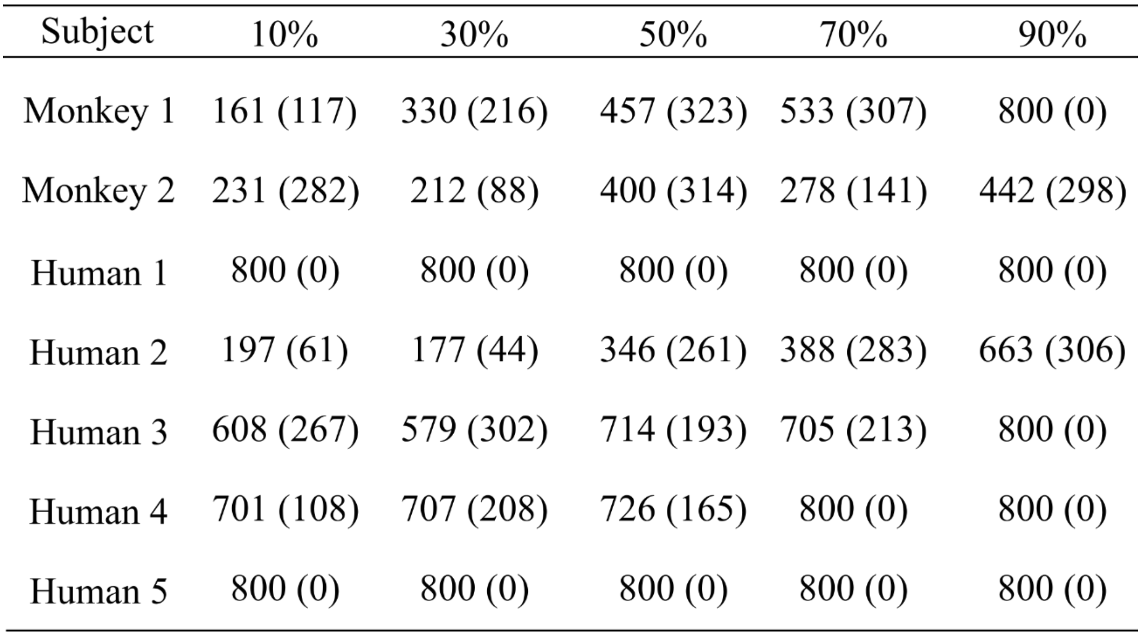
Percentages represent the morph proportions of F0 from Monkey B.



Subject	10%	30%	50%	70%	90%
Monkey 1	445 (291)	649 (233)	697 (253)	713 (213)	800 (0)
Monkey 2	160 (75)	181 (71)	135 (58)	233 (123)	373 (226)
Human 1	273 (52)	533 (246)	714 (192)	800 (0)	800 (0)
Human 2	219 (21)	184 (43)	196 (21)	482 (293)	574 (310)
Human 3	672 (285)	689 (289)	689 (247)	800 (0)	800 (0)
Human 4	291 (301)	385 (184)	595 (183)	614 (263)	800 (0)
Human 5	701 (221)	800 (0)	800 (0)	800 (0)	800 (0)

Table 3-3. Mean (SD) reaction times to VTC-morph stimuli in each subject.

Percentages represent the morph proportions of VTC from Monkey B.



Subject	10%	30%	50%	70%	90%
Monkey 1	161 (117)	330 (216)	457 (323)	533 (307)	800 (0)
Monkey 2	231 (282)	212 (88)	400 (314)	278 (141)	442 (298)
Human 1	800 (0)	800 (0)	800 (0)	800 (0)	800 (0)
Human 2	197 (61)	177 (44)	346 (261)	388 (283)	663 (306)
Human 3	608 (267)	579 (302)	714 (193)	705 (213)	800 (0)
Human 4	701 (108)	707 (208)	726 (165)	800 (0)	800 (0)
Human 5	800 (0)	800 (0)	800 (0)	800 (0)	800 (0)

Table 4-1. Median (interquartile range) of reaction times to training and test stimuli.

Subject	cooA	F0$_{cooB}$-VTC$_{cooA}$	cooB	F0$_{cooA}$-VTC$_{cooB}$
Subject 1	416 (351-558)	368 (276-592)	800 (800-800)	800 (800-800)
Subject 2	226 (108-321)	230 (161-499)	800 (581-800)	800 (391-800)

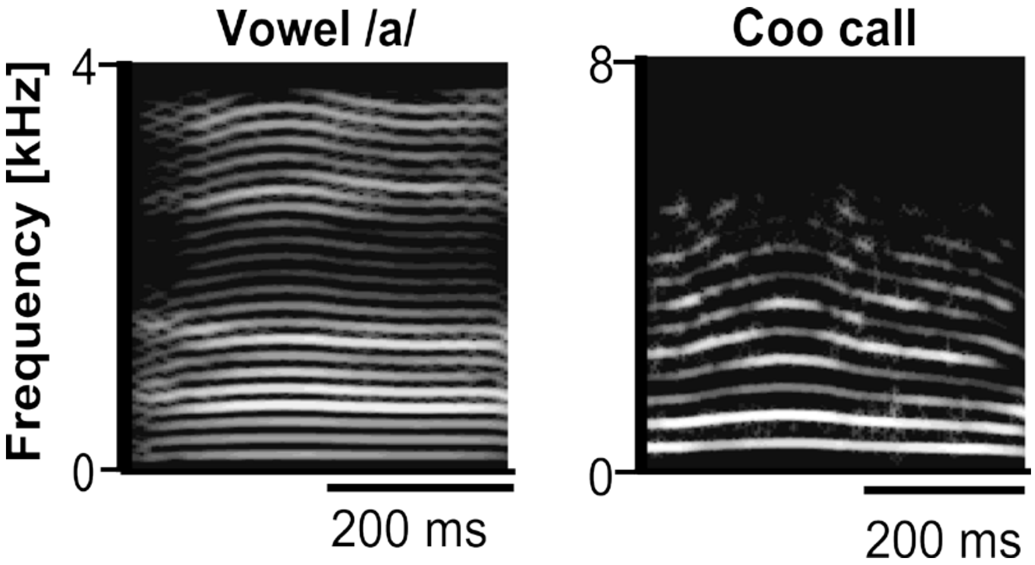


Figure 1-1. Spectrograms of a human vowel (/a/, left panel) and coo calls from monkeys (right panel). Monkeys often utter coo calls for greeting and locating other individuals. Acoustic energies in coo calls are harmonically structured as in a human vowel.

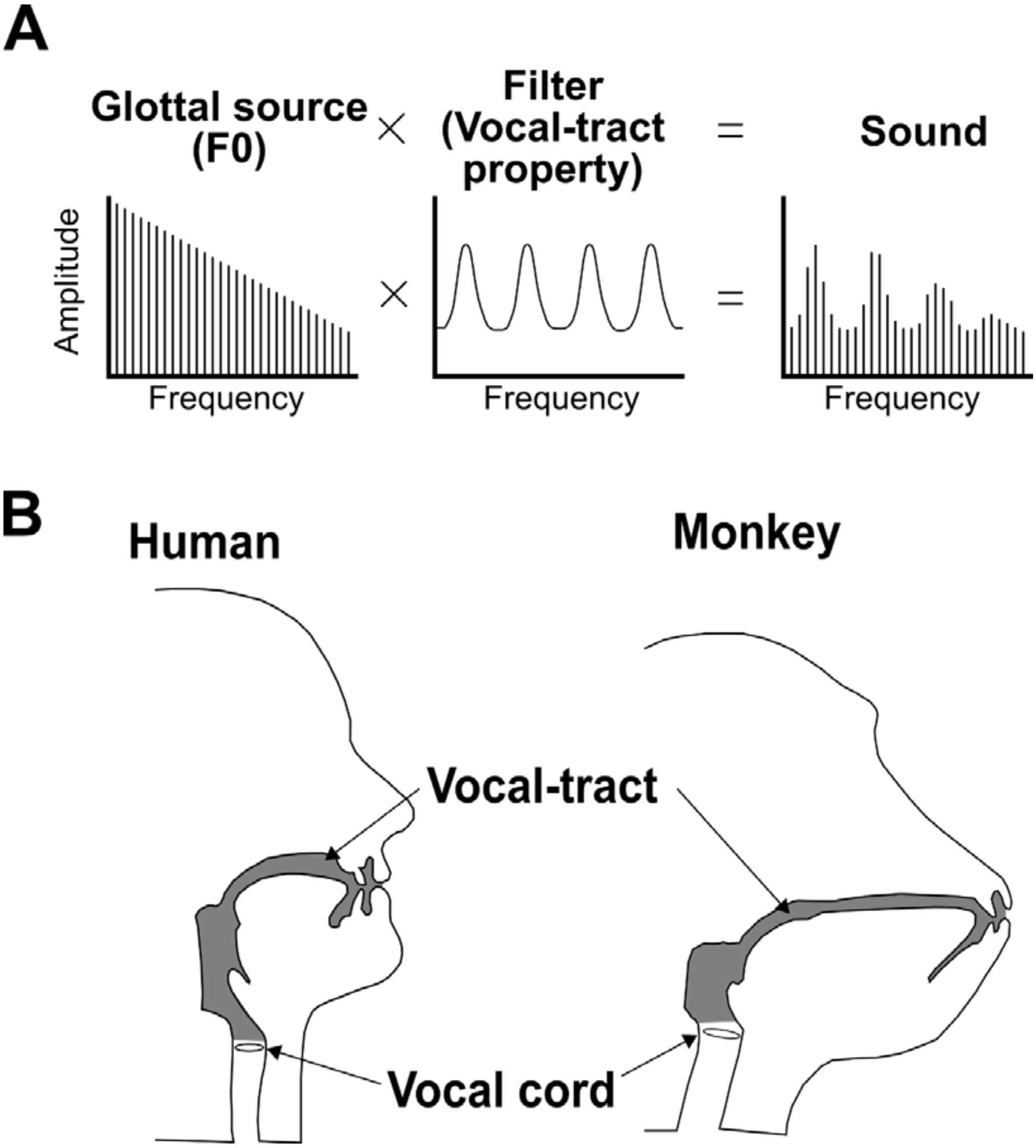


Figure 1-2. Vocal generation mechanism in primates. A: The source-filter theory. The periodic opening and closing of the vocal folds generate pulses for vocalizations. The repetition rates of these pulses (source) are used to determine the F0 of the vocalization and are perceived as pitch. As pulses created by vocal folds pass through the vocal tract, and the vocal tract properties (filter) produce resonances and enhance/dampen particular frequency bands; these are the formants. B: Sagittal views of the vocal tract anatomy. Formants are generated by the filter characteristics of vocal tract properties (the oral and nasal cavities above the vocal folds, gray area).

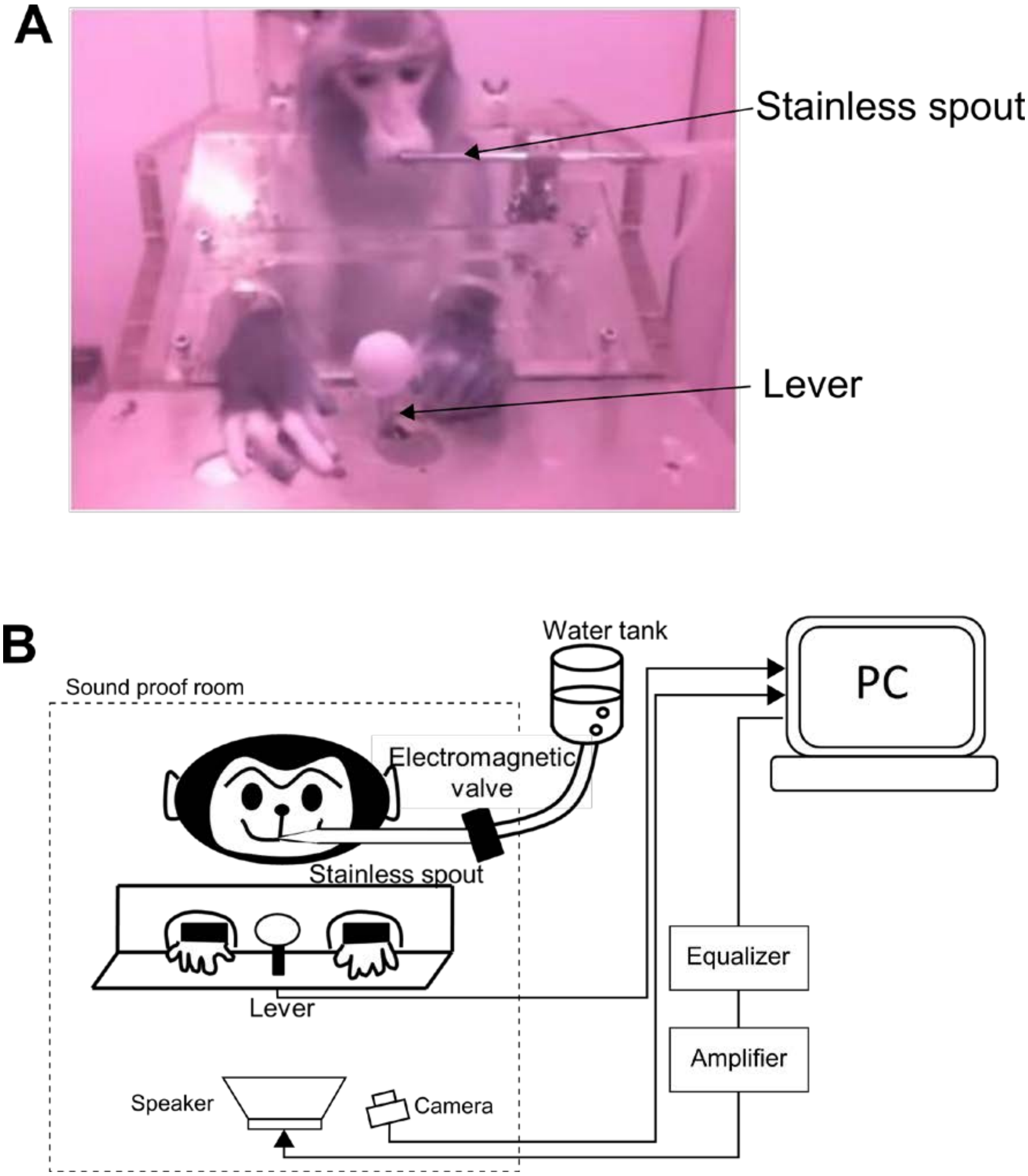


Figure 2-1. Experimental setting. (A) Photo image. (B) Schematized experimental setting. The monkeys were trained to sit in a monkey chair in a sound proof room. The loud speaker was fixed 68 cm in front of the subject's head. The animal was given juice from stainless spout when an electromagnetic valve opened.

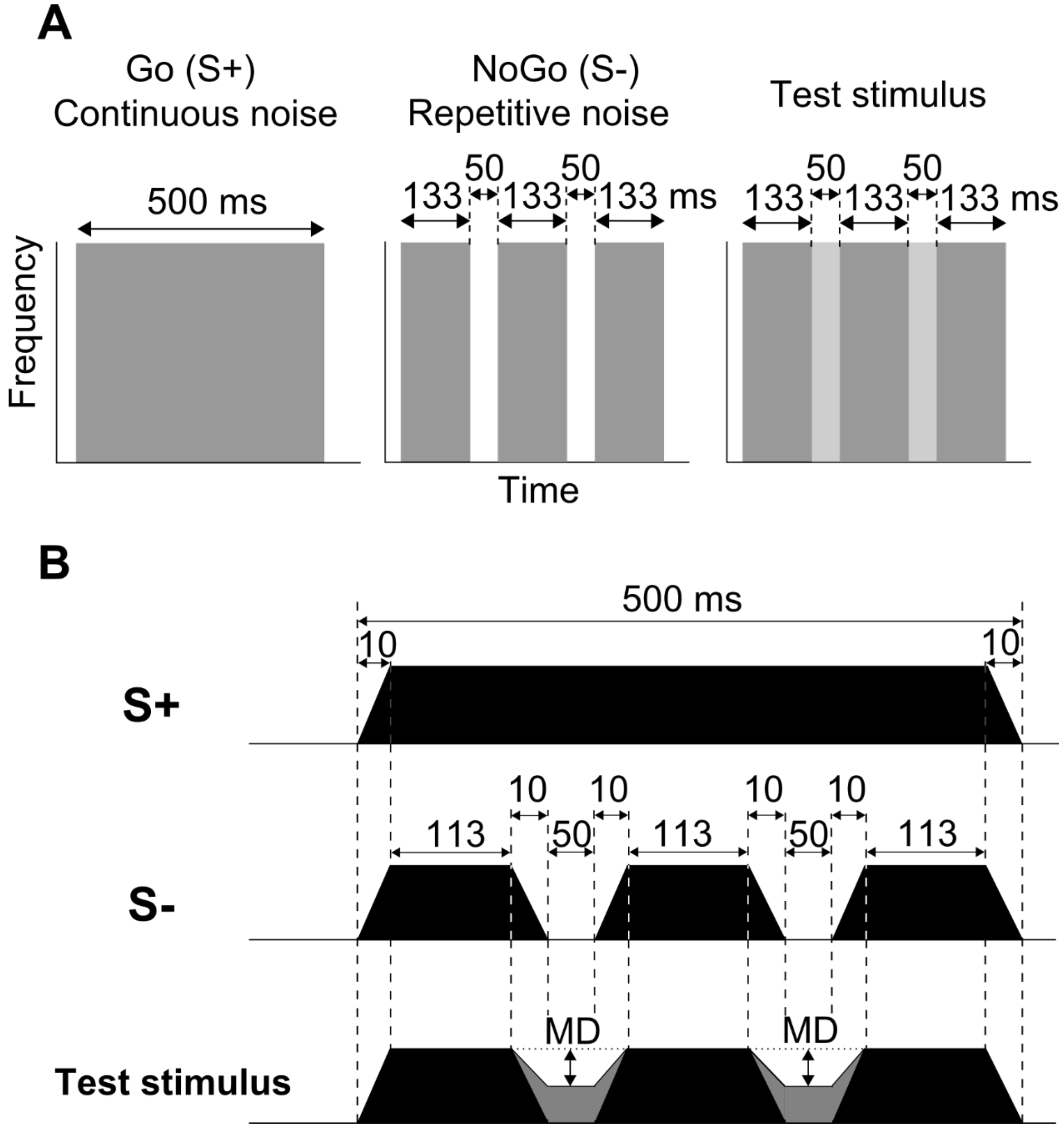


Figure 2-2. Schematized spectrograms (A) and amplitude envelopes of discriminative and test stimuli (B). (A) Training and test stimuli. In the spectrogram display, gray rectangles represent the white noise burst. Test stimuli were amplitude-modulated (AM) white noise bursts, in which the amplitude of time periods corresponding to the silent portion of S- varied. (B) Temporal structure of training and test stimuli. Only the positive portions of amplitude envelopes are shown. Gray areas were depicted amplitude difference between the S- and test stimulus.

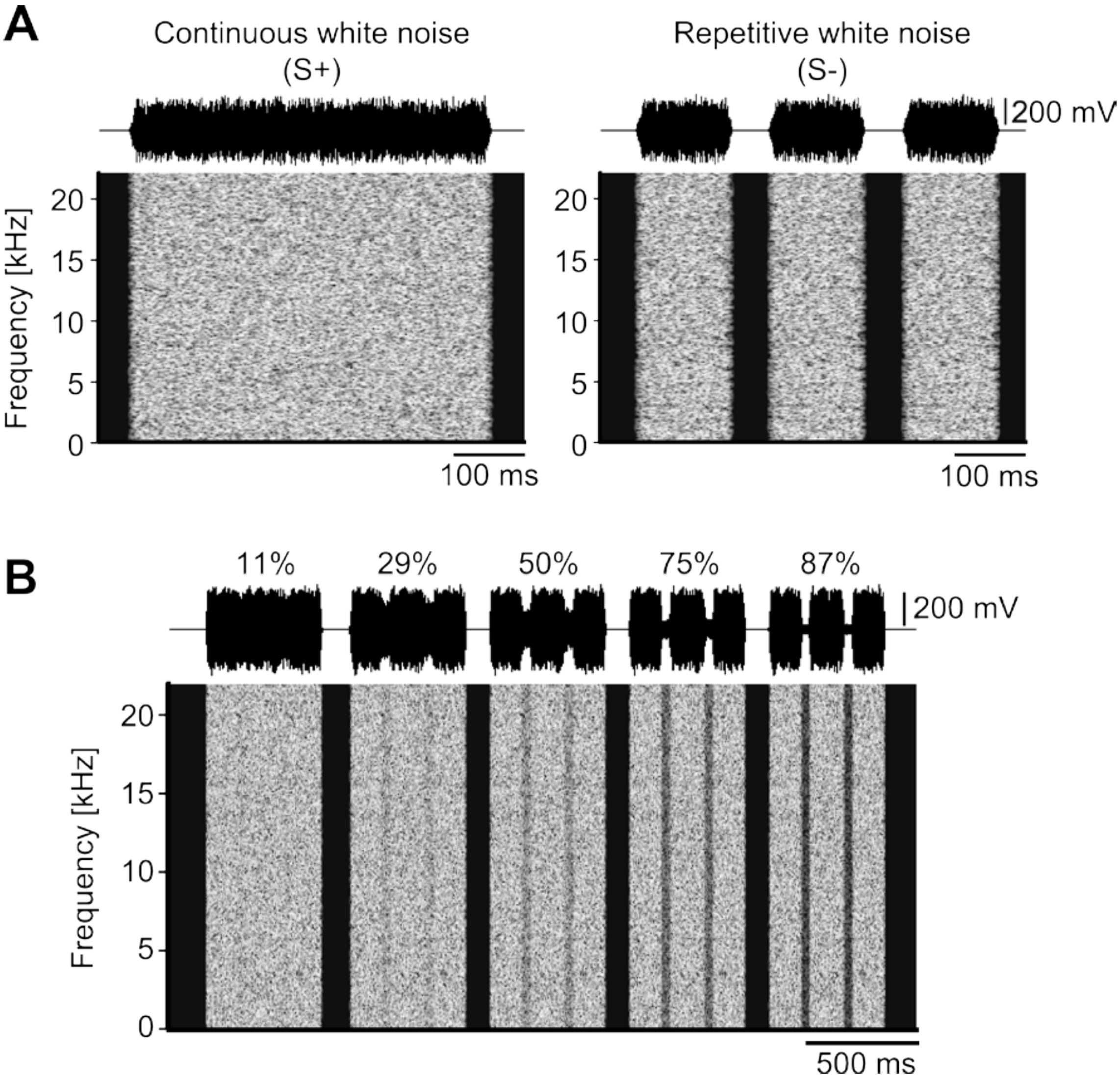


Figure 2-3. Spectrograms of training and test stimuli. A: Spectrograms of training stimuli (Left panel: Continuous white noise burst, right panel: repetitive white noise burst). Continuous white noise burst was used as Go (S+) stimulus, whereas repetitive white noise burst was used as NoGo (S-) stimulus. B: Spectrograms of test stimuli. Test stimuli were amplitude-modulated (AM) white noise bursts, in which the amplitude of time periods corresponding to the silent portion of S- varied. Five different modulation depths (11, 29, 50, 75, and 87 %) were provided. Each type of test stimuli was presented for 5 times.

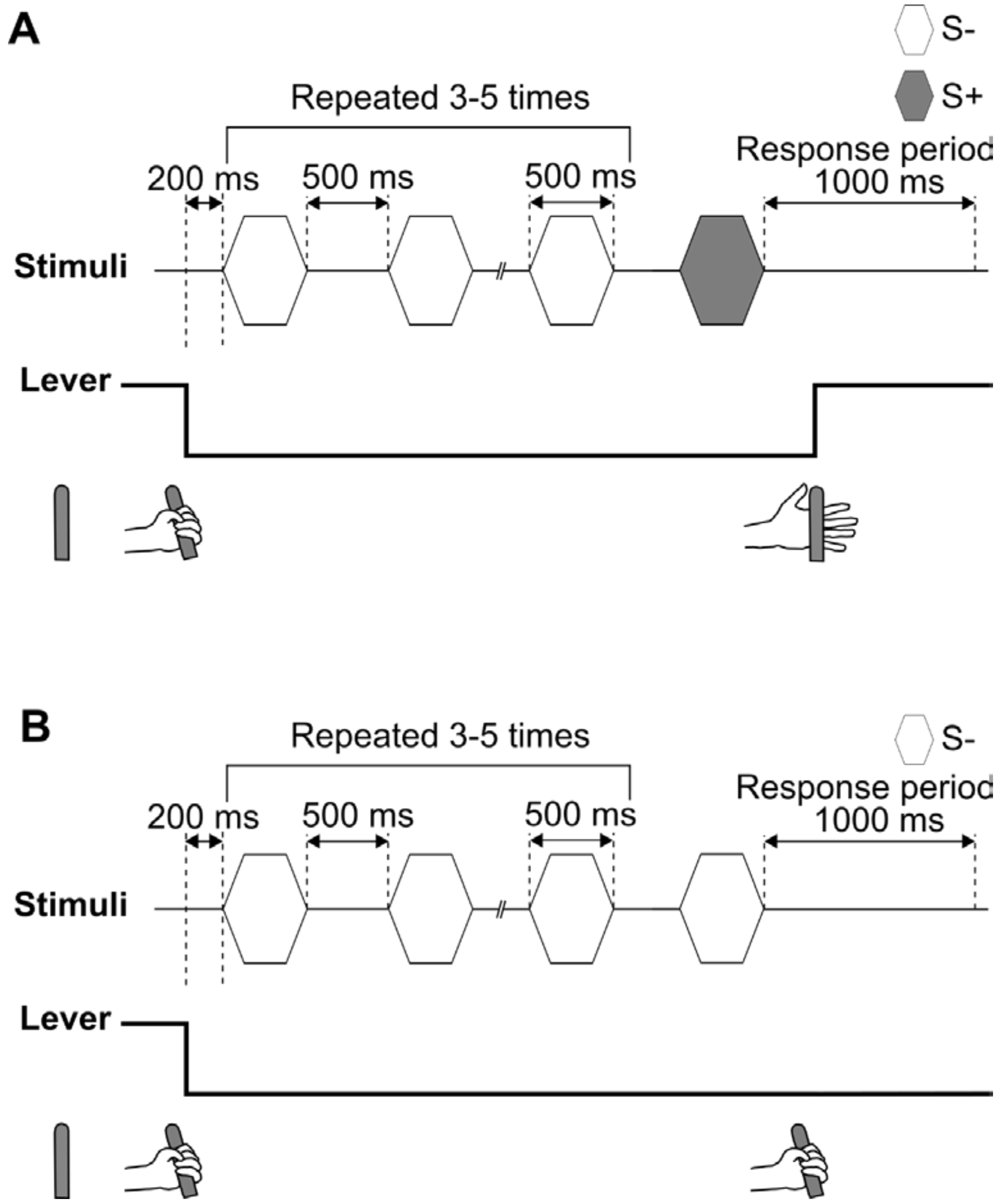


Figure 2-4. Schematized behavioral task. White hexagon: repetitive white noise burst (NoGo stimulus, S-), Gray hexagon: S-, continuous white noise burst (Go stimulus, S+) or test stimuli. In the training session, either S- or S+ was presented as a discriminative stimulus after presentation of S- 3–5 times. The inter-onset interval was 1000 ms. (A) When S+ was presented as a discriminative stimulus, the subjects had to release the lever within 1000 ms (response period) after the offset of S+. (B) When S- continued as the discriminative stimulus, the animal had to keep depressing the lever during the response period.

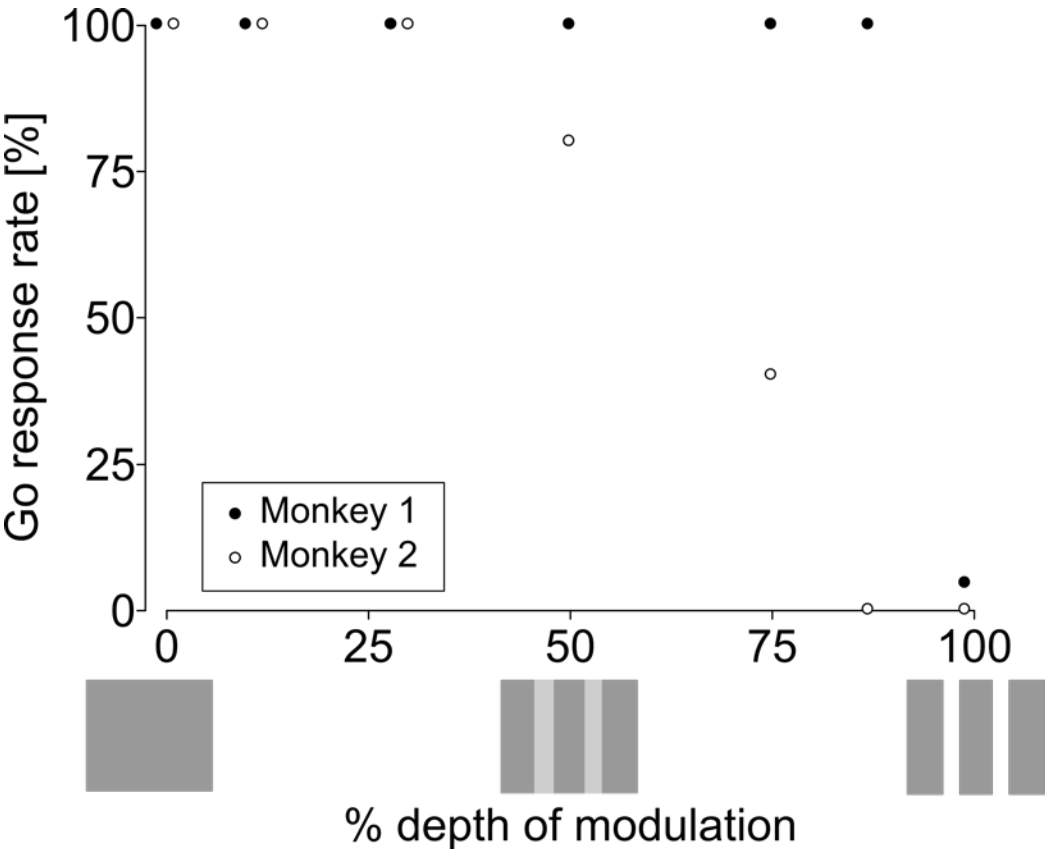


Figure 2-5. Go response rates to stimuli for two monkeys. Closed circle: Monkey 1, open circle: Monkey 2.

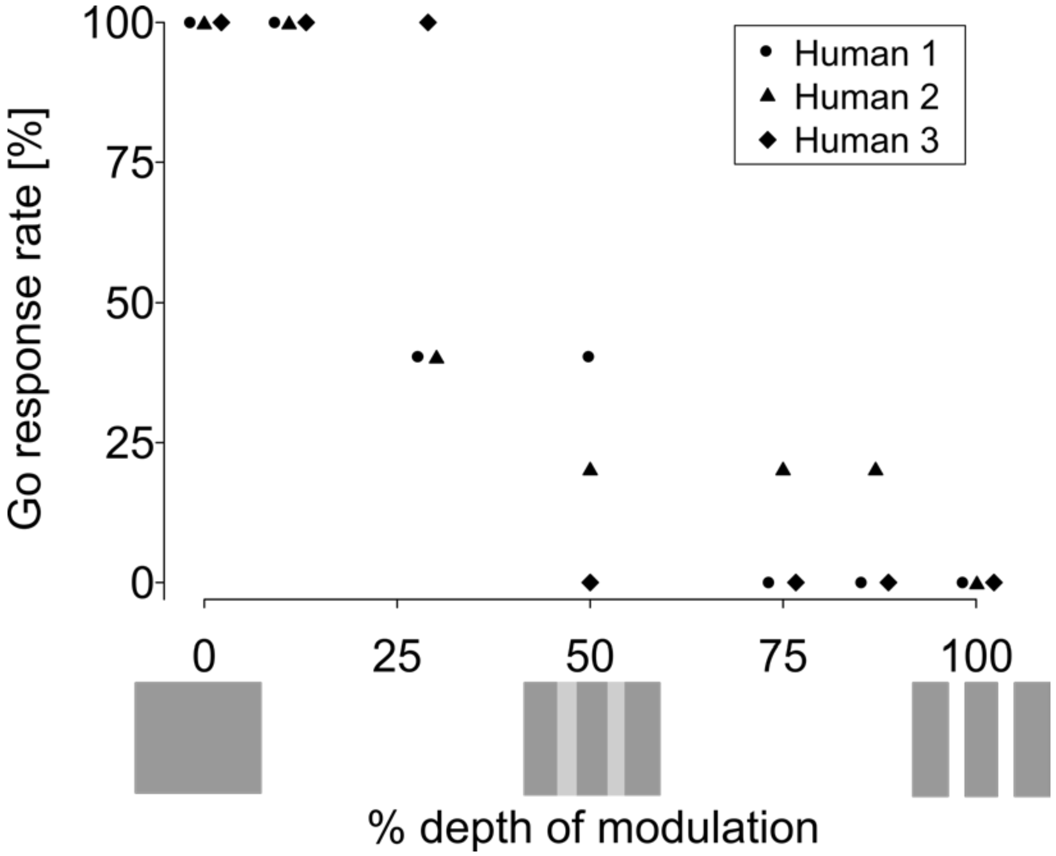


Figure 2-6. Go response rates to stimuli for three humans. Close circle: Human 1, triangle: Human 2, diamond: Human 3.

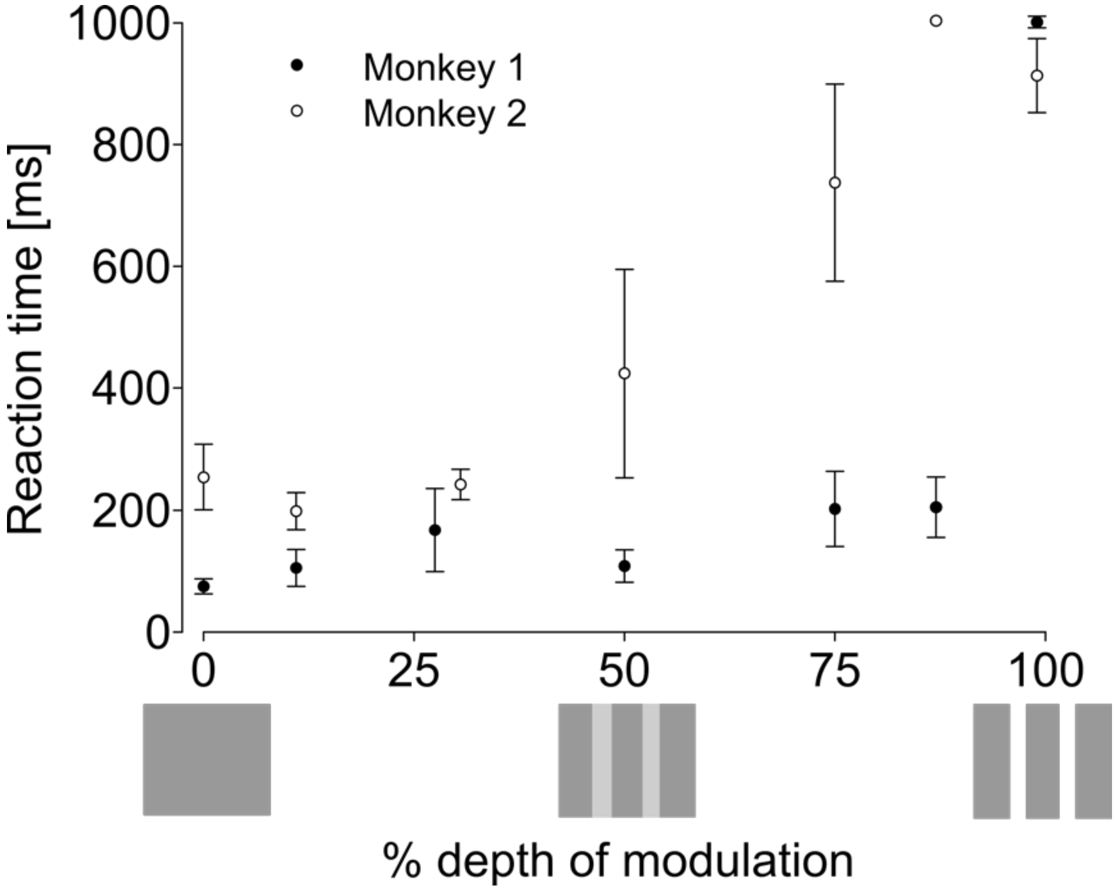


Figure 2-7. Reaction times to stimuli for two monkeys. Closed circle: Monkey 1, open circle: Monkey 2. Error bar: standard error of mean.

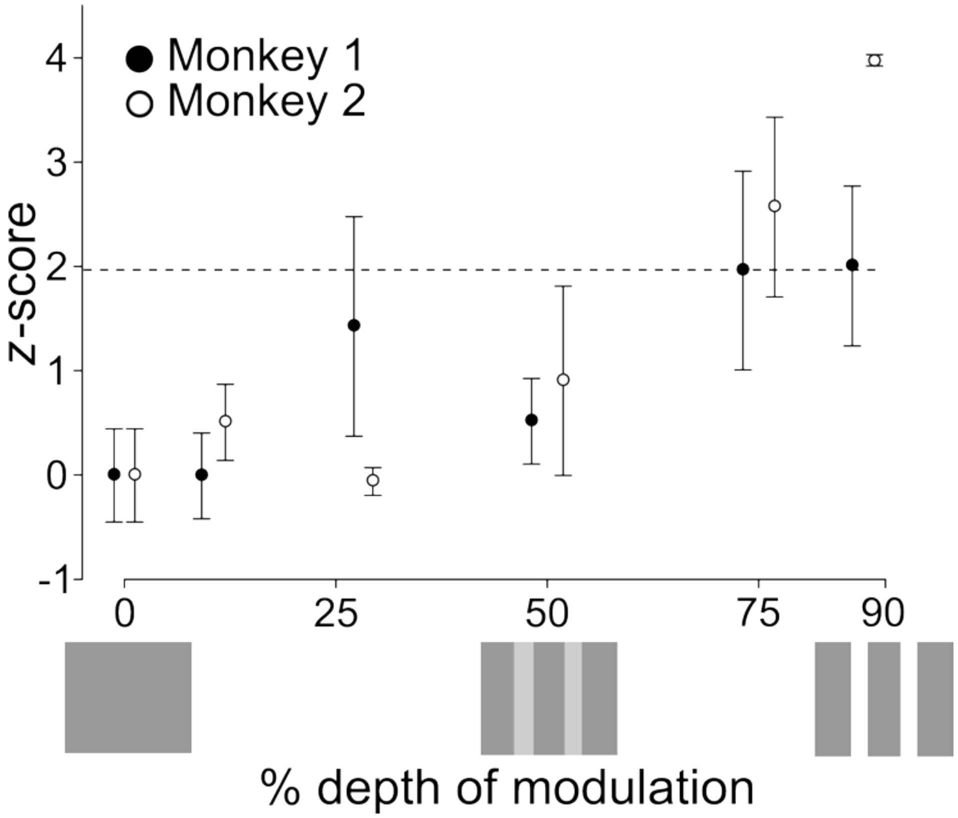


Figure 2-8. The z-scores of reaction times to test stimuli with different modulation depths in each monkey. Closed circle: Monkey 1, open circle: Monkey 2. Dashed line: z-score of 1.96 ($p = 0.05$). Error bars: standard errors of the mean. The horizontal axis: amplitude modulation depths of white noise bursts in percent (%). The reaction times to different stimuli (white noise burst with different modulations depths) were normalized into z-scores based on average reaction time to S+ (modulation depth = 0 %) by each monkey. The reaction time of Monkey 1 for S+ was 75 ± 65 ms (mean ± SD) while that of Monkey 2 was 254 ± 188 ms. The z-score exceeded the criterion of 1.96 when the modulation was greater than 75 % in both monkeys (at 75 %: Monkey 1: $z = 1.96$, $p = 0.05$, Monkey 2: $z = 2.57$, $p = 0.01$).

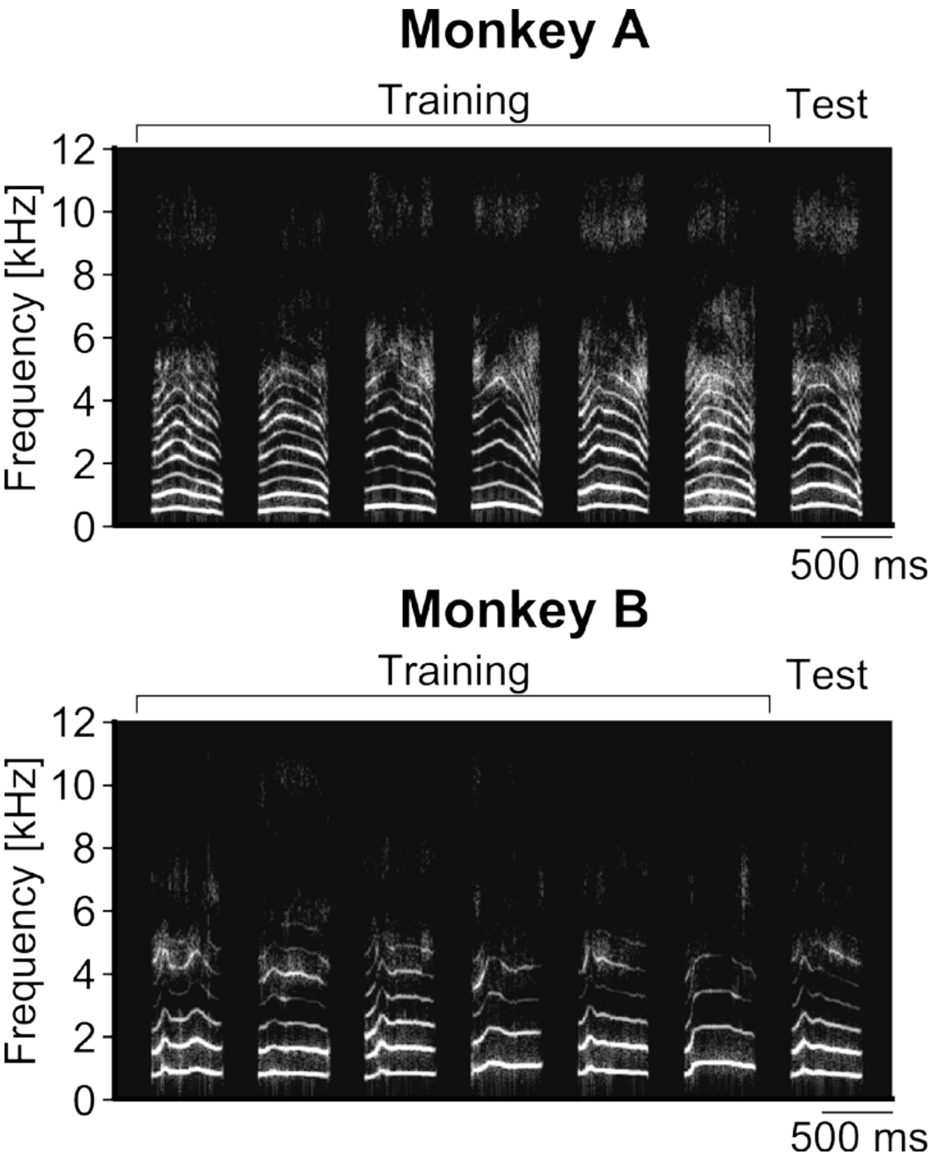


Figure 3-1. Spectrograms of coo calls in Monkey A (top) and Monkey B (bottom).

The right-most calls were used to synthesize the test stimuli.

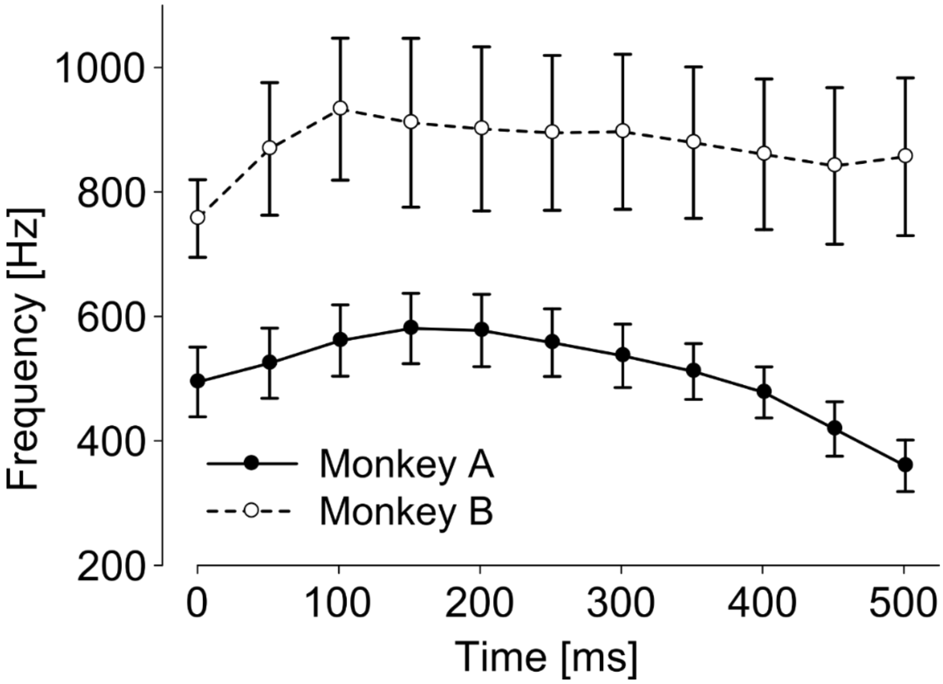


Figure 3-2. Temporal F0s of coo calls in two monkeys. Solid line: the mean F0 of Monkey A. Dashed line: the mean F0 of Monkey B. Mean F0 of cooA was 519±50 Hz [mean ± standard deviation], whereas mean F0 of cooB was 875±121 Hz.

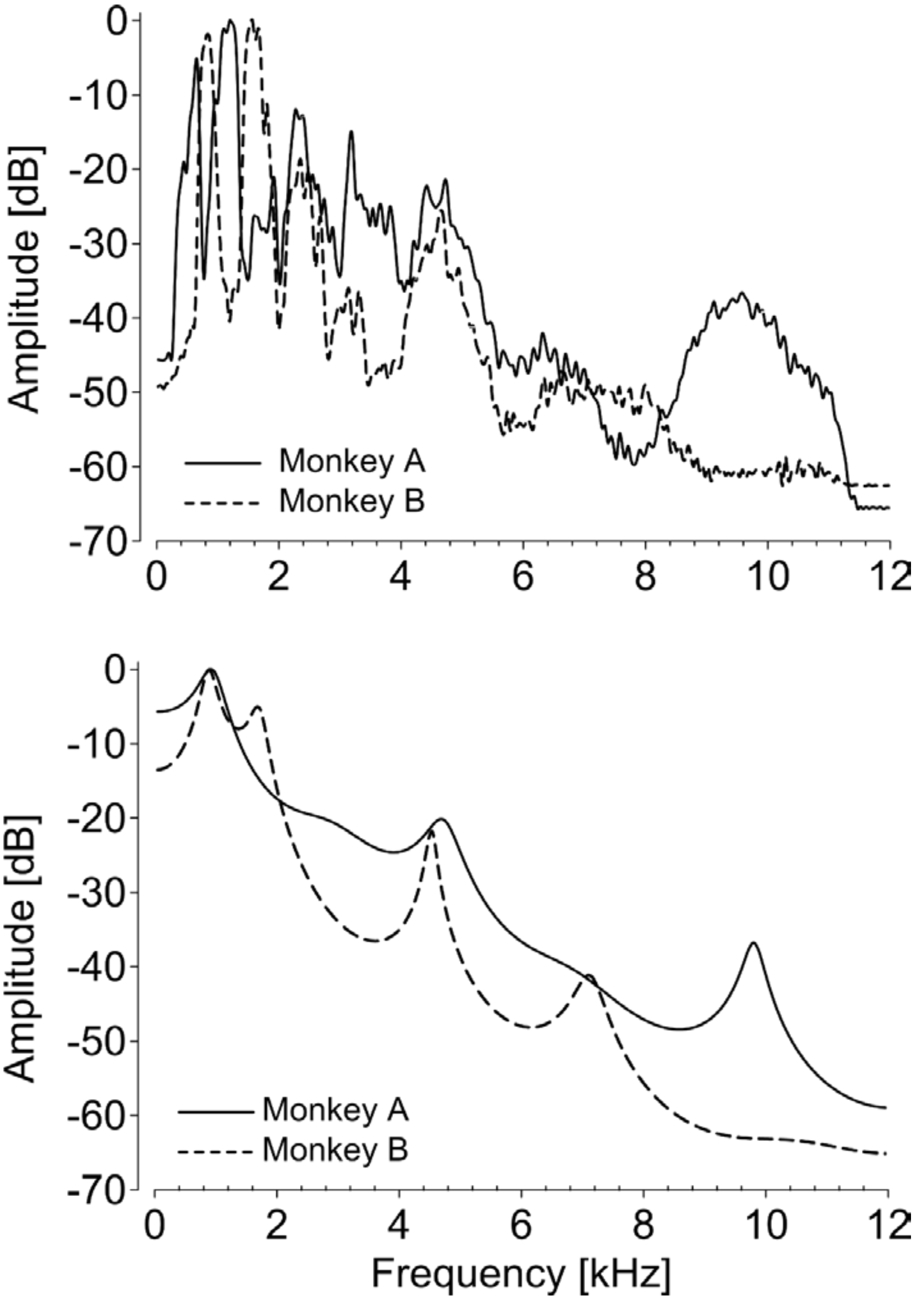


Figure 3-3. Power spectrograms (top) and linear predictive coding spectra (bottom) of vocalizations in two monkeys. Solid line: Spectrograms of Monkey A. Dashed line: spectrograms of Monkey B.

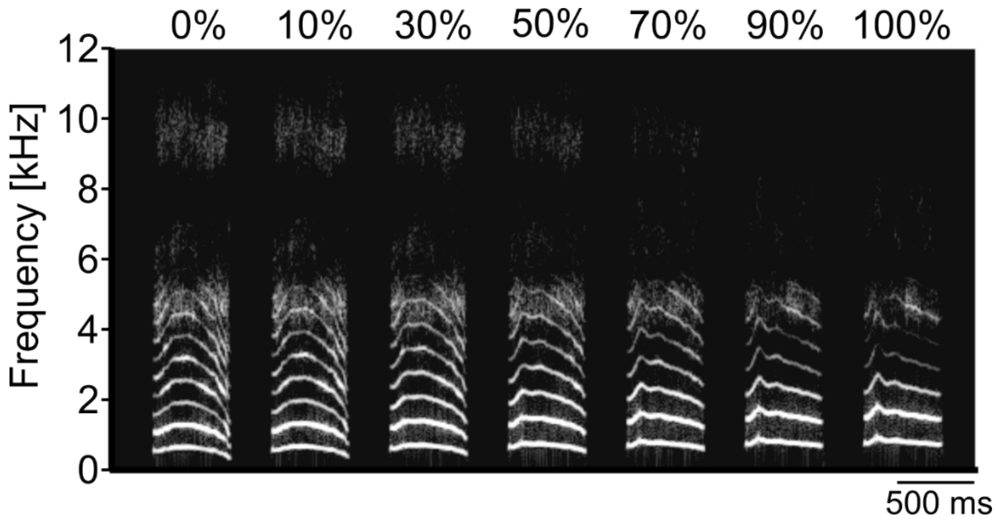


Figure 3-4. Spectrograms of continuum stimuli between the coo calls of two monkeys. The numbers above the spectrograms represent the percentages of vocalizations from Monkey B in the continuum stimuli.

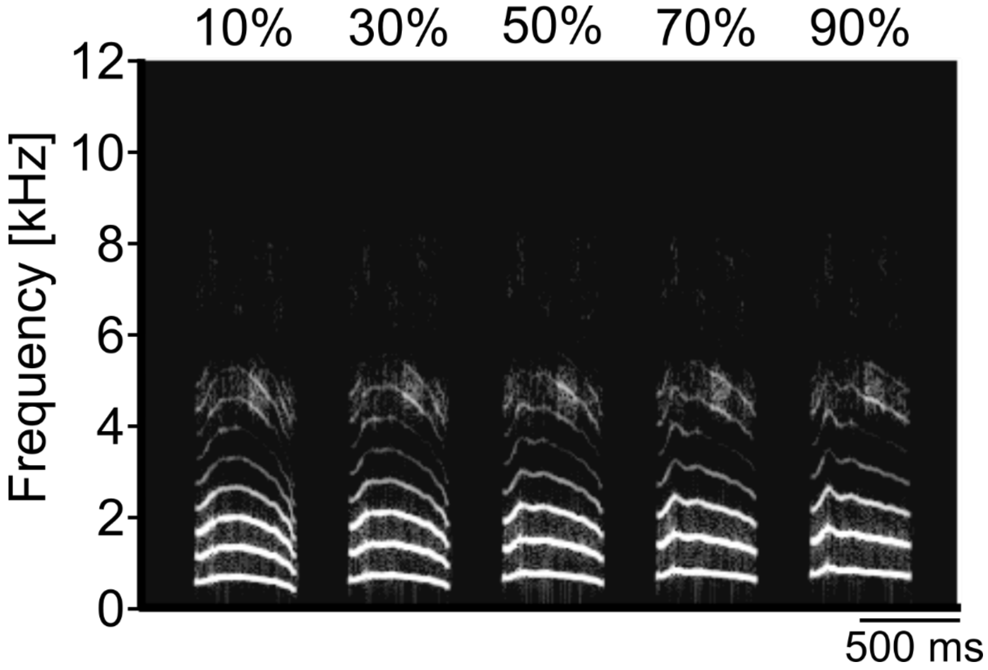


Figure 3-5. Spectrograms of F0-morph stimuli. The numbers above the spectrograms represent the percentages of vocalizations from F0 of Monkey B in the continuum stimuli.

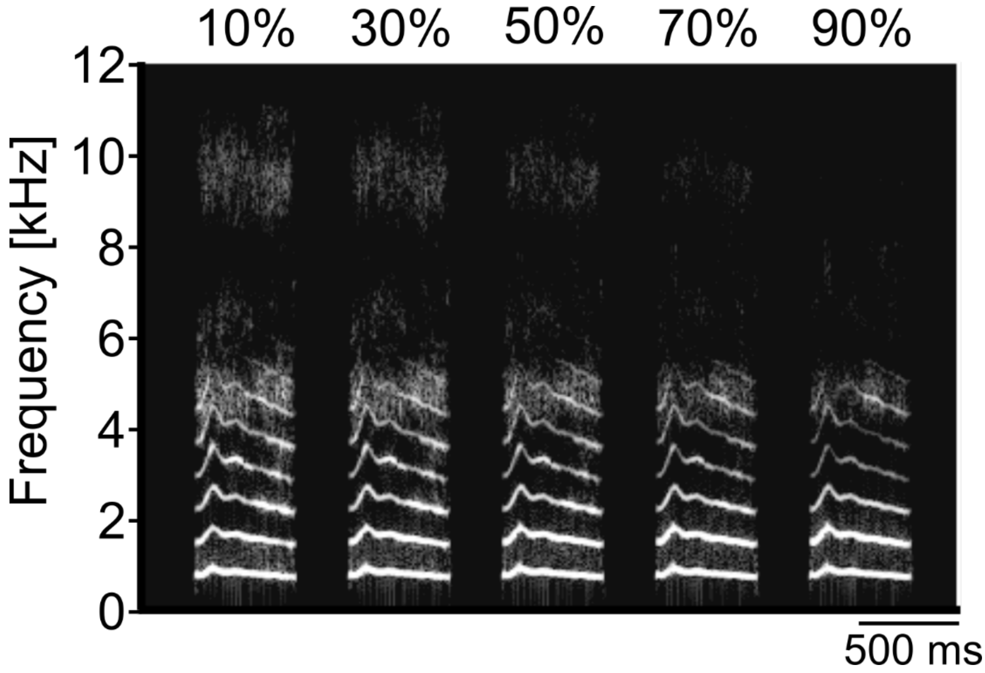


Figure 3-6. Spectrograms of VTC-morph stimuli. The numbers above the spectrograms represent the percentages of vocalizations from VTC of Monkey B in the continuum stimuli.

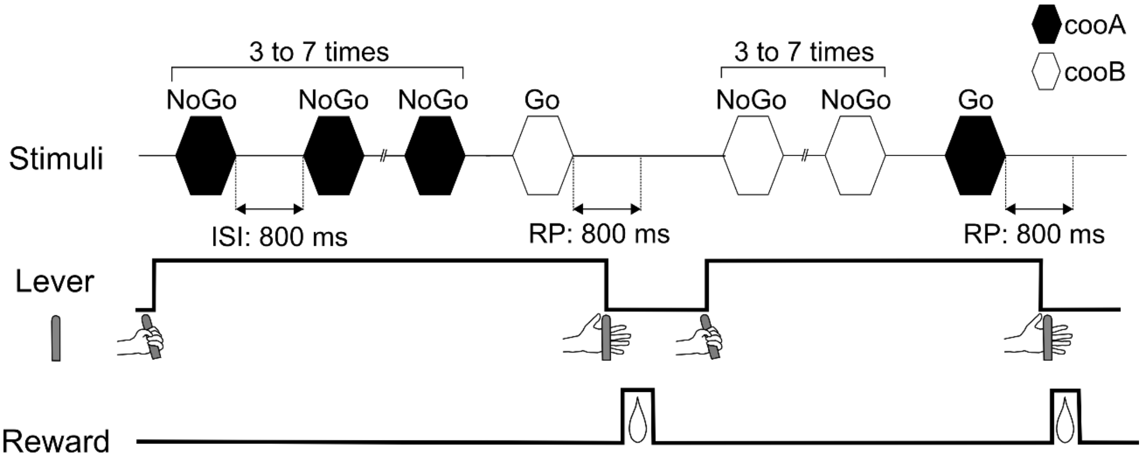


Figure 3-7. Schematized trial event sequence. Upper trace: timing of the stimulus. Middle trace: response of the animal. Lower trace: timing of the reward. Open hexagon: cooA; closed hexagon: cooB. The subjects were required to depress a lever switch to begin the trial. Then, cooA was presented three to seven times with an inter-stimulus interval of 800 ms. The subjects were required to continue depressing the lever while cooA was repeated. If cooB (Go stimulus) was presented, the subjects were required to release the lever within 800 ms after the offset of cooB to receive a reward. After a correct response to a Go stimulus, the stimulus contingencies were reversed in the next trial. That is, cooA became the Go stimulus, and cooB became the NoGo stimulus. In the test trials, cooA was replaced with a test stimulus, and the stimulus was presented after cooBs were repeated as the NoGo stimuli. Neither a reward nor a punishment followed the test trial.

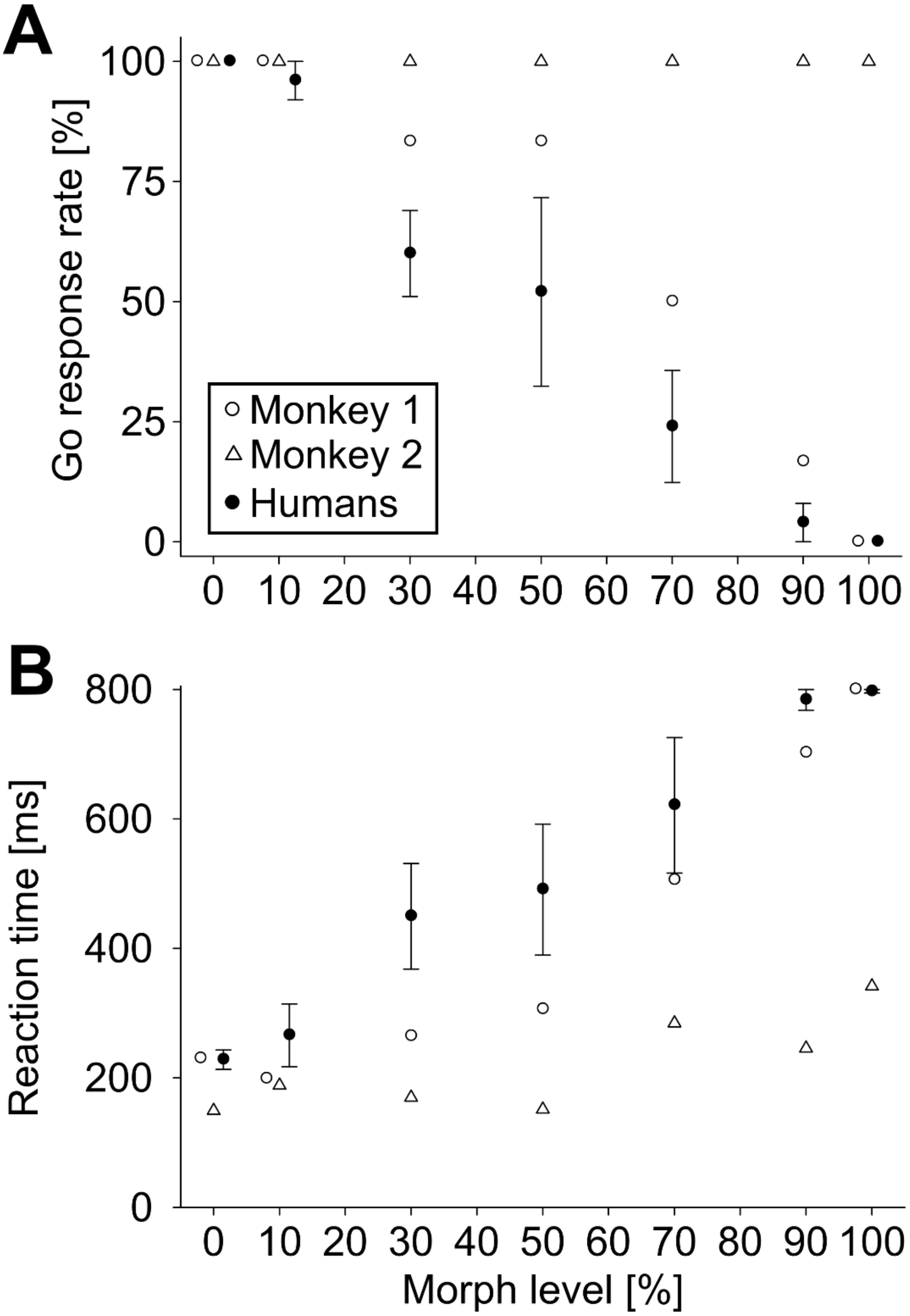


Figure 3-8. Go response rates (A) and reaction times (B) to whole-morph stimuli in the subjects. Open circle: Monkey 1; open triangle: Monkey 2; closed circle: humans. Error bar: standard error of the mean.

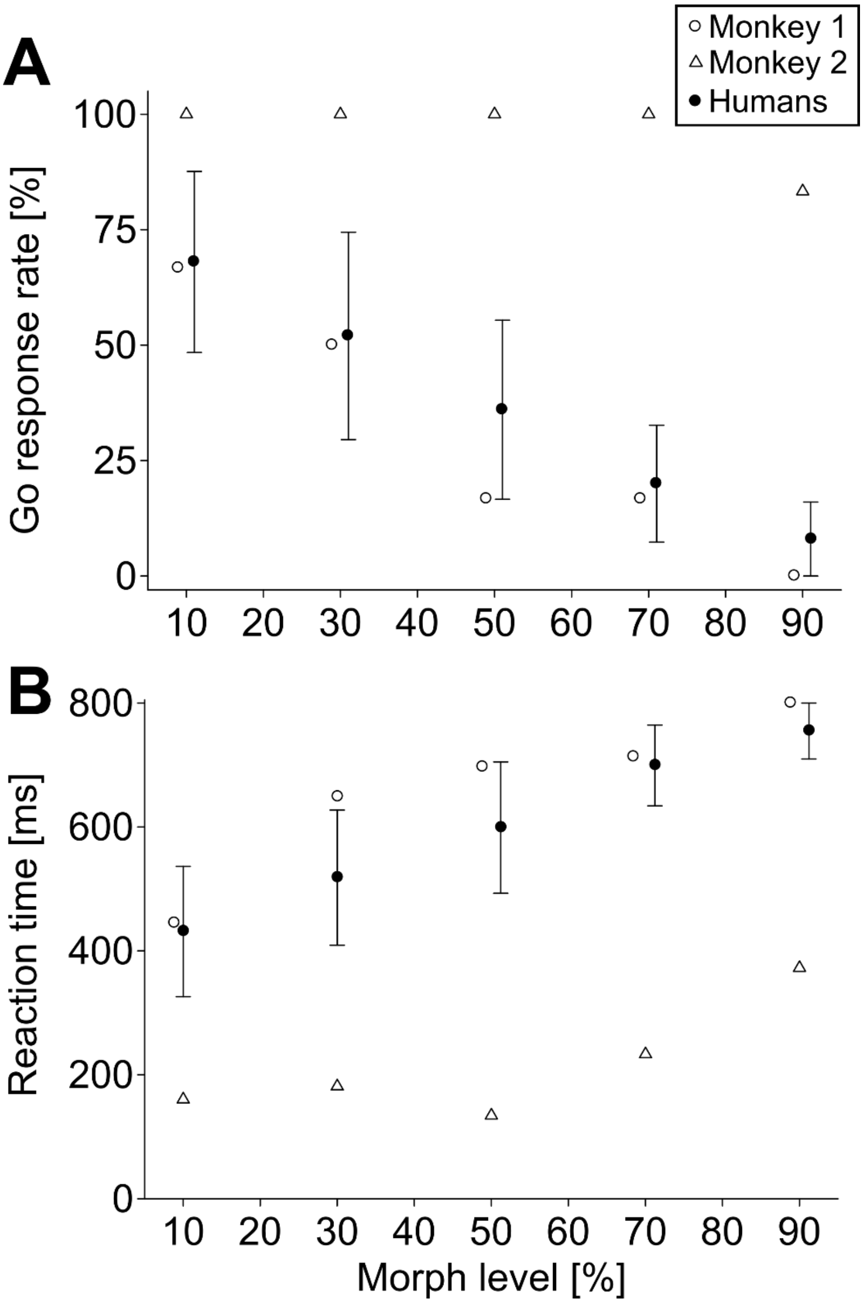


Figure 3-9. Go response rates (A) and reaction times (B) to F0-morph stimuli in the subjects. Open circle: Monkey 1; open triangle: Monkey 2; closed circle: humans. Error bar: standard error of the mean.

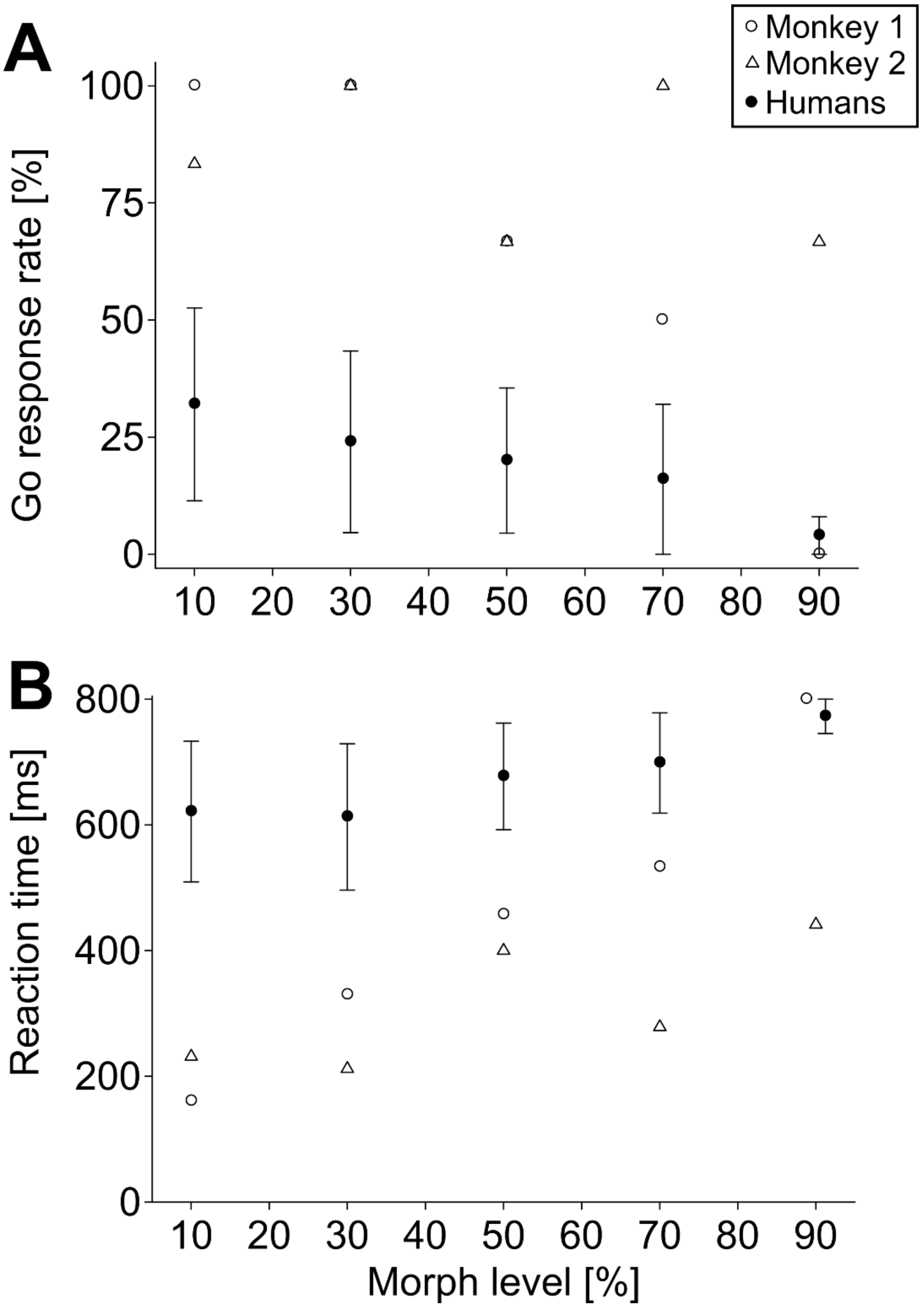


Figure 3-10. Go response rates (A) and reaction times (B) to VTC-morph stimuli in the subjects. Open circle: Monkey 1; open triangle: Monkey 2; closed circle: humans. Error bar: standard error of the mean.

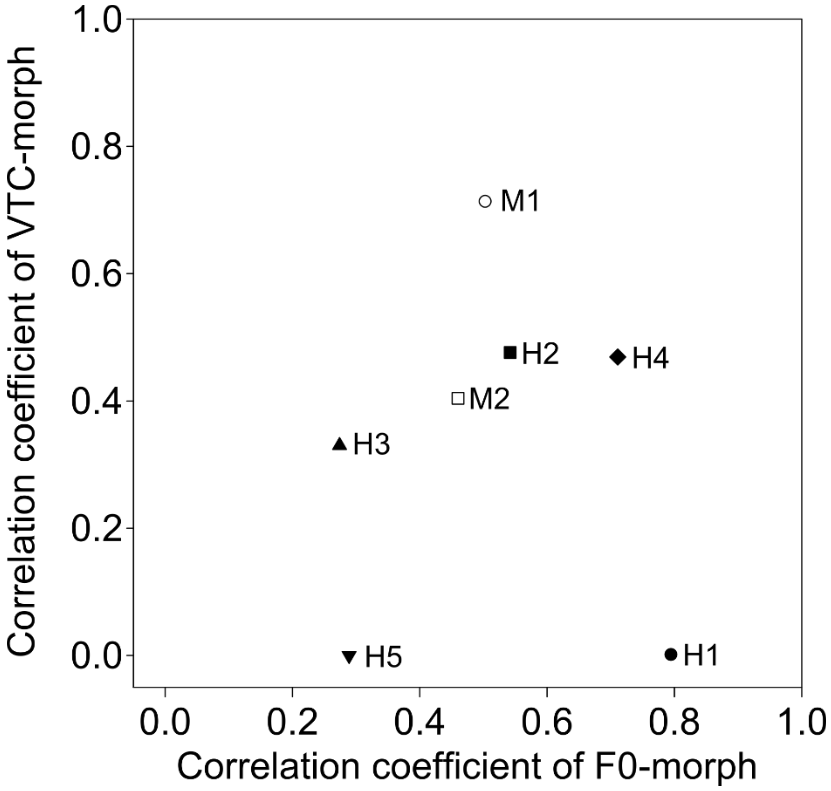


Figure 3-11. Distributions of correlation coefficients for F0-morph and VTC-morph stimuli in each subject. M1: Monkey 1; M2: Monkey 2; H1: Human 1; H2: Human 2; H3: Human 3; H4: Human 4; H5: Human 5.

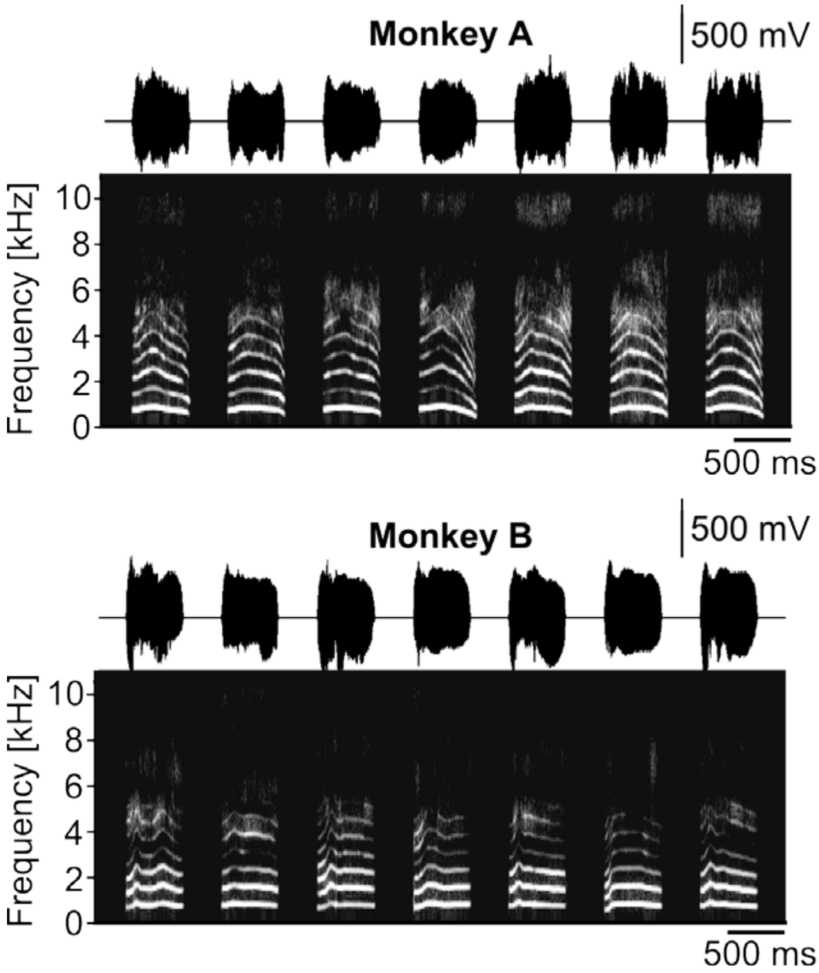


Figure 4-1. Spectrograms of the coo calls from the two monkeys. Top panel: the coo calls of Monkey A (cooA). Bottom panel: the coo calls of Monkey B (cooB). These monkeys were unfamiliar to the subjects, and the recorded calls were modified such that they had the same durations, amplitude envelopes, and average fundamental frequencies. The subjects were trained to discriminate between the cooAs and cooBs. The right-most calls were used to synthesize the test stimuli.

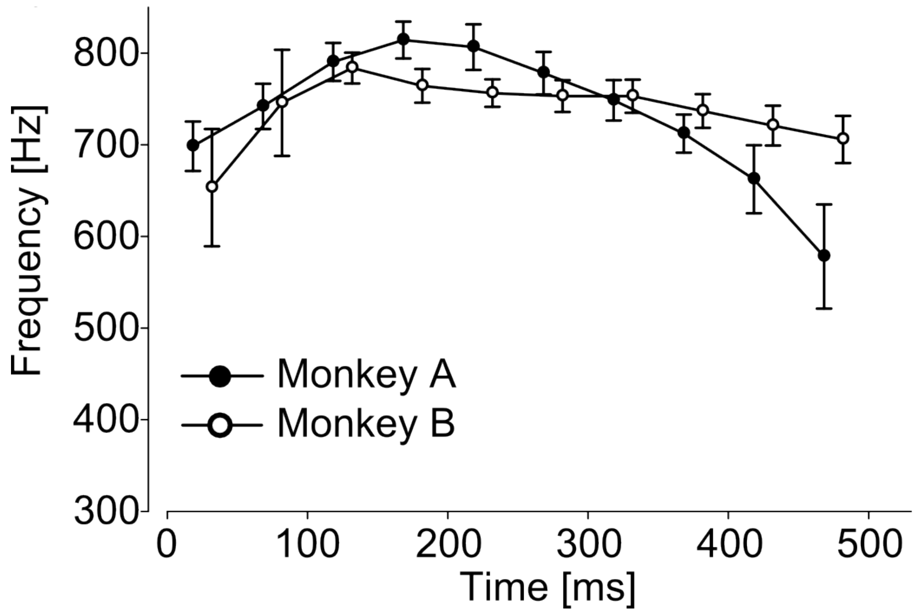


Figure 4-2. Temporal pitch patterns of the coo calls of the two monkeys. Closed circles: the mean temporal pitch pattern of the coo calls of Monkey A; open circles: those of Monkey B. Error bars: standard deviations. Although the fundamental frequencies (F0) were normalized, the two stimulus sets varied in terms of both the end frequency and the time of the F0 peak.

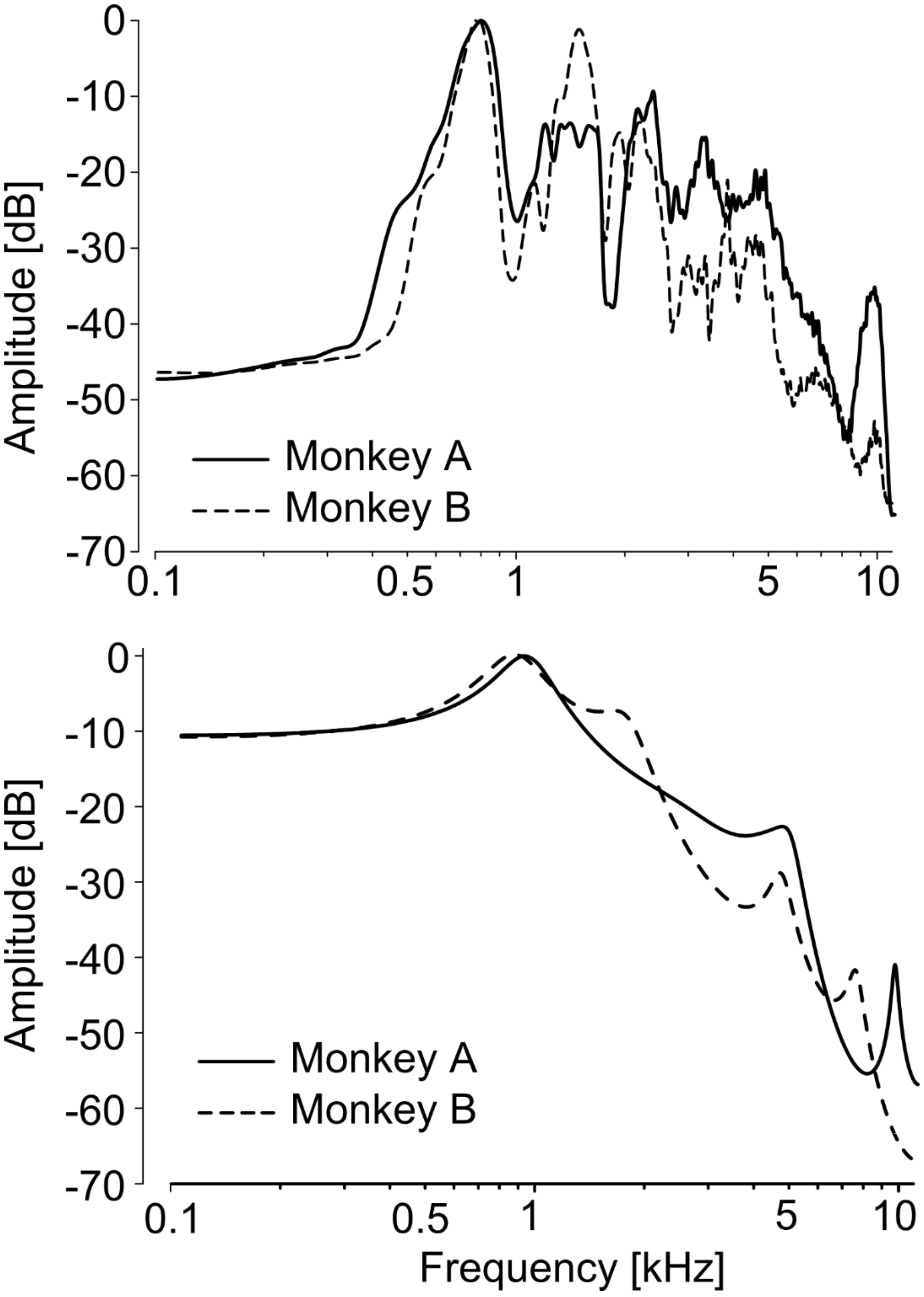


Figure 4-3. Power spectra and linear predictive coding spectra of the cooA (solid line) and cooB (dash line) stimuli. The data illustrate the differences in the vocal tract characteristics (VTCs) of the two monkeys.

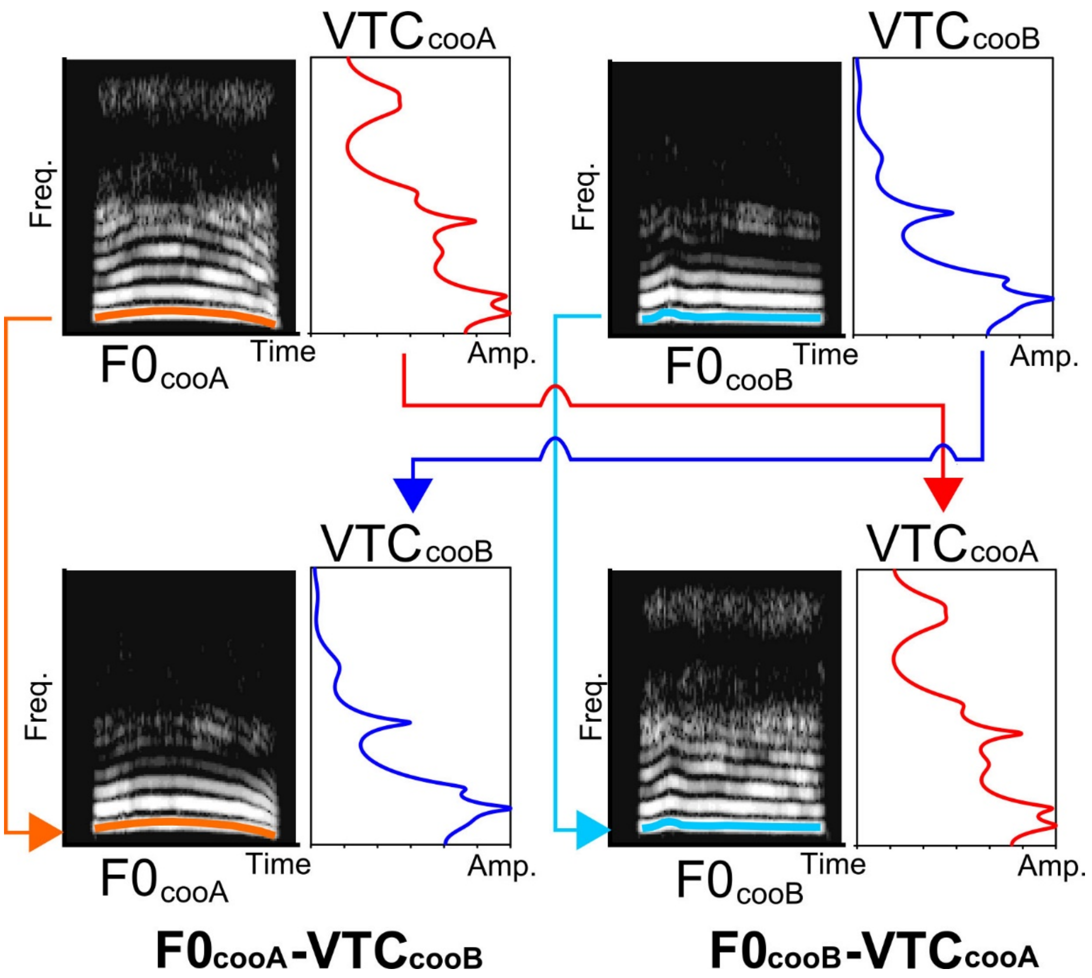


Figure 4-4. Methods for the synthesis of the stimuli. The test stimuli were synthesized by combining the F0s and the VTCs from different animals. Orange line: the F0 of Monkey A; light blue line: the F0 of Monkey B. Red line: the linear predictive coding spectrum of Monkey A; blue line: the linear predictive coding spectrum for Monkey B. $F0_{cooA}$-VTC_{cooB} (bottom left) was synthesized from the F0 of Monkey A (orange) and the VTC of Monkey B (blue), whereas $F0_{cooB}$-VTC_{cooA} (bottom right) was created from the F0 of Monkey B (light blue) and the VTC of Monkey A (orange).

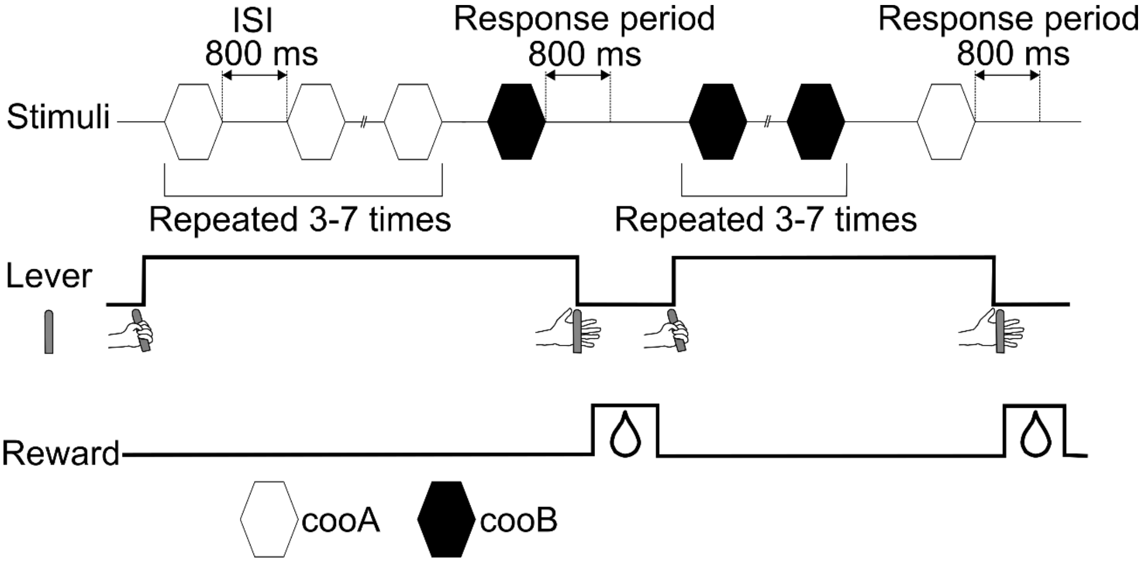


Figure 4-5. Schematized trial event sequence. Upper trace: the timing of the stimulus. Middle trace: the response of the animal. Lower trace: the timing of the reward. The subjects were required to depress a lever switch for 200 ms to begin the trial. Then, cooA (open hexagon: NoGo stimulus) was presented 3–7 times with an interstimulus interval (ISI) of 800 ms. During the repetitions, the type of cooA (out of the total of six, Fig. 4-1) and the intensity of the stimulus (57, 60, and 63 dB SPL) were randomly changed. The subjects were required to continue depressing the lever while cooA was repeated. If cooB (Go stimulus) was presented, the subjects were required to release the lever within 800 ms after the offset of the cooB to receive a reward. After a correct response to a Go stimulus, the stimulus contingencies were reversed in the next trial. That is, cooA became the Go stimulus, and cooB became the NoGo stimulus. In the test trials, cooA was replaced with a test stimulus, and the stimulus was presented after cooBs were repeated as the NoGo stimuli. Neither a reward nor a punishment followed the test trial.

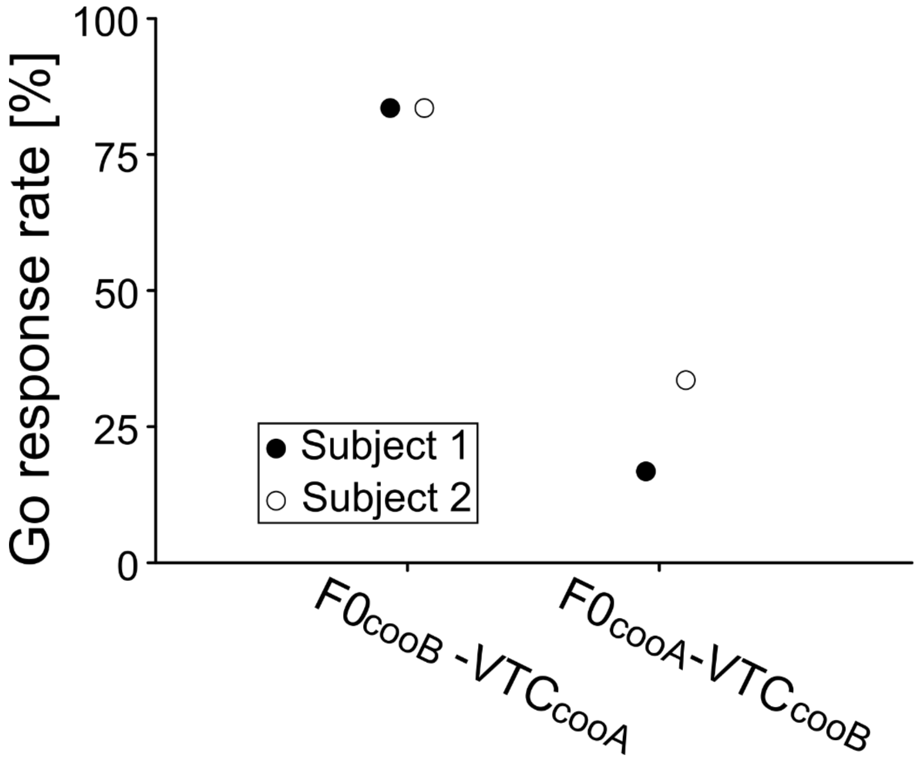


Figure 4-6. Go response rates to the test stimuli. The Go response rates of each monkey to the test stimuli. The Go response rates to $F0_{cooB}$-VTC_{cooA} (subject 1: 83.3%, subject 2: 83.3%) of each monkey were higher than the Go response rates to $F0_{cooA}$-VTC_{cooB} (subject 1: 16.7%; subject 2: 33.3%). Both monkeys responded to $F0_{cooA}$-VTC_{cooB} as they did to a coo call of Monkey A, whereas they responded to $F0_{cooB}$-VTC_{cooA} as they did to a coo call of Monkey B.

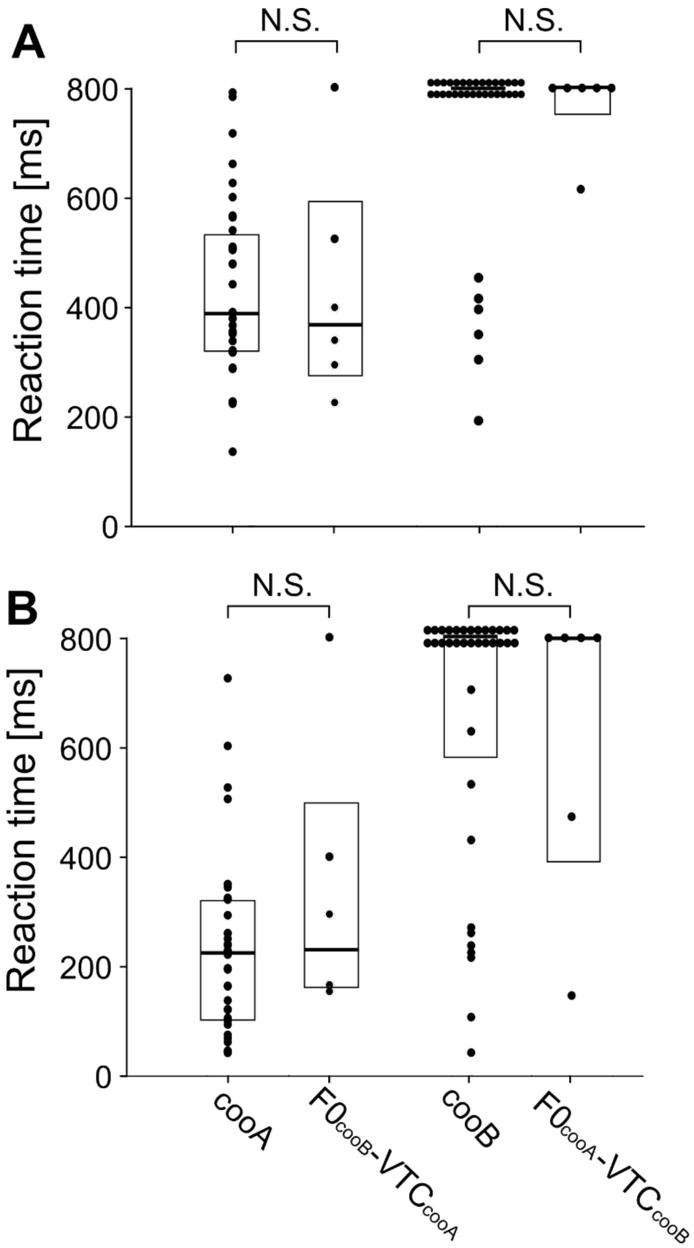


Figure 4-7. Comparisons of the reaction times to the training and test stimuli for the two subjects (A: Subject 1, B: Subject 2). Box plots represent the median (horizontal line) and interquartile range (box) of the indicated distribution. Each plot point represents the reaction time of each trial. N.S.: not significant.