frontiers in Neuroscience

ORIGINAL RESEARCH
published: 29 October 2021
doi: 10.3389/fnins.2021.728686

Formant-Based Recognition of Words and Other Naturalistic Sounds in Rhesus Monkeys

*Jonathan Melchor[1], José Vergara[2], Tonatiuh Figueroa[1], Isaac Morán[1] and Luis Lemus[1]**

[1] Department of Cognitive Neuroscience, Institute of Cell Physiology, Universidad Nacional Autónoma de México, Mexico City, Mexico, [2] Department of Neuroscience, Baylor College of Medicine, Houston, TX, United States

In social animals, identifying sounds is critical for communication. In humans, the acoustic parameters involved in speech recognition, such as the formant frequencies derived from the resonance of the supralaryngeal vocal tract, have been well documented. However, how formants contribute to recognizing learned sounds in non-human primates remains unclear. To determine this, we trained two rhesus monkeys to discriminate target and non-target sounds presented in sequences of 1–3 sounds. After training, we performed three experiments: (1) We tested the monkeys' accuracy and reaction times during the discrimination of various acoustic categories; (2) their ability to discriminate morphing sounds; and (3) their ability to identify sounds consisting of formant 1 (F1), formant 2 (F2), or F1 and F2 (F1F2) pass filters. Our results indicate that macaques can learn diverse sounds and discriminate from morphs and formants F1 and F2, suggesting that information from few acoustic parameters suffice for recognizing complex sounds. We anticipate that future neurophysiological experiments in this paradigm may help elucidate how formants contribute to the recognition of sounds.

Keywords: Psychophysics, long-term memory, formants, auditory discrimination, non-human primate (NHP)

OPEN ACCESS

Edited by:
Monita Chatterjee, Boys Town, United States

Reviewed by:
Iain DeWitt, Infoscitex Inc., United States
Lan Shuai, Haskins Laboratories, United States

***Correspondence:**
Luis Lemus
lemus@ifc.unam.mx

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *21 June 2021*
Accepted: *08 October 2021*
Published: *29 October 2021*

Citation:
Melchor J, Vergara J, Figueroa T, Morán I and Lemus L (2021) Formant-Based Recognition of Words and Other Naturalistic Sounds in Rhesus Monkeys. Front. Neurosci. 15:728686. doi: 10.3389/fnins.2021.728686

INTRODUCTION

Non-human primates (NHP) identify conspecific vocalizations (Rendall et al., 1996; Jovanovic et al., 2000; Ceugniet and Izumi, 2004; Belin, 2006) that inform troop members about food quality (Hauser, 1998; Slocombe and Zuberbühler, 2006) or nearby predators (Seyfarth et al., 1980b). These communication abilities are likely to rely on the activity of vocal recognition brain areas, homologous in humans and macaques (Petkov et al., 2008; Leaver and Rauschecker, 2010; Ortiz-Rios et al., 2015; Belin et al., 2018). However, how different acoustic parameters contribute to the recognition of sounds in NHP is not fully understood.

The literature points to periodicity (i.e., the fundamental and harmonic frequencies at which the vocal folds vibrate during phonation) and temporal envelope as possible cues for vocal recognition (Stevens, 1983; Chandrasekaran et al., 2011; Mesgarani et al., 2014; Brewer and Barton, 2016). Also important to recognition are the prominences in the spectral envelope, formant frequencies, that vary with changes in the shape of the supralaryngeal tract (e.g., jaw height and tongue protrusion)

Abbreviations: NHP, non-human primates; T, Target; NT, non-target; F1, first formant; F2, second formant; CR, correct rejections; FA, false alarms; GC, go-cue; RT, reaction time; PF, psychometric function; PSE, point of subjective equality; JND, just noticeable difference.

and the length of the individuals' vocal tract (Remez et al., 1981; Lieberman and Blumstein, 1988; Rendall et al., 2004; Ghazanfar and Rendall, 2008; Ackermann et al., 2014).

First formant (F1) and formant 2 (F2) have been shown to be important for the identification of vowels in human languages (Peterson and Barney, 1952; Remez et al., 1981; Lieberman and Blumstein, 1988; Hillenbrand et al., 1995). Behavioral studies on baboons (*Papio anubi*), vervet monkeys (*Chlorocebus pygerythrus*), and Japanese monkeys (*Macaca fuscata*) have shown that the monkeys can use formants to discriminate synthetic vowels (Hienz and Brady, 1988; Sinnott, 1989; Sinnott and Kreiter, 1991; Sommers et al., 1992; Hienz et al., 2004). In addition, evidence suggests that rhesus macaques (*Macaca mulatta*) spontaneously perceive changes in formants (Fitch and Fritz, 2006), possibly for recognizing individuals by body size, gender, or age (Sinnott, 1989; Fitch, 1997; Rendall et al., 1998; Bachorowski and Owren, 1999; Smith and Patterson, 2005; Ghazanfar et al., 2007; Furuyama et al., 2016).

However, it has not been tested whether formants contribute to the discrimination of complex sounds, including words in macaques. We trained two rhesus monkeys to discriminate sounds learned as target (T) or non-target (NT). After training, we challenged the monkeys to discriminate morphs of T and NT and F1, F2, or F1F2-pass filters. Our results show that macaques are not only capable of storing numerous sounds in their long-term memories but that they also discriminate sounds embedded in morphs or from formant-pass filters. We anticipate that future neural recordings in this paradigm may explain the neuronal mechanisms of acoustic recognition.

MATERIALS AND METHODS

Animals and Experimental Setup

Two adult rhesus macaques (*M. mulatta*; one 13 kg, 10-year-old male, and one 6 kg, 10-year-old female) participated in this study. The animals inhabited an enriched facility that allowed interactions with other monkeys. The macaques were restricted to water only for 3 h before experimental sessions. However, afterward, they received water *ad libitum*. The monkeys performed ∼1,000 trials for 3 h a day (4–5 days per week). Experiments took place in a soundproof booth where a macaque remained sitting on a primate chair, 60 cm away from a 21" LCD color monitor (1,920 × 1,080 resolution, 60 Hz refresh rate). A Yamaha MSP5 speaker (50 Hz–40 kHz frequency range) was set 15 cm above and behind the monitor to deliver sounds at ∼60 dB SPL measured at the monkeys' ear level. Additionally, a Logitech® Z120 speaker was situated directly below the Yamaha speaker to render white background noise at ∼50 dB SPL. Finally, a metal spring lever positioned at the monkeys' waist level captured their responses.

Behavioral Task

We trained two rhesus monkeys (V and X) to discriminate learned sounds from various categories (**Figure 1A**). Each trial began with a gray circle at the center of the screen, indicating the monkey to press and hold down the lever in order to start a sequence of 1–3 sounds. Each sound lasted 0.5 s and was followed by a 0.5 s delay and the delay by a 0.5 s green go-cue (GC; **Figure 1B**). The probability of a T in a trial was: p (T| position_1) = 1/3, p (T| position_2) = 1/2, and p (T| position_3) = 1 (**Figure 1C**). Thus, trials of 1–3 sounds were presented pseudorandomly and with the same probability. The four possible outcomes of the behavior are illustrated in **Figure 1D**. To obtain a juice reward, the animal was required to keep down the lever throughout 0–2 NT (i.e., correct rejections, CR) and release within 0.8 s of the onset of the T GC (Hit). Releases before this period counted as false alarms (FA), causing the trial to be aborted. On the other hand, to release after the T GC window computed as a Miss. The task was programmed in LabVIEW 2014 (SP1 64-bits, National Instruments®).

Acoustic Stimuli

The sounds were recorded in our laboratory or downloaded from free online libraries. They consisted of Spanish words (T = 6, NT = 10), monkey calls (T = 2, NT = 4), other animal's vocalizations (T = 1, NT = 6), and artificial sounds (T = 2, NT = 5; **Table 1**). We normalized sounds to last 0.5 s, and we then resampled them to 44.1 kHz (cutoff frequencies, 100 Hz to 20 kHz) and finally equalized them (RMS; Adobe Audition® version 6.0). The phonetic nomenclature for Spanish words was obtained using the automatic phonetic transcriptionist by Xavier López Morrás[1]. We also created the seven stimulus-morph-line continua (**Figure 2A**). In each morph-line, nine stimuli were spaced between an NT and a T. The morphs were created using the signal-processing software STRAIGHT (Speech Transformation and Representation based on Adaptive Interpolation of weighted spectrograms; Kawahara et al., 1999; http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_e), following the protocol described by Chakladar et al. (2008) for mixing two sounds by relating salient spectral modulations. The monkeys obtained a reward for releasing the lever at morphs >50% T. However, the reward was delivered pseudorandomly for half the trials at 50% T in order to prevent the learning of that sound, which provided no real decisional criteria.

Finally, we used a voice analysis app for Matlab (VoiceSauce version 1.36, http://www.phonetics.ucla.edu/voicesauce/; Shue et al., 2009) to generate formant-pass sounds (i.e., F1, F2, or F1F2). First, we derived F1 and F2 bandwidths in 25 ms windows every 1 ms. Then, we interpolated the bandwidths using Gaussian time-frequency representations (Elliott and Theunissen, 2009) and used an iterative inversion algorithm to synthesize the sounds[2].

[1] http://aucel.com/pln/transbase.html

[2] http://theunissen.berkeley.edu/software.html

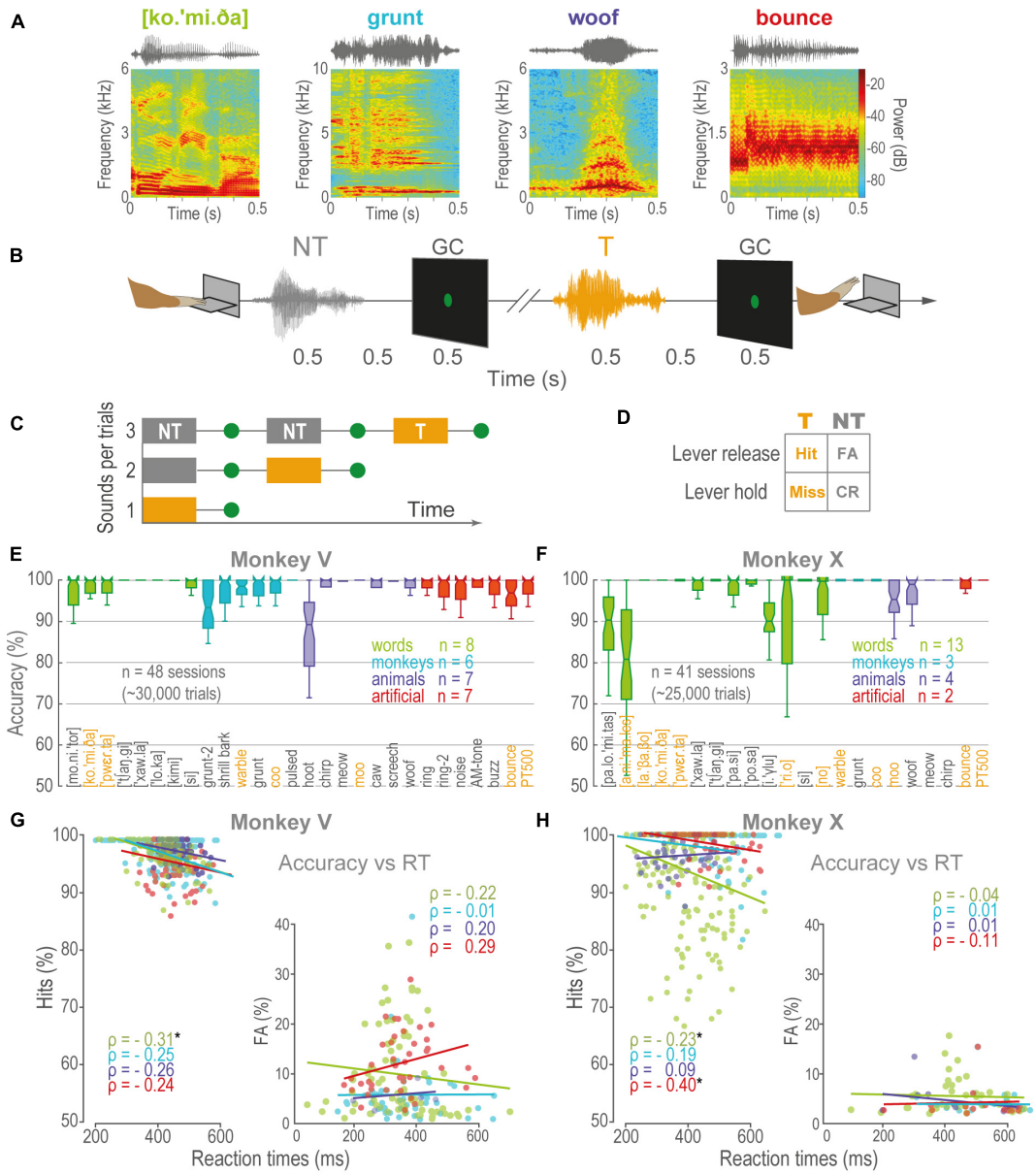

FIGURE 1 | Acoustic discrimination task and performance. **(A)** Spectrograms of four acoustic categories: Green, words, Cyan, conspecific monkey vocalizations, Purple, vocalizations of animals, Red, artificial sounds. **(B)** The sequence of events in a trial: First, the monkey pressed a lever to start. After a variable period (0.5–1 s), a playback of 1–3 sounds commenced. Each sound was followed by a 0.5 s delay and a 0.5 s go-cue (GC). The monkey obtained a liquid reward for releasing the lever within 0.8 s of GC of sounds learned as T, but not during NT sounds. **(C)** Trials consisted of 0–2 NT followed by a T. **(D)** Outcomes of behavior. **(E,F)** Boxplots of the performance of monkeys V and X, respectively, during the discrimination of learned sounds. Orange, T, Gray, NT, other colors follow the color code for categories in **(A)**. Boxplot edges correspond to the 25th and 75th percentiles, central lines, medians. The vertical lines cover ± 2.7 SD. **(G)** ρ, Spearman's Rho correlations between RT as a function of accuracy for Monkey V, during Hits (left panel), and FA (right panel). Linear regressions are visual comparisons of the correlations. Each dot is a session, same color code as in **(E,F)**. **(H)** Same as in **(G)**, but for monkey X. Asterisks are categories whose rho correlations were significant, $p < 0.005$.

Monkeys Training

We attempted diverse strategies to instruct the monkeys. Some details about instructions have been published elsewhere (Morán et al., 2021). However, some key elements were the following: First, the animals learned to press the lever in response to a gray circle and release it after a monkey coo vocalization, a 0.5 s delay, and a 0.5 s GC. Then, we introduced an NT, a delay, and a GC, and the monkeys had to wait and be still until T appearance. After learning a few T and NT, we introduced 0–2 NT to be presented before T. Once the monkeys learned the task sequence, they took only a few days to learn each new sound. The monkeys

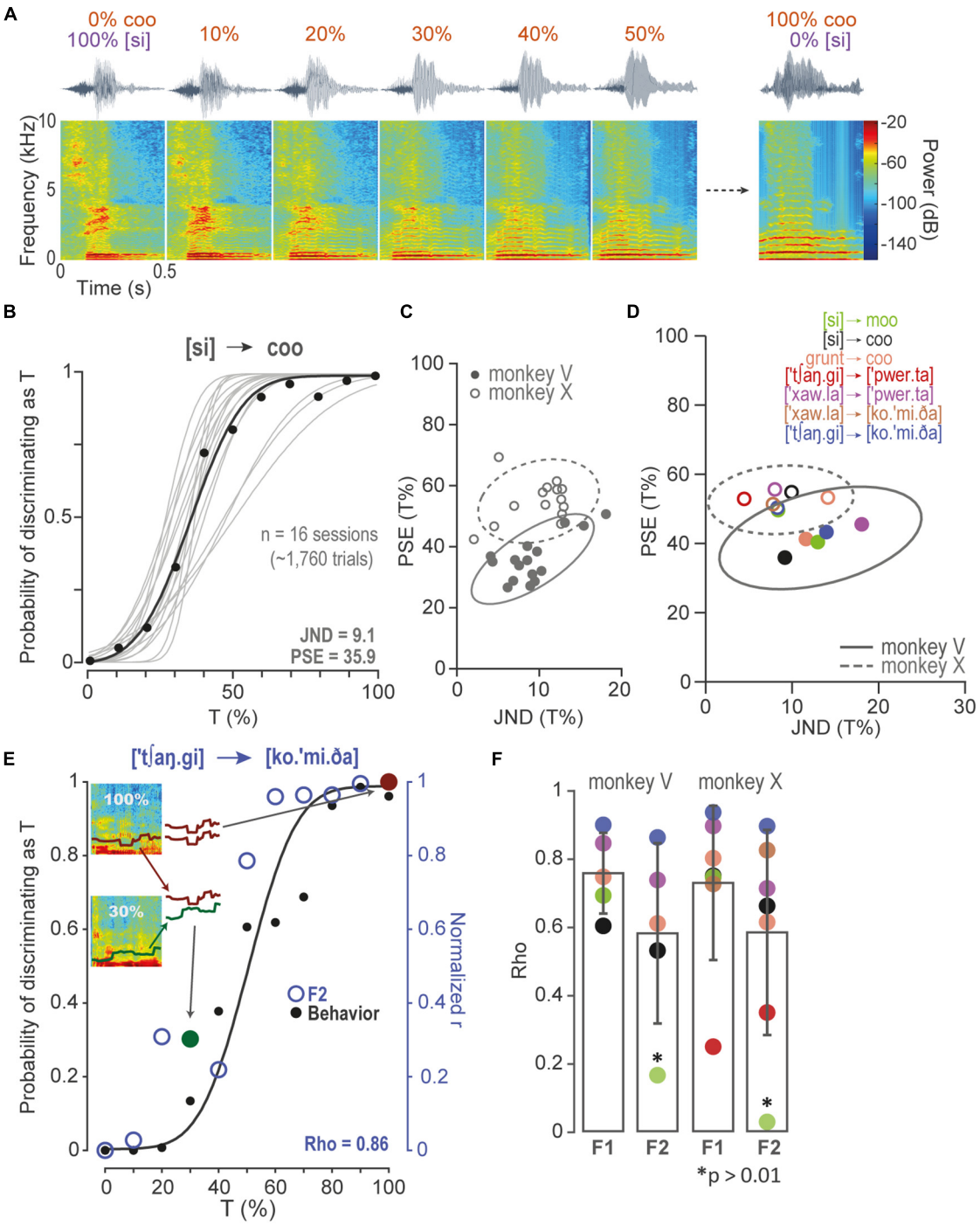

FIGURE 2 | Discrimination of morphs and correlations between performance and formants. **(A)** Some spectrograms of the [si] to coo morph-line continua. The NT [si], i.e., the Spanish word for "yes," morphed gradually in steps of 10 to 100% T "coo" monkey call. **(B)** PF of the probability of recognizing a morph as T. Gray lines, PFs of sessions of monkey V performing in set in **(A)**. Black dots, mean performance of all sessions in each morph. Black line, overall PF performance. **(C)** 2D-Gaussian fits of JND as a function of PSE, for PFs in **(B)**, and for monkey X. **(D)** 2D-Gaussian fits of centroids in **(C)** and the other morphing sets. **(E)** Spearman's Rho correlation of monkey V performance (black dots) in a morphing set, and the distribution of Pearson's *r* correlations (blue open circles). In this example, each open circle resulted from correlating F2 modulated in 100% T vs. F2 modulated in the other morphs. The red closed circle corresponds to the Pearson's *r* of F2 in 100% T vs., again, F2 in 100% T (i.e., $r = 1$). Similarly, the green closed circle corresponds to the Pearson's *r* of F2 in 100% T and F2 in 30% T. Notice that Pearson correlations are normalized in order to compare performance and F2 correlation probabilities directly. **(F)** Spearman's Rho coefficients were obtained as in **(E)** for all morphing sets, monkeys, and F1 and F2. Same color code as **(D)**. Asterisks are non-significant rho correlations, $p > 0.01$.

TABLE 1 | Description of sounds.

	Acoustic category	Sound ID	Description
Target	Monkey	**coo**	Conspecific vocalization
		warble	Conspecific vocalization
	Words	**[ko.ˈmi.ða]**	Spanish word for food
		[ˈpwɛɾ.ta]	Spanish word for door
		[a.ni.ˈma.les]	Spanish word for animals
		[ˈři.o]	Spanish word for river
		[no]	Spanish word for not
		[la.ˈβa.βo]	Spanish word for sink
	Animal	**moo**	Vocal sound of a cow
	Artificial	**bounce**	Bouncing tone
		PT500	Pure tone (500 Hz)
Non-target	Monkey	**grunt**	Conspecific vocalization
		grunt2	Conspecific vocalization
		shrill bark	Conspecific vocalization
		Pulsed	Conspecific vocalization
	Words	[ˈlo.ka]	Spanish word for crazy
		[kimi]	Spanish pseudoword
		[ˈtʃaŋ.gi]	Spanish pseudoword
		[si]	Spanish word for yes
		[ˈxaw.la]	Spanish word for cage
		[mo.ni.ˈtor]	Spanish word for monitor
		[ˈpo.sa]	Spanish pseudoword
		[ˈpa.si]	Spanish pseudoword
		[i.ˈɣlu]	Spanish word for igloo
		[pa.lo.ˈmi.tas]	Spanish word for popcorn
	Animal	meow	Cat vocalization
		chirp	Bird vocalization
		screech	Parrot vocalization
		caw	Crow vocalization
		woof	Dog vocalization
		hoot	Owl vocalization
	Artificial	AM-tone	1 kHz
		buzz	Mosquito whine
		ring	Cellphone ring tone
		ring2	Ring bell
		noise	Passband noise (1–4 kHz)

Sounds in bold were selected for generating morphs and formant-pass sounds.

were not trained in the discrimination of morphs nor formant pass sounds; they were only exposed to those sounds at sessions reported here.

Experimental Sessions

Each daily session consisted of one or two different experiments (e.g., the discrimination of learned sounds, morphs, or formants-pass filters). The morphs experiment consisted of one morph-line-continua set (e.g., [si]-moo or moo-coo). Each sound was presented randomly across trials and positions until repeated at least 10 times. The morphs were presented in the first position, where the probability of encountering a T was the lowest. However, the formant-pass sounds were presented in the first and second positions to achieve enough repetitions per sound. Each set was presented in a block so that trials of different experiments were not intermingled.

Analysis

After exposing the animals to diverse sounds, we arbitrarily selected 5 T and 5 NT to perform most experiments (**Table 1**, bold fonts). We used non-parametric tests (Kruskal–Wallis, Mann–Whitney, and Wilcoxon) to evaluate performance and reaction times (RT) as a function of categories, positions, and subjects. We created psychometric functions (PF) by fitting Gaussian cumulative distribution functions to performance at morphing sets in order to quantify perceptual biases.

$$P\left(\text{release}\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(\text{T}\% - \mu)^2}{2\sigma^2}\right)$$

Where T% corresponds to T proportion in a morph, "μ," is the point of subjective equality (PSE, or the morphing value at 50% chance of perceiving a T), and "σ" (STD) or just noticeable difference (JND, or the proportion to differentiate NT from T 84% of the times; σ = 1; Duarte and Lemus, 2017; Duarte et al., 2018). For all PF, $Q > 0.05$, $Q = \Gamma(0.5 \bullet \chi 2, 0.5 \bullet v)$; where Γ = upper incomplete gamma function, χ2 = chi-square, and v = degrees of freedom (Press et al., 2007).

To evaluate performance throughout sessions of morphs, we fitted a 2D-gaussian of all PSE vs. their corresponding JND. **Figure 2C** compares both monkeys performing in all [si]-coo sessions. **Figure 2D** shows 2D-Gaussians to the centroids of all the other sets (**Supplementary Figure 2B**).

To quantify the contribution of each formant to the discrimination of morph-line stimuli, we calculated the similarity of each formant (F1 and F2) at each morph step to the same formant for the 100%-T stimulus. Similarity was quantified as Pearson's *r*. These values were then correlated, Spearman's rho, with the observed probability of identifying each stimulus in the morph line as a T (see **Figures 2E,F**).

We analyzed data using customized algorithms in MATLAB® version 8.5.0.1, R2015a, The Mathworks, Inc.

RESULTS

The monkeys performed in a task consisting of discriminating as T or NT numerous sounds (n = 36, T = 11, NT = 25; **Figures 1E,F**). After instruction, we did three independent experiments: (1) the discrimination of learned sounds, (2) morphs, and (3) formant-pass filters.

Rhesus Monkeys Learn and Discriminate Complex Sounds

The monkeys V and X discriminated the learned sounds above 50 % chance (V: n = 28; X: n = 22; Hits median: V = 0.97, X = 0.96; CR median: V = 0.98, X = 0.96; one-sample Wilcoxon signed-rank test, median = 0.75, Z [V_Hits] = 10.41, Z [V_CR] = 8.51, Z [X_Hits] = 9.63, Z [X_CR] = 7.87; $p < 0.001$). The animals did not show significant biases for any sound or category (**Supplementary Figures 1A,B**; pairwise Wilcoxon rank-sum test, false discovery rate corrected for multiple comparisons using the Benjamini-Hochberg procedure; q-value = 0.01). Despite the differences between the monkeys (V, X), the categories (T, NT), and the stimulus position (1st, 2nd, 3rd), mean performance was consistently above 90% accuracy (**Supplementary Figures 1C,D**). In general, monkey X was faster than V. However, there were only significant correlations between accuracy and RT for monkey X, with discriminating synthetic sounds and both monkeys discriminating words (**Figures 1G,H** and **Supplementary Figures 1E,F**). Overall, these results indicate the monkeys could learn and discriminate sounds of different categories.

The Discriminations of Morphs Correlated With First Formant and Second Formant Modulations

To measure the monkeys' capacity to discriminate sounds, we tested them in seven sets consisting of morphs of T and NT in different proportions. **Figure 2A** illustrates the NT [si] (i.e., the Spanish word for "yes") gradually morphing to a T monkey "coo" call. **Figure 2B** shows PFs of all sessions (n = 16) in which monkey V performed at [si] to coo set (see also **Supplementary Figure 2A**). To compare their behaviors, we fitted a 2D-gaussian to all JND vs. PSE derived from each PF (**Figure 2C** and **Supplementary Figure 2B**). Similarly, we fitted 2D-Gaussians to the centroids obtained from the 2D-gaussian distributions of all sets (**Figure 2D**). The mean of centroids of monkey V was 19.7 ± 8.7, 41.5 ± 7.5 (JND ± SD, PSE ± SD), and of monkey X, 12.9 ± 6.3, 52.7 ± 4.9 (JND ± SD, PSE ± SD). Monkey V showed some bias to discriminate morphs as T (pairwise Wilcoxon rank-sum test, Benjamini-Hochberg FDR correction, q-value = 0.01, **Supplementary Figures 2C,D**). Nevertheless, both monkeys discriminated morphs proficiently.

To further study the contribution of formants to the monkeys' discriminations, we calculated Spearman's rho correlations between performance and F1 and F2 modulations to test the hypothesis that the probability of discriminating a morph as T was proportional to the correlation between the formants of the morphs and of 100% T. **Figure 2E** presents a PF and the distribution of the normalized Pearson's r correlations along the morph-line continua. In this example, F2 correlated significantly to the probability of recognizing sounds as T (Spearman's Rho, $p < 0.01$; see **Supplementary Figure 2E** for all morphing set). **Figure 2F** shows that F1 correlated with both of the monkeys' performance in all morphing sets, whereas F2 correlated in 4 out of 5 sets for monkey V and 6 out of 7 for monkey X (Spearman's Rho, $p < 0.01$).

The Monkeys Discriminated Sounds Comprised of First Formant and Second Formant-Pass Filters

We presented the monkeys with F1, F2, and F1F2-pass filters synthesized from the learned sounds (**Figure 3A**). **Figures 3B,C** shows that both animals discriminated above chance most of the sounds, i.e., F1, 70.1% ± 14 (mean ± SD), F2, 72.6 ± 21, and F1F2, 79.2 ± 12.2. However, performance was significantly lower than during the discrimination of the learned sounds: Learned > F1F2 > F2 > F1 (Benjamini-Hochberg and FDR correction for multiple Wilcoxon signed-rank test comparisons; q-value = 0.01; **Figure 3D**). These results suggest that formants F1 and F2 provide relevant information about sounds.

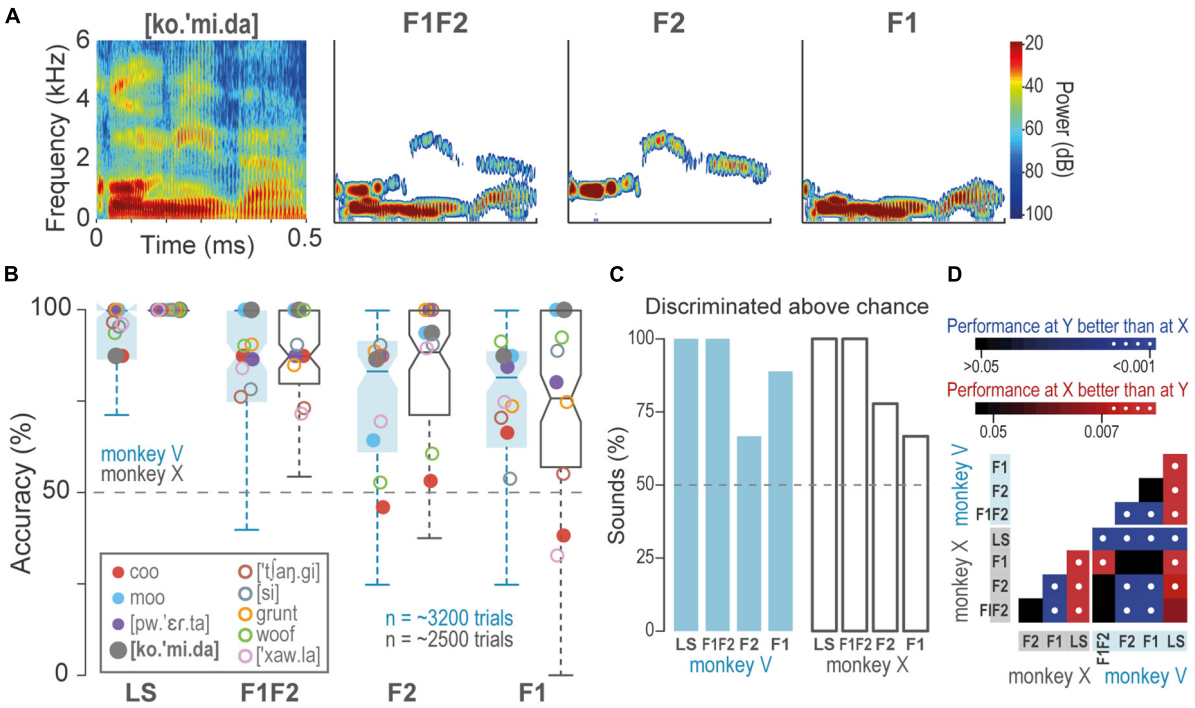

FIGURE 3 | Discrimination of formant-pass sounds. **(A)** Spectrograms of T [ko.'mi.da] and F1F2, F2, and F1-pass sounds. **(B)** Boxplots of accuracy during the discrimination of formant-pass filters compared to the discrimination of the learned sounds (LS). Dash line, performance at chance = 50%. Closed circles, T, open circles NT. **(C)** Percentage of learned and formant-pass sounds discriminated above chance (Wilcoxon signed-rank test, $q = 0.01$). Dash line, performance at the chance. **(D)** p-values of multiple pairwise comparisons of sounds and monkeys. Color gradients indicate p-values, white dots, significant differences.

DISCUSSION

We have presented evidence of the capacity of rhesus monkeys to learn and discriminate sounds from a broad range of frequencies and temporal modulations and corroborated that they are capable of discriminating morphs between pairs of sounds (Tsunada et al., 2011).

Rhesus Macaques Have Long-Term Memories of Complex Sounds

Evidence of long-term memory of ethological sounds in monkeys is restricted to conspecific vocalizations (Seyfarth et al., 1980a). In the present study, we demonstrate that rhesus macaques can discriminate non-conspecific vocalizations and other naturalistic sounds. This perceptual ability may depend on circuits of acoustic categories, whose projections to motor areas could serve as feedback for vocal learning in species such as NHP and birds (Takahashi et al., 2017; Moore and Woolley, 2019; Zhao et al., 2019). It has been proposed that the learning of sounds in NHP is genetically determined (Brockelman and Schilling, 1984; Owren et al., 1992; Zador, 2019). In such a scenario, genetically programmed circuits should admit inclusions of non-ethological sounds as those that our monkeys learned.

In our task, learning consisted of associating two behaviors with diverse sounds, including conspecific vocalizations that may have had stereotyped responses. Similar associations to sounds have been reported previously for other communicating animals (Town et al., 2018; Saunders and Wehr, 2019; Yu et al., 2020). An important open question here is whether storing new sounds in long-term memory is achieved by nesting them to homophones (Chomsky, 1959). Consistent with previous reports, the training of our monkeys was more tenuous and prolonged than in visual or tactile tasks (Colombo and D'Amato, 1986; Colombo and Graziano, 1994; Wright, 1999, 2007; Fritz et al., 2005; Lemus et al., 2009a; Scott et al., 2012; Rajalingham et al., 2015). Therefore, acoustic learning based on nesting is unlikely since it would be possible to incorporate new sounds into existing circuits quickly. Alternatively, learning may depend on context (e.g., sentences), which, compared to humans, may be limited in macaques.

Did the monkeys learn whole sounds or only some segments? A possibility is that the animals learned only a chunk of sounds rather than all spectrotemporal modulations. Functional magnetic resonance imaging and electrocorticography studies in humans suggest that the representations of sounds start by phonetic relationships at the lateral bank of the auditory cortex (Chang et al., 2010; Obleser et al., 2010; Mesgarani et al., 2014). In macaques, neurons of the lateral belt respond to "monosyllabic" conspecific vocalizations of various broadband frequencies (Rauschecker et al., 1995) processed hierarchically along the superior temporal gyrus (Leaver and Rauschecker, 2010; Ortiz-Rios et al., 2015; Belin et al., 2018) up to the prefrontal cortex (Romanski et al., 1999; Rauschecker and Romanski, 2011).

In our task, the animals were exposed to multisyllabic words, which were arguably learned in only the first or last portions. This possibility would concur with the idea of macaques being only capable of processing single units of sound, such as their vocalizations. Previous reports suggest that macaques use all available information to discriminate acoustic flutter (Lemus et al., 2009a,b). Those sounds consisted of periodic trains of pulses that might not have required the monkeys to listen entirely in order to discriminate. In our paradigm, sounds also lasted 0.5 s; however, sounds consisted of dynamical spectral modulations that the monkeys likely attended to in order to accumulate evidence and to improve performance (Brunton et al., 2013).

Ng et al. (2009) exposed macaques to complex sounds similar to ours in a match-to-sample task. In contrast to our results, they found that the animals performed better for conspecific calls than for other categories. This inconsistency may derive from differences between the short-term memory they tested and the long-term memory explored in our task. Similarly, in a delayed match-to-sample task (Scott et al., 2012), performance depended on presenting 0–2 distractors in a trial (i.e., 91, 73, and 39%, respectively). The authors concluded that this detriment was due to the number of distractors interfering with working memory. Again, performance was not affected in our study despite the position of sounds in a trial or ethological relevance. Future studies may determine differences in mechanisms and anatomical representations of short- and long-term memory in NHP (Munoz-Lopez et al., 2010; Muñoz-López et al., 2015; Fritz et al., 2016).

Rhesus Monkeys Discern Categories From Acoustic Mixtures

We exposed the monkeys to acoustic morphs of T and NT to explore their discrimination thresholds. Our results are consistent with previous reports in humans categorizing monkey calls (e.g., coos, grunts, and harmonic arches; Furuyama et al., 2017; Jiang et al., 2018) and the /a/ vowel (Chakladar et al., 2008), suggesting that macaques possess an acoustic perception similar to that of humans. Similarly, Tsunada et al. (2011) trained macaques to discriminate morphs of the syllables /bad/ and /dad/ to study the neuronal correlates of acoustic categorization. They found that the neurons of the auditory belt area presented categorical responses to the graded mixtures, meaning that those neurons correlated with decisions rather than the perception of acoustic parameters. Therefore, to explore the impact on acoustic perception of parameters such as F1 and F2 formants, related to the recognition of vowels in humans (Peterson and Barney, 1952; Remez et al., 1981; Lieberman and Blumstein, 1988; Hillenbrand et al., 1995), we computed correlations between the psychometric curves in monkeys and those features. Our results show that F1 and F2 indeed correlated with behavior. Something noteworthy to mention is that regardless of the fact that the animals learned only some sounds, they nevertheless could discriminate morphs to which they were exposed on only a few occasions. In other words, the monkeys discriminated from modified information of learned sounds, suggesting that perception is invariant. In any case, this result cannot rule out that other acoustic features contribute to perception (Stevens, 1983; Brewer and Barton, 2016).

Monkeys Discriminate Complex Sounds Based on Formant Frequencies

To test whether formants sufficed for discriminations, we presented the monkeys with formant-pass sounds. We found that formants indeed sufficed. Furthermore, F1 and F2 combined improved performance as compared to F1 and F2 alone. However, to further understand how formants participate in acoustic perception, an exciting control would be to present only the complementary information to F1- and F2-pass filters.

Since formants constitute the most energetic modulations in sounds, they may significantly shape neuronal circuits representing sounds. Here the hypothesis is that salient signals excite neurons in higher probability than other signals (at least in primary sensory areas). For instance, formants simultaneously activate neurons at different frequency bands of the auditory cortex. Those cells, in turn, could activate upstream neurons, creating circuits of acoustic representations (Hebb, 1949). Our findings suggest that formants contribute to the discrimination of complex sounds in macaques, perhaps like for humans in the perception of communication sounds (Remez et al., 1981; Fitch and Fritz, 2006; Ghazanfar et al., 2007; Furuyama et al., 2016, 2017).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Mexican Official Standard Recommendations for the Care and Use of Laboratory Animals (NOM-062-ZOO-1999) and the Internal Committee for the Use and Care of Laboratory Animals of the Institute of Cell Physiology, UNAM (CICUAL; LLS80-16).

AUTHOR CONTRIBUTIONS

JM and IM performed experiments. JM, JV, and LL analyzed data and prepared the figures. JM, TF, JV, and IM revised the manuscript. TF programmed the task. LL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We are grateful for the financial support provided by CONACYT CB-256767, and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica [Support Program for Research Projects and Technological Innovation (PAPIIT) IN207919].

ACKNOWLEDGMENTS

Jonathan Melchor Hernández is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas (Doctoral program in biomedical sciences), Universidad Nacional Autónoma de México (UNAM) and has received CONACyT fellowship 229866. The data in this work are part of his doctoral dissertation. We wish to thank Francisco Pérez, Gerardo Coello, and Ana Escalante of the computing department of the IFC, and Gabriel Pérez Ruelas for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.728686/full#supplementary-material

REFERENCES

Ackermann, H., Hage, S. R., and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav. Brain Sci.* 72, 1–84.

Bachorowski, J.-A., and Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J. Acoust. Soc. Am.* 106, 1054–1063. doi: 10.1121/1.427115

Belin, P. (2006). Voice processing in human and non-human primates. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2091–2107. doi: 10.1098/rstb.2006.1933

Belin, P., Bodin, C., and Aglieri, V. (2018). A "voice patch" system in the primate brain for processing vocal information? *Hear. Res.* 366, 65–74. doi: 10.1016/j.heares.2018.04.010

Brewer, A. A., and Barton, B. (2016). Maps of the auditory cortex. *Annu. Rev. Neurosci.* 39, 385–407. doi: 10.1146/annurev-neuro-070815-014045

Brockelman, W. Y., and Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature* 312, 634–636. doi: 10.1038/312634a0

Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* 340, 95–98. doi: 10.1126/science.1233912

Ceugniet, M., and Izumi, A. (2004). Vocal individual discrimination in Japanese monkeys. *Primates* 45, 119–128. doi: 10.1007/s10329-003-0067-3

Chakladar, S., Logothetis, N. K., and Petkov, C. I. (2008). Morphing rhesus monkey vocalizations. *J. Neurosci. Methods* 170, 45–55. doi: 10.1016/j.jneumeth.2007.12.023

Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., and Ghazanfar, A. A. (2011). Monkeys and humans share a common computation for face/voice integration. *PLoS Comput. Biol.* 7:e1002165. doi: 10.1371/journal.pcbi.1002165

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641

Chomsky, N. (1959). On certain formal properties of grammars. *Inf. Control* 2, 137–167. doi: 10.1016/S0019-9958(59)90362-6

Colombo, M., and D'Amato, M. R. (1986). A comparison of visual and auditory short—term memory in monkeys (*Cebus apella*). *Q. J. Exp. Psychol. Sect. B* 38, 425–448.

Colombo, M., and Graziano, M. (1994). Effects of auditory and visual interference on auditory-visual delayed matching to sample in monkeys (*Macaca fascicularis*). *Behav. Neurosci.* 108, 636–639. doi: 10.1037/0735-7044.108.3.636

Duarte, F., Figueroa, T., and Lemus, L. (2018). A two-interval forced-choice task for multisensory comparisons. *J. Vis. Exp.* 141:e58408. doi: 10.3791/58408

Duarte, F., and Lemus, L. (2017). The time is up: compression of visual time interval estimations of bimodal aperiodic patterns. *Front. Integr. Neurosci.* 11:17. doi: 10.3389/fnint.2017.00017

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5:1000302. doi: 10.1371/journal.pcbi.1000302

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* 102, 1213–1222. doi: 10.1121/1.421048

Fitch, W. T., and Fritz, J. B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J. Acoust. Soc. Am.* 120, 2132–2141. doi: 10.1121/1.2258499

Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206, 159–176. doi: 10.1016/j.heares.2005.01.015

Fritz, J. B., Malloy, M., Mishkin, M., and Saunders, R. C. (2016). Monkey's short-term auditory memory nearly abolished by combined removal of the rostral superior temporal gyrus and rhinal cortices. *Brain Res.* 1640, 289–298. doi: 10.1016/j.brainres.2015.12.012

Furuyama, T., Kobayasi, K. I., and Riquimaroux, H. (2016). Role of vocal tract characteristics in individual discrimination by Japanese macaques (*Macaca fuscata*). *Sci. Rep.* 6:32042. doi: 10.1038/srep32042

Furuyama, T., Kobayasi, K. I., and Riquimaroux, H. (2017). Acoustic characteristics used by Japanese macaques for individual discrimination. *J. Exp. Biol.* 220, 3571–3578. doi: 10.1242/jeb.154765

Ghazanfar, A. A., and Rendall, D. (2008). Evolution of human vocal production. *Curr. Biol.* 18, R457–R460. doi: 10.1016/j.cub.2008.03.030

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., and Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in Rhesus monkeys. *Curr. Biol.* 17, 425–430. doi: 10.1016/j.cub.2007.01.029

Hauser, M. D. (1998). Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. *Anim. Behav.* 55, 1647–1658. doi: 10.1006/anbe.1997.0712

Hebb, D. O. (1949). *The Organisation of Behaviour: A Neuropsychological Theory*. New York, NY: Science Editions.

Hienz, R. D., and Brady, J. V. (1988). The acquisition of vowel discriminations by nonhuman primates. *J. Acoust. Soc. Am.* 84, 186–194. doi: 10.1121/1.396963

Hienz, R. D., Jones, A. M., and Weerts, E. M. (2004). The discrimination of baboon grunt calls and human vowel sounds by baboons. *J. Acoust. Soc. Am.* 116, 1692–1697. doi: 10.1121/1.1778902

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099–3111. doi: 10.1121/1.411872

Jiang, X., Chevillet, M. A., Rauschecker, J. P., and Riesenhuber, M. (2018). Training humans to categorize monkey calls: auditory feature- and category-selective neural tuning changes. *Neuron* 98, 405–416.e4. doi: 10.1016/j.neuron.2018.03.014

Jovanovic, T., Megna, N. L., and Maestripieri, D. (2000). Early maternal recognition of offspring vocalizations in rhesus macaques (*Macaca mulatta*). *Primates* 41, 421–428. doi: 10.1007/BF02557653

Kawahara, H., Masuda-Katsuse, I., and De Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun.* 27, 187–207. doi: 10.1016/S0167-6393(98)00085-5

Leaver, A. M., and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612. doi: 10.1523/JNEUROSCI.0296-10.2010

Lemus, L., Hernández, A., and Romo, R. (2009a). Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9471–9476. doi: 10.1073/pnas.0904066106

Lemus, L., Hernández, A., Romo, R., Hernandez, A., Romo, R., Hernández, A., et al. (2009b). Neural encoding of auditory discrimination in ventral premotor cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14640–14645. doi: 10.1073/pnas.0907505106

Lieberman, P., and Blumstein, S. E. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9781139165952

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994

Moore, J. M., and Woolley, S. M. N. (2019). Emergent tuning for learned vocalizations in auditory cortex. *Nat. Neurosci.* 22, 1469–1476. doi: 10.1038/s41593-019-0458-4

Morán, I., Perez-Orive, J., Melchor, J., Figueroa, T., and Lemus, L. (2021). Auditory decisions in the supplementary motor area. *Prog. Neurobiol.* 202:102053. doi: 10.1016/j.pneurobio.2021.102053

Muñoz-López, M., Insausti, R., Mohedano-Moriano, A., Mishkin, M., and Saunders, R. C. (2015). Anatomical pathways for auditory memory II: information from rostral superior temporal gyrus to dorsolateral temporal pole and medial temporal cortex. *Front. Neurosci.* 9:158. doi: 10.3389/fnins.2015.00158

Munoz-Lopez, M. M., Mohedano-Moriano, A., and Insausti, R. (2010). Anatomical pathways for auditory memory in primates. *Front. Neuroanat.* 4:129. doi: 10.3389/fnana.2010.00129

Ng, C. W., Plakke, B., and Poremba, A. (2009). Primate auditory recognition memory performance varies with sound type. *Hear. Res.* 256, 64–74. doi: 10.1016/j.heares.2009.06.014

Obleser, J., Leaver, A. M., Vanmeter, J., and Rauschecker, J. P. (2010). Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Front. Psychol.* 1:232. doi: 10.3389/fpsyg.2010.00232

Ortiz-Rios, M., Kuśmierek, P., DeWitt, I., Archakov, D., Azevedo, F. A. C., Sams, M., et al. (2015). Functional MRI of the vocalization-processing network in the macaque brain. *Front. Neurosci.* 9:113. doi: 10.3389/fnins.2015.00113

Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (1992). 'Food' calls produced by adult female Rhesus (*Macaca Mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour* 120, 218–231. doi: 10.1163/156853992X00615

Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374. doi: 10.1038/nn2043

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3rd Edn. Cambridge: Cambridge University Press.

Rajalingham, R., Schmidt, K., and DiCarlo, J. J. (2015). Comparison of object recognition behavior in human and monkey. *J. Neurosci.* 35, 12127–12136. doi: 10.1523/JNEUROSCI.0573-15.2015

Rauschecker, J. P., and Romanski, L. M. (2011). "Auditory cortical organization: evidence for functional streams," in *The Auditory Cortex*, eds J. Winer, and C. Schreiner (Boston, MA: Springer), 99–116. doi: 10.1007/978-1-4419-0074-6_4

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114. doi: 10.1126/science.7701330

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–950. doi: 10.1126/science.7233191

Rendall, D., Owren, M. J., and Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J. Acoust. Soc. Am.* 103, 602–614. doi: 10.1121/1.421104

Rendall, D., Owren, M. J., Weerts, E., and Hienz, R. D. (2004). Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners. *J. Acoust. Soc. Am.* 115, 411–421. doi: 10.1121/1.1635838

Rendall, D., Rodman, P. S., and Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* 51, 1007–1015. doi: 10.1006/anbe.1996.0103

Romanski, L. M., Bates, J. F., and Goldman-Rakic, P. S. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403, 141–157. doi: 10.1002/(SICI)1096-9861(19990111)403:2<141::AID-CNE1>3.0.CO;2-V

Saunders, J. L., and Wehr, M. (2019). Mice can learn phonetic categories. *J. Acoust. Soc. Am.* 145, 1168–1177. doi: 10.1121/1.5091776

Scott, B. H., Mishkin, M., and Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12237–12241. doi: 10.1073/pnas.1209685109

Seyfarth, R. M, Cheney, D., and Marler, P. (1980a). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801–803. doi: 10.1126/science.7433999

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980b). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* 28, 1070–1094. doi: 10.1016/S0003-3472(80)80097-2

Shue, Y.-L., Keating, P., and Vicenik, C. (2009). VoiceSauce: a program for voice analysis. *J. Acoust. Soc. Am.* 126:2221. doi: 10.1121/1.3248865

Sinnott, J. M. (1989). Detection and discrimination of synthetic English vowels by Old World monkeys (*Cercopithecus*, *Macaca*) and humans. *J. Acoust. Soc. Am.* 86, 557–565. doi: 10.1121/1.398235

Sinnott, J. M., and Kreiter, N. A. (1991). Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans. *J. Acoust. Soc. Am.* 89, 2421–2429. doi: 10.1121/1.400974

Slocombe, K. E., and Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or relative food value? *Anim. Behav.* 72, 989–999. doi: 10.1016/j.anbehav.2006.01.030

Smith, D. R. R., and Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J. Acoust. Soc. Am.* 118, 3177–3186. doi: 10.1121/1.2047107

Sommers, M., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (1992). Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *J. Acoust. Soc. Am.* 91, 3499–3510. doi: 10.1121/1.402839

Stevens, K. N. (1983). Acoustic properties used for the identification of speech sounds. *Ann. N. Y. Acad. Sci.* 405, 2–17. doi: 10.1111/j.1749-6632.1983.tb31613.x

Takahashi, D. Y., Liao, D. A., and Ghazanfar, A. A. (2017). Vocal learning *via* social reinforcement by infant marmoset monkeys. *Curr. Biol.* 27, 1844–1852.e6. doi: 10.1016/j.cub.2017.05.004

Town, S. M., Wood, K. C., and Bizley, J. K. (2018). Sound identity is represented robustly in auditory cortex during perceptual constancy. *Nat. Commun.* 9:4786. doi: 10.1038/s41467-018-07237-3

Tsunada, J., Lee, J. H., and Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *J. Neurophysiol.* 105, 2634–2646. doi: 10.1152/jn.00037.2011

Wright, A. A. (1999). Auditory list memory and interference processes in monkeys. *J. Exp. Psychol. Anim. Behav. Process.* 25, 284–296. doi: 10.1037/0097-7403.25.3.284

Wright, A. A. (2007). An experimental analysis of memory processing. *J. Exp. Anal. Behav.* 88, 405–433. doi: 10.1901/jeab.2007.88-405

Yu, K., Wood, W. E., and Theunissen, F. E. (2020). High-capacity auditory memory for vocal communication in a social songbird. *Sci. Adv.* 6, 440–453. doi: 10.1126/sciadv.abe0440

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* 10:3770. doi: 10.1038/s41467-019-11786-6

Zhao, L., Rad, B. B., and Wang, X. (2019). Long-lasting vocal plasticity in adult marmoset monkeys. *Proc. R. Soc. B Biol. Sci.* 286:20190817. doi: 10.1098/rspb.2019.0817

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.