A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*)

James A. Agamaite, Chia-Jung Chang, Michael S. Osmanski, and Xiaoqin Wang[a]
Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205, USA

(Received 20 June 2015; revised 16 September 2015; accepted 29 September 2015; published online 11 November 2015)

The common marmoset (*Callithrix jacchus*), a highly vocal New World primate species, has emerged in recent years as a promising animal model for studying brain mechanisms underlying perception, vocal production, and cognition. The present study provides a quantitative acoustic analysis of a large number of vocalizations produced by marmosets in a social environment within a captive colony. Previous classifications of the marmoset vocal repertoire were mostly based on qualitative observations. In the present study a variety of vocalizations from individually identified marmosets were sampled and multiple acoustic features of each type of vocalization were measured. Results show that marmosets have a complex vocal repertoire in captivity that consists of multiple vocalization types, including both simple calls and compound calls composed of sequences of simple calls. A detailed quantification of the vocal repertoire of the marmoset can serve as a solid basis for studying the behavioral significance of their vocalizations and is essential for carrying out studies that investigate such properties as perceptual boundaries between call types and among individual callers as well as neural coding mechanisms for vocalizations. It can also serve as the basis for evaluating abnormal vocal behaviors resulting from diseases or genetic manipulations.
© *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4934268]

[AMS] Pages: 2906–2928

I. INTRODUCTION

Non-human primates are the closest evolutionary relatives to humans and exhibit many complex behaviors, including the extensive usage of acoustically diverse vocal signals for intra-species communication. Primates use their species-specific vocalizations to identify specific external referents such as predators or food and to convey biologically important information such as gender, talker identity, and emotional state (Seyfarth and Cheney, 2003). Primates also exhibit many of the perceptual and cognitive phenomena observed in human speech, such as categorical perception of vocalizations, robust vocal perception in noisy environments, and learning proper usage of different call types (Moody *et al.*, 1990; Seyfarth and Cheney, 2003). A prerequisite for understanding vocal communication mechanisms in primates, however, is an understanding of the nature of the acoustic signals that primates use in their intra-species communication. In the present study, we have attempted to address this issue in a highly vocal primate species, the common marmoset (*Callithrix jacchus*), that has emerged in recent years as a promising non-human primate model for behavioral, neurophysiological, and anatomical studies of brain mechanisms underlying perception, motor functions, and cognition.

The marmoset is an arboreal New World primate that is particularly well suited for captive studies of vocal communication mechanisms for several reasons. First, these animals have a relatively complex social system (Digby, 1995; Smith, 2006; Bezerra *et al.*, 2007). Vocalizations are of great importance in marmoset social behavior, and they have a rich vocal repertoire that is produced across a large number of social and emotional states (Epple, 1968; Rylands, 1993). Also, unlike most primate species, marmosets remain highly vocal in captivity, and vocalizations produced in captivity share similarities with those produced in the wild (Bezerra and Souto, 2008). Marmosets are small-bodied (adults weigh approximately 300–500 g) and easy to house in a socially interactive colony setting. Marmosets breed well in captivity, giving birth twice a year to either twins or triplets, which allows for the analysis of vocal behavior across all stages of the marmoset life cycle. Finally, marmosets have been used in a number of neurophysiological and anatomical studies of various brain regions in the past two decades (e.g., auditory: Lu *et al.*, 2001; Wang *et al.*, 2005; Kajikawa *et al.*, 2005; visual: Rosa and Tweedale, 2000; Mitchell *et al.*, 2014; imaging: Liu *et al.*, 2013; Belcher *et al.*, 2013). With the recent breakthrough in creating transgenic marmosets (Sasaki *et al.*, 2009), these animals are poised to become a major non-human primate model for neuroscience research.

A deep understanding of vocal communication mechanisms in the marmoset requires a thorough quantitative analysis of its entire vocal repertoire. Although species-specific vocalizations have been described for the marmoset in various previous studies such as the earlier work in adult marmosets by Epple (1968) and the recent work in developing marmosets by Pistorio *et al.* (2006), a full evaluation of the marmoset vocal repertoire using rigorous, quantitative methods has yet to be conducted. In addition, most previous studies largely focused on an acoustically simple call type, an

[a] Electronic mail: xiaoqin.wang@jhu.edu

isolation call termed phee (Jones *et al.*, 1993; Norcross and Newman, 1993; Norcross *et al.*, 1994). By recording vocalizations from marmosets housed in and adapted to a social environment within a captive colony, we were able to study a much wider range of vocalizations than previous studies, including other call types such as twitters and trills.

Several recent studies have used quantitative methods to examine the vocal repertoire of a diverse array of species, including primates (Hedwig *et al.*, 2014; Fuller, 2014), rodents (Kobayasi and Riquimaroux, 2012; Soltis *et al.*, 2012), birds (Giret *et al.*, 2011), and frogs (Pettit *et al.*, 2012). These kinds of analyses are critical to understanding the types of vocalizations produced by a particular species and for carrying out behavioral studies that investigate properties such as perceptual boundaries between vocalization types as well as neurophysiological studies that investigate neural coding mechanisms for vocalizations.

Our goal in this study is to provide a detailed, quantitative analysis of the marmoset's vocal repertoire in a captive, but socially interactive, environment. A detailed, quantitative description of this species' vocalizations is necessary as a basis for future studies to investigate, for example, behavioral correlates to each call type, the ontogeny and plasticity of vocalizations, as well as physiological mechanisms underlying the perception and production of vocalizations in this species. Furthermore, a quantitative description of vocalizations is critical for understanding how the acoustic structure of a call may vary when uttered in different situations (e.g., by different animals or by the same animal in different emotional and/or behavioral states) as well as to determine which features of a vocalization are critical for conspecifics to accurately interpret a call's intended communicative content. Quantitative measures of vocalizations are essential to synthesize stimuli for psychophysical and electrophysiological experiments designed to investigate the cognitive and neural mechanisms involved in vocalization processing (DiMattina and Wang, 2006). Finally, a quantitative description of marmoset vocalizations can serve as the basis for evaluating abnormal vocal behaviors resulting from diseases or genetic manipulations.

II. METHODS

All experimental procedures were approved by the Johns Hopkins University Animal Care and Use Committee.

A. Vocalization recording

We recorded the vocal activity of captive adult marmosets while they were housed in a socially rich colony environment, including animals of all ages. This arrangement ensured that we sampled across different types of marmoset calls. This marmoset colony has been maintained in the animal facility at The Johns Hopkins University School of Medicine since 1995. Marmosets in our colony are highly active and vocal exchanges between members of the colony occur regularly throughout the day. The colony is housed in a climate-controlled room (80 °F–85 °F, >50% humidity). Housing cages were furnished with branches, various forms of behavioral enrichment (e.g., toys, mirrors, etc.), and nesting boxes. Each cage could house a family (a breeding pair with offspring), a pair of adults, or an individual adult. Every cage has both auditory and visual contact with the rest of the colony. Each animal we recorded from was individually housed during recording sessions within a large, socially rich colony room that allowed for both acoustic and visual contact with all the other animals in the colony. This arrangement was made to ensure that the caller identity of each captured vocalization could be reliably determined. Importantly, at no time during the study were any animals isolated from the other members of the colony. The test subjects were able to communicate acoustically and visually with the other animals in the colony at all times. The fact that, not only phee calls, but other types of marmoset vocalizations (twitter, trill, trillphee, etc.) were also observed in all of our recording sessions is a reflection of the interactive social environment in which this study was conducted.

The results reported in this study are based on vocalizations recorded from two different groups of marmosets in our colony. Figure 1 shows a schematic of the general recording paradigm employed for both groups. The recordings of the first group occurred over a 15-month period during 1995–1996, shortly after our colony was established. The colony at this time contained 14 adult marmosets in total, including several breeding pairs that were housed together with their offspring. Recordings at this time were made from a focal group of eight adult marmosets within the colony including four males and four females (referred to hereafter as Population 1). These subjects were housed in individual cages within the colony while their vocalizations were recorded to ensure that the caller identity of each captured and analyzed vocalization could be determined (see below). This is necessary for us to measure individual variation for each call type in a future report. Some of these eight animals were from breeding pairs without young offspring sharing their cage and were occasionally individually housed within the colony to allow recordings to be made and returned to their pair-housed cages after the recordings were completed. Recordings were typically conducted three days a week for four hours at a time.

The recordings of the second group occurred over a 7-month period in 2013. By then our colony had expanded to include ∼100 animals housed in two large rooms, including individually and pair-housed animals in addition to several breeding pairs that were housed with both parents and their offspring. Recordings at this time were made from a focal group of 14 individual marmosets within the colony including 6 males and 8 females (referred to hereafter as Population 2). We sampled vocalizations from the second group of marmosets so that we could examine if there were significant differences in the acoustic properties of animals in our colony over this long time period of 18 yrs. However, this study was not designed to track changes in vocalizations over time. The two groups of animals are not genetically related based on our records.

Recordings of Population 1 were made using directional microphones (AKG Acoustics, Vienna, Austria, C1000S) aimed toward a specific animal to allow vocalizations from that subject to be uniquely identified during audio replay through headphones. Microphone output signals were

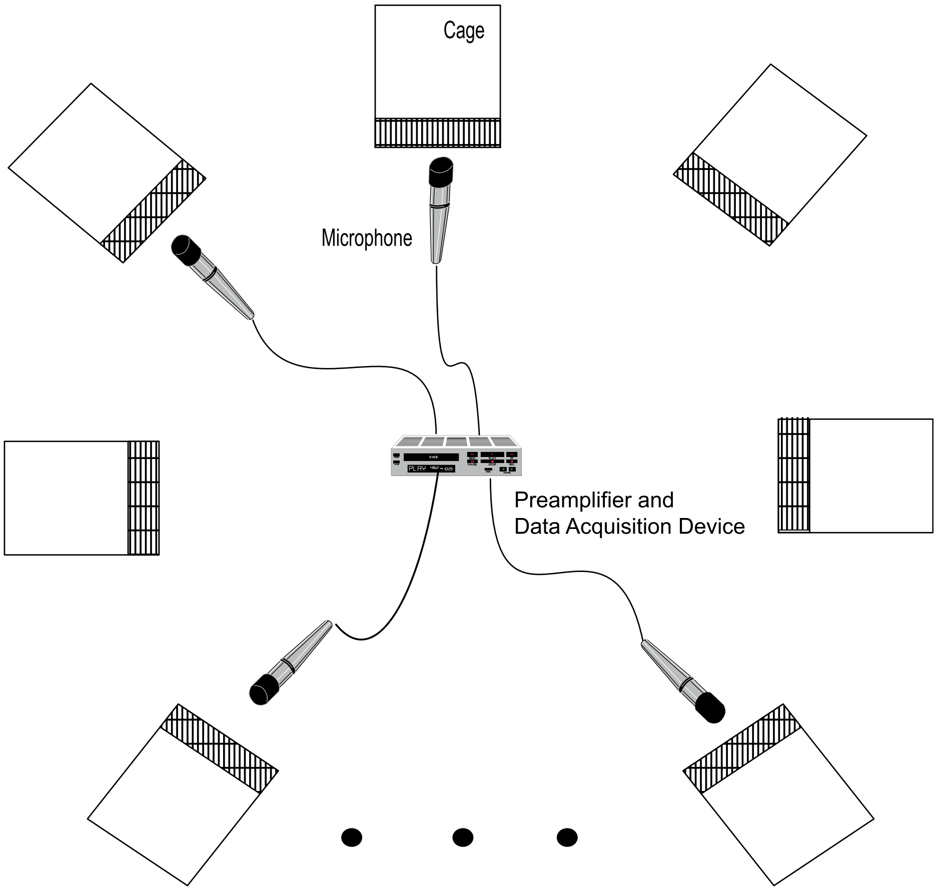

FIG. 1. (Color online) Schematic of the setup used to record vocalizations from individually housed adult marmosets in both populations. Directional microphones were pointed at individual monkeys so that the calls from that monkey could be traced during audio replay.

amplified using a dual microphone preamplifier (Symetrix Inc., Mountlake Terrace, WA, SX202) and subsequently recorded using two 2-channel professional digital audio tape (DAT) recorders (Panasonic Co., Osaka, Japan, SV-3700) sampled at 48 kHz. The two tapes used during four channel recordings were synchronized using a single remote controller to start and stop recordings on both DAT recorders. Recordings of Population 2 were also made with directional microphones (Sennheiser, Wedemark, Germany, ME66) that were placed ∼15 cm in front of the target cage. Microphone output signals were amplified with two dual microphone amplifiers (Symetrix Inc., Mountlake Terrace, WA, 302) and recorded onto a computer via a 4-channel data acquisition interface (National Instrument, Austin, TX, NI 9237) sampling at 50 kHz.

B. Vocalization screening

The goals of this procedure were to segment vocalizations with a sufficient signal-to-noise ratio from background signals and to identify the source (caller) of each captured vocalization. Recorded vocalizations from Population 1 were re-sampled at 50 kHz (from the original 48 kHz) and manually screened via a real-time spectrographic analyzer (RTS, Engineering Design, Bedford, MA) running on a computer concurrent with audio replay through headphones. Vocalizations from specific individuals were identified based on perceived interaural intensity differences reflecting the aim of the directional microphones during recordings. Because of a large amount of vocalizations recorded from Population 2, we developed a custom automated program using MATLAB (Mathworks, Natick, MA) to detect vocalizations based on intensity and duration criteria. To localize the signal source, an algorithm using time differences of arrival (TDOA) and a maximum likelihood estimator (MLE) was applied to identify the speaker source and assign the call to the nearest channel. Technical aspects of these algorithms are described in the Appendix. Captured vocalizations along with caller identity were stored on a computer hard disk with preceding and following silent intervals for further analyses.

C. Vocalization classification and feature measurements of Population 1 recordings

Vocalizations recorded from Population 1 were first manually classified primarily based on visual inspection of spectrogram patterns to establish a corpus of marmoset vocalization types in our colony. Our classification scheme was then confirmed by quantitative analyses of acoustic feature measurements. We adopted Epple's classification system (Epple, 1968) as a starting point for establishing call types. An observed vocalization distinctly dissimilar to all previously defined call types was identified as a new call type if it was uttered by at least two animals and observed during at least two recording sessions. These call type templates had distinctly separable spectrograms. Apart from being given a unique call type identifier, each call was further classified as being a *simple* or *compound* call. Simple calls were defined as basic acoustic elements uttered either as a complete vocalization or as a discrete component (syllable) in a compound call.

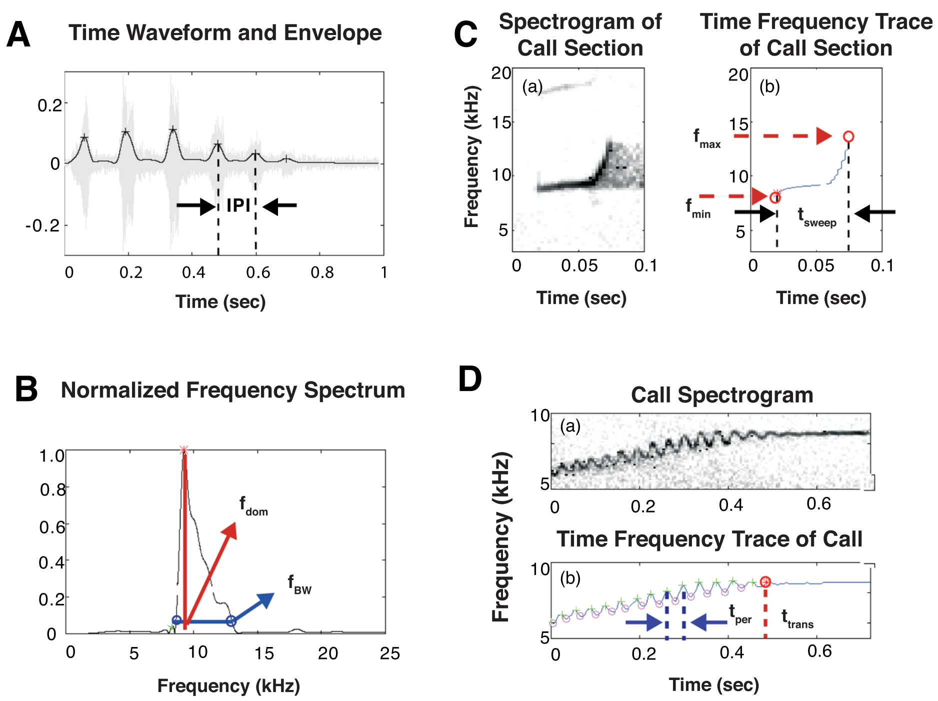

FIG. 2. (Color online) Signal representations used to measure the acoustic features described in Table I, with representative feature measurements for each signal representation shown. (A) Time waveform (gray) and envelope (black) of a twitter call, with detected envelope peaks marked with "+" symbols. (B) Smoothed magnitude of the frequency spectrum for the beginning phrase of a twitter call. The "*" symbol marks the detected spectral peak. (C) Spectrogram and time-frequency trace for the beginning phrase of a twitter call. The minimum and maximum detected frequencies are shown along with the sweep time. (D) Spectrogram and time-frequency trace for a trillphee call. The + markers indicate detected peaks in the FM sinusoid segment of the call, the "O" markers indicate detected troughs in the FM sinusoid segment of the call, and the O marker indicates where the transition point from the FM sinusoidal to tonal segment of the call was detected. The markers in all signal representations were generated using automated feature detection software.

Compound calls were combinations or sequences of simple calls uttered in such a way that the interval between syllables was less than 0.5 s and did not overlap with a complete vocalization from another animal.

For each simple call type, between 7 (peep and tsik calls) and 18 (twitter calls), spectro-temporal features were extracted using custom MATLAB software developed in our laboratory. These features were physically intuitive parameters such as amplitude and frequency modulations, temporal and spectral features, and transition points, etc. Similar approaches have been taken in analyzing the vocalizations of other animal species (e.g., Gamba and Giacoma, 2007, Pettitt *et al.*, 2012). Figure 2 illustrates the analysis methods used along with example features. Simple call features were measured from three signal representations: the time waveform envelope [Fig. 2(A)], frequency spectrum [Fig. 2(B)], and spectrogram [Figs. 2(C) and 2(D)] (see the Appendix for further explanations). Table I lists all measured features for each of the four major call types (e.g., twitters, phees, trills, and trillphees). For phees, trills, and trillphees, features were

TABLE I. Synopsis of measured features from all major call types.

Name	Description	Measured From
Dur (s)	Length of the vocalization	*All Calls*
F_{dom} (kHz)	Frequency corresponding to the maximum in the spectrum	*All Calls*
F_{min} (kHz)	Minimum frequency within a call	*All Calls*
F_{max} (kHz)	Maximum frequency within a call	*All Calls*
F_{start} (kHz)	Starting frequency within a call	*All Calls*
F_{end} (kHz)	Ending frequency within a call	*All Calls*
F_{BW} (kHz)	Frequency bandwidth across a call	*All Calls*
Tf_{min}	Time to minimum frequency	*Phee*, *Trill*, *Trillphee*
Tf_{max}	Time to maximum frequency	*Phee*, *Trill*, *Trillphee*
FM_{rate} (Hz)	Modulation rate of a sinusoidal frequency segment: the average number of successive peaks per second [derived from t_{per} in Fig. 1(D)]	*Trill*, *Trillphee*
Max FM_{depth} (Hz)	Maximum difference between a trough and successive peak in a call's sinusoidal FM segment	*Trill*, *Trillphee*
FM_{depth} (Hz)	Average difference between a trough and successive peak in a call's sinusoidal FM segment	*Trill*, *Trillphee*
T_{trans} (s)	Time of transition from sinusoidal to linear FM	*Trillphee*
N_{phr}	Number of discernible voicing segments in a call	*Twitters*
IPI (ms)	Inter-phrase Interval: the average time interval between consecutive peaks in the envelope	*Twitters*
T_{phr} (ms)	Phrase sweep time taken as the length of time between the minimum and maximum frequencies in a call phrase	*Twitters*

measured across the entire vocalization. Because many twitter features varied over time (e.g., minimum and maximum frequency), each call was divided into three sections (beginning, middle, and ending), and identical measurements were applied to each section. Many marmoset vocalizations have significant energy in harmonics that were highly correlated with the fundamental frequency component. Therefore, analyses of each call type's spectro-temporal structure were only based on the characteristics of that call type's fundamental frequency component.

Due to the noise below 2 kHz in the Population 1 recordings resulting from background sounds in the colony room, all vocalization samples were filtered before further processing was performed. The filters used were sixth-order zero-phase Butterworth filters (sixth-order zero phase filtering is achieved by passing a signal first forward and then backward through the same third order filter). For narrowband vocalizations (phees, trills, and trillphees), a bandpass filter was used with 3-dB cutoff frequencies of 3 and 15 kHz, respectively. For wideband vocalizations (twitters), a high-pass filter was used with a 3-dB cutoff frequency of 3 kHz. No features were measured from the ock and egg call types because most of their energy was concentrated in the same 0–2 kHz band as the background noise.

D. Vocalization feature measurements of Population 2 recordings

Vocalizations from Population 2 were analyzed and compared with those from Population 1 to illustrate how statistically representative the features measured from Population 1 were for each call type, and to explore whether marmosets in our colony exhibited significant differences in the acoustic structures of their vocalizations between populations. Because of the much larger overall sample size of Population 2, however, we chose to adopt an automated rather than manual classification scheme (see below). Prior to classification, each vocalization sample was high-pass filtered at 3 kHz and its harmonics were removed. Spectrograms for all calls were generated using a fixed 512-point fast Fourier transform (FFT) with a 75% overlapping Hamming window, which provided 2.6 msec temporal resolution and 97 Hz frequency resolution. The magnitude traces (defined by the maximum intensity in each windowed spectrum of the signal) were extracted from the spectrogram and smoothed using a trajectory prediction algorithm using the position of each pixel on the spectrogram image. From this smoothed trace representation, along with the envelope and spectrum, we measured those acoustic features previously defined for Population 1. For Population 2, we chose to analyze only the four primary call types in the marmoset repertoire (twitters, phees, trills, and trillphees) due to a relatively low sample size of non-primary calls and a lower signal-to-noise ratio in this dataset. As mentioned above, we implemented an automated support vector machine (SVM) model (Vapnik, 1998; Chang and Lin, 2011) using custom software in MATLAB to automatically classify vocalizations due to the large sample size of Population 2. Classified vocalizations were then analyzed using the acoustic features established for each call type from Population 1. While unsupervised clustering such as the *K*-means method could reveal the hidden structure of unlabeled data, the tendency to produce equal-sized clusters could lead to counterintuitive results. Moreover, the parameter cluster number *K* is difficult to choose without giving external constraints. Therefore, we decided to apply SVM model for categorization. SVM is a discriminative classifier defined by a hyperplane constructed with supervised learning. In other words, given labeled training data, the algorithm outputs an optimal hyperplane that separates labeled data into categories, in a way that the distance from it to the nearest data points on each side is maximized. After the training period, new data points will be classified based on which side they are relative to the hyperplane. After the calls from Population 2 were segmented from raw recordings in the colony, we first randomly selected 10 sessions from the 150 total recording sessions and manually labeled these calls using criteria established from Population 1 to train the SVM classifier. We then randomly split the labeled data into two sets: A training set to construct the model and a validation set to test the model's classification accuracy. This process was repeated ten times for each of the ten recording sessions. Details of this classification procedure are described in the Appendix. Classification accuracy was judged by manually inspecting each call sample of a random subset of calls (∼30%) from the resulting categories. Results of this manual inspection showed 86.63% cross-validated accuracy for our SVM classification.

E. Quantitative comparisons of vocalization features between two populations

In order to accurately compare quantitative measures of the four major marmoset call types between the two populations, we re-classified the data for the four primary call types from Population 1 using the automatic classifier and re-analyzed their acoustic features using the same measurement algorithm developed for the Population 2 dataset. We examined statistical differences between the population distributions using a Mann-Whitney U-test. However, due to the large sample sizes of the two populations, significance testing could produce small *p*-values that do not necessarily indicate practically meaningful differences. Therefore, we further examined the distance between the population means with a measure of effect size (Hedges' *g*), which quantifies the strength of the population differences, where $g = 0.2$ equals a small effect size, $g = 0.5$ means a moderate effect size, and $g = 0.8$ is a large effect size (Hedges and Olkin, 1985).

III. RESULTS

Results are based on 34 629 individually identified vocalizations from the two different marmoset populations (Population 1: $N = 9772$; Population 2: $N = 24\,857$, see Table II). On the basis of the analysis of extensively recorded vocalizations from the 8 focal animals in Population 1, we were able to reliably identify 25 call types. These include 12 types of simple calls and 13 types of

TABLE II. Number of samples for each call type.

Call Type	Population 1	Population 2	Total
Twitter	1312	7963	**9275**
Phee	763	10 700	**11 463**
Trill	2000	2676	**4676**
Trillphee	1635	1098	**2733**
p-Peep	98	*	**98**
t-Peep	274	*	**274**
sa-Peep	21	*	**21**
sd-Peep	103	*	**103**
dh-Peep			
Tsik	117	*	**117**
Egg	147	*	**147**
Ock	45	*	**45**
Compound	3257	2420	**5677**
Total	**9772**	**24 857**	**34 629**

TABLE III. Feature measurements for the Twitter-class call type (mean ± std).

Feature	Twitter—Population 1 (963)	Twitter—Population 2 (1672)
Dur (s)	1.01 ± 0.36	1.34 ± 0.45
N_{phr}	7.72 ± 2.55	9.44 ± 2.90
IPI (ms)	139.80 ± 18.23	143.83 ± 19.40
T_{phr}^{B} (ms)	46.90 ± 20.68	35.59 ± 22.24
T_{phr}^{M} (ms)	39.26 ± 9.24	41.47 ± 11.70
T_{phr}^{E} (ms)	42.37 ± 10.22	40.31 ± 12.88
F_{dom}^{B} (kHz)	9.72 ± 1.47	8.32 ± 1.96
F_{dom}^{M} (kHz)	7.47 ± 0.68	7.17 ± 0.60
F_{dom}^{E}(kHz)	6.79 ± 0.60	6.65 ± 1.39
F_{min}^{B} (kHz)	8.39 ± 1.24	7.14 ± 1.71
F_{min}^{M} (kHz)	6.10 ± 0.62	5.63 ± 0.69
F_{min}^{E}(kHz)	6.02 ± 0.60	5.59 ± 1.13
F_{max}^{B} (kHz)	12.87 ± 1.92	10.54 ± 2.44
F_{max}^{M} (kHz)	12.23 ± 1.59	11.40 ± 1.46
F_{max}^{E} (kHz)	9.67 ± 1.44	9.39 ± 1.78
F_{BW}^{B} (kHz)	4.48 ± 1.73	3.40 ± 2.00
F_{BW}^{M} (kHz)	6.19 ± 1.74	5.85 ± 1.65
F_{BW}^{E}(kHz)	3.64 ± 1.49	3.80 ± 1.86

compound calls (see Sec. II). As noted earlier, four call types comprise the majority of vocalizations produced by marmosets in captivity (twitter, phee, trill, and trillphee), and together these constitute 81.28% of all calls collected across both populations. We describe below the defining characteristics of each call type. For ten of the simple call types for which a sufficiently large number of samples were available from at least one population, quantitative analyses of their acoustic parameters were performed (Tables III–V). No quantitative analyses were applied to the compound calls due to insufficient numbers of samples. In addition to the 25 call types identified in this study, we also observed a number of vocalizations that do not belong to any of the classified call types. These calls were unclassified due to their limited occurrence.

A. Simple call types

Table II lists the number of samples for each of the 12 simple call types. Defining characteristics of simple call types are described below in four groups according to their overall acoustic structures: *twitter*-class, *phee*/*trill*-class, *peep*-class, and others. Twitter-class and phee/trill-class are long duration calls whereas peep-class includes short duration calls. Tables III–V provide mean and standard deviation values for all features measured for simple call types. Unless specified, the statistics cited in the text are mean and standard deviation (std) (mean ± std). The range of numerical values provided in the text refers to 80%–90% of observed cases.

1. Twitter-class

One of the most commonly observed call types in our colony is the “twitter.” This call type forms its own class because no other call types resemble the unique features of this wide-band call type. Several examples of twitter calls are shown in Fig. 3. Twitter is most often observed as a complete call and is frequently uttered in vocal exchanges among two or more marmosets. Usually, marmosets in the colony responded to twitters from other conspecifics with either a twitter of their own or sometimes another type of vocalization (e.g., a trill). Twitter calls are characterized by a sequence of upward frequency modulated (FM) sweeps (“phrases”) uttered at regular intervals (Fig. 3). These

TABLE IV. Feature measurements for the Phee/trill-class call types (mean ± std).

Feature	Phee Pop 1 (2246)	Phee Pop 2 (10 595)	Trill Pop 1 (1740)	Trill Pop 2 (547)	Trillphee Pop 1 (1528)	Trillphee Pop 2 (844)
Dur (s)	1.15 ± 0.50	1.21 ± 0.35	0.45 ± 0.16	0.47 ± 0.19	0.95 ± 0.31	1.09 ± 0.36
F_{dom} (kHz)	7.16 ± 0.48	7.16 ± 0.50	6.64 ± 0.82	6.66 ± 0.83	7.17 ± 0.53	7.43 ± 0.72
F_{min} (kHz)	6.89 ± 0.43	6.83 ± 0.55	5.97 ± 0.86	5.99 ± 0.79	6.78 ± 0.55	7.01 ± 0.72
F_{max} (kHz)	8.19 ± 0.75	8.14 ± 0.58	7.70 ± 0.91	7.78 ± 0.81	8.00 ± 0.53	8.30 ± 0.61
T_{fmin}	0.07 ± 0.22	0.07 ± 0.22	0.29 ± 0.39	0.27 ± 0.38	0.11 ± 0.29	0.19 ± 0.32
T_{fmax}	0.77 ± 0.24	0.82 ± 0.20	0.69 ± 0.27	0.71 ± 0.26	0.68 ± 0.30	0.66 ± 0.34
F_{start} (kHz)	7.01 ± 0.50	6.95 ± 0.53	6.33 ± 0.91	6.34 ± 0.86	6.33 ± 0.91	6.34 ± 0.86
F_{end} (kHz)	7.81 ± 0.71	7.88 ± 0.63	6.96 ± 1.03	7.03 ± 1.01	6.96 ± 1.03	7.03 ± 1.01
F_{BW} (kHz)	1.30 ± 0.60	1.30 ± 0.65	1.73 ± 0.89	1.79 ± 0.81	1.22 ± 0.55	1.29 ± 0.75
FM_{rate} (Hz)	*	*	29.96 ± 5.38	29.69 ± 5.89	25.78 ± 5.94	32.17 ± 11.52
Max FM_{depth} (kHz)	*	*	0.92 ± 0.36	0.87 ± 0.42	0.42 ± 0.23	0.40 ± 0.40
FM_{depth} (kHz)	*	*	0.50 ± 0.18	0.49 ± 0.19	0.19 ± 0.09	0.09 ± 0.07
T_{trans} (s)	*	*	*	*	0.40 ± 0.22	0.29 ± 0.24

TABLE V. Feature measurements for the Peep-class and Tsik call types (mean ± std).

Feature	p-Peep (490)	t-Peep (396)	Sa-Peep (60)	Sd-Peep (590)	Dh-Peep (34)	Tsik (287)
Dur (s)	0.15 ± 0.08	0.12 ± 0.05	0.05 ± 0.05	0.07 ± 0.03	0.19 ± 0.07	0.06 ± 0.01
F_{min} (kHz)	6.53 ± 0.91	6.26 ± 1.22	6.93 ± 1.56	6.34 ± 0.95	6.68 ± 1.09	5.20 ± 1.70
F_{max} (kHz)	7.36 ± 0.90	7.92 ± 1.08	9.00 ± 2.50	8.54 ± 1.36	9.00 ± 1.05	18.3 ± 1.59
T_{fmin}	0.02 ± 0.04	0.05 ± 0.06	0.01 ± 0.03	0.05 ± 0.02	0.11 ± 0.07	0.06 ± 0.01
T_{fmax}	0.08 ± 0.07	0.07 ± 0.04	0.04 ± 0.04	0.01 ± 0.02	0.01 ± 0.02	0.04 ± 0.01
F_{start} (kHz)	6.75 ± 0.90	6.79 ± 1.31	7.10 ± 1.52	8.19 ± 1.40	8.48 ± 1.60	13.9 ± 1.67
F_{end} (kHz)	7.08 ± 0.93	6.99 ± 1.22	8.46 ± 2.48	6.89 ± 1.12	7.07 ± 0.94	5.38 ± 1.87

phrases appear as periods of high amplitude in the time-amplitude waveform, separated by unvoiced silence periods.

Table III shows statistics of acoustic features of twitter calls measured in the two populations. Twitters in both populations were on average ∼1 s in duration [(Population 1: 1.01 ± 0.36, Population 2: 1.34 ± 0.45, Table III and Fig. 4(A)] and typically consisted of 3–15 phrases [Population 1: 7.72 ± 2.55, Population 2: 9.44 ± 2.90, Table III and Fig. 4(B)] with the inter-phrase interval (IPI) (from the start of one phrase to the start of the next) being 120–160 ms [Population 1: 139.8 ± 18.23, Population 2: 143.83 ± 19.40, Table III and Fig. 4(C)]. There were highly significant differences between the population distributions of these features, and the effect sizes were large for duration ($g = 0.797$) and medium for the number of phrases ($g = 0.620$). Therefore, these results indicate that the twitters of Population 2 showed a longer duration and a greater number of phrases compared to Population 1. On the other hand, the small effect size for IPI ($g = 0.219$) suggests the two populations shared similar average IPIs.

Twitter phrases are approximately piecewise linear ascending FM sweeps that vary in starting frequency and bandwidth depending on a phrase's position in the call. For example, the middle phrases that comprise most of the call generally start at 5–7 kHz and sweep through a bandwidth of 3–10 kHz in 20–60 ms [Figs. 4(E), 4(H), and 4(K)]. The ending phrases tend to start at the similar frequency as the

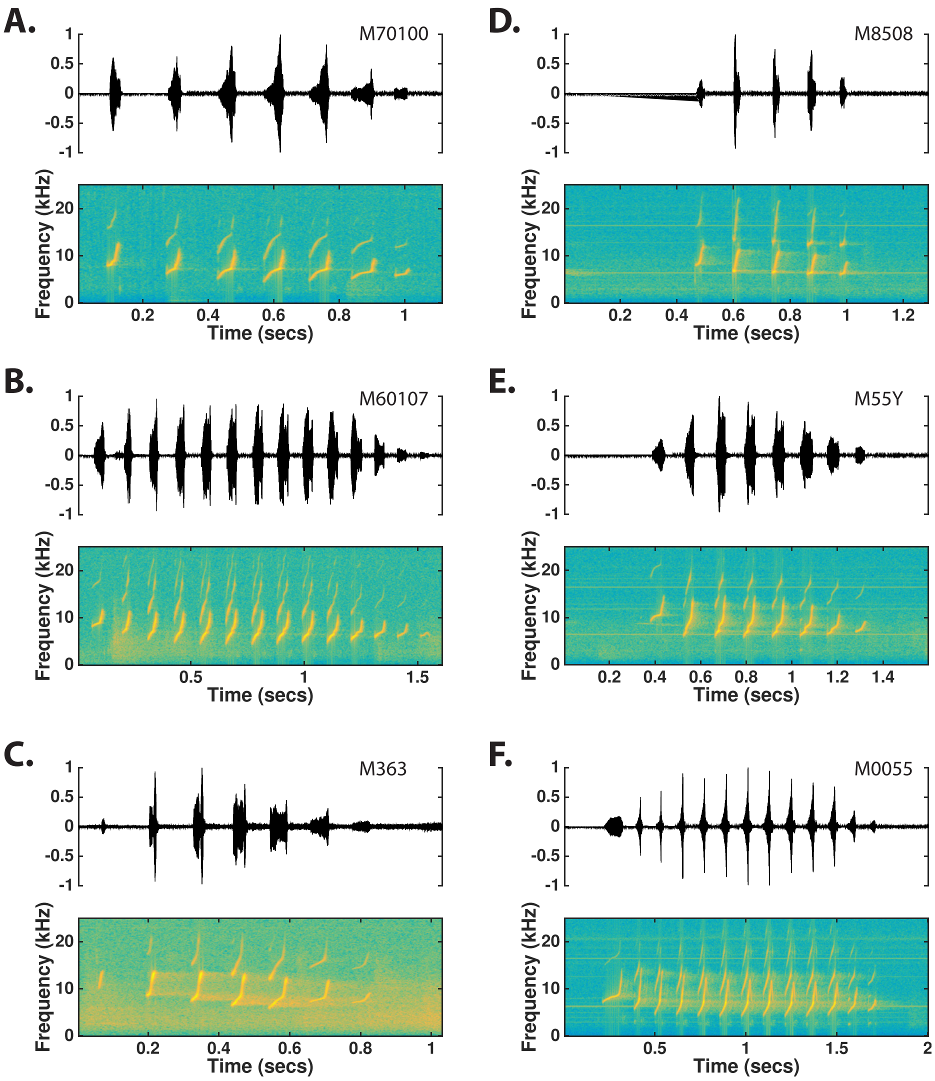

FIG. 3. (Color online) Time waveforms and spectrograms of twitter calls observed from six different monkeys. Although the phrased nature and upward tendency of the FM sweeps comprising the phrases makes the twitter call easily recognizable, twitters from different monkeys show a large degree of variation in the interval between phrases, the number of phrases typically uttered, and the specific time-frequency structure of the FM sweeps.

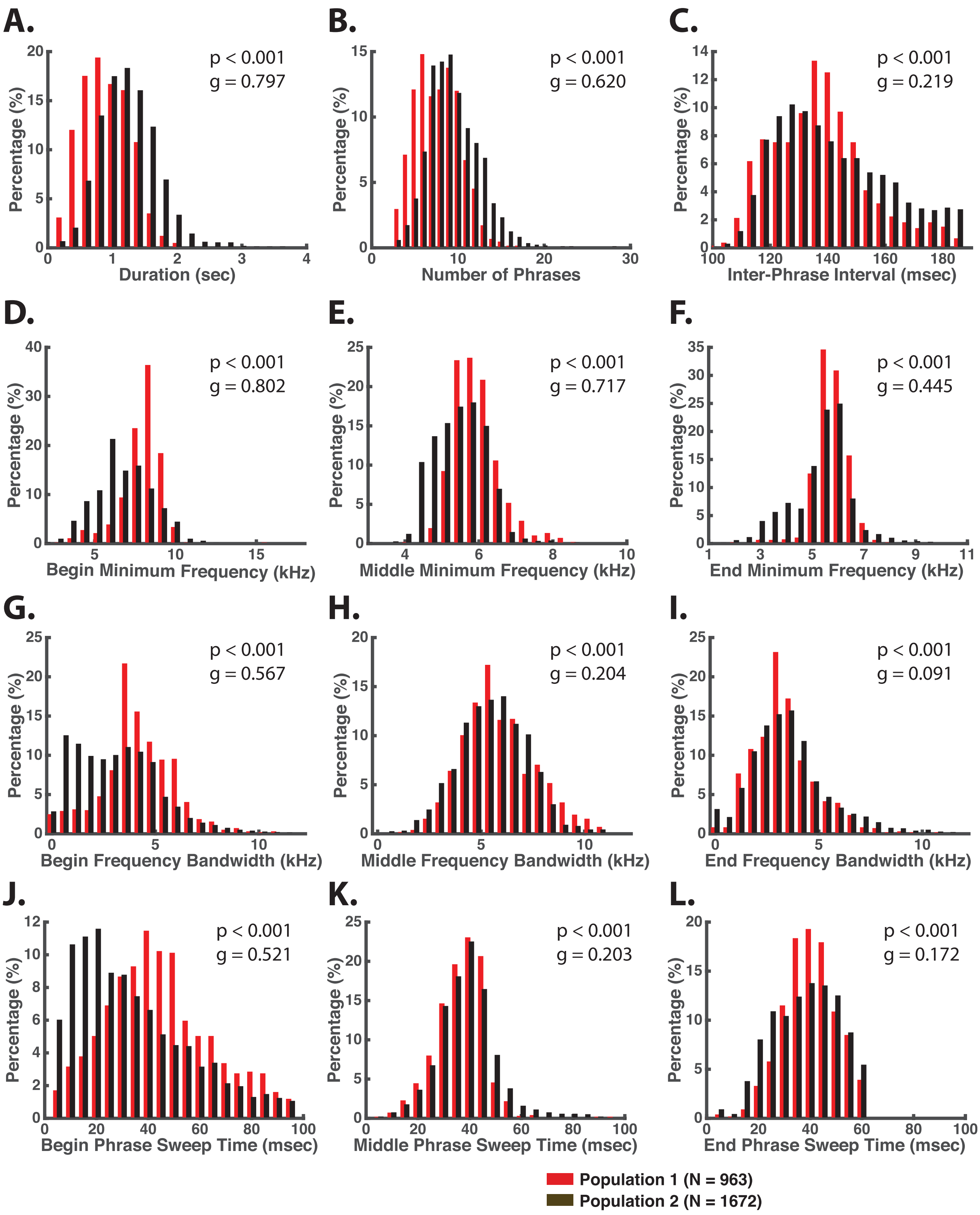

FIG. 4. (Color online) Examples of observed twitter call features are based on measurements made from both populations.

middle phrases, but frequently sweep through a much narrower bandwidth (0.5–6 kHz) in roughly the same amount of time [Fig. 4(F), 4(I), and 4(L)]. The beginning phrases in a twitter also sweep through a narrow bandwidth (0.5–6 kHz) compared to the middle phrases, but typically start at a significantly higher frequency (7–10 kHz) than either the middle or ending phrases and usually take longer in their FM sweep (20–80 ms) [Figs. 4(D), 4(G), and 4(J)]. Furthermore, the beginning phrases showed medium to large effect sizes ($g > 0.5$) between the two populations in minimum frequency, bandwidth, and sweep time. These data indicate that Population 2 showed lower starting frequencies, narrower bandwidths, and shorter sweep times in the beginning phrases of a twitter compared to Population 1. Effect sizes for the ending phrases were small to medium ($g < 0.5$ for IPI and minimum frequency, $g < 0.2$ for bandwidth and sweep time), suggesting small differences between the two populations in the ending phrases of twitters.

2. Phee/trill-class

This class contains three types of narrowband calls: *phee*, *trill*, and *trillphee* (Table IV). These call types have a relatively simple acoustic structure compared to twitters, essentially comprised of a single long duration narrowband fundamental frequency component.

a. Phee. Phee calls were by far the most commonly produced vocalization by marmosets in the colony (Fig. 5). Marmosets usually make this type of call when they are not in physical contact or are separated from other marmosets, but they also frequently produce phees in a social environment such as our colony, perhaps trying to communicate with other conspecifics that are not immediately visible due to the cage arrangement. Marmosets have been shown to produce phees in antiphonal vocal exchanges between pairs of individuals (Miller *et al.*, 2010). The acoustic structure of phees was highly conserved between the two populations (see Table IV). Although most of the features showed statistical significance when comparing between the two populations, the effect sizes were universally small ($g < 0.2$), which suggests that phee call structure was largely similar between the two populations. Phees are 0.5–2.0 s long [Population 1: 1.15 ± 0.50, Population 2: 1.21 ± 0.35, Table IV and Fig. 6(A)] and are typically uttered between 6 and 8 kHz [Population 1: 7.16 ± 0.48, Population 2: 7.16 ± 0.50, Table IV and Fig. 6(C)]. Phee calls are often produced at high intensity although sometimes they could be heard as faint whistles. Phees could be produced as either a single simple call or as part of a compound call. Phees generally began with a short upward FM sweep that transitioned to a long flat or gradually ascending FM sweep (Fig. 5). While there was a large degree of variability in how phee calls ended, they most commonly end with either an abrupt cessation of the long flat FM sweep [Fig. 5(A)] or with a rapid descending FM sweep [Figs. 5(C) and 5(F)]. Note that although phees exhibit a highly regular frequency-time structure, they show no such regularity in their amplitude-time characteristics (Fig. 5).

b. Trill. Trills are primarily distinguished from phees by their characteristic sinusoidal FM structure (Fig. 7, Table IV). Trill calls most often occur as a complete simple call and frequently occur in vocal exchanges among two or more

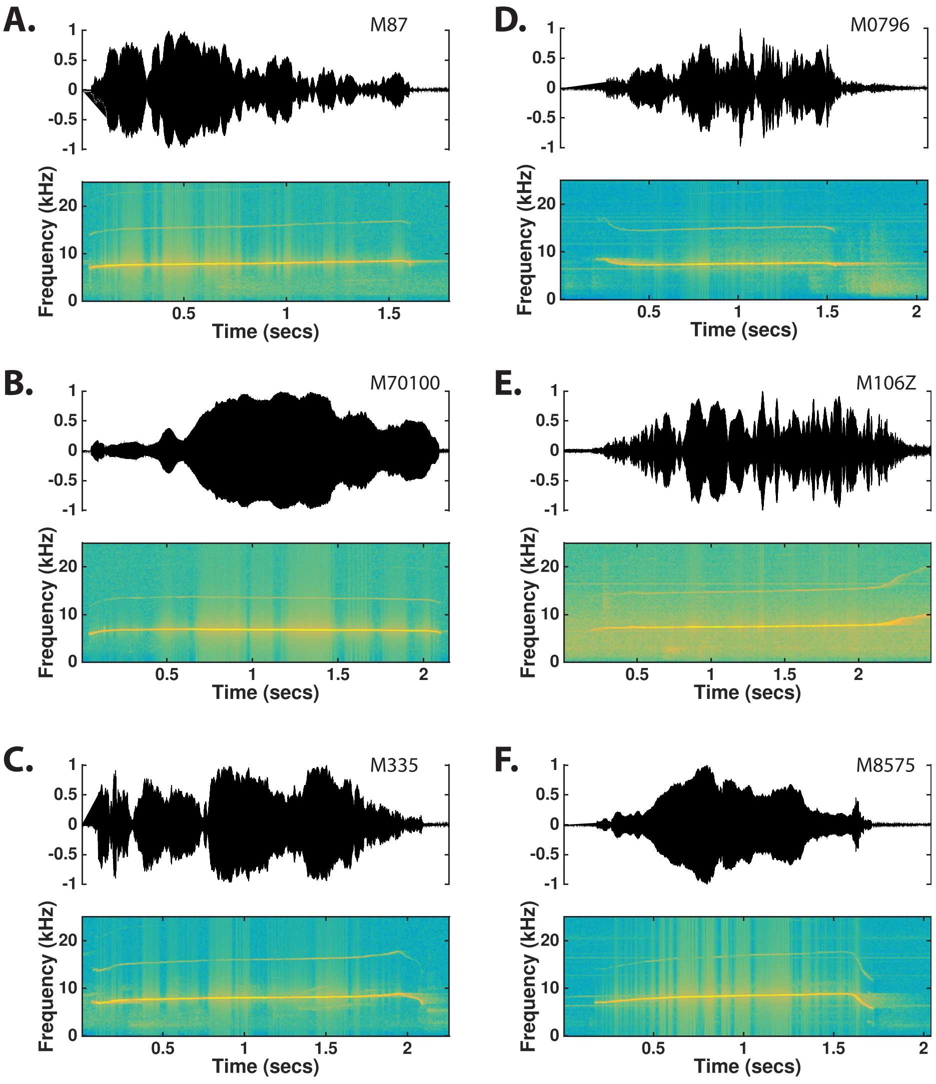

FIG. 5. (Color online) Time waveforms and spectrograms of phee calls observed from six different monkeys. Contrasting the time-waveforms of (A), (B), and (C) clearly shows the broad dynamic range used in phee calls. The time-frequency characteristics of phee calls are highly stereotyped and variation between utterances is largely limited to the end of the call.

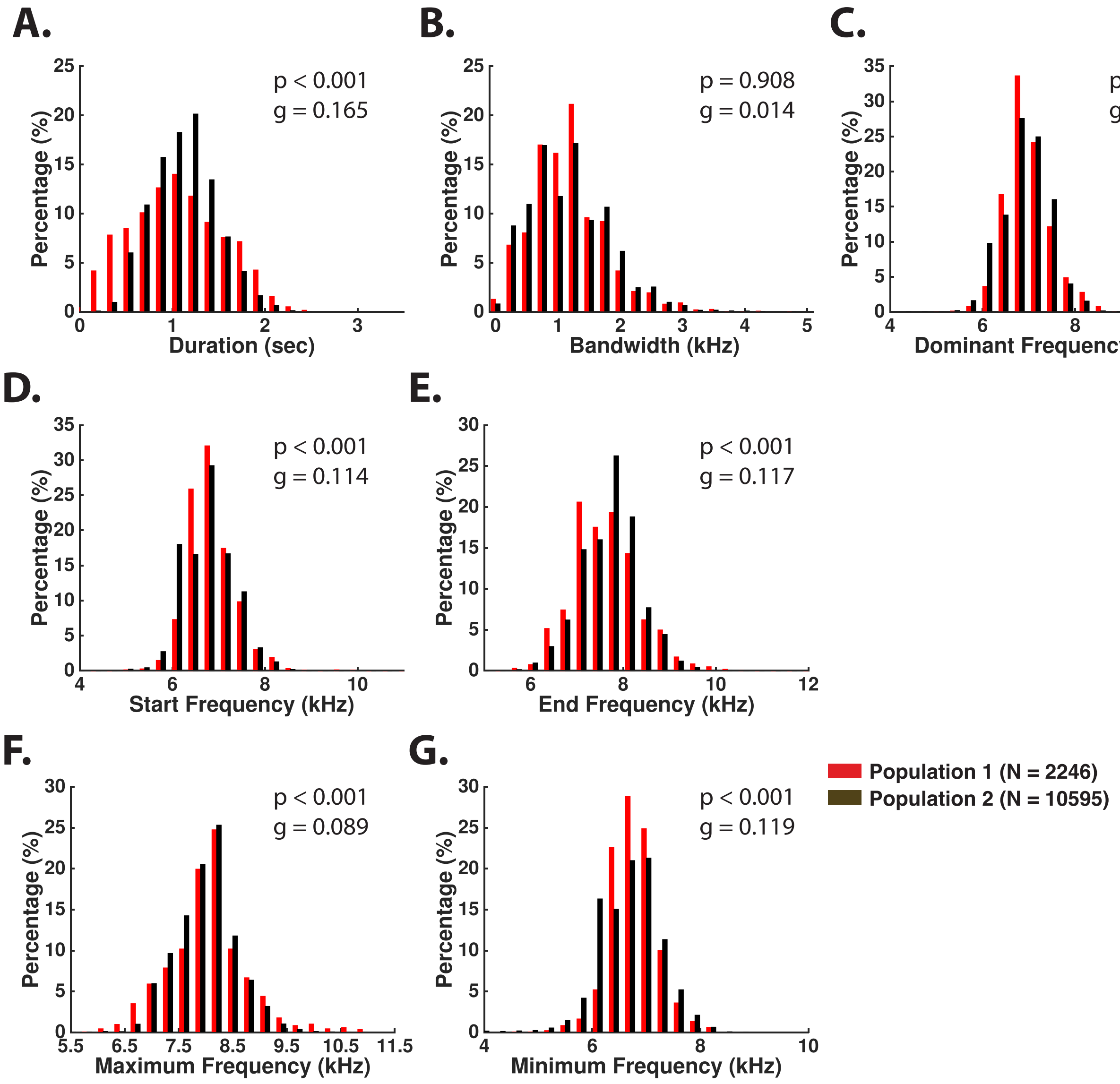

FIG. 6. (Color online) Phee call spectro-temporal characteristics are observable in the distributions of measurements made from calls across the two populations.

marmosets. Trills are largely used by marmosets within a short distance of each other, in contrast to phees. As with the phee calls, the acoustic structure of trills appeared highly conserved between the two populations (Table IV). All of the features showed small effect sizes ($g < 0.2$), suggesting that trill call structure was largely the same between the two populations. Similar to phee calls, a trill's fundamental component is narrowband in nature and is typically uttered between 5 and 8 kHz [Population 1: 6.64 ± 0.82 s, Population 2: 6.66 ± 0.83, Table IV and Fig. 8(B)]. Trills are typically 250–600 ms long [Population 1: 0.45 ± 0.16 s, Population 2: 0.47 ± 0.19, Table IV and Fig. 8(A)] and are produced at relatively low intensities; both characteristics are markedly different from those of phee calls. The sinusoidal modulation seen in the spectrogram of trills has a modulation frequency of 25–33 Hz (i.e., a cycle period of 30–40 ms) and a modulation depth of 0.2–1.2 kHz. Trills generally exhibit sinusoidal amplitude modulation (AM) in their amplitude-time characteristic, corresponding to sinusoidal FM in their spectrum [Fig. 7(A)].

c. Trillphee. The trillphee is an intermediate call type between phee and trill calls. Trillphees are identified as beginning with a sinusoidal FM segment that dampens into a slowly rising linear FM segment (Fig. 9). As with trills, trillphees are usually observed as complete calls. The duration of a trillphee is typically 0.5 to 1.5 s [Population 1: 0.95 ± 0.31, Population 2: 1.09 ± 0.36, Table IV and Fig. 10(A)], similar to phees. The average modulation rate of the sinusoidal FM component is 20–40 Hz [Population 1: 25.78 ± 5.94, Population 2: 32.17 ± 11.52, Table IV and Fig. 10(E)]. The intensity range of trillphees tends to be in between phees and trills. Although the transition point from sinusoidal to linear FM usually occurs within 60% of the call's duration, it is also not uncommon for the transition point to occur in the latter 40% of the call [Fig. 10(D)]. Both populations showed a high degree of overlap across all measured feature distributions, although there were medium effect sizes ($g \sim 0.3$–0.7) for mean FM rate, time to transition, and minimum and maximum frequency. These effect sizes suggest that Population 2 had higher FM rates, shorter

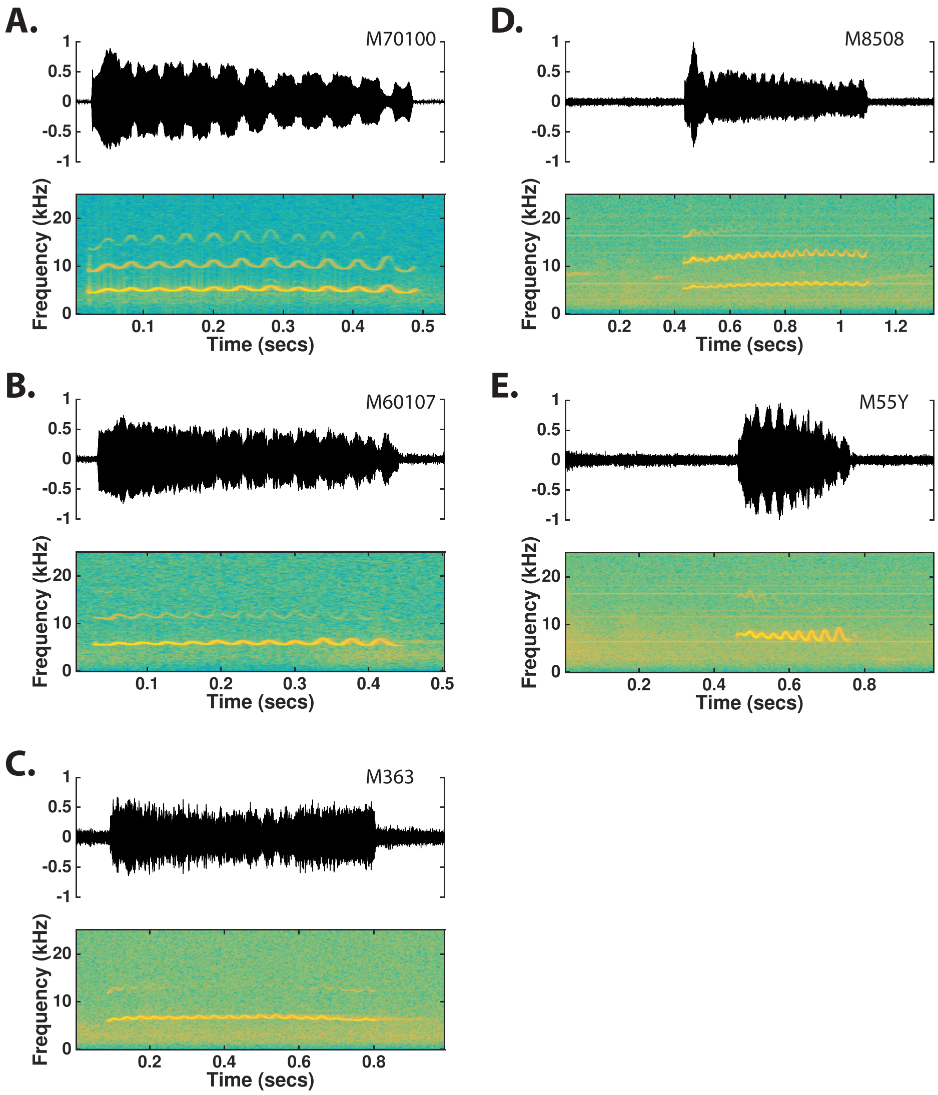

FIG. 7. (Color online) Time waveforms and spectrograms of trill calls observed from five different monkeys. The trill is distinguished based on the characteristic sinusoidal FM that comprises the call.

times to transition to linear FM, and higher overall fundamental frequencies compared to Population 1 (Table IV).

3. Peep-class

There are five call types in peep-class (Table V and Fig. 11). These call types all have short durations and are classified mainly based on their frequency-time characteristics because their amplitude-time characteristics are highly variable. Peeps were seen as both simple calls and as components in compound calls with the exception of dh-Peeps, which were exclusively seen in compound calls. The presence of background noise in our Population 2 recordings rendered our automatic detection algorithm unable to capture short duration calls, and so all feature measurements from these calls are taken solely from recordings from Population 1.

a. Phee-like peep (*p-Peep*). *P-peeps* are distinguished from phee calls based on their short duration (0.15 ± 0.08 s, Table V). An example is given in Fig. 11(A). These calls are uttered at low intensity levels. In all other regards, p-peeps share the same characteristics as phee calls (Fig. 5). P-peeps are generally uttered as a component in a compound call.

b. Trill-like peep (*t-Peep*). *T-peeps* are distinguished from trill calls based on their short duration [Fig. 11(B)]. T-peeps are 30–200 ms long (Table V) and are uttered at low intensity levels. In all other regards, t-peeps share the same characteristics as trill calls (Fig. 7). T-peeps are usually observed as a complete call.

c. Sharply ascending peep (*sa-Peep*). *Sa-peeps* are rapidly ascending FM sweeps [Fig. 11(C)]. Sa-peeps are 10–80 ms long (Table V) and are uttered at relatively low intensity levels. These peeps generally start at 4–9 kHz (7.10 ± 1.52) and pass through a bandwidth of 0.2–5 kHz. The shape of the FM sweep is highly variable, but it is usually either linear or piecewise linear. Sa-peeps have no obvious structure in their time-amplitude characteristics. Sa-peeps are usually uttered as a component in a compound call type.

d. Sharply descending peep (*sd-Peep*). *Sd-peeps* are rapidly descending FM sweeps [Fig. 11(D)]. Sd-peeps are 30–120 ms long (Table V) and are uttered at relatively low intensity levels. These peeps generally sweep through a bandwidth of 0.5–4 kHz terminating at 4–8 kHz (6.89 ± 1.12). The shape of the FM sweep is highly variable and may be either linear, piecewise linear, or slightly curved. Furthermore, the sd-peep may begin with a brief ascending

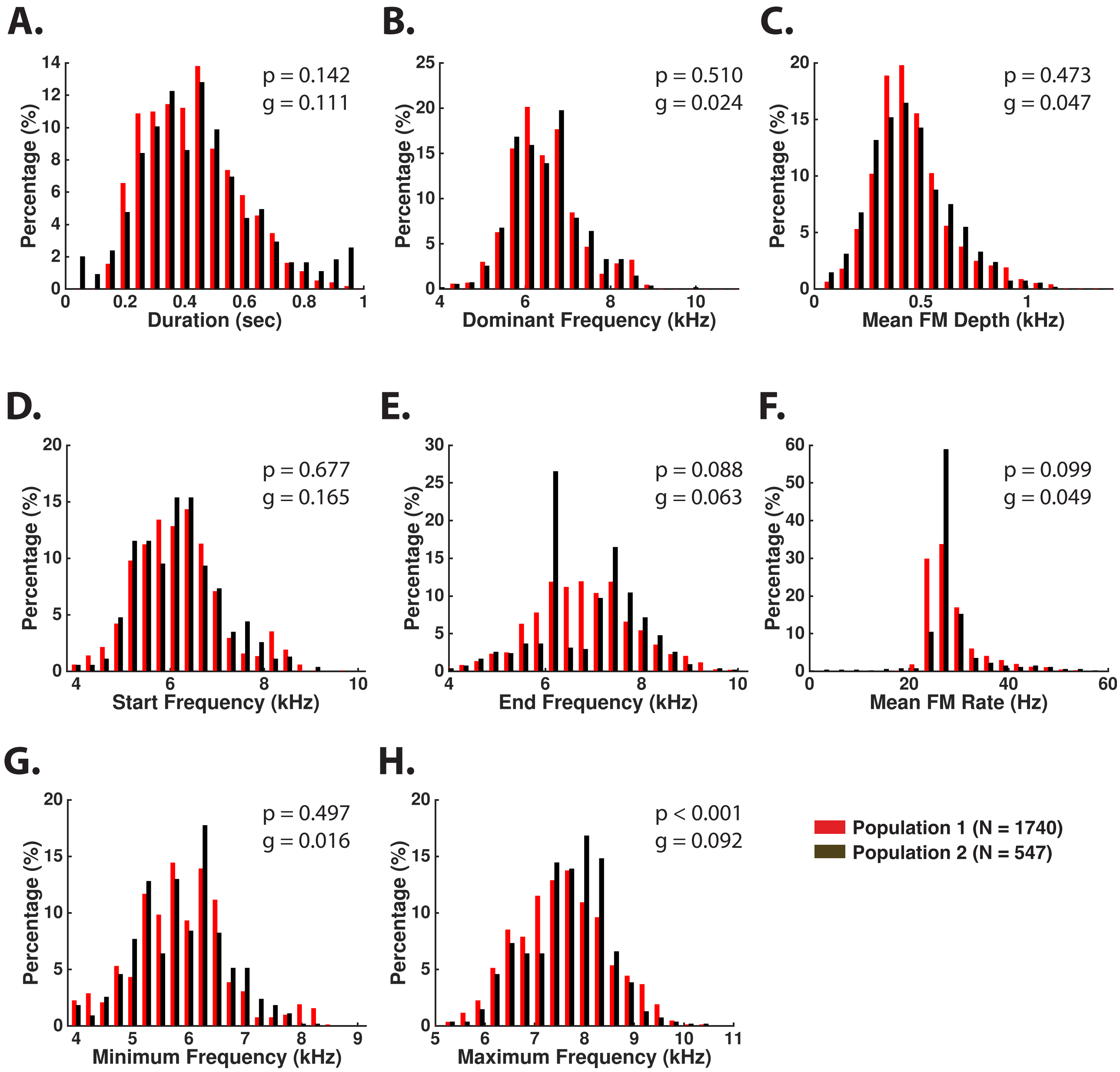

FIG. 8. (Color online) Examples of observed trill call features are based on measurements made from both populations.

FM segment before transitioning to the characteristic downward FM sweep. Sd-peeps have no obvious structure in their time-amplitude characteristics. Sd-peeps are usually uttered as a component in a compound call type.

e. Descending to hump peep (dh-Peep). Dh-peeps are characterized by a descending FM segment that transitions into an FM "arch" [Fig. 11(E)]. Dh-peeps are 50–250 ms long (Table V) and are uttered at low intensity levels. Dh-peeps utilize a bandwidth of 0.5–3.5 kHz. Usually this bandwidth is traversed by the linear descending portion of the call that ends at 4–8 kHz. The arched portion of the call varies in bandwidth from being almost completely flat to using the same bandwidth as the descending portion. Typically the arched call segment is only slightly longer in duration than the descending segment. Dh-peeps have no obvious structure in their time-amplitude characteristic. Dh-peeps are almost exclusively uttered as a component in a compound call type.

4. Other call types

a. Tsik. The *tsik* is a broadband call consisting of a linearly ascending FM sweep that transitions directly into a sharply descending linear FM sweep [Fig. 12(A)]. Tsiks are extremely short calls (0.06 ± 0.01 s, Table V) that are typically uttered at high intensity levels. The tsik is the broadest band simple call uttered by the marmoset, generally starting at the frequency of about 14 kHz and traversing a bandwidth of 9–16 kHz in the descending FM segment of the call (Table V). The ascending portion of the call is more gradual in its slope and generally occupies a narrower bandwidth than does the descending segment. No regular structure was observed in the tsik's time-amplitude characteristic. Tsiks were most frequently uttered as a component in a compound call.

b. Egg and ock. Eggs and *ocks* are easily recognizable as the lowest frequency vocalizations the marmoset produces [Figs. 12(B) and 12(C)]. Eggs and ocks were both

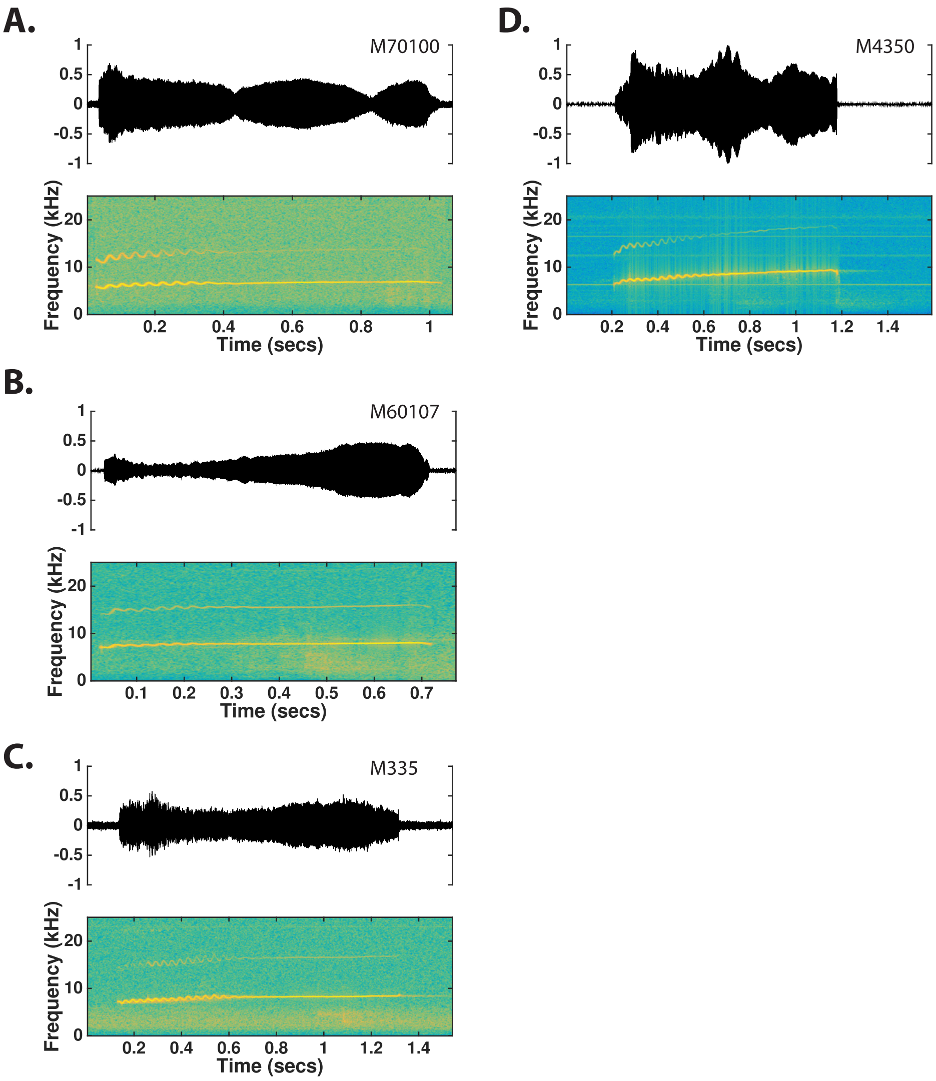

FIG. 9. (Color online) Time waveforms and spectrograms of trillphee calls observed from four different monkeys. Trillphees are a hybrid form of the trill and phee, containing both a sinusoidal FM segment and a flat tonal segment.

estimated to be ∼20–80 ms long and uttered at low intensity levels. Eggs are tonal in structure with a fundamental frequency estimated to be ∼0.8–1.6 kHz and a clear harmonic structure that may extend up to approximately 10 kHz. Ocks are generally characterized by fundamental frequencies of a few hundred Hz with a significant amount of signal energy in noisy components that may extend as high as 20 kHz. Neither eggs nor ocks reveal any regular structure in their time amplitude characteristics. Eggs and ocks were most frequently uttered as a component in a compound call.

B. Compound call types

Thirteen compound call types were identified. The following paragraphs describe typical sequences of the simple call combinations that make up compound calls. Quantitative features were not measured from compound calls due to limited samples, therefore syllable intervals stated were derived from empirical estimates.

(1) Phee-string: Phee-strings are concatenations of 2–6 phee calls uttered in succession with intervening silent intervals less than 500 ms in duration [Figs. 13(A)–13(C)]. Occasionally, the initial component may actually be a trill or a trillphee. Phee-strings containing three or more individual phees may show a steady decrease in the duration of each component. In these cases, the final phee may actually be classified as p-peeps, but the call as a whole is still classified as a phee-string.

(2) Peep-phee: Peep-phees consist of a single phee or a phee-string preceded by 1–6 *sd-peeps* and/or *fd-peeps* uttered with intervening silent intervals of less than 300 ms [Fig. 13(D)]. The combination of sd-peeps and fd-peeps used in the peep-phee is highly variable from one utterance to the next.

(3) Phee-peep*:* Phee-peeps consist of a single phee followed by one or more *p-peeps* or *sa-peeps* of extremely short duration [Figs. 14(A) and 14(B)]. The interval between the phrases is typically 20–200 ms. Occurrences of these calls were rare.

(4) Peep-Trill: Peep-trills consist of a trill preceded by a single *sd-peep* or *p-peep* of extremely short duration [Fig. 14(C)]. The interval between phrases is typically 20–60 ms.

(5) Peep-Trillphee: Peep-trillphees consist of a trillphee preceded by a single sd-peep or p-peep [Fig. 14(D)]. The interval between phrases is typically 20–60 ms. Occurrences of these calls were rare.

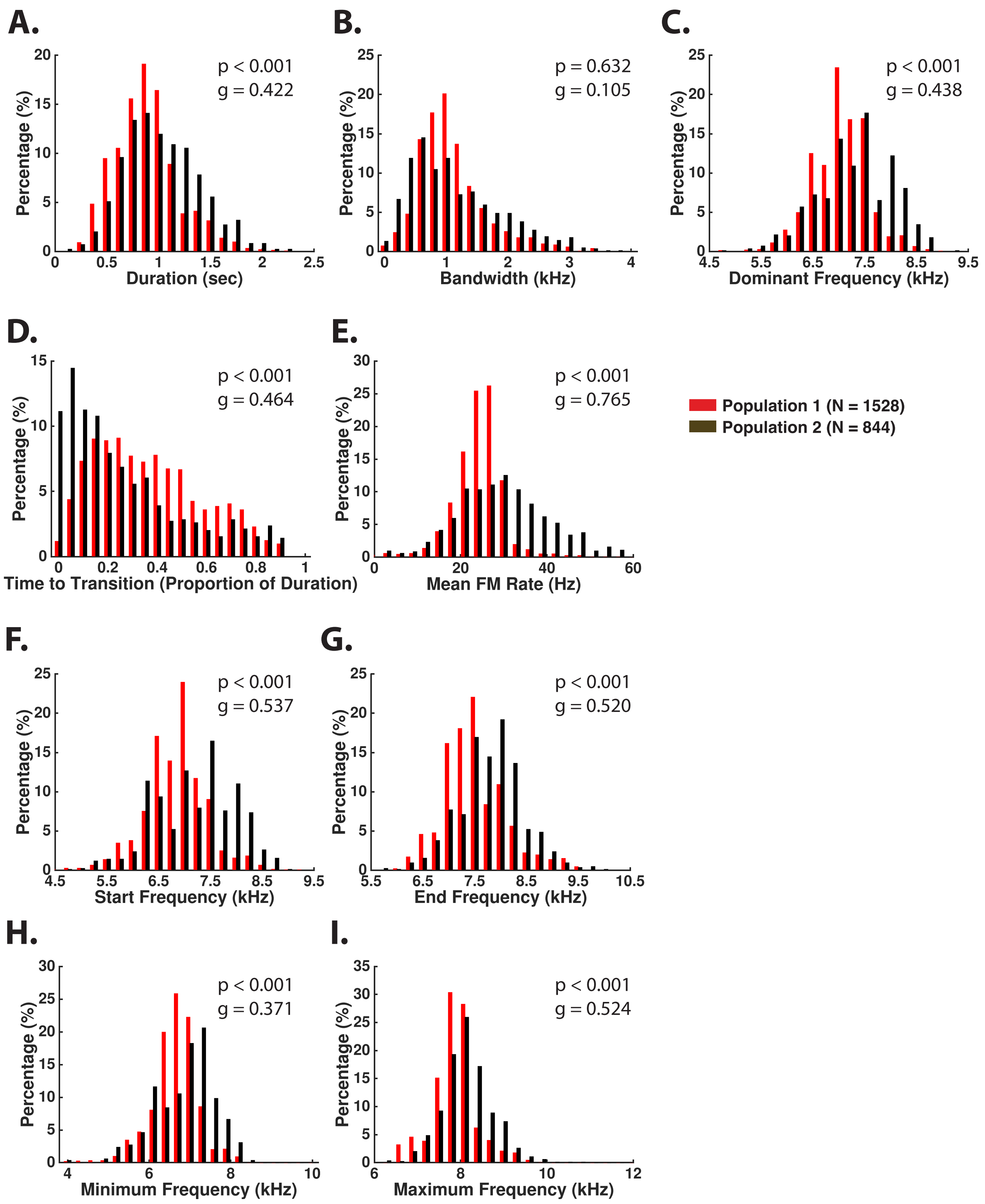

FIG. 10. (Color online) Examples of observed trillphee call features based on measurements made from both populations.

(6) Peep-String: Peep-strings are essentially concatenations of *p-peeps*, *t-peeps*, *sd-peeps*, *fd-peeps*, *sa-peeps*, and *dh-peeps* with intervening silent intervals less than 500 ms [Figs. 15(A) and 15(B)]. The composition of peep strings is highly variable with respect to the simple types that constitute the call, the number of syllables in the call, and the interval between components in the call. All such combinations of the simple peep types were classified as a single call type because of the inability to find any common structures within the general class that could be observed on a regular basis. Although numerous peep-strings are predominantly composed of either sd-peeps or

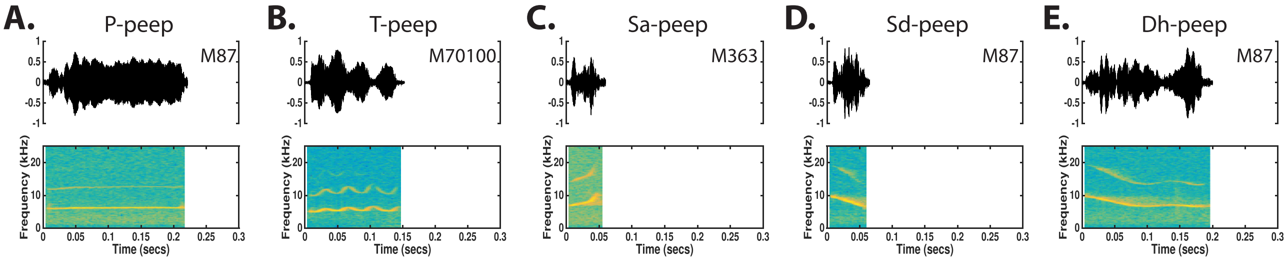

FIG. 11. (Color online) The marmoset utters a variety of short duration calls classified as "peeps" which were divided into five types. Time waveforms and spectrograms of the five observed simple peep types are shown. In general, the p-peep resembles a very short phee, the t-peep resembles a very short trill, the sa-peep is characterized by a steeply rising FM sweep, the sd-peep is characterized by a steeply falling FM sweep, and the dh-peep is a declining FM sweep that blends into an FM arch. Unlike other calls such as the phee and twitter, there is a high degree of variability in time-frequency characteristics within each of the peep types.

fd-peeps, the variation in the appearance of these peeps as well as the irregular inclusions of other p-peeps, t-peeps, and dh-peeps in these strings prevented them from being considered as separate call types.

(7) Trill-Peep: Trill-peeps are phrased compound calls that consist of a single trill followed by 1–8 *t-peeps*, *p-peeps*, or *sa-peeps* uttered with intervening silent intervals of less than 300 ms [Fig. 15(C)]. While trill-peeps with strings of peeps are observed, single peep trill-peeps were by far the most commonly encountered.

(8) Tsik-Egg*:* Tsik-eggs consist of a single tsik followed by 1–6 eggs separated by silent intervals less than 100 ms [Figs. 16(A) and 16(B)]. Occasionally, an ock will be substituted as the last syllable of a tsik-egg, but these occurrences were rare.

(9) Tsik-String: Tsik-strings are concatenations of tsiks, eggs, ocks, and/or tsik-eggs separated by silent intervals less than 500 ms in duration [Figs. 16(C) and 16(D)]. Tsik strings tend to be highly variable in the simple calls comprising each syllable and the interval between syllables. Tsik strings vary in duration from 500 ms to several minutes and are apparently uttered without acknowledging the vocalizations of other colony conspecifics or the expectation of a vocal response.

(10) Trill-Twitter: Trill-twitters consist of a *trill* or *t-peep* that either precedes a twitter with a short (less than 20 ms) silent interval or actually blends into the beginning phrase of the twitter syllable [Fig. 17(A)]. When the *trill* blends into the beginning phrase of the twitter, it tends to be more highly inclined in its time-frequency characteristic than ordinary trill calls. Other than this distinction, the trill segment of the call and the twitter segment of the call are characteristic of their simple call variants.

(11) Twitter-Peep*:* Twitter-peeps are twitter syllables followed by 1–2 sd-peeps or fd-peeps, with occasional occurrences of p-peeps [Figs. 17(B) and 17(C)]. Although the peep immediately following the twitter generally occurs with a silent interval equivalent to the phrasing interval of the twitter call, noticeably longer intervals have been observed. For twitter-peeps with two peeps, the interval between the first and second peep is highly irregular and will frequently vary from the phrasing interval of the twitter call.

(12) Trill-Twitter-Peep: Trill-twitter-peeps are trill-twitters followed by 1–2 sd-peeps or fd-peeps [Fig. 17(D)]. The relative spacing of the peeps in trill-twitter-peeps is the same as that described above for twitter-peeps. Occurrences of these call types were rare.

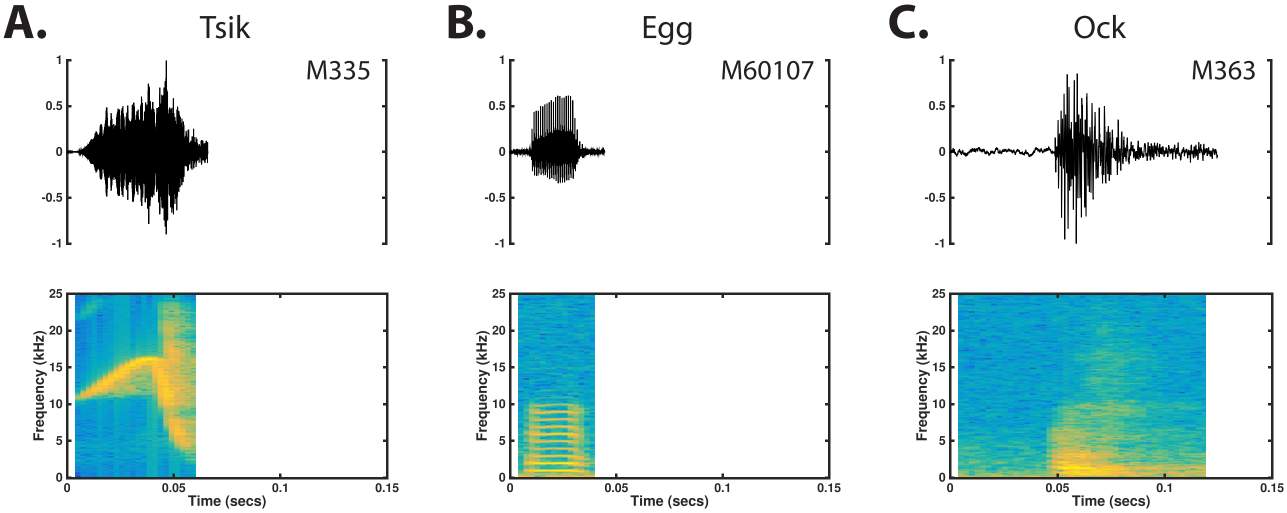

FIG. 12. (Color online) Tsiks (A), eggs (B), and ocks (C) are simple calls uttered in a mobbing response to a predator (Epple, 1968). Within our colony, these calls were primarily observed only when a human observer was in close proximity to the marmoset's cage.

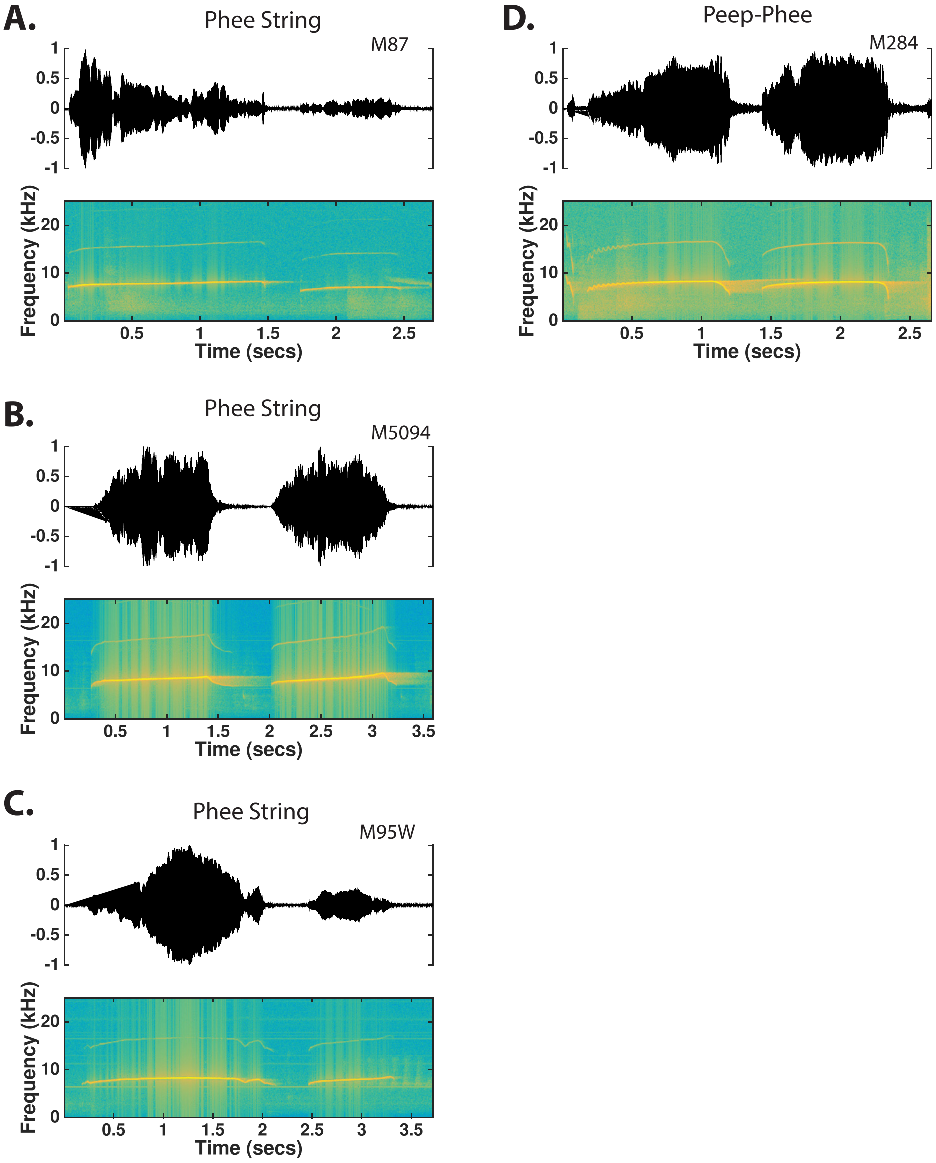

FIG. 13. (Color online) Marmosets frequently concatenate several phee calls in a compound call type as shown. Occasionally, the first phrase in a concatenation of phee calls is actually either a trill or a trillphee, as shown in (D). Phee strings and peep-phees are distinguishable based on whether or not the series of phee calls is preceded by one or more peep types.

(13) Twitter-Phee: Twitter phees are twitter syllables followed by a single phee or a phee-string [Figs. 17(E) and 17(F)]. As with the twitter-peeps, the first syllable in the phee-string segment generally starts after a silent interval equivalent to the phrasing interval of the twitter segment.

IV. DISCUSSION

A. Summary of findings

It is clear from the data presented in this study that the common marmoset produces a complex repertoire of calls in captivity. We have provided here a classification scheme for the marmoset's vocal repertoire consisting of 12 simple call types and 13 compound call types based on salient acoustic features and distinct patterns among marmoset vocalizations. To support the classification of vocalizations, a set of features designed to accurately capture acoustic structure were measured from a majority of the simple call types. These feature measures quantitatively provide the natural range of variation in the vocalizations of this primate species. Furthermore, examining vocalizations from two populations of animals at two widely separated time points provides a means of gauging which call features were conserved over time and which were more variable in our captive colony. These data form an essential basis for studying the marmoset's vocal production mechanisms and for properly synthesizing calls for use in behavioral and psychophysical studies of vocal perception and in electrophysiological experiments investigating their underlying neural representation in the brain. Comparison of measured features also provides a robust means of justifying the separation of similar vocalizations into distinct call types on the basis of their acoustic characteristics. To date, this study represents the most comprehensive investigation into the marmoset's vocal repertoire using quantitative approaches to differentiate call types. It is left to future studies to determine if these acoustically distinct calls are behaviorally distinct as well.

B. Comparison with previous studies

Previous work has examined the marmoset vocal repertoire in both captive (Epple, 1968; Rylands, 1993) and wild (Bezerra and Souto, 2008) environments, and our current

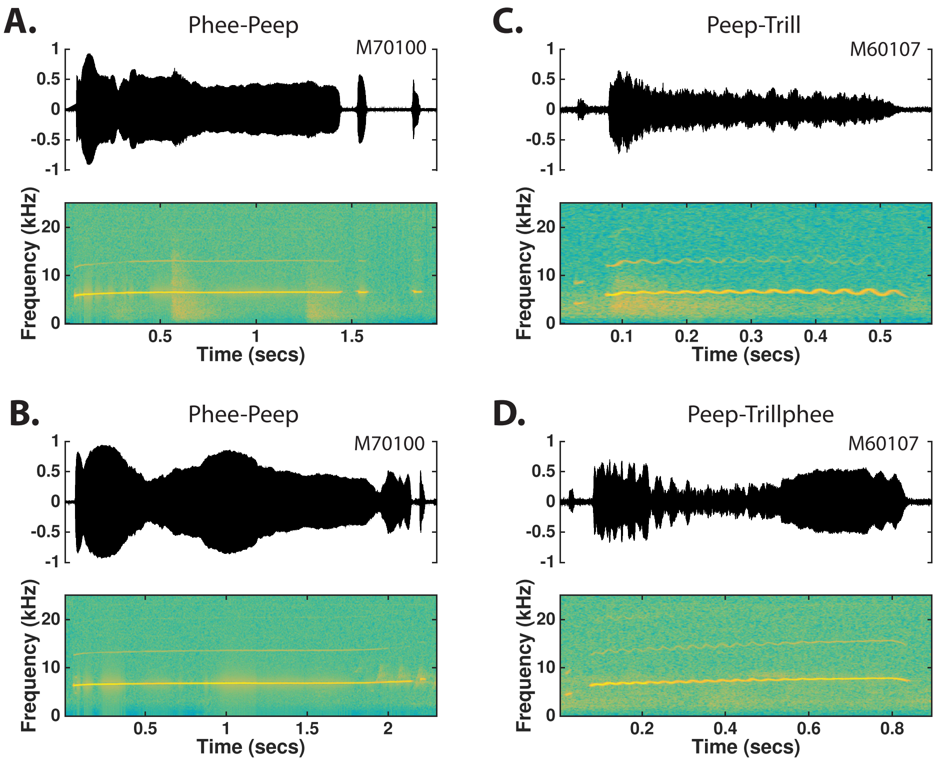

FIG. 14. (Color online) The marmoset utters several compound calls consisting of a single peep and a single trill, phee, or trillphee. In the phee-peep [(A) and (B)], the peep follows the longer call, whereas in the peep-trill and peep-trillphee [(C) and (D)] the peep precedes the longer call.

data agree well with these previous descriptions of adult vocalizations. These previous studies have variously described several of the major call types (e.g., twitter, phee, trill) in addition to other simple call types (e.g., tsik, egg) and compound calls (e.g., phee-strings). Furthermore, descriptions of the primary structural differences separating the call types (e.g., broadband, phrased twitters; narrowband phees; FM trills) and each call's basic acoustic features (e.g., frequency range, duration) are essentially identical throughout these various studies, suggesting that these call types are stable across marmoset populations.

Although there is considerable support for the idea that the vocal repertoire of many primate species is fixed and shows little change over time (Winter *et al.*, 1973; Symmes *et al.*, 1979; Butynski *et al.*, 1992), other work has shown that many callitrichid species exhibit evidence for some degree of vocal plasticity and social modification of their vocalizations. For example, pygmy marmosets showed convergence over

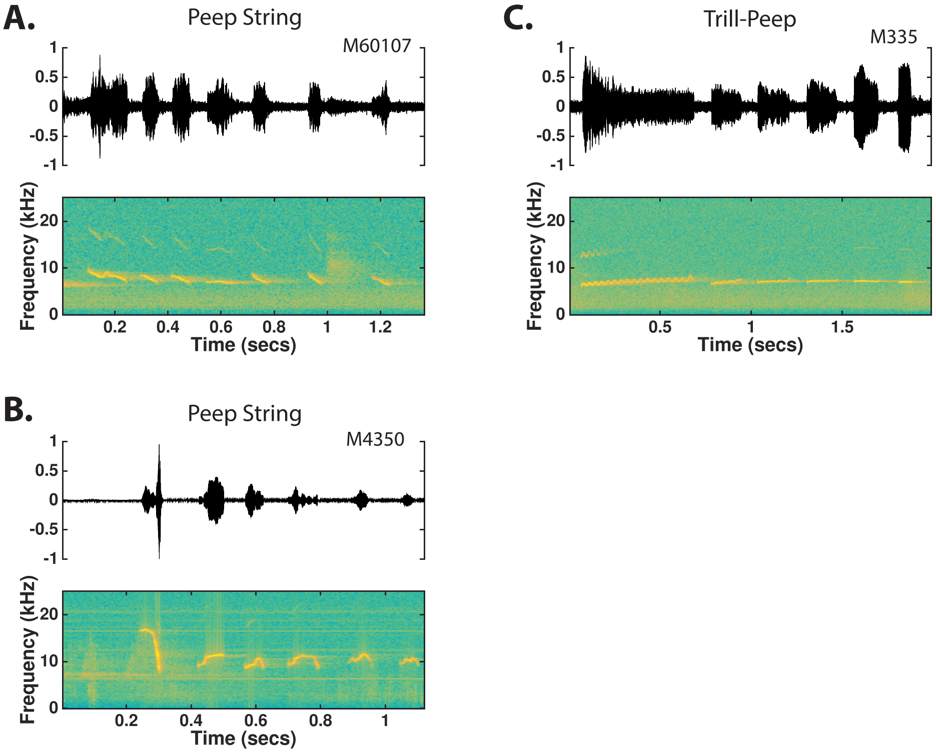

FIG. 15. (Color online) Sequences of various simple peep calls are often observed in compound call types. These sequences may compose the compound call entirely, as in the peep string [(A) and (B)], or they may be uttered in conjunction with another simple call type, as shown with the trill-peep (C).

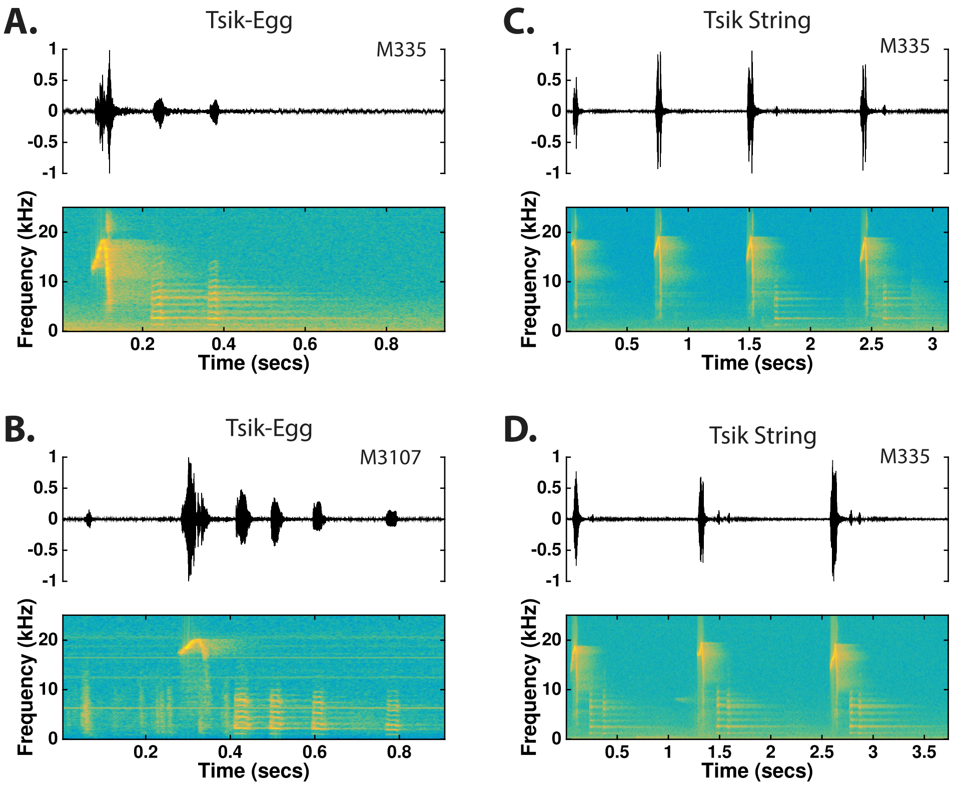

FIG. 16. (Color online) In a mobbing response to a predator, marmosets frequently combine tsik, egg, and ock simple call types into compound call types. The tsik strings are typically highly variable in the order of tsiks, eggs, and ocks that comprise them. However, the tsik-egg combination surfaced frequently enough that it was considered a separate call type.

time in various acoustic features of trill calls after two different populations were housed together (Elowson and Snowdon, 1994) and similar changes in acoustic structure were described in newly paired pygmy marmosets (Snowdon and Elowson, 1999). Furthermore, there is evidence that wild pygmy marmosets have regional dialects, with significant population differences measured in both *J*-calls and long calls (de la Torre and Snowdon, 2009). Other studies examining Wied's black tufted-eared marmosets have described changes in frequency and temporal parameters over time (Jorgensen and French, 1998), and that these kinds of changes were most pronounced for animals that were housed with new neighbors (Ruckstalis *et al.*, 2003). Similarly, a study by Norcross and Newman (1993) showed changes in the duration and the fundamental frequency parameters of common marmoset phee calls over a three year period, and Jones *et al.* (1993) showed significant variation in the average fundamental frequency of phee calls over a 12 month period. There is also evidence that marmoset vocalizations undergo both qualitative and quantitative ontogenetic changes (Pistorio *et al.*, 2006).

We observed in the present study differences in certain acoustic features of two major call types of the marmoset (twitter and trillphee) for two populations of marmosets sampled 18 yrs apart. There was little difference between acoustic features of the other two major types of calls (phee and trill). Specifically, we found that twitters were longer in duration (accompanied by an increase in the number of phrases), had lower starting frequencies, narrower bandwidths, and slower phrase sweep times in Population 2 compared to Population 1. Trillphees from Population 2 had higher FM rates, shorter sinusoidal-to-linear FM transition times, and higher fundamental frequencies compared to Population 1. Phees and trills, on the other hand, were highly similar across all measured acoustic features between the two populations. Thus, we have shown that certain acoustic features in particular, calls show greater variability between our two marmoset populations compared with other features.

There is some evidence that the acoustic structure of primate vocalizations can signal important information such as identity or emotional state (e.g., Bradbury and Vehrencamp, 1998; Owren and Rendell, 2001). For example, several acoustic features have been shown to elicit agonistic behaviors (including arousal and attention) in primates, including rapid, short duration pulses or FM sweeps, sounds with broadband or noisy spectra, and those with rapid AM fluctuations. Alternatively, tonal, harmonic, and continuous sounds typically elicit more affiliative behaviors and tend to convey more individual distinctiveness. Thus, it is possible that structural differences measured in a call type (as in twitters and trillphees) may signify different social environments (e.g., Newman *et al.*, 1983) whereas a highly conserved structure (as in phees and trills) could convey important individual identity information (Jones *et al.*, 1993; Miller *et al.*, 2010). Previous studies suggested changes in the acoustic structure of phee calls (e.g., Norcross and Newman, 1993), but the reported changes were typically in the range of several hundred Hz, which were well within the range of variation we describe here at the population level.

C. Implications for perception

Apart from determining an appropriate number of features, it is also important to measure features that are likely relevant for call perception. From studies on the perception of vowels in human speech, we know that clear classification boundaries based purely on acoustic considerations (first 2–3

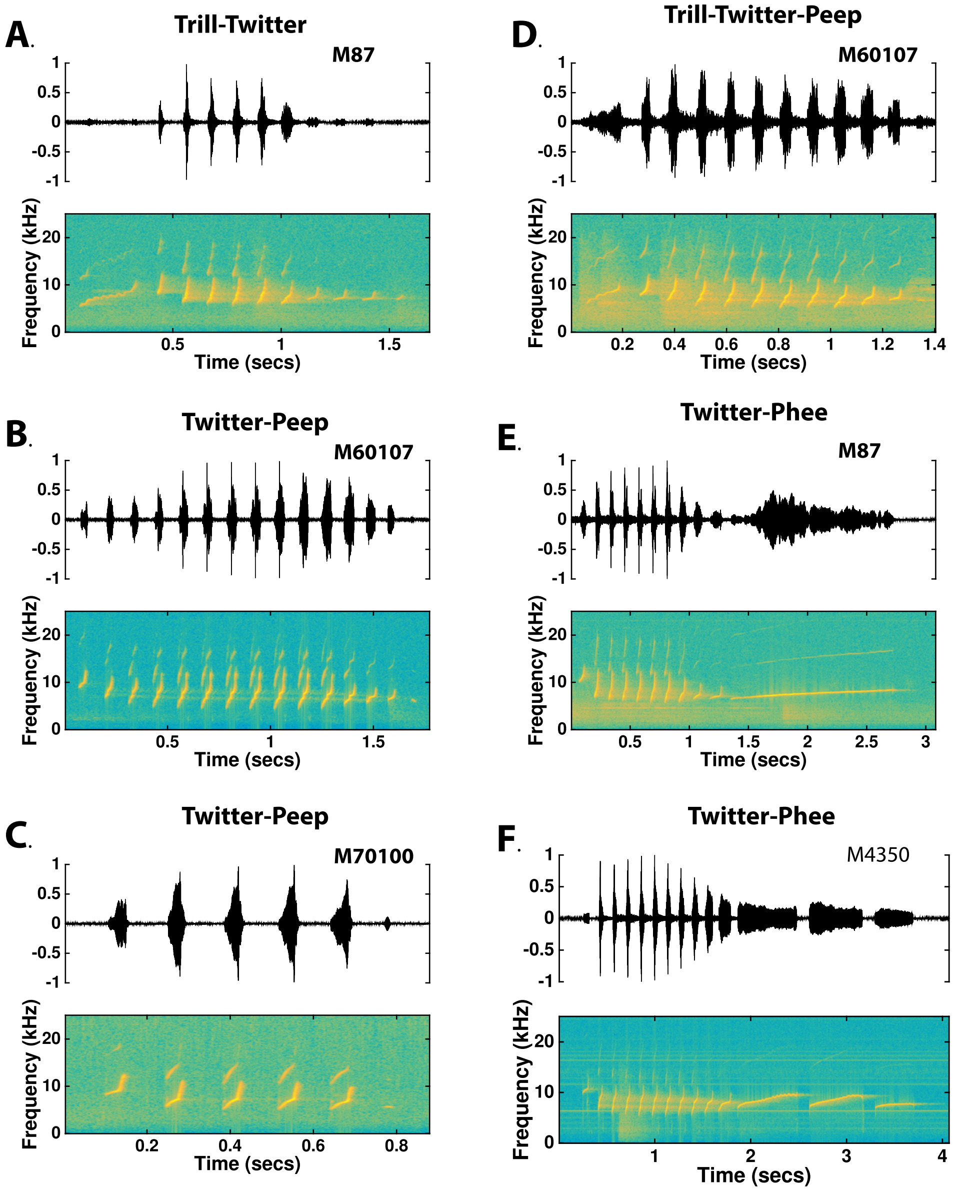

FIG. 17. (Color online) Although twitter calls are phrased, they are often observed as a distinct syllable in a compound call type. In compound call types involving twitter calls, the twitter may be preceded by a trill which often blends into the beginning twitter phrase [(A) and (D)], and/or followed by one or more peeps [(B), (C), and (D)], or followed by one or more phees [(E) and (F)].

formants of vowels) tend to be indicative of perceptual classification boundaries (Peterson and Barney, 1952). For marmoset vocalizations, differentiation based on selected acoustic features is obvious between some call types (e.g., between phee and twitter) and less obvious, but no less distinct, for others types (e.g., between phee, trill, and trill-phee). Because the call types we characterized are distinct and reproducible by marmosets, they are likely to be perceived as discrete entities by these animals, although direct proof still awaits behavioral testing.

D. Representation of complex vocalizations using a limited number of parameters

When representing a complex signal on a multidimensional space, one must be sure that a sufficient set of features is used to accurately represent the signal. In this study, up to approximately 18 distinct features were measured from each call, which we believe is a sufficiently large number. Using too few features could fail to capture the fine acoustic structure of a call type and prevent calls from being differentiated based on the measured features. Because discriminating call types is of primary interest to this study, it was critical for us to measure the features that potentially capture differences in acoustic structure among call types. This consideration was reflected in the selection of measured features.

In Sec. IV C we alluded to the fact that the clear boundaries between vocalizations in an acoustic feature space might be suggestive of perceptual distinctions that would be made by the marmoset. Verifying the correspondence between an acoustic and perceptual partitioning of call types requires behavioral analysis. The importance of our findings is that they provide a solid basis for future psychophysical studies to determine which acoustic features are perceptually significant to the marmoset in differentiating call types, and for future neurophysiological studies to reveal how vocalizations are represented by the brain. These psychophysical and neurophysiological studies would utilize synthesized calls, based on the currently established feature sets, instead of natural calls to eliminate the possibility of animals using subtle cues for differentiating test signals and to allow the manipulation of acoustic features within and beyond the statistical boundaries of particular call types. Although synthetic calls are generally spectrographically simplified versions of natural calls, it has been shown that carefully synthesized calls based on a quantitative understanding of call variation can evoke behavioral (Norcross *et al.*, 1994) or neural (DiMattina and Wang, 2006) responses similar to natural vocalizations.

E. Limitations of the present study

It must be pointed out that studies of captive animals are inherently limited in that some subsets of a species' natural vocal repertoire may not be present in captive animals due to a lack of certain social and behavioral conditions in a captive environment (e.g., more restricted movements, fewer opportunities for physical contact, the absence of natural prey and predators). However, it is also important to point out that housing animals in individual cages does not necessarily prohibit the production of a diverse range of normal social calls as long as these individually housed animals are placed within a socially interactive colony—which was the case in our study. For example, phees function as the marmoset isolation call (Jones *et al.*, 1993; Norcross and Newman, 1993; Norcross *et al.*, 1994), yet the fact that our recordings contained many other call types is a clear indication that these animals were in a socially diverse environment and that we elicited more than simply isolation calls.

Despite the limitations of a captive environment, studies of captive animals do serve well as a valuable and often irreplaceable complementary method to those conducted in the field when they are designed to answer appropriate questions. For example, in order to understand inherent acoustic variations and stability in primate vocalizations, a large number of vocalizations have to be recorded and analyzed with reference to their callers. Tasks like this are at present not feasible in field studies, but can be well accomplished in captive studies. An integration of both approaches shall give us the best chance to fully understand a primate species' complete capacity for vocal communication.

Although we now have an extensive, quantitative description of the marmoset vocal repertoire, we know little about what these different call types may mean to these animals or what information they may impart. There is considerable evidence that marmosets use phee calls as a contact call when separated by distance (Epple, 1968) while other studies suggest a role in territorial defense (Hubrecht, 1985). Tsik and egg calls appear to be aggressive vocalizations while ocks likely function as a mobbing call (Epple, 1968). However, the function of twitters, trills, trillphees, and the other simple call types uttered by this species is largely unknown, and we know little about the functional significance of combining simple calls into compound calls.

We have alluded to the fact that understanding the complete vocal repertoire requires a behavioral analysis to accompany the acoustic analysis. Likewise, the complete repertoire may only be observed if all behavioral conditions that elicit vocalizations are created during recording sessions. Nonetheless, a quantitative description of a species' vocal repertoire based on a large sample size such as the one provided by this study would facilitate the analysis of vocal communication behaviors. In our study, we wanted to maintain the identity of individual callers to insure that individual differences in vocal production did not cause us to fallaciously create distinct call types and to analyze vocalizations for possible vocal signatures. To achieve this, we recorded only from individually housed adults (within the colony), and thus calls not collected in this study include infant calls, "huddling" calls, calls uttered in aggressive encounters (e.g., the "chutter" observed when marmosets fight or when they are being handled by the veterinary staff), and calls uttered when in close proximity or in physical contact with one another (Epple, 1968). Further investigation is required to quantitatively describe these call types.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Health Grant No. R01-DC005808 to X.W. and Grant No. R03-DC013150 to M.S.O. We thank A. Pistorio, J. Estes, and N. Sotuyo for assistance with animal care. J.A.A. and C.-J.C. contributed equally to this work.

APPENDIX

1. TDOA procedure

There were four microphones distributed in three-dimensional space. Let the observations at microphone *i* be

$$u_i(k) = s(k - T_i) + n_i(k), \quad i = 1, 2, 3, 4,$$

where $s(k)$ is the source signal at the microphone, T is the time delay associated with the receiver, and $n(k)$ is noise—assumed to be a zero mean stationary Gaussian random process, which makes the noise covariance the same as the time delay covariance.

Relative time delay of arrivals between a target microphone and the other three microphones can be computed with the delay estimation error n as

$$d_{i,4} = T_i - T_4, \quad i = 1, 2, 3,$$

$$d_{i,4} = d^0_{i,4} + n_{i,4}, \quad i = 1, 2, 3.$$

Let the speaker source be located at an unknown location (x, y, z) and the microphones located at locations (x_i, y_i, z_i). The squared distance between the source and microphone i is computed as

$$\boldsymbol{r_i^2 = (x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2, \quad i = 1, 2, 3, 4.}$$

Distance difference between the target microphone and the other microphones can be further computed using the speed of sound propagation c

$$\boldsymbol{r_{i,4} = c \cdot d_{i,4} = r_i - r_4, \quad i = 1, 2, 3.}$$

The source location would be the intersection of these hyperbolic surfaces.

2. MLE procedure

In the presence of noise, the above equations will not meet at the same point. To find the best fit location, we applied a non-iterative realization of the MLE method (Chan and Ho, 1994) to solve these nonlinear equations. Let the proper answer be (x, y, z) and the distance to the target microphone be r

$$\boldsymbol{v} = [x, y, z, r]^T.$$

To estimate $\mathbf{v}$, we first assume that the positions and the distance are independent of each other, and then they can be solved by least-square (LS) error. The second step applies the known relationship between positions and distance with another LS error.

This two-step process is a MLE approximation, which can be written as

$$\boldsymbol{v} = (\boldsymbol{G}_a^T \psi^{-1} \boldsymbol{G}_a^T)^{-1} \boldsymbol{G}_a^T \psi^{-1} \boldsymbol{h},$$

$$\psi = \boldsymbol{h} - \boldsymbol{G}\boldsymbol{v}^0,$$

$$\boldsymbol{G} = -\begin{bmatrix} x_{1,4} & y_{1,4} & z_{1,4} & r_{1,4} \\ x_{2,4} & y_{2,4} & z_{2,4} & r_{2,4} \\ x_{3,4} & y_{3,4} & z_{3,4} & r_{3,4} \end{bmatrix},$$

$$\boldsymbol{h} = \frac{1}{2}\begin{bmatrix} r_{1,4}^2 - K_1 + K_4 \\ r_{2,4}^2 - K_2 + K_4 \\ r_{3,4}^2 - K_3 + K_4 \end{bmatrix},$$

$$K_i = x_i^2 + y_i^2 + z_i^2, \quad i = 1, 2, 3, 4.$$

By applying this algorithm, we can assign the call signal to the nearest channel so that each call signal only appears on one channel.

3. Definitions of signal representations for initial Population 1 analysis

a. Time waveform envelope

The time waveform envelope (hereafter envelope) was approximated by low-pass filtering the absolute value of a vocalization [Fig. 2(A)]. For all simple calls, the envelope was used to remove the intervals of silence preceding and following the vocalization. For twitter calls, the envelope was also used to isolate individual phrases based on the location of troughs and to measure the time interval between phrases [Fig. 2(A)]. For all vocalizations the low-pass filter was a sixth-order zero-phase Butterworth filter. For twitter calls, a cutoff frequency of 15 Hz was used to ensure that each phrase had a single peak, which is a critical criterion for accurately isolating phrases and measuring the IPI. For all other call types a cutoff frequency of 45 Hz was used because it was more critical in these vocalizations to accurately capture the sharp amplitude transitions marking the beginning and end of the call.

b. Frequency spectrum

The frequency spectrum (hereafter spectrum) was calculated for each vocalization using a modulo 2 FFT [Fig. 2(B)]. A Hanning window was applied to the zero-padded signal before calculating the FFT. The magnitude spectrum was derived from the absolute value of the complex spectrum and smoothed using a low-pass filter (sixth-order zero-phase Butterworth) designed to ensure the spectrum would have only one clearly defined peak without causing unnecessary broadening of the spectral mode. To meet these requirements, a separate filter was designed for each call type which resulted in cutoff frequencies ranging from 75 to 1000 Hz. In each spectrum, the frequency of the peak position and the spectrum bandwidth are measured [i.e., f_{dom}, f_{BW}, Fig. 2(B)].

c. Spectrogram

Spectrograms were used for making both section and whole call measurements [Figs. 2(C) and 2(D)]. Spectrograms were calculated using Hanning windows and 50% overlap. The length of the window used depended on the call type and was chosen to maximize the resolution in time and frequency for analyzing each vocalization. A 5.1 ms window (256 point FFT) was used for short duration calls (i.e., tsiks, p-peeps, t-peeps, sd-peeps, fd-peeps, sa-peeps, and dh-peeps). Long duration calls with rapid time-frequency transients (i.e., trills, trillphees, and twitters) were processed with a 10.2 ms window (512 point FFT). Phee calls, characterized by their long duration and slow time-frequency transients, were analyzed using 40.8 ms windows (2048 point FFT). All measurements were actually made using the magnitude trace of the spectrogram. The traces provided a reliable representation of time-frequency characteristics from which measurements could easily be made.

4. SVM procedure

Separation between two call types was achieved by a hyperplane that had the largest distance to the nearest training data point (margin). While a one-against-all strategy is widely used in these kinds of classification algorithms, a pairwise one-against-one strategy has been shown to be more stable (Wu *et al.*, 2004). We thus applied a C-SVM method (Vapnik, 1998; Chang and Lin, 2011) with a pairwise one-against-one strategy.

Given training feature vectors $\mathbf{x}$ with length m and a label vector $\mathbf{y}$ such that y_i belongs to one of the labeled call types, C-SVM solves the optimization problem below:

$$\min_{\omega,b,\xi} \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + C \cdot \sum_{i=1}^{m} \xi_i,$$

$$\text{subject to } y_i(\boldsymbol{\omega}^T + \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi \geq 0,$$

$$i = 1, 2, \ldots, m,$$

where $f(x)$ maps x into a high-dimensional space and C is the hyper-parameter. Due to the high dimensionality of the vector **w**, we solved the following problem with LIBSVM tool (Chang and Lin, 2011).

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T Q\boldsymbol{\alpha} - \boldsymbol{v}^T\boldsymbol{\alpha},$$

$$Q_{i,j} = y_i y_j \cdot K(x_i, x_j) = y_i y_j \cdot \phi(x_i)^T \phi(x_i)$$
$$= y_i y_j \cdot \mathrm{e}^{-(\gamma\cdot|u-v|^2)},$$

$$\text{subject to } \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \quad 0 \le \alpha_i \le C, \quad i = 1, 2, ..., m,$$

where $\mathbf{v} = [1,\ldots, 1]T$ is the vector of all ones, Q is a positive semi-definite matrix, and K is the radial basis kernel function.

After the above equation was solved, the optimal $\boldsymbol{\omega}$ satisfied

$$\boldsymbol{\omega} = \sum_{i=1}^{m} y_i\alpha_i\phi(x_i).$$

Belcher, A. M., Yen, C. C., Stepp, H., Gu, H., Lu, H., Yang, Y., Silva, A. C., and Stein, E. A. (**2013**). "Large-scale brain networks in the awake, truly resting marmoset monkey," J. Neurosci. **33**, 16796–16804.

Bezerra, B. M., and Souto, A. (**2008**). "Structure and usage of the vocal repertoire of Calltihrix jacchus," Int. J. Primatol. **29**, 671–701.

Bezerra, B. M., Souto, A. S., and Schiel, N. (**2007**). "Infanticide and cannibalism in a free-ranging plurally breeding group of common marmosets (*Callithrix jacchus*)," Am. J. Primatol. **69**, 945–952.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA), 882 pp.

Butynski, T. M., Chapman, C. A., Chapman, L. J., and Weary, D. M. (**1992**). "Use of male blue monkey 'pyow' calls for long-term individual identification," Am. J. Primatol. **28**, 183–189.

Chan, Y. T., and Ho, K. C. (**1994**). "A simple and efficient estimator for hyperbolic location," IEEE Trans. Signal Process. **42**, 1905–1915.

Chang, C. C., and Lin, C. J. (**2011**). "LIBSVM: A library for support vector machines," ACM Trans. Intell. System Tech. **2**(3), 1–27.

de la Torre, S., and Snowdon, C. T. (**2009**). "Dialects in pygmy marmosets? Population variation in call structure," Am. J. Primatol. **71**, 333–342.

Digby, L. (**1995**). "Infant care, infanticide, and female reproductive strategies in polygynous groups of common marmosets (*Callithrix jacchus*)," Behav. Ecol. Sociobiol. **37**, 51–61.

DiMattina, C., and Wang, X. (**2006**). "Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations," J. Neurophysiol. **95**, 1244–1262.

Elowson, A. M., and Snowdon, C. T. (**1994**). "Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment," Anim. Behav. **47**, 1267–1277.

Epple, G. (**1968**). "Comparative studies on vocalization in marmoset monkeys (hapalidae)," Folia Primatol. **8**, 1–40.

Fuller, J. L. (**2014**). "The vocal repertoire of adult male blue monkeys (*Cercopithecus mitis stulmanni*): A quantitative analysis of acoustic structure," Am. J. Primatol. **76**, 203–216.

Gamba, M., and Giacoma, C. (**2007**). "Quantitative acoustic analysis of the vocal repertoire of the crowned lemur," Ethol. Ecol. Evol. **19**(4), 323–343.

Giret, N., Roy, P., Albert, A., Pachet, F., Kreutzer, M., and Bovet, D. (**2011**). "Finding good acoustic features for parrot vocalizations: The feature generation approach," J. Acoust. Soc. Am. **129**, 1089–1099.

Hedges, L. V., and Olkin, I. (**1985**). *Statistical Methods for Meta-Analysis* (Academic Press, Orlando, FL), 369 pp.

Hedwig, D., Robbins, M. M., Mundry, R., Hammerschmidt, K., and Boesch, C. (**2014**). "Acoustic structure and variation in mountain and western gorilla close calls: A syntactic approach," Behaviour **151**, 1091–1120.

Hubrecht, R. C. (**1985**). "Home-range size and territorial behavior in the common marmoset *Callithrix jacchus jacchus*, at the Tapacura Field Station, Recife, Brazil," Int. J. Primatol. **6**, 533–550.

Jones, B. S., Harris, D. H. R., and Catchpole, C. K. (**1993**). "The stability of vocal signature in phee calls of the common marmoset, *Callithrix jacchus*," Am. J. Primatol. **31**, 67–75.

Jorgensen, D. D., and French, J. A. (**1998**). "Individuality but not stability in marmoset long calls," Ethol. **104**, 729–742.

Kajikawa, Y., de La Mothe, L., Blumell, S., and Hackett, T. A. (**2005**). "A comparison of neuron response properties in areas A1and CM of the marmoset monkey auditory cortex: Tones and broadband noise," J. Neurophysiol. **93**, 22–34.

Kobayasi, K. I., and Riquimaroux, H. (**2012**). "Classification of vocalizations in the Mongolian gerbil, *Meriones unguiculatus*," J. Acoust. Soc. Am. **131**, 1622–1631.

Liu, J. V., Hirano, Y., Nascimento, G. C., Stefanovic, B., Leopold, D. A., and Silva, A. C. (**2013**). "fMRI in the awake marmoset: Somatosensory-evoked responses, functional connectivity, and comparison with propofol anesthesia," Neuroimage **78**, 186–195.

Lu, T., Liang, L., and Wang, X. (**2001**). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates," Nat. Neurosci. **4**, 1131–1138.

Miller, C. T., Mandel, K., and Wang, X. (**2010**). "The communicative content of the common marmoset phee call during antiphonal calling," Am. J. Primatol. **72**, 974–980.

Mitchell, J. F., Reynolds, J. H., and Miller, C. T. (**2014**). "Active vision in marmosets: A model system for visual neuroscience," J. Neurosci. **34**, 1183–1194.

Moody, D. B., Stebbins, W. C., and May B. J. (**1990**). "Auditory perception of communication signals by Japanese monkeys," in *Comparative Perception—Volume II: Complex Signals*, edited by W. C. Stebbins and M. A. Berkley (John Wiley & Sons, Inc., New York), pp. 311–343.

Newman, J. D., Smith, H. J., and Talmagge-Riggs, G. (**1983**). "Structural variability in primate vocalizations and its functional significance: An analysis of squirrel monkey chuck calls," Folia Primatol. **40**, 114–124.

Norcross, J. L., and Newman, J. D. (**1993**). "Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls," Am. J. Primatol. **30**, 37–54.

Norcross, J. L., Newman, J. D., and Fitch, W. (**1994**). "Responses to natural and synthetic phee calls by common marmosets (*Callithrix jacchus*)," Am. J. Primatol. **33**, 15–29.

Owren, M. J., and Rendall, D. (**2001**). "Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling," Evol. Anthropol. **10**, 58–71 (2001).

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–194.

Pettitt, B. A., Bourne, G. R., and Bee, M. A. (**2012**). "Quantitative acoustic analysis of the vocal repertoire of the golden rocket frog (*Anomaloglossus beebei*)," J. Acoust. Soc. Am. **131**, 4811–4820 (2012).

Pistorio, A. L., Vintch, B., and Wang, X. (**2006**). "Acoustical analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*)," J. Acoust. Soc. Am. **120**, 1655–1670.

Rosa, M. G. P., and Tweedale, R. (**2000**). "Visual areas in lateral and ventral extrastriate cortices of the marmoset monkey," J. Comp. Neurol. **422**, 621–651.

Ruckstalis, M., Fite, J. E., and French, J. A. (**2003**). "Social change affects vocal structure in a Callitrichid primate (*Callithrix kuhlii*)," Ethol. **109**, 327–340.

Rylands, A. B. (ed.). (**1993**). *Marmosets and Tamarins: Systematics, Behavior, and Ecology* (Oxford University Press, Oxford), 396 pp.

Sasaki, E., Suemizu, H., Shimada, A., Hanazawa, K., Oiwa, R., Kamioka, M., Tomioka, I., Sotomaru, Y., Hirakawa, R., Eto, T., Shiozawa, S., Maeda, T., Ito, M., Ito, R., Kito, C., Yagahashi, C., Kawai, K., Miyoshi, H., Tanioka, Y., and Tamaoki, N. (**2009**). "Generation of transgenic non-human primates with germline transmission," Nature **459**, 523–527.

Seyfarth, R. M., and Cheney, D. L. (**2003**). "Signalers and receivers in animal communication," Ann. Rev. Psychol. **54**, 145–173.

Smith, T. (**2006**). "Individual olfactory signatures in common marmosets (*Callithrix jacchus*)," Am. J. Primatol. **68**, 585–604.

Snowdon, C. T., and Elowson, A. M. (**1999**). "Pygmy marmosets modify call structure when paired," Ethol. **105**, 893–908.

Soltis, J., Alligood, C. A., Blowers, T. E., and Savage, A. (**2012**). "The vocal repertoire of the key largo woodrat (*Neotoma floridana smalli*)," J. Acoust. Soc. Am. **132**, 3550–3558.

Symmes, D., Newman, J. D., Talmage-Riggs, G., and Lieblich, A. K. (**1979**). "Individuality and stability of isolation peeps in squirrel monkeys," Anim. Behav. **27**, 1142–1152.

Vapnik, V. (**1998**). *Statistical Learning Theory* (Wiley, New York, NY), 768 pp.

Wang, X., Lu, T., Snider, R. K., and Liang, L. (**2005**). "Sustained firing in auditory cortex evoked by preferred stimuli," Nature **435**, 341–346.

Winter, P., Handley, P., and Ploog, D. (**1973**). "Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation," Behaviour **47**, 230–239.

Wu, T. F., Lin, C. J., and Weng, R. C. (**2004**). "Probability estimates for multi-class classification by pairwise coupling," J. Mach. Learn. **5**, 975–1005.