Appendix A

Calls coded in the current study with their: behavioral category; description for visual identification in live coding; parameters for audio and spectrogram coding and the definition of a single call.

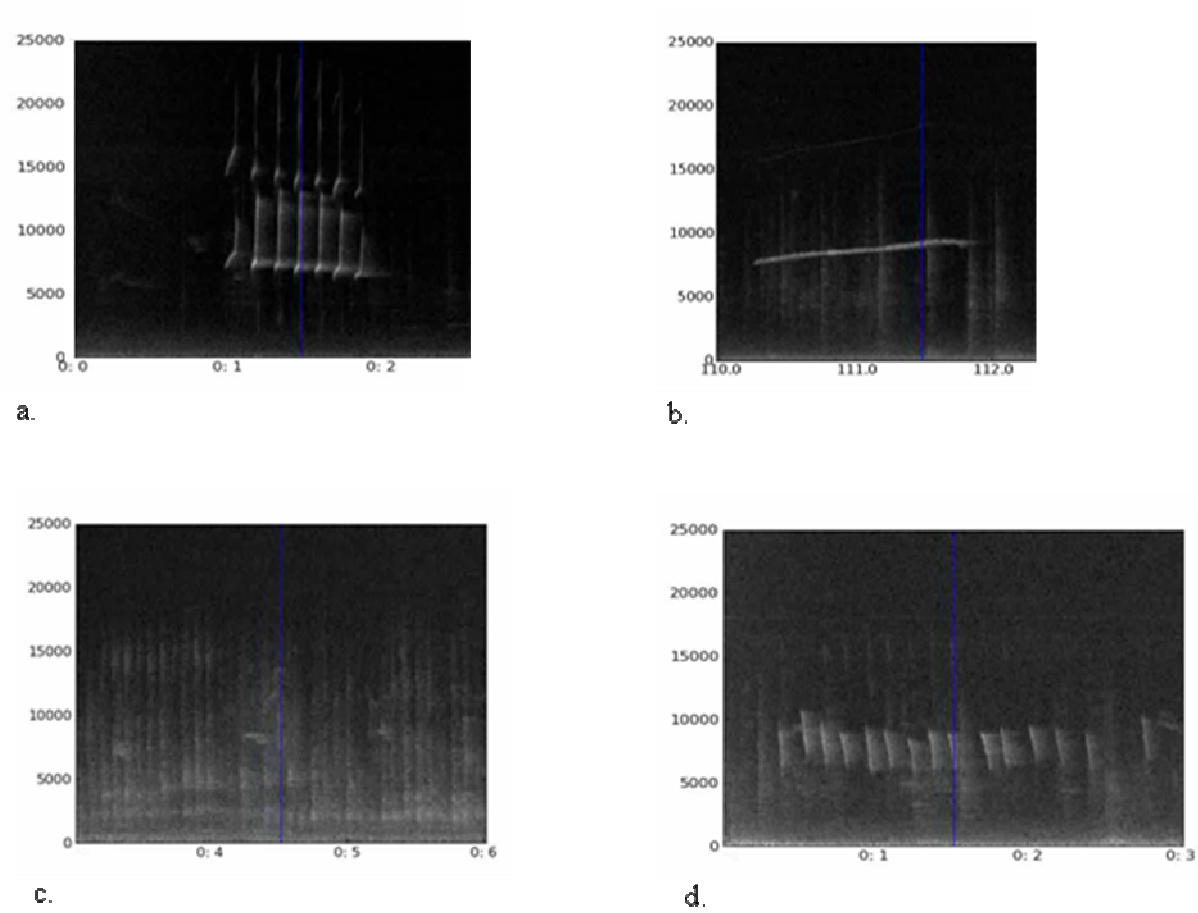
Behavioral category	**Vocalization type**	**Live coding visual/ audio description**	**Spectrographic description**	**Definition of a single call***
Intergroup agonistic	twitter	Mouth continuously open throughout, with lips slightly puckered and slight abdominal contractions visible.	A series of regularly spaced seep-like elements (but at lower frequency) characterized by a rapidly increasing fundamental frequency.	Sequence of elements: a phrase consists of several regularly spaced syllables. At least two elements present. Minimum inter-bout interval: 0.5 seconds (s). Overlapping calls coded separately.
	loud shrill	Wide open mouth. Very loud, piercing whistle-like call (loudest element in vocal repertoire).	Fundamental frequency tends to increase smoothly across the call. Distinguish from whirr due to absence of cyclic frequency fluctuations. Distinguish from long phee calls by the high amplitude and longer duration of elements. Only calls with first/only syllable 1.3 s in length or over were counted as loud shrill (to exclude non-open mouth calls).	Single syllable or sequence of syllables. Classified as 1, 2 or 3 and more syllables. It was fairly easy to determine which elements belonged in the same call (since the syllables are made at the same frequency). Where overlap occurred best judgment was used.
Intragroup agonistic	chatter	Body vibrates with abdominal contractions.	Series of low-pitched harsh elements similar to ek but distinguish since the elements are of shorter duration. Distinguish from cough due to the lack of 'noise' character.	Sequence of elements. Minimum inter-bout interval: 0.5 s (observed inter-element interval: about 0.05 s).
Affiliative	chirp	Made with mouth closed or slightly open.	Series of elements uttered in rapid series characterized by constant descent in frequency over a range of 8-5 kHz.	Sequence of elements: at least two in series. Minimum inter-bout interval of 0.5 s (observed inter- element interval is around 0.15 s).

* Vocalizations can be divided into individual elements and many call types are made up of sequences of these individual elements in series. The inter-bout interval is defined for each call type.

(Call descriptions were taken mainly from Pook [1976] since he divided calls into open and non-open mouth variants but also from: [Epple, 1968; Stevenson and Rylands, 1988; Jones, 1993; Goldman, 2000].)

Appendix B

Spectrograms of the four types of *Callithrix jacchus* vocalizations coded in the current study: a. twitter; b. loud shrill; c. chatter; d. chirp (x-axis shows frequency in kHz and y-axis shows time in seconds).



Appendix C

Common marmoset (*Callithrix jacchus*) call names in previous studies of vocalizations equivalent (or roughly equivalent) to the vocalization types coded in the present study (call definitions for the current study given in Appendix B).

Vocalization name (used here)	Equivalent call names (given in other vocalization studies)
twitter	'twitter' [Epple, 1968; Pook, 1976; Stevenson & Rylands, 1988; Jones 1993; Goldman, 2000]
loud shrill	roughly equivalent to 'monosyllabic calls given in isolation' [Epple, 1968] (but are non-isolation calls); equivalent to the wide open mouth calls/ top end of 'loud shrill' [Pook, 1976]; equivalent to top end of 'open-mouthed phee calls' [Goldman, 2000]; equivalent to top end of 'home phee calls' [Norcross and Newman, 1993]
chatter	'chatters given when angry' [Epple, 1968]; 'cackle' [Pook, 1976]; 'cackle' ('erh erh') [Stevenson & Rylands, 1988]
chirp	'chirping' [Goldman, 2000]; 'chirruping' [Pook, 1976]; 'rhythmical contact calls given in close visual and bodily contact' [Epple, 1968]

Appendix D

Additional considerations for audio and spectrogram coding.

Issue	Explanation and description of guidelines used in the current study
Loud shrill calls	The difference between the categories of long phee and loud shrill created by Pook [1976] is one of degree only given that the division was drawn arbitrarily according to whether or not the phee call was made with mouth closed or open, for the purposes of his experimental study. He includes all calls made with the mouth open, whether partially or wide open (loud shrills at the upper end of the spectrum represent the wide open mouth calls). However, many other studies have investigated solely open-mouth phee calls [e.g. Norcross & Newman, 1993]. In the current study it was possible to distinguish between focal individual calls made with mouth closed or open during live observation but it was not possible to make the same distinction for neighboring calls from the element structure in the spectrogram alone, in audio coding, therefore a criterion index was required. One distinguishing feature of loud shrill calls is their relatively high amplitude; however precise measurement of amplitude requires controlled recording conditions [Jones, 1993]. Loud shrill call elements are also, on average, much longer in duration than phee calls made with a closed mouth [e.g. Pook, 1976]. In order to ensure that all the calls coded as neighboring loud shrill calls did in fact represent open mouth calls, and therefore to ensure that all such calls made by the focal individual would be reliably detected, the minimum element length was set fairly high; at 1.3s (at just above the mean for loud shrill and significantly above the mean for long phee calls, 0.89s). All phee calls with at least one syllable of 1.3 seconds or more duration were coded as loud shrill calls. Focal individual loud shrill calls detected in the live observation (those under 1.3 s in length) were thus discarded from the analysis. Thus the total number of loud shrill calls coded was reduced but with the benefit that the loud shrill was reliably and consistently coded both in the live and in the audio and spectrogram coding.
Infant vocalizations	Infant marmosets make certain infant-specific calls as well as calls approximating to those of adults. Infant specific calls include: cry calls; compound cry calls; and twitter hook. The cry is a long duration call covering a broad frequency range [Pistorio et al., 2006]. The cry is equivalent to ‘infant squeal’ [Epple, 1968] and ‘caw’ [Pook, 1976]. Compound cry calls [calls joined to cry calls: Pistorio et al., 2006] and twitter calls with descents (‘twitter hook’) are infant-specific calls. Infant specific calls were not coded in the audio coding in the current study. Infants also make highly variable call types and these were not coded since they did not clearly fit any one particular adult call description. Because infants marmosets make calls approximating those of adults and call frequently but not in usual adult call contexts [Pistorio et al., 2006] it is important to try to exclude such calls from the audio and spectrogram coding since they unlikely to have an equivalent effect on focal adults as would similar calls made by mature adults. Pistorio et al. [2006] carried out a comprehensive analysis of the ontogeny of calls in infants. They showed that the parameters of the calls changed from 5 – 25 weeks (becoming more and more adult-like in particular parameters) however, the process is gradual rather than absolute so it is difficult to absolutely exclude these calls. At

	3-4 months, marmosets begin to make calls in appropriate contexts and respond appropriately to adult calls [Pook, 1976]. It was not possible to distinguish sub-adult calls made in isolation of cry calls. It is estimated that these calls will not greatly affect the results of the analysis. Infants are observed to make long strings of tsik calls, cry calls and twitter, phee and whirr calls to signal distress, for example when they are put down from an adult's back [Epple, 1968; Pook, 1976]. Only groups of adults with young infants respond to playback of these calls [Epple, 1968]. The infants also make other adult-like calls but much less frequently (e.g. ek). Pistorio et al. [2006] found that infant cry and compound cry calls disappeared entirely by 10-11 weeks in all individuals in their study. There were infants (aged less than 11 weeks/ 2.5 months) present in each of the four study colony rooms throughout the entire duration of the study and therefore the issue of identifying and excluding infant calls applied across all colony rooms. The duration of bouts of infant calls, as identified by cry calls and tsiks in sequence from cry calls were coded in the audio analysis. The calls known to be most frequently uttered by infants were coded as possibly infant calls when they occurred within 2 seconds either side of known infant calls. Twitter calls within infant strings (and 2 seconds either side) and those that appeared to be a continuation of twitter calls were coded as possibly infant. Twitter calls clearly overlapping infant calls when there was clearly only one infant vocalising were coded as adult calls. Infants make a large number of such calls so they were considered likely to impact on the analysis of adult calls if they had been coded as such. All analyses were carried out both including and excluding the calls coded as possibly infant to check that they did not impact on the result.
Transitional calls	Transitional calls are those that constitute a combination of two different call types e.g. twitter-phee [Jones, 1993] and thus do not fit the descriptive criteria for a particular individual call. Overall, transitional calls are rare in adults [e.g. Jones, 1993]. Transitional calls were not coded in the audio coding in the current study. Any focal individual calls coded during the live observation that were found to be transitional on examination of the spectrogram were discarded from the analysis.
Ultrasonic calls	The high-frequency cut-off for the hearing ability of the common marmoset is around 28 kHz [Coleman, 2009]. Pook [1976] stated that he detected no calls of a purely ultrasonic nature (that is, calls with a fundamental frequency starting above the audible range for humans). The equipment for the current study was carefully chosen in order to record frequencies of up to 27 kHz. Initial pilot studies showed the only call type that the marmosets occasionally made in the ultrasonic range (relative to human listeners) was the seep call [starting at around 20 kHz; concurrent with Stevenson & Rylands, 1988]. In the current study, only seep calls audible to the human ear were coded (the practical upper hearing ability threshold for the observer in the current study was determined to be 17 kHz).

Additional references for supplementary information

Coleman, MN. 2009. What do primates hear? A meta-analysis of all known nonhuman primate behavioral audiograms. Int J Primatol 30:55-91.

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

Pistorio AL, Vintch B, Wang XQ. 2006. Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). J Acoust Soc Am 120:1655-1670