UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

INSTITUTO DE FISIOLOGÍA CELULAR

RESPUESTAS PSICOFÍSICAS Y NEUROFISIOLÓGICAS DURANTE LA DISCRIMINACIÓN DE OBJETOS ACÚSTICOS EN MONOS RHESUS (*Macaca mulatta*).

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

P R E S E N T A:

M.C. JONATHAN MELCHOR HERNÁNDEZ

DIRECTOR DE TESIS

DR. LUIS ALONSO LEMUS SANDOVAL

INSTITUTO DE FISIOLOGÍA CELULAR, UNAM

MIEMBROS DEL COMITÉ TUTOR

DR. LUIS CONCHA LOYOLA
INSTITUTO DE NEUROBIOLOGÍA, UNAM

DR. VÍCTOR DE LAFUENTE FLORES
INSTITUTO DE NEUROBIOLOGÍA, UNAM

CIUDAD DE MÉXICO, ENERO 2022

AGRADECIMIENTOS

Al Dr. Luis Lemus por permitirme continuar y contribuir a mi formación académica en su laboratorio, por todo el apoyo y las enseñanzas brindadas, muchas gracias.

A los doctores: Victor de Lafuente y Luis Concha por sus valiosos comentarios y aportaciones durante el desarrollo del proyecto de investigación.

Al Dr. José Vergara por compartir sus conocimientos y fundamental colaboración, pero también por su fraternal apoyo.

A mis compañeros de laboratorio por sus diferentes contribuciones y todas las experiencias compartidas: Isaac, Tonatiuh, Miguel, Elizabeth y Fabiola. A Ivonne y Gabriel por su invaluable asistencia técnica y muy especialmente a Vaporel y Xomara.

Al Dr. Enrique Soto y la Dra. Angélica Almanza, por su guía durante mis estudios de licenciatura y maestría, en la Benemérita Universidad Autónoma de Puebla.

A mis padres, Reveca y Emilio, por su amor e incondicional apoyo.

A mi familia y amigos (ustedes saben porque): Nashelly, Brenda, Daniel, Emiliano, Gael, David, Lea, Epifanio, Ozue, Marco Antonio, Miguel, Verónica, Amado, Luis, Belén, Audrey, Alfredo, Elizabeth, Zuriel, Johanna, Pedro, Gerardo, Salvador, Fernanda, René, Amandita, Itzel, Diego y Zeus.

A Janett, por cada minuto juntos.

Agradezco a la Universidad Nacional Autónoma de México, al Instituto de Fisiología Celular, al Posgrado en Ciencias Biomédicas, a la DGAPA, al COMECYT y CONACYT por las becas, el financiamiento y las facilidades otorgadas para la realización de este proyecto.

ÍNDICE

ABREVIATURAS

A1, corteza auditiva primaria

AMS, área motora suplementaria

CC, células ciliadas

cm, centímetro

CPFdl, corteza prefrontal dorsolateral

CPFvl, corteza prefrontal ventrolateral

dB, decibelio

F0, frecuencia fundamental

F1, primer formante

F2, segundo formante

GTS, giro temporal superior

Hz, hercio (*Hertz*)

kHz, kilohercio

m, metro

N, newton

OA, objetos auditivos

Pa, pascal

PNH, primates no humanos

R, rostral

RT, rostrotemporal

s, segundo

SA, sistema auditivo

SPL, nivel de presión sonora (*sound pression level*)

STS, surco temporal superior

1. RESUMEN

Desde un punto de vista filogenético, los primates no humanos (PNH) son los animales más próximos a nuestra especie. Derivado de estudios anatómicos, electrofisiológicos y de neuroimagen (Ahveninen et al., 2013; Alain et al., 2001; Romanski et al., 1999; Tian et al., 2001) se ha propuesto que la representación de contenido, identidad y significado de los sonidos ocurre a lo largo de una vía ventral de procesamiento acústico (Bizley y Cohen, 2013; Rauschecker, 2018). Sin embargo, aún no se ha establecido cómo se procesan e integran a lo largo de esta vía, distintas características acústicas para formar objetos auditivos (OA). Por lo tanto, para estudiar los correlatos psicofísicos y neuronales del reconocimiento de OA, desarrollamos un modelo de discriminación auditiva en monos rhesus (*Macaca mulatta*). Los resultados obtenidos de nuestros experimentos sugieren que los monos son capaces de aprender y discriminar sonidos de distintas categorías (p. ej. sonidos artificiales y palabras), y que lo hacen de forma similar a lo reportado en humanos. Adicionalmente, encontramos que las frecuencias de mayor energía de los OA (formantes F1 y F2), son importantes para la discriminación de OA en macacos; tal y como ocurre con el reconocimiento de vocales en humanos. Finalmente, descubrimos que la actividad neuronal del área motora suplementaria (AMS) correlaciona con la selección de respuestas asociadas con cada OA, y que lo hace invariantemente a distintos emisores. Todo lo anterior nos permite plantear experimentos futuros en nuestro modelo, que ayuden a entender los mecanismos corticales del procesamiento acústico, particularmente, de cómo los formantes contribuyen a la percepción invariante de OA.

ABSTRACT

From a phylogenetic point of view, non-human primates are the closest animals to our species. Derived from anatomical, electrophysiological and, neuroimaging studies (Romanski et al., 1999; Tian, 2001; Alain et al., 2001; Ahveninen et al., 2013) it has been proposed that the representation of content, identity, and meaning of sounds occurs along a ventral acoustic processing pathway (Bizley and Cohen, 2013; Rauschecker, 2018). However, it has not yet been established how different acoustic characteristics are processed and integrated along this pathway to form auditory objects (AO). Therefore, to study the psychophysical and neural correlates of AO recognition, we developed a model of auditory discrimination in rhesus monkeys (*Macaca mulatta*). The results obtained suggest that monkeys are capable of learning and discriminating sounds of different categories (e.g. artificial sounds and words) and they do so similarly to that reported in humans. Additionally, we found that the higher energy frequencies of the AO (formants) are important for the discrimination of AO in macaques; just as it happens with the recognition of vowels in humans. Finally, we discovered that the neuronal activity of the supplementary motor area (SMA) correlates with the selection of responses associated with each OA and that it does so invariably to different emitters. All the above allows us to propose future experiments in our model, which help to understand the cortical mechanisms of acoustic processing, particularly how formants contribute to the invariant perception of AO.

2. INTRODUCCIÓN

La capacidad de emitir, reconocer y localizar los sonidos es una habilidad fundamental para la interacción social de los primates. Sabemos que nuestro cerebro construye representaciones de objetos y conceptos a partir de la información sensorial y de la experiencia (Romo et al., 2012). Sin embargo, los mecanismos cerebrales subyacentes al reconocimiento auditivo solo se conocen de forma parcial. En esta sección se revisan algunas propiedades físicas de los sonidos, y distintos aspectos de la estructura y función del sistema auditivo que se ha propuesto como principales responsables de la generación del fenómeno de la percepción de objetos auditivos (OA).

2.1 Propiedades físicas y perceptuales del sonido

Físicamente, el sonido es una onda mecánica longitudinal que se propaga por un medio elástico. También puede ser definido como oscilaciones de presión generadas por la propagación del movimiento de las moléculas del medio. Se necesita una fuente de vibración mecánica para que se produzca, y su velocidad de propagación está determinada por el medio, aumentando con la temperatura y la altitud. En el aire la velocidad del sonido es de 343 m/s a 20 °C, a nivel del mar.

El sonido constituye un flujo de energía a través de la materia y como onda, puede ser caracterizado en términos de su amplitud, frecuencia y duración (Figura 1A). La frecuencia hace alusión al número de ciclos por unidad de tiempo en el que se repite el cambio de la presión; se mide en unidades llamadas hercios (Hz), donde 1 Hz equivale a un ciclo por segundo. El periodo es el tiempo transcurrido entre un ciclo y el siguiente, y la amplitud o intensidad representa la energía que transporta la onda sonora y se expresa como variaciones de presión (Pa/s) o de energía (Watt/m^2). Donde un Pascal (Pa) corresponde a la presión que ejerce una fuerza de un Newton (N) sobre una superficie de un metro cuadrado (m^2).

El oído humano puede percibir ondas sonoras de entre los 20 y 20 000 Hz, pero su mayor sensibilidad se encuentra en el rango de 1 a 4 kHz (Goldstein y Brockmole, 2016). Los sonidos simples o tonos puros son ondas sinusoidales de una sola frecuencia, pero son prácticamente inexistentes en la naturaleza. En cambio, los sonidos naturales poseen modulaciones espectro-temporales, que, aunque sean complejas, les

caracterizan y, por tanto, nos permiten diferenciarlos. Por ejemplo, la voz humana resulta de la combinación de frecuencias generadas por la vibración de las cuerdas vocales (Figura 1B-C). A la frecuencia más baja se le conoce como frecuencia fundamental (F0) y a sus múltiplos enteros, armónicos. La F0 de la voz masculina se encuentra en el rango de 100 a 200 Hz, mientras que la voz femenina es más aguda, típicamente va de 150 a 300 Hz (Peterson y Barney, 1952; Kent y Vorperian, 2018).

Los sonidos de comunicación como las palabras se caracterizan por concentrar la mayor parte de su energía en bandas de frecuencias llamadas formantes. Los formantes se enumeran de forma subsecuente de acuerdo con el incremento de frecuencias (Figura 1D). De hecho, somos capaces de distinguir vocales a partir de los formantes (Peterson y Barney, 1952; Frank et al., 2020).

Existen cuatro cualidades perceptuales del sonido: tono o altura, timbre o color, duración y la sonoridad. El tono puede ser agudo (alto), o grave (bajo), y se relaciona directamente con la F0. El timbre nos permite diferenciar sonidos de igual F0 e intensidad. El timbre depende de la cantidad de armónicos y de la intensidad de cada uno de ellos, de la variación de amplitudes en el tiempo, y de los formantes. También se ha propuesto que permite la identificación de la fuente sonora. Por analogía con la textura, puede ser considerado como áspero, aterciopelado o metálico. La duración es el intervalo de tiempo trascurrido entre el inicio y término del estímulo acústico y puede ser considerado como corto o largo.

La sonoridad es la sensación perceptual relacionada a la intensidad de una onda sonora y se mide en fonios, de escala es arbitraria y nos permite calificar sonidos como fuertes o débiles. En esta escala, el cero corresponde al umbral de audición o amplitud mínima a la que un humano puede escuchar, y corresponde a 10^{-16} Watt/cm^2 a una frecuencia de 1000 Hz (Tippens, 2005). Dado que en el rango de intensidades que el oído humano puede detectar sin dolor hay grandes diferencias en el número de cifras empleadas en una escala lineal, es habitual utilizar una escala logarítmica de intensidad cuya unidad es el decibelio (dB). El nivel de presión sonora (SPL, del inglés "*sound pression level"*) es el indicador más utilizado de la intensidad acústica y se define como:

$$SPL(dB) = 10log_{10}\left[\left(\frac{p}{P_{ref}}\right)^2\right]$$

Donde, p es la presión sonora instantánea y P_{ref} es la presión de referencia, 20 µPa. Para ser detectados, los sonidos deben superar el umbral auditivo (0 dB o 20 µPa), mientras que el umbral de dolor va de los 130 a los 140 dB, o 20 pascales. Si bien la sonoridad depende principalmente de la intensidad de la señal acústica, también es afectada por sus frecuencias y duración.

2.1.1 Métodos descriptivos de las señales acústicas

El análisis descriptivo de señales sonoras puede realizarse con diferentes métodos dependiendo del objetivo. Por ejemplo, el oscilograma es una representación gráfica de las modulaciones de amplitud a lo largo del tiempo. Dado que la amplitud de un sonido es variable, si se unen las amplitudes de ciclos sucesivos se obtiene una señal continua que se denomina envolvente y que es característico para cada uno (Figura 1B).

El análisis de Fourier es la herramienta matemática empleada para transformar señales entre el dominio del tiempo y el dominio de la frecuencia. El teorema de Fourier demuestra que cualquier forma de onda periódica puede representarse como la suma de una serie de ondas sinusoidales, cuyos múltiplos enteros, conforman armónicos que pueden variar en amplitud. Esta descomposición simplifica el estudio de sonidos complejos ya que permite analizar cada componente de frecuencia de forma independiente. De manera similar, el espectro de poder nos permite determinar las frecuencias y respectivas amplitudes en un sonido, y se puede representar gráficamente como un espectrograma donde se observan las variaciones de energía en cada frecuencia a lo largo del tiempo (Figura 1C).

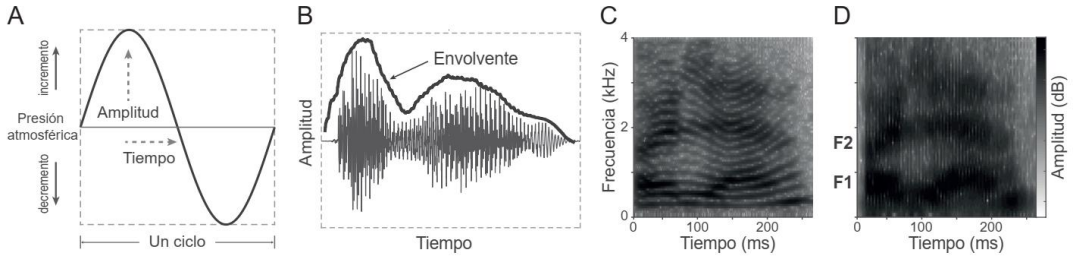


Figura 1. Propiedades físicas de los sonidos. **A**, Onda sinusoidal. En el eje horizontal se representa la duración y en el eje vertical los cambios de presión. **B**, Oscilograma o forma de onda de la palabra */hola/*. La línea gruesa es el envolvente que describe los cambios de amplitud en el tiempo. **C**, Espectrograma de la palabra mostrada en **B**. **D**, Las bandas horizontales más oscuras corresponden a los dos primeros formantes (F1 y F2) de la misma palabra.

2.2 Estructura y función del Sistema Auditivo (SA)

La percepción auditiva es fundamental para comunicarnos e identificar peligros potenciales del entorno. El sistema auditivo consiste en un conjunto de órganos que transforman la presión oscilatoria del aire (ondas sonoras) en impulsos eléctricos que viajan desde el oído hasta el cerebro a través de una serie de relevos neuronales. Es en la corteza cerebral donde las señales eléctricas se integran para formar representaciones perceptuales auditivas. A continuación, revisaremos los principales componentes estructurales y funcionales del SA.

2.2.1 Anatomía del SA

Para su estudio, el SA se ha dividido en dos subsistemas: periférico (oído) y central (del núcleo coclear a la corteza auditiva). La estructura principal del SA periférico es el oído, localizado en la base del hueso temporal del cráneo. El oído se encarga de descomponer los sonidos en sus componentes frecuenciales y los comunica en forma de potenciales eléctricos a las siguientes instancias de procesamiento. Para su estudio, generalmente es subdividido en tres partes: oído externo, medio e interno (Figura 2A).

El oído externo está conformado por el pabellón auricular y el conducto auditivo externo, su función principal es reunir las ondas sonoras y conducirlas hacia la membrana timpánica o tímpano. Dentro del oído medio hay tres huesecillos (martillo, yunque y estribo) que unidos amplifican las vibraciones del tímpano comunicándolas a una fina membrana llamada ventana oval que separa el oído medio del oído interno. Además, el oído medio se conecta con el tracto nasofaríngeo a través de la trompa de Eustaquio, de manera que exista una entrada de aire al oído medio para mantener la presión constante. Finalmente, el oído interno se conforma del órgano de descomposición de frecuencias, la cóclea, y del sistema de percepción de posición de la cabeza (vestíbulo y canales semicirculares), indispensable para mantener el equilibrio. La cóclea o caracol, es una estructura en forma de espiral que contiene una membrana que oscila en fase con las ondas sonoras para mover los cilios de las células de transducción acústica (células ciliadas, CC). El movimiento ciliar despolariza a las CC, convirtiendo así, la energía mecánica de los sonidos en impulsos eléctricos que son

llevados a través del nervio vestíbulo-coclear (VIII par craneal) al núcleo coclear del tallo cerebral, y de ahí a otras instancias acústicas dentro del sistema nervioso central.

Los somas de las neuronas sensoriales primarias están localizados en el ganglio espiral, mientras que sus axones periféricos hacen sinapsis con la región basal de las CC, las prolongaciones centrales proyectan a los núcleos cocleares (dorsal y ventral) situados en el bulbo raquídeo. Las células de estos núcleos envían proyecciones a través de tres estrías acústicas (Middlebrooks, 2015). La estría dorsal o de von Monakow está constituida en su mayoría por fibras de las neuronas del núcleo dorsal que decusan y ascienden por el fascículo del lemnisco lateral contralateral hasta el colículo inferior. La estría intermedia o de Held está conformada por fibras procedentes del núcleo posteroventral que atraviesan la línea media y suben por el lemnisco lateral hasta su núcleo. La estría ventral o fascículo del cuerpo trapezoidal, está formada principalmente por fibras provenientes del núcleo anteroventral, que envían sus axones a las neuronas de los núcleos del complejo olivar superior de ambos lados. Las fibras de estas neuronas olivares ascienden por el lemnisco lateral y contactan con las neuronas de su núcleo y con las del colículo inferior. La información auditiva continua su trayecto hacia los núcleos geniculados mediales del tálamo para finalmente llegar a la corteza auditiva situada en el lóbulo temporal (Figura 2B).

En los humanos, la corteza auditiva primaria (A1) está ubicada en el giro de Heschl (área 41 de Brodmann), mientras que la secundaria (cinturón y paracinturón, área 42 de Brodmann) está localizada en algunas regiones del giro de Heschl y se extiende hacia el plano temporal abarcando gran parte de la zona posterior del giro temporal superior (GTS, Figura 2C; Javitt y Sweet, 2015; Moerel et al., 2014). Con base en las características citoarquitectónicas, anatómicas y funcionales, la corteza auditiva de los monos rhesus (*Macaca mulatta*) ha sido dividida en tres regiones principales: núcleo (subdividido en 3 áreas: A1, rostral [R], rostrotemporal [RT]), cinturón (con 4 regiones mediales y 4 laterales) y el paracinturón con solo dos divisiones (rostral y caudal, Kaas y Hackett, 1999; Hackett, 2011). Un reciente estudio de conectividad determinó que las proyecciones cortico-corticales de la A1 ocurren principalmente a través de dos ejes: el eje mediolateral que va del núcleo hacia el cinturón y después al paracinturón, y el eje caudorostral que va de las regiones R y RT al área polar

rostrotemporal y posteriormente al polo temporal dorsal (Scott et al., 2017). Tanto la región rostral del cinturón como la del paracinturón proyectan a las áreas prefrontales rostral y orbital, mientras que sus regiones caudales envían proyecciones a la corteza periarcuato (área 8a) y la corteza prefrontal dorsolateral (área 46, Romanski et al., 1999; Bizley y Cohen, 2013).

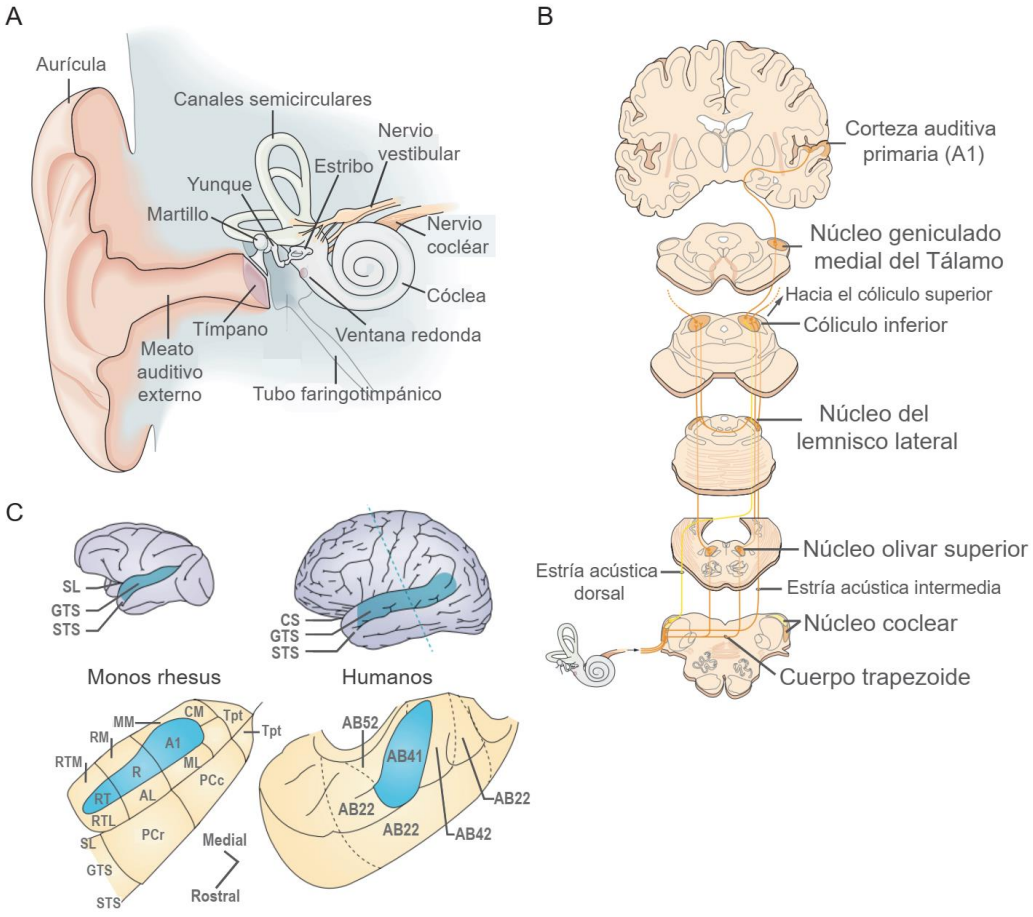


Figura 2. Anatomía del sistema auditivo. **A**, Principales estructuras del oído externo, medio e interno. **B**, La vía auditiva ascendente va desde la cóclea hasta la corteza auditiva del lóbulo temporal (Tomado y modificado de Kandel, 2013). **C**, Localización y divisiones de la corteza auditiva en monos rhesus y humanos. A1, corteza auditiva primaria; CS, cisura de Silvio; GTS, giro temporal superior; R, rostral; RT, rostrotemporal; SL, surco lateral; STS, surco temporal superior. Regiones del cinturón: AL, anterorateral; CM, caudomedial; ML, mediolateral; MM, mediomedial; RM, rostromedial; RTL, rostrotemporal lateral; RTM, rostrotemporal medial. PCc, paracinturón caudal; PCr, paracinturón rostral; Tpt, corteza de asociación del lóbulo temporal; AB, Área de Brodmann (Tomado y modificado de Bizley y Cohen, 2013; Javitt y Sweet, 2015).

2.2.2 Fisiología del SA

Las ondas sonoras amplificadas por los huesecillos del oído medio son transmitidas a la ventana oval, al inicio del oído interno, propagando las ondas sonoras a través del líquido contenido dentro de la cóclea. En la Figura 3A se presenta una sección trasversal de la cóclea donde se puede observar que al interior de este órgano hay dos membranas, la basilar y la vestibular o de Reissner. Las membranas dividen longitudinalmente a la cóclea para formar tres compartimentos llenos de líquido: la escala vestibular, del tímpano y la media o conducto coclear. El líquido (perilinfa) de las escalas vestibular y del tímpano tiene una composición similar a la del líquido cefalorraquídeo, mientras que, la endolinfa de la escala media presenta una composición iónica más parecida al líquido intracelular (alta concentración en potasio y baja en sodio). En el interior de la escala media, localizado por encima de la membrana basilar, se encuentra el órgano espiral o de Corti que contiene células ciliadas, de soporte y una membrana que lo rodea denominada membrana tectoria. Existen tres filas de CC externas y una fila de CC internas. Cuando las ondas en la perilinfa provocan el desplazamiento de las membranas de la cóclea, los estereocilios de las CC internas se inclinan, lo que inicia la entrada de iones potasio a través de canales mecanosensibles y la consiguiente despolarización (potencial receptor). Este cambio en el potencial de membrana de la CC incrementa la apertura de canales calcio, lo que permite la liberación del neurotransmisor y la resultante generación de potenciales de acción en la neurona aferente.

Tonotopía

La membrana basilar no es uniforme, presenta diferentes propiedades mecánicas a lo largo de su extensión, es decir, su región basal que es estrecha y gruesa responde a frecuencias altas mientras que en el ápex donde es ancha y de menor grosor responde a las frecuencias bajas (Fettiplace, 2020). Por lo tanto, la frecuencia de los sonidos es representada por el sitio de la cóclea donde se originan las neuronas que éste excita, mientras que su amplitud, por la intensidad de descarga y el número total de células activadas. La respuesta de las neuronas aferentes es entonces una función de la intensidad y de las magnitudes relativas de las diferentes frecuencias que componen un sonido. Esta organización espacial de frecuencias es llamada tonotopía (o cocleotopía, Figura 3B) y se mantiene a lo largo de la vía ascendente hasta la corteza auditiva tanto

en monos (Kajikawa et al., 2015; Brewer y Barton, 2016) como en humanos (Saenz y Langers, 2014; Leaver y Rauschecker, 2016). Además de la representación de frecuencias, también se ha reportado que las neuronas de A1 codifican otras características de los estímulos sonoros, como las que a continuación se explican.

Periodotopía

El contenido temporal del sonido también es parte integral del procesamiento auditivo (Brewer y Barton, 2016). La periodotopía hace referencia a la organización topográfica de neuronas sensibles a la periodicidad. Ésta organización ha sido reportada en la corteza auditiva de humanos, gatos y monos (Langner et al., 1997; Barton et al., 2012; Herdener et al., 2013; Langner et al., 2009; Baumann et al., 2015) (Figura 3C). Se encontraron respuestas sintonizadas a las modulaciones periódicas de la amplitud de los sonidos (envolvente temporal) en células organizadas de manera ortogonal a las respuestas sintonizadas a las frecuencias. Sin embargo, otros reportes descartan la existencia de una organización periodotópica (Leaver y Rauschecker, 2016).

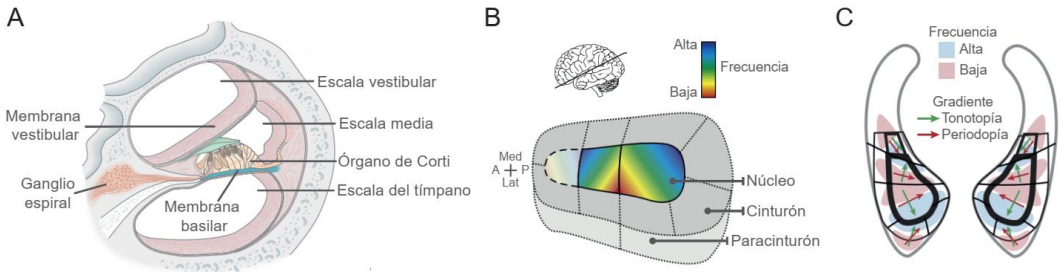


Figura 3. Mecanismos fisiológicos del sistema auditivo. **A**, Sección trasversal de la cóclea. Este corte nos permite identificar los diferentes elementos involucrados en el proceso de transducción mecanoeléctrica (Tomado y modificado de Kandel, 2013). **B**, Tonotopía de la corteza auditiva primaria del mono (Tomado y modificado de Saenz y Langers, 2014). **C**, Comparación entre los gradientes de tonotopía y periodotopía en A1 (Tomado y modificado de Baumann et al. 2015).

Percepción del tono y timbre

La percepción del tono de los sonidos harmónicos complejos es esencial para su identificación y segregación, particularmente en el contexto de la comunicación y la música (Oxenham, 2018; Xiaoqin Wang, 2018). El tono es el atributo subjetivo asociado a la F0, está determinado por la regularidad temporal (periodicidad) y la tasa de repetición promedio de la forma de onda. Mediante estudios de imagenología en humanos se ha identificado una pequeña área de la corteza auditiva secundaria selectiva

al tono (Patterson et al., 2002; Penagos et al., 2004), mientras que en los monos se ha reportado que las neuronas de la región del borde anterolateral de la A1 son selectivas al tono (Bendor y Wang, 2005, 2006). Por lo tanto, la percepción del tono podría existir también en primates no humanos (Tomlinson y Schwarz, 1988; Song et al., 2016). Sin embargo, dado que en los estudios previos han usado principalmente sonidos harmónicos artificiales, aún no es claro cómo las neuronas de la corteza auditiva codifican el tono de sonidos complejos naturales (p. ej., vocalizaciones).

El timbre es el atributo perceptual que nos permite distinguir sonidos con el mismo tono, volumen y duración (Town y Bizley, 2013). A nivel fonético, el timbre es crucial para determinar la identidad de las vocales y consonantes. Muchas especies de animales, han mostrado tener la capacidad de discriminar vocales: primates (Kojima y Kiritani, 1989; Sinnott et al., 1997), aves (Hienz et al., 1981), hurones (Bizley et al., 2013) y roedores (Saunders y Wehr, 2019). Los registros electrofisiológicos en modelos animales y de imagenología funcional en humanos, han permitido establecer que la actividad neuronal subyacente a la percepción del timbre está extensamente distribuida a través del núcleo y el cinturón de la corteza auditiva (Warren et al., 2005; Bizley y Cohen, 2013).

Procesamiento cortical jerárquico

De forma análoga al funcionamiento del sistema visual, se ha sugerido que en los primates el procesamiento auditivo está organizado a lo largo de dos diferentes vías paralelas e independientes: la vía dorsal y la vía ventral (Kaas y Hackett, 1999; Rauschecker y Scott, 2009; Ahveninen et al., 2013). En los monos rhesus, la vía dorsal inicia en la región caudal del cinturón, proyecta hacia la corteza parietal y termina en la región dorsolateral de la corteza prefrontal; se encarga de procesar la información espacial y de la preparación de acciones motoras relacionadas con los sonidos. Por otra parte, la vía ventral está involucrada en el reconocimiento de los estímulos acústicos, inicia en el núcleo (A1 y R), proyecta a la región anterolateral del cinturón y continúa a lo largo del giro temporal superior (paracinturón), y de ahí se envían proyecciones a la corteza prefrontal ventrolateral (Romanski et al., 1999; Bizley y Cohen, 2013; Cohen et al., 2016). Durante las últimas dos décadas se han recabado evidencias anatómicas, electrofisiológicas y de imagenología que sugieren que este modelo de procesamiento

cortical, tanto en humanos como macacos, es plausible (Figura 4, Alain et al., 2001; Tian et al., 2001; Woods et al., 2006; Rauschecker y Romanski, 2011).

El procesamiento sonoro en diferentes regiones corticales y subcorticales ha sido investigado a través de algunos estudios neurofisiológicos en monos. Por ejemplo, se encontró que las neuronas de A1 responden principalmente a tonos puros, mientras que las neuronas del cinturón lateral responden también a sonidos complejos, como son los ruidos paso-banda y las vocalizaciones (Rauschecker et al., 1995; Rauschecker y Tian, 2000). Tian y colaboradores (2001) estudiaron las respuestas neuronales del cinturón lateral de macacos para diversos tipos de vocalizaciones conespecíficas presentadas en siete ubicaciones distintas. Encontraron que las neuronas de la región caudolateral se activan preferentemente a ubicaciones espaciales sin importar el tipo de vocalización. En cambio, las neuronas de la región anterolateral fueron más selectivas a las vocalizaciones y no tanto a su ubicación espacial. Estas evidencias llevaron a pensar que el cinturón representa vocalizaciones. Sin embargo, registros realizados posteriormente mostraron que, pese a que los macacos son capaces de diferenciar entre vocalizaciones presentadas de manera normal, y presentadas temporalmente invertidas (Ghazanfar et al., 2001), las neuronas del cinturón no son capaces de hacerlo, ya que responden de manera similar para ambos tipos de estímulos (Recanzone, 2008).

Los registros neuronales del paracinturón son escasos, debido en parte, a las dificultades técnicas que implica registrar su ubicación anatómica, y en parte porque siendo un área involucrada en procesos cognitivos, encontrar señales de interés de registros unitarios requiere monos entrenados, lo cual ha sido prácticamente inexistente. Por consiguiente, solo se han reportado algunas aproximaciones experimentales en condiciones no activas, y carente de un estudio sistemático de la participación del paracinturón en la codificación de estímulos acústicos (Camalier et al., 2012; Kajikawa et al., 2015; Tani et al., 2018; Heelan et al., 2019).

En la corteza prefrontal ventrolateral (CPFvl, áreas 12 y 45) se han descrito neuronas responsivas a vocalizaciones que son similares a las descritas en el cinturón, con quien está conectada recíprocamente (Romanski y Goldman-Rakic, 2002; Romanski y Averbeck, 2009; Russ et al., 2008). Sin embargo, mientras que algunos reportan que la actividad neuronal de la CPF es similar para los sonidos con propiedades acústicas

parecidas (Romanski et al., 2005), otros sugieren que codifica categorías semánticas más que puramente acústicas (Gifford et al., 2005; Cohen et al., 2007).

Los OA son considerados la unidad perceptual fundamental de la audición y aunque no existe una definición comúnmente aceptada, son el resultado computacional de la capacidad que tiene el sistema auditivo para detectar, extraer, segregar, y agrupar regularidades espectro-temporales del medio ambiente (Griffiths y Warren, 2004; Bizley y Cohen, 2013). Basados en el modelo de procesamiento jerárquico, suponemos que debiera existir un cambio paulatino en la codificación acústica a lo largo de la vía ventral, hasta formar OA; comenzando con la representación de las propiedades físicas de los sonidos, luego de categorías acústicas, y finalmente representaciones semánticas y asociaciones conductuales en las regiones corticales más superiores de la vía.

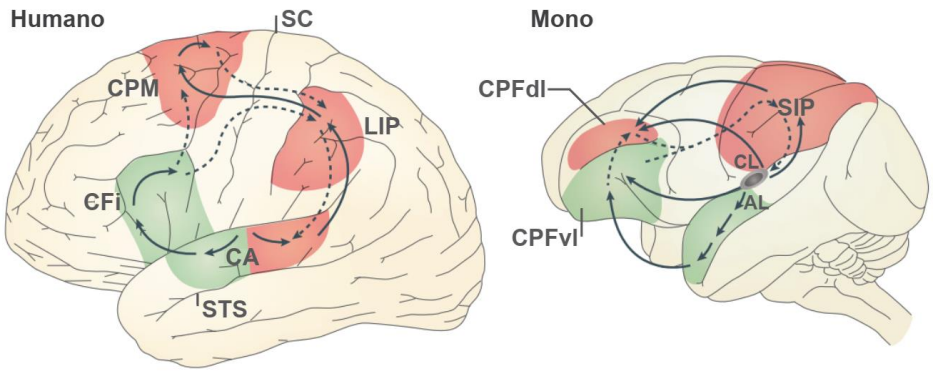


Figura 4. Vías corticales de procesamiento auditivo. Conexiones principales de la vía dorsal (rojo) y ventral (verde), en humanos, y monos. AL, región anterolateral del cinturón; CA, corteza auditiva; CL, región caudolateral del cinturón; CFi, corteza frontal inferior; CPFdl, corteza prefrontal dorsolateral; CPFvl, corteza prefrontal ventrolateral; CPM, corteza premotora; LIP, lóbulo intraparietal; SC, surco central; SIP, surco intraparietal; STS, surco temporal superior (Tomado y modificado de Bizley y Cohen, 2013).

2.3 Reconocimiento de los objetos auditivos

Los primates utilizan las vocalizaciones principalmente para comunicarse con los miembros de su especie. Estos sonidos tienen propiedades espectrotemporales variables que se ha sugerido proporcionan información del sexo, tamaño corporal, estatus social y reproductivo del emisor (Fitch, 1997; Ghazanfar et al., 2007; Ey et al., 2007; Honorof y Whalen, 2010; Bowling et al., 2017). En los monos, la clase de vocalización emitida puede indicar diferentes tipos de comida, o contingencias del

entorno como la presencia de un depredador (Hauser y Marler, 1993; Hauser, 1998; Seyfarth et al., 1980). También se ha propuesto que los monos reconocen vocalmente a los miembros de su tropa, e incluso responden a señales de alarma provenientes de otras especies de primates (Rendall et al., 1996, 1998; Zuberbühler, 2000).

Algunas regiones del lóbulo temporal de humanos han mostrado más preferencia por vocalizaciones conespecíficas, que a otros tipos de sonidos (Belin et al., 2000; Fecteau et al., 2004; Kriegstein y Giraud, 2004; Lewis et al., 2009; Leaver y Rauschecker, 2010), ocurriendo lo mismo en monos (Wang et al., 1995; Petkov et al., 2008; Perrodin et al., 2011, 2015). Específicamente, se han reportado regiones preferentes a voces en la porción media de la corteza temporal superior, del GTS y del surco temporal superior (STS, Belin et al., 2018). El polo temporal también responde a vocalizaciones y sus neuronas son principalmente selectivas a voces individuales que a categorías de vocalizaciones (Petkov et al., 2008; Perrodin et al., 2011). Sin embargo, es importante considerar que la mayoría de estos estudios fueron realizados utilizando resonancia magnética funcional y mediante protocolos experimentales de escucha pasiva.

Tanto el habla humana como las vocalizaciones de primates no humanos (PNH) se producen por los movimientos coordinados de los pulmones, la laringe y el tracto vocal supralaríngeo (Ghazanfar y Rendall, 2008). Las cuerdas vocales determinan la F0 y los correspondientes harmónicos, mientras que los formantes de la envolvente espectral dependen del tamaño y forma del tracto vocal (Ey et al., 2007). Sin embargo, cómo las diferentes propiedades acústicas contribuyen al reconocimiento de sonidos no ha sido completamente determinado.

Estudios pioneros en humanos mostraron la relevante participación del primer y segundo formante (F1 y F2, respectivamente) en la discriminación de vocales (Peterson y Barney, 1952; Remez et al., 1981; Lieberman y Blumstein, 1988; Hillenbrand et al., 1995). Existen evidencias que sugieren que los babuinos (*Papio anubi*), monos vervet (*Chlorocebus pygerythrus*) y macacos japoneses (*Macaca fuscata*) son capaces de discriminar vocales artificiales a partir de los formantes (Hienz y Brady, 1988; Hienz et al., 2004; Sinnott, 1989; Sinnott y Kreiter, 1991). Recientemente también se reportó que los monos rhesus (*Macaca mulatta*) perciben de manera espontánea cambios en los

formantes (Fitch y Fritz, 2006). Sin embargo, hasta el presente trabajo, la contribución de los formantes durante el reconocimiento activo de sonidos complejos como las palabras y otras vocalizaciones en PNH no había sido estudiada.

2.3.1 Invarianza perceptual

Otra característica fundamental de la percepción auditiva es la constancia o invarianza perceptual, que es la habilidad de reconocer y categorizar sonidos de cierto rango de variación física (Nusbaum y Magnuson, 1997; Town y Bizley, 2013). Por ejemplo, enunciados idénticos muestran una considerable variabilidad acústica entre emisores, y pese a ello, los oyentes invariantemente logran reconocer las palabras (Peterson y Barney, 1952; Smith et al., 2005; Johnson y Sjerps, 2021).

La invarianza perceptual requiere representaciones neurales selectivas a la identidad, y que toleren variaciones físicas. Actualmente conocemos algunas bases psicofísicas y neurales de la constancia perceptual de objetos visuales (p. ej., caras) durante modificaciones de parámetros como son tamaño, iluminación y orientación (Rolls y Baylis, 1986; Logothetis y Pauls, 1995; Leopold et al., 2006; Roy et al., 2014). En el sistema auditivo también se han descrito respuestas invariantes a sonidos, tanto en corteza auditiva primaria de pinzones (Billimoria et al., 2008), titíes (Sadagopan y Wang, 2008) y hurones (Mesgarani et al., 2014), como en áreas secundarias de estorninos (Meliza y Margoliash, 2012), pinzones (Moore et al., 2013), ratas (Carruthers et al., 2015) y humanos (Kell y McDermott, 2019). Sin embargo, esta habilidad no ha sido demostrada en PNH discriminando sonidos heteroespecíficos, por lo que decidimos evaluarla mediante nuestro modelo experimental de discriminación auditiva.

3. PLANTEAMIENTO DEL PROBLEMA

Una de las funciones principales del sistema auditivo es transformar las señales acústicas en representaciones perceptuales abstractas que determinen la conducta. Sin embargo, las señales neuronales responsables de la percepción de OA es un tema pendiente en neurociencias. Para abordar el problema es necesario estudiar los correlatos psicofísicos y neuronales de la discriminación de OA, usando un modelo de discriminación auditiva en monos rhesus. El modelo nos permitiría evaluar la capacidad de los macacos para aprender diversos OA, y de discriminarlos a partir de mezclas de OA que aquí llamaremos "*morphs*". De manera importante, nos permitiría estudiar la contribución de las frecuencias formantes en su identificación. Finalmente, podríamos estudiar el procesamiento neural de la percepción invariante sonidos en áreas corticales como el área motora suplementaria (AMS), durante la transformación de OA en comandos pre-motores.

4. HIPÓTESIS

Dado que los formantes 1 y 2 son los grupos de frecuencias de más energía en los sonidos de comunicación como las palabras, entonces también serán suficientes para que los monos rhesus discriminen objetos auditivos.

5. OBJETIVOS

1. Entrenar un par de monos rhesus (*Macaca mulatta*) en una tarea de discriminación de OA.
2. Realizar pruebas psicométricas de los monos rhesus mientras se presentan sonidos tipo "*morphs*".
3. Determinar si los formantes 1 y 2 bastan para discriminar OA.
4. Establecer si los monos rhesus reconocen de manera invariante OA generados por distintos emisores.
5. Estudiar la actividad extracelular de neuronas del área motora suplementaria de los monos rhesus durante la discriminación de OA.

6. MÉTODOS Y RESULTADOS

Los métodos, resultados y parte de la discusión se encuentran en los artículos, Melchor *et al.,* 2020 y 2021, que se anexan a continuación:

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

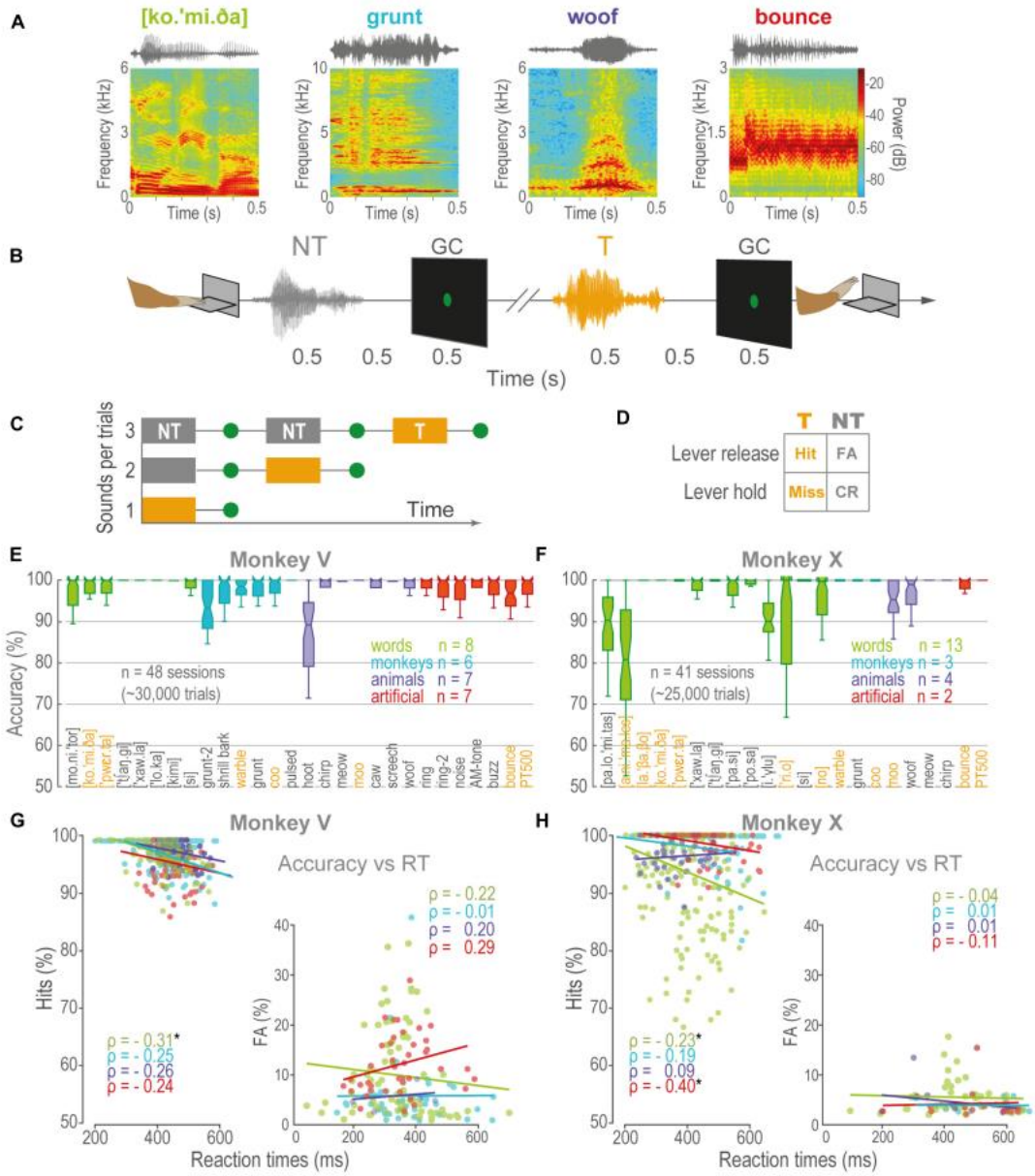


FIGURE 1 | Acoustic discrimination task and performance. **(A)** Spectrograms of four acoustic categories: Green, words, Cyan, conspecific monkey vocalizations, Purple, vocalizations of animals, Red, artificial sounds. **(B)** The sequence of events in a trial: First, the monkey pressed a lever to start. 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

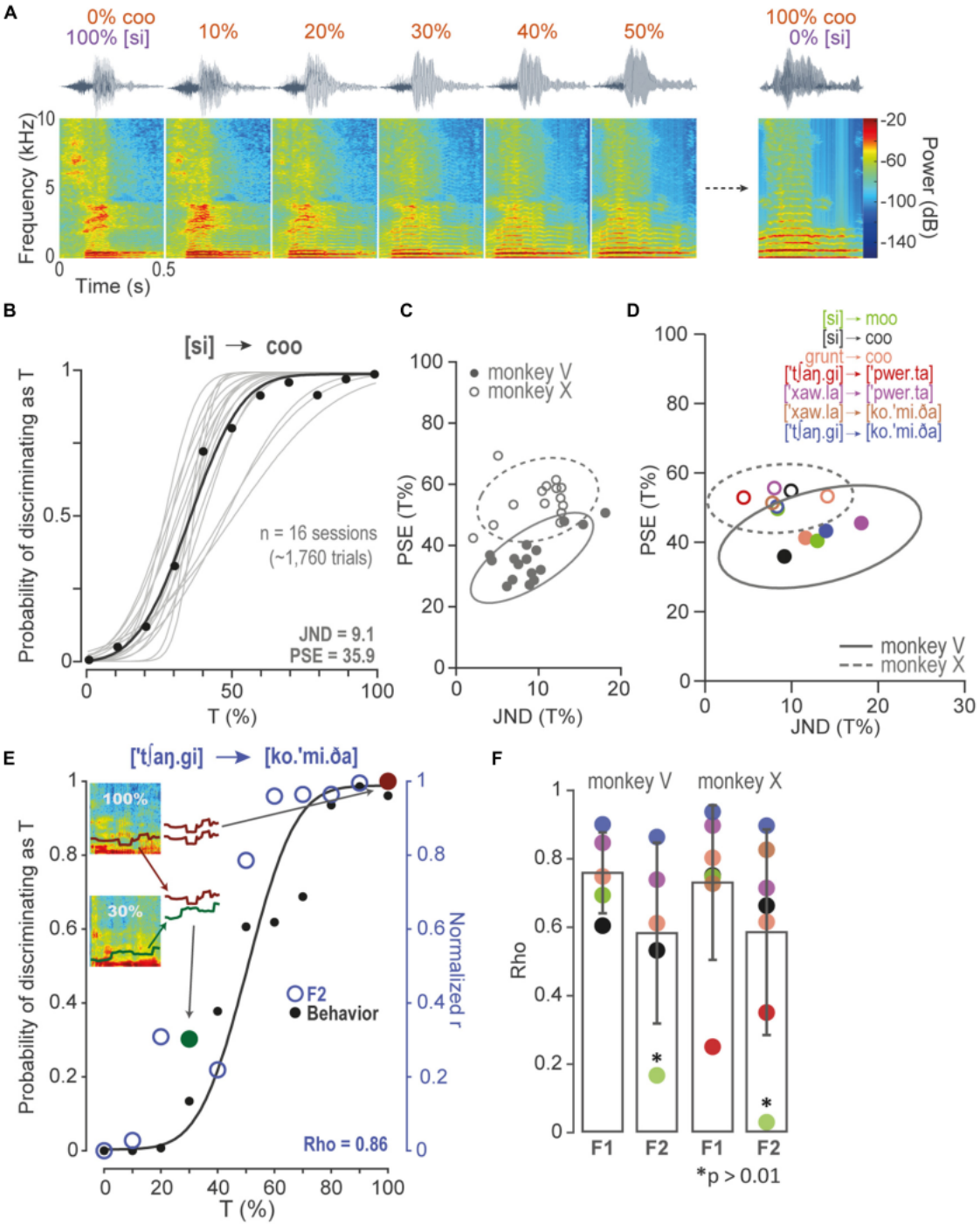


FIGURE 2 | Discrimination of morphs and correlations between performance and formants. **(A)** Some spectrograms of the [si] to coo morph-line continua. The NT [si], i.e., the Spanish word for "yes," morphed gradually in steps of 10 to 100% T "coo" monkey call. **(B)** PF of the probability of recognizing 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.

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

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.

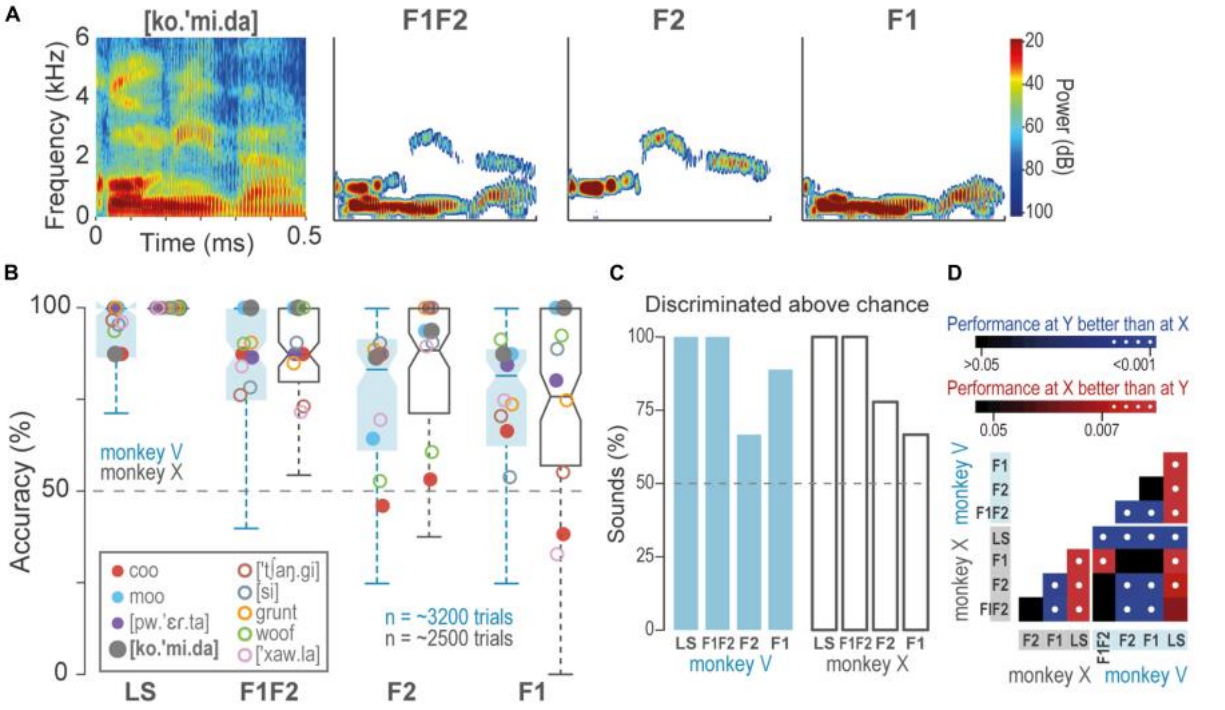


FIGURE 3 | Discrimination of formant-pass sounds. **(A)** Spectrograms of T [ko.'mi.da] and F1F2, F2, and F1-pass sounds. **(B)** Boxplots of accuracy during the discrimination of formant-pass filters compared to the discrimination of the learned sounds (LS). Dash line, performance at chance = 50%. Closed 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

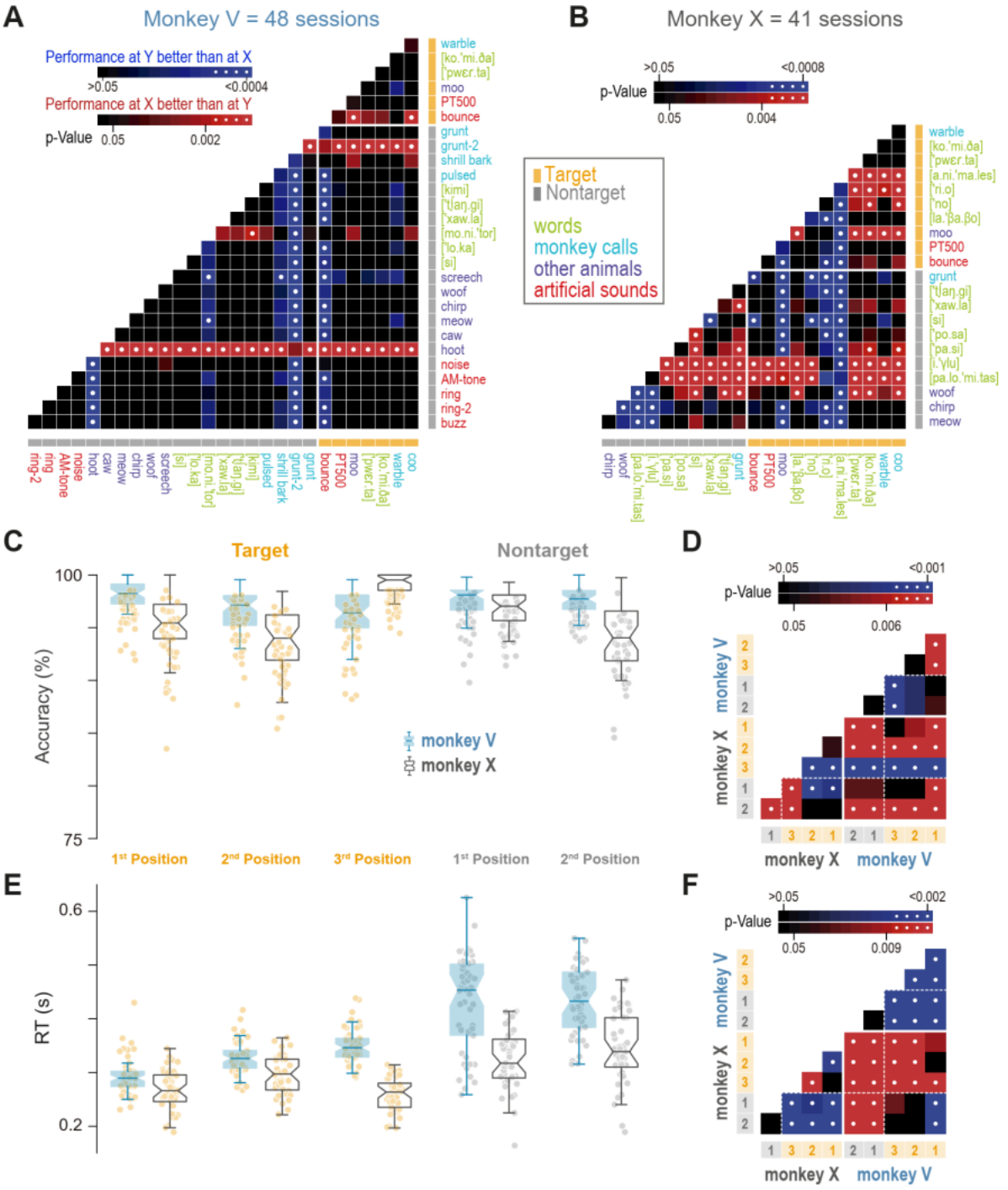
Wright, A. A. (2007). An experimental analysis of memory processing. *J. Exp. Anal. Behav.* 88, 405–433. doi: 10.1901/jeab.2007.88-405

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

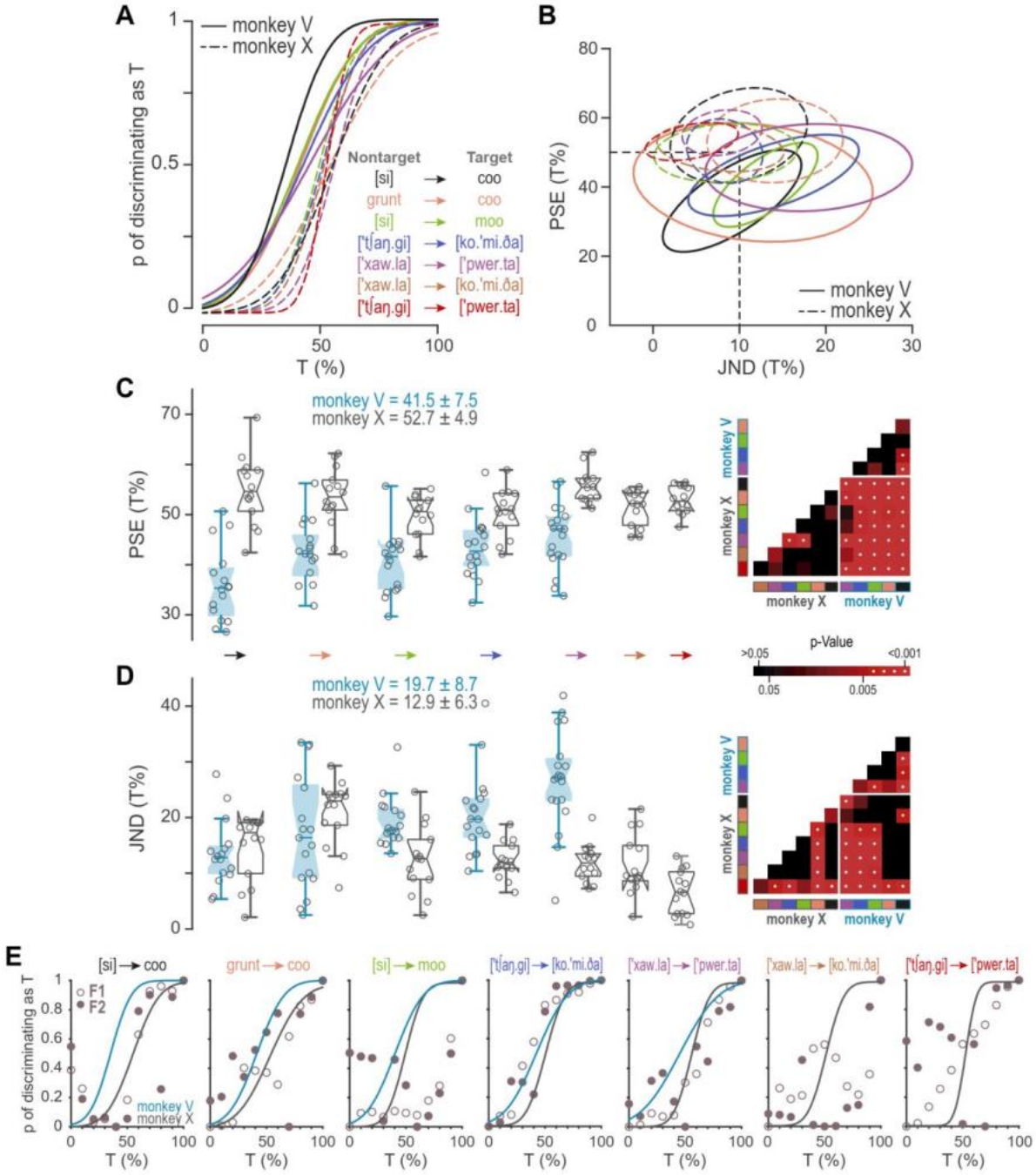
Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* 10:3770. doi: 10.1038/s41467-019-11786-6

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.



SUPPLEMENTARY FIGURE 1. Accuracy and RT comparisons across acoustic categories and monkeys. (A) Confusion matrix of p-values of multiple pairwise comparisons of monkey V accuracy during the discrimination of different acoustic categories. The intensity at the upper color bar is proportional to the p-value. White circles indicate significant differences. **(B)** Same as **A**, but for monkey X. **(C)** Monkeys' accuracy during the discrimination of T and NT presented in different positions. Each dot represents the mean performance in a session. Box plot edges correspond to the 25th and 75th percentiles, the central line to the median, and vertical lines ± 2.7 SD. **(D)** Confusion matrix of p-values of multiple comparisons of performance between monkeys, categories, and position. **(E)** Same as in **C** but for RT. **(F)** p-values of multiple pairwise comparisons of RT, same labels as panel **D**.



SUPPLEMENTARY FIGURE 2. Psychometric analysis of the discrimination of sounds. **(A)** PFs of both monkeys discriminating in different morphing sets. Each PF corresponds to the probability of discriminating as T the morphs at the abscissas. **(B)** 2D-Gaussian fits of PSE as a function of JND of both monkeys performing in all morphing sets. **(C)** Boxplots of both monkeys' PSE of all morphing sets. Box edges and lines follow the same convention as in Supplementary Figure **1C, E**. Same color code as **A.** Right panel, p-values of multiple pairwise comparisons of PSE. Color gradients indicate p-values, white dots, significant differences. **(D)** Same as in **C**, but for JND. **(E)** PFs and the distribution of the normalized Pearson's r correlations for formants along the morph-line continua.

Neuronal Correlates of the Perceptual Invariance of Words and Other Sounds in the Supplementary Motor Area of Macaques

Jonathan Melchor,[1] Isaac Morán,[1] José Vergara,[2] Tonatiuh Figueroa,[1] Javier Perez-Orive,[3] and Luis Lemus[1]*

[1] Department of Cognitive Neuroscience, Institute of Cell Physiology, Universidad Nacional Autónoma de México (UNAM). 04510. Mexico City, Mexico.

[2] Department of Neuroscience, Baylor College of Medicine, Houston, Texas.

[3] Luis Guillermo Ibarra Ibarra National Rehabilitation Institute. Mexico City, Mexico.

* Corresponding author

lemus@ifc.unam.mx

Telephone: (+52) 55 5622 5675

KEYWORDS: Perceptual constancy, macaques, acoustic recognition, frontal cortex, psychophysics.

Abstract

How does the brain identify words in spite of the diversity of speakers? It has been suggested that the words stored in circuits of the temporal and parietal lobes connect to frontal cortices to orchestrate behaviors such as language. Notably, recent functional MRI studies have shown that the supplementary motor area (SMA) becomes activated during the invariant perception of words pronounced by different speakers and therefore plays a role in the phenomenon of invariant recognition. We recorded SMA neural activity in two rhesus monkeys trained to recognize numerous sounds including multisyllabic words, and we tested whether the monkeys were capable of recognizing novel versions. Our results show that the animals proficiently recognized novel versions while SMA activity was correlated with their invariant perception at the single and the population level. Our findings suggest that perceptual invariance comprises premotor representations in regions other than those that are commonly known such as speech areas. We conclude that the recognition of complex sounds such as words is assisted by behavioral programs coded in premotor cortical areas.

1 Introduction

We typically learn our first words from the voice of our mother. Notably, after that, we are capable of recognizing the same words spoken by multiple speakers. Rhesus monkeys and chimpanzees acknowledge vocalizations for foods of different values and from various troop members [1,2]. Similarly, in songbirds, ferrets, and mice, the ability to identify versions of learned sounds (vLS) has been reported [3–5]. Although this ability is vital for communication, the neural basis for the invariant recognition of sounds is not well understood.

Current knowledge of how neuronal activity produces the invariant perception of vocal sounds is also scant [3,4,6,7]. For instance, "voice-sensitive" and "vocalization-sensitive" cortical areas have been identified using neurophysiological and imaging studies in humans and nonhuman primates [8–12]. Additionally, ferrets generalizing vowel identities (/u/ and /ϵ/) throughout variations of fundamental frequencies, sound levels, and locations has been correlated with the activity of the auditory cortex [4]. In zebra finches, reports suggest invariant responses for different categories of conspecific vocalizations [3]. A recent study found that mice are able to discriminate the consonants (/g/ and /b/) and recognize them when combined with different vowels or when emitted by different speakers [5]. However, perceptual phenomena correlate with the activity of frontal cortices rather than with primary sensory cortices.

For instance, auditory processing reaches the frontal lobe via the ventral auditory stream [13,14], which, in humans, groups syllables into words [15,16]. The prefrontal cortex, in turn, projects to premotor cortices in order to orchestrate behaviors [17]. It is noteworthy that in macaques, the ventral premotor cortex is homologous to Broca's human speech production area, and participates in acoustic discrimination [18]. Additionally, the SMA, which is a premotor cortex and participates in voluntary movement control [19–21], working memory, and decision making [22–25], has recently been involved in acoustic imagery of humans [17]. Therefore, the SMA is a likely candidate area for linking sounds and the invariant perception thereof [26].

Here we present the results of extracellular recordings of SMA neurons in rhesus monkeys trained to recognize sounds. During the task, the macaques reported the appearance of a target sound (T) presented randomly after 0 to 2 non-target sounds (nT). We evaluated the neuronal responses to various vLS of T and nT to which the monkeys had had no previous exposure (e.g. a sound uttered by different emitters). We observed that the monkeys perceived vLS as the originally learned sounds (LS), and that the SMA's neuronal responses were

correlated with the animals' invariant perceptions. Our results suggest that the SMA associates incoming unexperienced sounds with the closest known category.

2 Materials and Methods

2.1 Ethics statement

All experimental protocols were performed in compliance with the Official Mexican Standard Recommendations for the Care and Use of Laboratory Animals (NOM-062-ZOO-1999) and approved by the Internal Committee for the Use and Care of Laboratory Animals at the Institute of Cell Physiology, UNAM (CICUAL; LLS80-16).

2.2 Animals and experimental setup

Two adult rhesus macaques (*Macaca mulatta*; one 13-kg, ten-year-old male, and one 6-kg, ten-year-old female) participated in this study. The animals inhabited an enriched facility that allowed physical interactions with other monkeys. The night before experimental sessions, the monkeys were restricted to water only. However, after the sessions, they received water *ad libitum*. Monkeys performed ~1000 trials during three-hour sessions, one session per day, four to five days per week. Experiments took place in a soundproof booth in which a macaque sat on a primate chair, 60 cm away from a 21" LCD color monitor (1920 x 1080 resolution, 60 Hz refresh rate). A Yamaha MSP5 speaker (50 Hz–40 kHz frequency range) set 15 cm above and behind the monitor delivered acoustic stimuli at ~65 dB SPL measured at the monkey's ear level. Additionally, a Logitech® Z120 speaker was positioned directly below the Yamaha speaker in order to render white background noise at ~55 dB SPL. Finally, a metal spring-lever at the monkeys' waist level captured their responses.

2.3 Behavioral task

We trained two rhesus macaques in an acoustic recognition task that consisted of categorizing sounds presented in trials of 1 to 3 sounds as either target (T) or non-target (nT) sounds. In each trial, a 3°-aperture gray circle appeared in the center of the monitor, after which the monkey pressed the lever. After a variable delay of 0.5 to 1 s, a 0.5 s sound was heard, followed by a 0.5 s delay and a 0.5 s period when the gray circle turned green. If the sound was a T, the macaque released the lever within a 0.8 s response window commencing with the green cue. However, if the sound was an nT, the monkey kept the lever down and waited for the next sound. We presented the T at the 1st, 2nd, or 3rd position with equal probability, i.e. $p(T \mid position) = 1/3$. Releases before a T resulted in false alarms, leading to different new trials. We required that the monkeys perform above an 80% hit rate before electrophysiological recordings were made. The task's programming was in LabVIEW 2014 (64-bit SP1, National Instruments®).

2.4 Acoustic stimuli

The sounds were recorded in our laboratory or downloaded from free online libraries (https://freesound.org/). They consisted of words, monkey vocalizations, other animal vocalizations, and artificial sounds (n = 37, Extended Data Table 1-1). The sample rate was 44.1 kHz; cutoff frequencies, 100 Hz to 20 kHz, compressed or extended to 0.5 s, and equalized to the same root-mean-square (RMS) amplitude value (Adobe Audition® version 6.0). We selected 4 T and 5 nT from the pool of learned sounds in order to perform statistical repetitions (ten times each sound in a set). The phonetic nomenclature of Spanish words originated from the automatic phonetic transcriptionist created by Xavier López Morrás (http://aucel.com/pln/transbase.html). To assess the monkeys' perceptual invariance to unheard sounds, we presented them with 5 versions of each learned sound.

2.5 Neuronal recordings

We performed extracellular recordings of single SMA neurons in monkey 2. We positioned a 20 mm diameter recording chamber above the stereotaxic coordinates of the SMA (Paxinos G 2009) compared to monkey 2 MRI (IA = 27 mm, left hemisphere: lateral 6 mm). We used an array of 5 independently movable microelectrodes (1-3 MΩ; Thomas Recording®) inserted at different locations in each session. We sampled extracellular membrane potentials at 40 kHz and performed offline sorting using Plexon sorter software (Plexon®).

2.6 Statistical analysis

To assess the monkeys' capacity to identify each sound, we computed the probability of hits [p(release | T), T] and correct rejections [CR = p(no-release | nT), nT] and we performed a one-sample Wilcoxon signed-rank test. To verify the difference between LS and vLS, we calculated the number of vLS that were different from LS (pairwise multiple Wilcoxon rank-sum comparison tests, Benjamini-Hochberg FDR correction, q value = 0.01).

To evaluate what aspects of the task were coded by SMA neurons, e.g. T and nT categories or acoustic identities, we calculated *F*-statistics in a one-way ANOVA of the T vs. nT and each sound. We calculated the mean firing rate in 200-ms windows in steps of 20 ms. In order to avoid biases in the *F*-statistic due to the number of trials, we only included the same number in all classes. We repeated the analysis 1000 times to create an *F*-statistic distribution. We compared the actual distribution against a random *F*-statistic distribution obtained by shuffling the stimulus labels 1000 times to determine the significant *F*-statistic bins. A neuron was significant if confidence intervals (95%) did not overlap in at least one time-bin.

We performed demixed principal component analysis (dPCA) [27] in order to observe the effects of the task components and of acoustic identities on the population responses. During the supervised stage, dPCA decomposes the neural activity for each variable in covariance matrices of different marginalizations. Then, for each marginalization matrix, unsupervised analysis is similar to a PCA. We marginalized the population activity for T, nT, and time, from the beginning of the sounds to the monkeys' responses, in 200 ms windows every 20 ms. dPCA differentiates the matrices from the decoder (D) and the encoder (F) for each parameter by minimizing the loss function (1):

$$L_\phi = \|\underline{X}_\phi - F_\phi D_\phi \underline{X}\| \quad (1)$$

Where ϕ denotes the marginalization for each parameter and X, the mean-centered population matrix. Each component represents an amount of firing rate variance. The axes obtained by the decoder and the encoder allow data reduction in a few features, capturing the most significant variance of each parameter.

To determine statistically decoded bins, we organized the data into training and testing sets. The testing sets consisted of random inclusions of 1 trial from each category (T and nT). The training set was the mean of the remaining tests of each class. We performed dPCA with the training set in order to obtain the decoding axes. Then, we projected the testing set on these axes and classified them according to the closest category mean (T or nT). We repeated this procedure 1000 times and measured the proportion of correct classification. We shuffled 100 times for significance.

To test whether the neuronal responses were correlated with behavior, we computed Spearman correlations between the monkeys' performance and neuronal firing rates of LS and vLS, in 200 ms windows, every 20 ms. finally, we calculated a Pearson correlation between the resulting mean Spearman's rho of LS and vLS in order to assess for similarities.

3 Results

3.1 Rhesus monkeys learned numerous complex sounds including multisyllabic words

We trained two rhesus monkeys to categorize numerous sounds as either T or nT during an acoustic recognition task (Fig. 1*A*; see Material and methods). The monkeys released a lever for a T in the first, second, or third position of a sequence (Fig. 1*B*). Trials with one, two, or three stimuli had an equal probability of appearing in a session (p = 1/3). To assess the monkeys' capacity to identify each sound, we computed the probability of hits p (release | T) and correct rejections CR = p (no-release | nT) (Fig. 1*B*, inset). Overall, each monkey performed above chance for more than 20 sounds [Hits median: monkey1 = 0.97, monkey2 = 0.96; CR median: monkey1 = 0.98, monkey2 = 0.96; one-sample Wilcoxon signed-rank test higher than 0.75: Z (monkey1_Hits) = 10.41, Z (monkey1_CR) = 8.51, Z (monkey2_Hits) = 9.63, Z (monkey2_CR) = 7.87; all p-values < 0.001]. Overall, performance for all T and nT categories ranked above 85% for both monkeys. The monkeys learned 36 sounds (T = 11, nT = 25). The sounds consisted of seven artificial sounds (T = 2, nT = 5), six monkey vocalizations (T = 2, nT = 4), sixteen human words (T = 6, nT = 10), and seven other animal vocalizations (T = 1, nT = 6; see [28], for further behavioral details).

3.2 Rhesus monkeys recognized novel versions of the learned sounds

To test for invariant recognition of sounds, we presented the monkeys with several vLS. Fig. 1*C* shows the T ['pwɛɾ.ta] and the nT 'grunt' spectrograms, together with three vLS. To reduce the possibility of learning the vLS, we presented each intercalated with LS no more than fifteen times per session. Fig. 1-1 shows how the recognition of vLS required few trials. Ultimately, the monkeys recognized significantly above chance: 84.4% of vLS (n = 45; Wilcoxon signed-rank test; Benjamini-Hochberg FDR corrected q-value = 0.01). This result means that the animals succeeded in categorizing vLS as either T or nT. However, to verify whether the monkeys perceived vLS as an invariant of LS, we looked for performance differences between each of the nine LS and their corresponding five vLS (Fig. 1*D*). Overall, the monkeys recognized 71% vLS of LS (pairwise multiple Wilcoxon rank-sum comparison tests, Benjamini-Hochberg FDR correction, q value = 0.01). It is noteworthy that the 'woof' category was highly biased towards false alarms since the monkeys only recognized 1 out of 5 vLS. Overall, with the only few exceptions such as 'woof' and ['xaw.la], the monkeys invariantly perceived most vLS as LS.

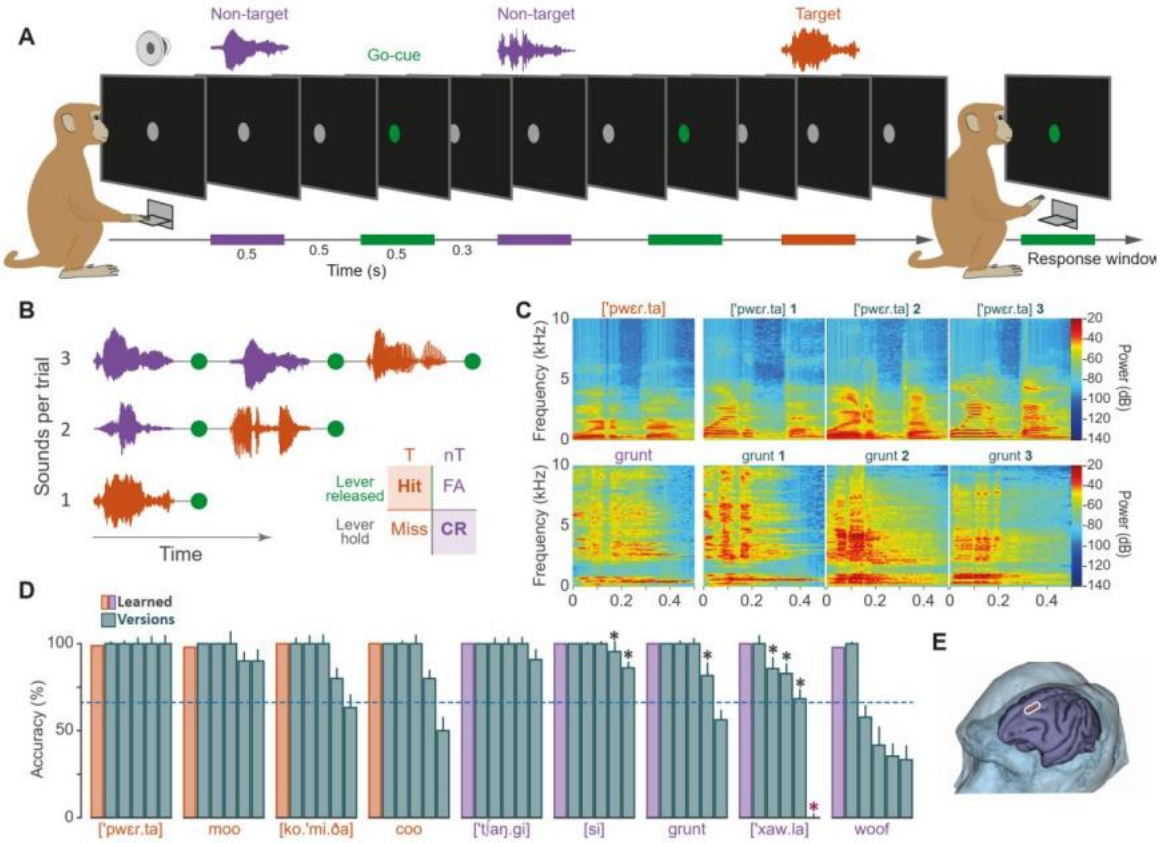


Figure 1. Auditory recognition task and behavioral performance. (A) Events in a trial. First, a visual cue (gray circle) appears at the center of the screen, indicating that the monkey should press and hold down the lever. After a variable period (0.5 to 1 s), a playback of 1-3 sounds commenced. Each sound continued with a 0.5 s delay and a 0.5 s green cue that replaced the gray circle. The monkey received a reward for releasing the lever within 0.7 s after the beginning of the green cue. Color code: orange (T), purple (nT), green (go-cue). **(B)** Types of trials. The T could appear in the first, second, or third position during a trial. Inset: Behavior as a function of the lever release in response to T and nT. FA, false alarms; CR, correct rejections. **(C)** Spectrograms of T and nT sounds and some acoustic versions (upper and lower rows, respectively). **(D)** Mean hit rates for learned T and CR for learned nT (orange and purple bars, respectively) and vLS (gray bars). The central horizontal line demarcates acoustic recognition by chance at 50%. Black asterisks: performance different from LS. Red asterisk: an exceptional sound that was recognized as a T. **(E)** Structural MRI of monkey 1; SMA indicated.

3.3 SMA neuronal activity correlated with behavior

Once the monkeys were performing the task consistently, we recorded 65 SMA neurons (Fig. 1*E*) in order to determine whether the neural responses showed invariance to vLS. Fig. 2*A* shows the raster plots of a neuron's responses to LS and vLS (left and right panels, respectively). Here, the peristimulus time histograms (PSTHs) in Fig. 2*B* describe similar response patterns of increased firing rates in all T regardless of whether they were LS or vLS. Interestingly, the neuron fired after the vLS 'woof' up to the activity level of the T (FA, dashed line). Then, at the beginning of the visual cue, the firing rate fell to nT levels. Neither the monkeys nor the neurons correctly classified the 'woof' sound as nT.

During that experimental session, we also recorded the activity of other neurons, e.g. the neuron in Fig. 2*C*. As opposed to the neuron described above, this neuron inhibited its

responses to all T to almost zero spikes/second. However, it was unaffected by nT, including vLS of 'woof,' which means that in this case the excitatory and not the inhibitory activity was biased during the false alarms.

So that we might evaluate the extent to which the neuronal population also coded T and nT as invariant, i.e. regardless of their particular identities, we performed an *F*-statistic analysis. Fig. 3*A* shows the number of neurons coding for sounds or categorical decisions during LS or vLS. Fifty-three out of the 65 recorded neurons (81.5%) showed categorical responses to LS and 49 to vLS (75.4%). However, 2 neurons (3%) coded for a particular T or nT LS and 6 (9%) for nT vLS. Fig. 3*B* presents the periods in which the neuronal population coded for T and nT in LS and vLS. Interestingly, the *F*-statistic values for LS categorical responses were higher

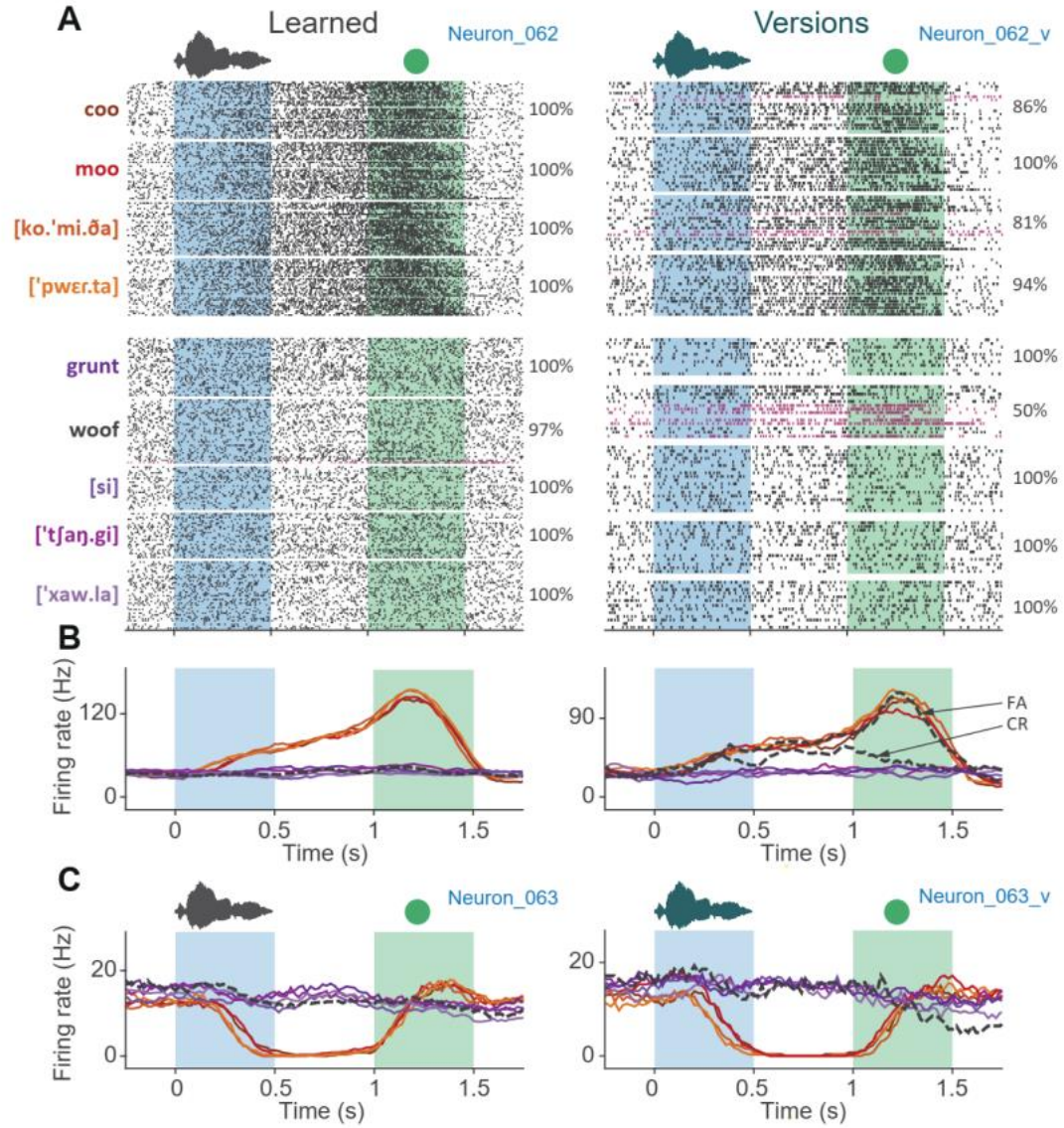


Figure 2. Responses of SMA neurons during the recognition of learned and version sounds. **(A)** Raster plot of a neuron's responses during trials of LS and vLS (left and right panels, respectively). Each line is a trial aligned to the beginning of sounds. Each dot is an action potential. Trials in red dots indicate misses for T, and false alarms for nT. The sounds were presented randomly during the experiment but shown here in blocks of T and nT sounds (upper and lower groups, respectively. **(B)** PSTHs of the neuron in a. CR, correct rejections, and FA, false alarms of vLS 'woof.' **(C)** PSTHs of a second neuron recorded in the same session.

than vLS, showing a more robust difference between T and nT rather than an effect of the number of neurons which was almost the same for LS and vLS conditions.

We also performed dPCA (see *Methods*) in order to verify whether the neuronal population can code for acoustic identities within categorical representations. The population activity observed with dPCA showed similar patterns for LS and vLS (Fig. 3*C*). Here, the three principal components reflect categorical coding of T. However, this analysis did not show tuning to particular T or nT sounds. One exception occurred at the third component of vLS where it was observed that the 'woof' stands out from the nT distribution. As described in Fig. 2*B*, the neuronal population seems to explain the monkey's false alarms. These results suggest that invariant perception in individual and population neuronal activity does not depend on recognition of timbre or modulations of particular emitters. Instead, invariant perception relies on a more general representation of each category.

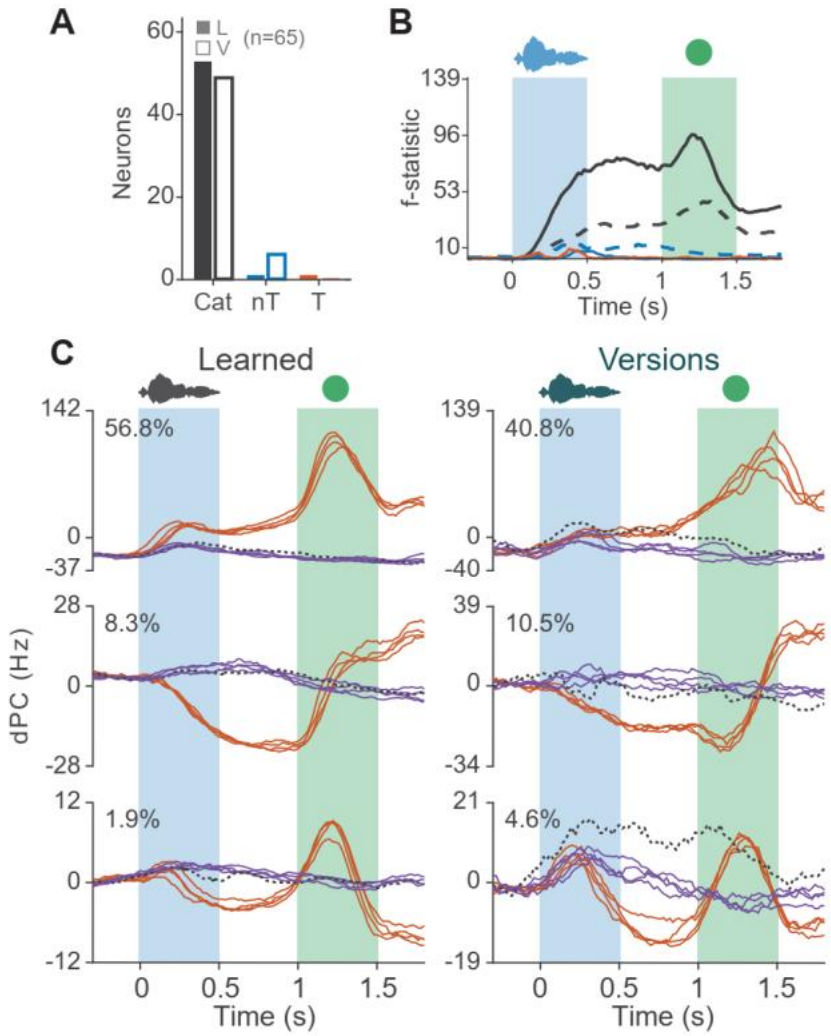


Figure 3. Categorical responses of the SMA population. (A) Neurons with categorical responses to T or nT groups or particular identities during LS and vLS computed by *F*-tests. **(B)** SMA mean *F*-statistic values as a function of time. Black lines: categorical responses. Orange: T identities. Blue: nT identities. Continuous lines: LS. Dashed lines: vLS. **(C)** Each row: first three demixed principal components of the SMA neuronal population during LS and vLS. Dashed line: 'woof' sound. Orange: T. Purple: nT.

In this regard, to test whether the neuronal responses were correlated with the monkeys' behavior, we calculated Spearman correlations between behavior and the firing rates of LS and vLS in 200 ms bins calculated every 20 ms (Fig. 4*A*). Fig. 4*B* shows the significant Spearman correlations throughout the task's periods and each neuron. The mean population's rho values varied similarly over time (i.e. Pearson's R = 0.90, p < 0.001). This implies that the neuronal population relates the task's events with the monkey's behavior for both LS and vLS. Moreover, the neurons that engaged across the events showed similar LS and vLS dynamics (Fig. 4*C*; Pearson's R = 0.86, p < 0.001). Overall, these results suggest that individual and population neuronal activity correlates with invariant perception in rhesus monkeys.

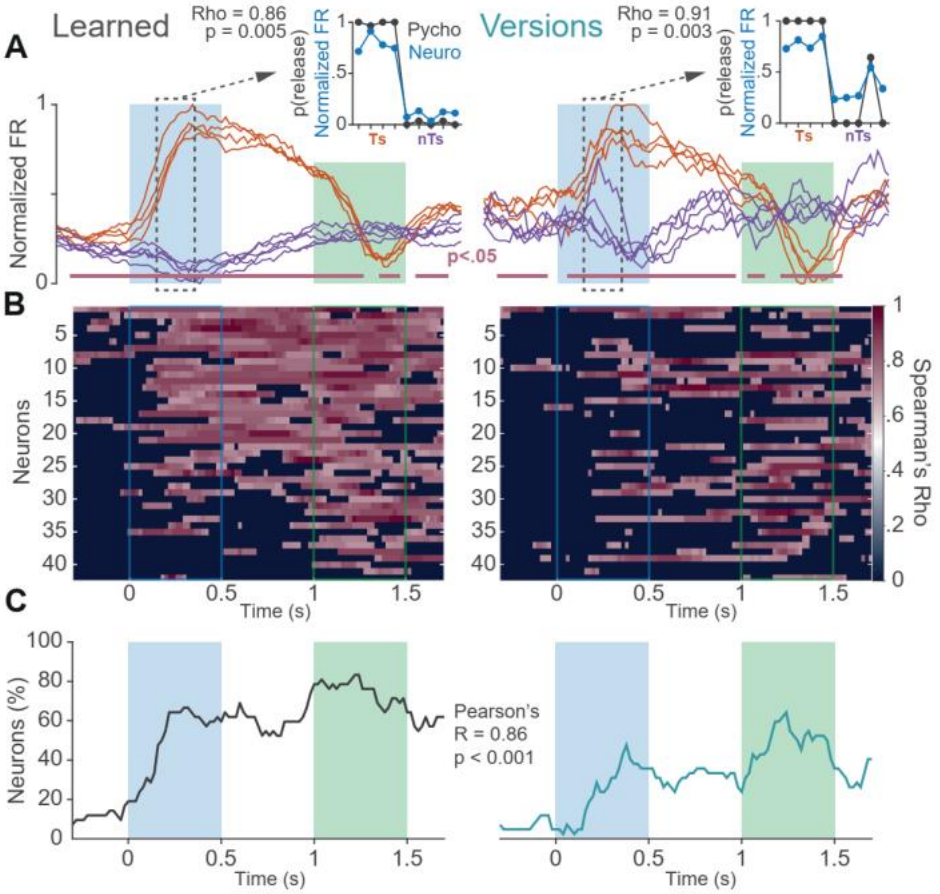


Figure 4. Correlation between the monkeys' performance and SMA neuronal response. (A) Responses of 1 neuron to LS and vLS. Red horizontal lines indicate periods of significant categorical differences. Insets: a time window indicated by the dashed box, where the probability of the monkey releasing the lever after each T and nT (black line) is similar to neuronal responses to LS and vLS (blue line). Rho: The Spearman's correlation and p significance for that temporal window. **(B)** Significant Spearman's rho along with the task for each neuron during LS and vLS. **(C)** Occasions on which the neuronal population recorded during LS and vLS were correlated with the monkey's performance.

4 Discussion

Our results demonstrate that the rhesus monkeys correctly categorized sounds to which they had had no previous exposure, such as different versions of learned sounds. This result is significant because in nonhuman primates, vocal production is highly stereotyped and probably genetically determined [29–33]. Numerous studies report limited acoustic recognition in

nonhuman primates [34–36]; however, emerging evidence suggests long-term plasticity in vocal learning in adult monkeys exposed to vocalizations during development [37,38]. For example, in a match-to-sample task, macaques were better at conspecific calls than other vocalizations or sounds [39]. Although we used similar sounds, we did not find performance biases towards conspecific vocalizations. One possible explanation is that in the experiments of Ng et al. discrimination of monkey vocalizations was better because those sounds had existed in long-term memory since birth. In contrast, working memory alone was probably more demanding and required different brain circuits. In our task, there was no working memory involved. After the monkeys learned numerous sounds that may or may not be ethologically relevant, they practiced for many days until they consistently achieved performance above 90%. Therefore, we expect different processing of short- and long-term memories in primates [40–42]. Future studies may offer insight into those differences.

Although there are no reports of perceptual invariance of sounds in non-human primates, our results show that behavior and SMA responses to LS and vLS were invariant. In other words, SMA activity supports the constant perception of sounds. This result is similar to experiments in primary auditory areas that demonstrate perceptual invariance during the recognition of vowels [4]. However, our work does not show that the SMA represents particular sounds, but rather that it directs actions based on acoustic recognition. Notably, those sounds are recognized when they relate to the motor actions triggered by known sounds. In other words, the SMA relates sounds to behaviors [43]. Thus, it is reasonable to suggest that the SMA is able to sort various types of sensory information regardless of whether or not it is stored in long term circuits.

In our task, different acoustic categories lead to the release or the holding down of the lever. Therefore, each group of heterogeneous sounds produces 1 of 2 learned motor outputs. Arguably, many sounds become synonyms of "release the lever" and many others of "hold the lever." In support of this argument, SMA neuronal responses were correlated with two orthogonal motor plans emerging from the recognition of T or nT sounds. Responses were invariant to T or nT groups and not to particular acoustic categories within each group. Nonetheless, the neuronal activity shows some sensory modulation to stimuli generated from the combination of T and nT sounds in different proportions (morphed sounds) (Fig. 4-1, for more details, see [22]). Experiments in visual areas such as the inferotemporal cortex have demonstrated that single neurons represent visual categories regardless of sensory modulations or visual perspective [44,45]. Therefore, instances in the hierarchical processing of sensory information create invariant representations of images. However, in those experiments, the neurons were not tested during active recognition. Thus, perceptual invariance was not tested. In contrast, experiments with monkeys that actively categorized visual [46–48] and auditory stimuli [49–51] have shown that neuronal activity in the prefrontal cortex is correlated with the animals' performance.

Our results suggest that the macaque SMA participates in the perceptual invariance of sounds by associating un-experienced sounds with the responses related to known categories. In other words, the brain accomplishes acoustic recognition by linking novel sounds with existing motor programs. Further experiments in behaving monkeys may unravel the mystery as to whether or not perceptual invariance of complex sounds relies on invariant representations in the parietal and temporal lobes.

5 Author contributions

JM, IM, TF, and LL performed experiments, JM, IM, JV, and JP analyzed data, LL designed the paradigm, TF programmed the task, while JM and LL wrote the paper.

6 Data and materials availability

Upon request.

7 Competing interests

The authors declare no competing financial interests.

8 Acknowledgements

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**. Jonathan Melchor Hernández is a doctoral student in the Programa de Doctorado en Ciencias Biomédicas [Doctoral program in biomedical sciences], at Universidad Nacional Autónoma de México (UNAM) and he was supported by CONACYT 229866. The data in this work are part of his doctoral dissertation. We thank Emilio Salinas for comments on the manuscript, Gerardo Coello and Ana Escalante of the computing department of the IFC, and Patrick Weill for reviewing the manuscript.

9 References

1. Hauser MD. Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. Anim Behav. 1998;55: 1647–1658. doi:10.1006/anbe.1997.0712
2. Slocombe KE, Zuberbühler K. Food-associated calls in chimpanzees: Responses to food types or relative food value? Anim Behav. 2006;72: 989–999.
3. Elie JE, Theunissen FE. Meaning in the avian auditory cortex: Neural representation of communication calls. Eur J Neurosci. 2015;41: 546–567. doi:10.1111/ejn.12812
4. Town SM, Wood KC, Bizley JK. Sound identity is represented robustly in auditory cortex during perceptual constancy. Nat Commun. 2018;9. doi:10.1038/s41467-018-07237-3
5. Saunders JL, Wehr M. Mice can learn phonetic categories. J Acoust Soc Am. 2019;145: 1168–1177. doi:10.1121/1.5091776
6. Daniel Meliza C, Margoliash D. Emergence of selectivity and tolerance in the avian auditory cortex. J Neurosci. 2012;32: 15158–15168. doi:10.1523/JNEUROSCI.0845-12.2012
7. Jiang X, Chevillet MA, Rauschecker JP, Riesenhuber M. Training Humans to Categorize Monkey Calls: Auditory Feature- and Category-Selective Neural Tuning Changes. Neuron. 2018;98: 405-416.e4. doi:10.1016/j.neuron.2018.03.014
8. Belin P, Bodin C, Aglieri V. A "voice patch" system in the primate brain for processing vocal information? Hear Res. 2018;366: 65–74. doi:10.1016/j.heares.2018.04.010
9. Belin P, Zatorre R, Pike BG. Voice-selectiveareas in humanauditorycortex Modulating top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music View project Neural Mechanisms of Voice Processing in Marmosets. View project. 2000. Available: https://www.researchgate.net/publication/235653597
10. Leaver AM, Rauschecker JP. Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. J Neurosci. 2010;30: 7604–7612. doi:10.1523/JNEUROSCI.0296-10.2010
11. Ortiz-Rios M, Kuśmierek P, DeWitt I, Archakov D, Azevedo FAC, Sams M, et al. Functional MRI of the vocalization-processing network in the macaque brain. Front Neurosci. 2015;9. doi:10.3389/fnins.2015.00113
12. Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. Nat Neurosci. 2008;11: 367–374. doi:10.1038/nn2043
13. Bizley JK, Cohen YE. The what, where and how of auditory-object perception. Nat Rev Neurosci. 2013;14: 693–707. doi:10.1038/nrn3565
14. Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the

prefrontal cortex in the rhesus monkey. J Comp Neurol. 1999;403: 141–157. doi:10.1002/(SICI)1096-9861(19990111)403:2<141::AID-CNE1>3.0.CO;2-V
15. Leonard MK, Chang EF. Dynamic speech representations in the human temporal lobe. Trends Cogn Sci. 2014;18: 472–479. doi:10.1016/j.tics.2014.05.001
16. Repp BH. Categorical Perception: Issues, Methods, Findings. SPEECH AND LANGUAGE: Advances in Basic Research and Practice. ACADEMIC PRESS, INC.; 1984. doi:10.1016/b978-0-12-608610-2.50012-1
17. Lima CF, Krishnan S, Scott SK. Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery. Trends Neurosci. 2016;39: 527–542. doi:10.1016/j.tins.2016.06.003
18. Lemus L, Hernández A, Romo R. Neural encoding of auditory discrimination in ventral premotor cortex. Proc Natl Acad Sci U S A. 2009;106: 14640–14645. doi:10.1073/pnas.0907505106
19. Lara AH, Cunningham JP, Churchland MM. Different population dynamics in the supplementary motor area and motor cortex during reaching. Nat Commun. 2018;9. doi:10.1038/s41467-018-05146-z
20. Nachev P, Kennard C, Husain M. Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci. 2008;9: 856–869. doi:10.1038/nrn2478
21. Shima K, Tanji J. Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. J Neurophysiol. 2000;84: 2148–2160.
22. Morán I, Perez-Orive J, Melchor J, Figueroa T, Lemus L. Auditory Decisions in the Supplementary Motor Area. bioRxiv. 2020; 2020.10.20.347864. doi:10.1101/2020.10.20.347864
23. Hernández A, Zainos A, Romo R. Temporal evolution of a decision-making process in medial premotor cortex. Neuron. 2002;33: 959–972. doi:10.1016/S0896-6273(02)00613-X
24. Lemus L, Hernández A, Luna R, Zainos A, Nácher V, Romo R. Neural correlates of a postponed decision report. Proc Natl Acad Sci U S A. 2007;104: 17174–17179. doi:10.1073/pnas.0707961104
25. Romo R, Merchant H, Zainos A, Hernández A. Categorical perception of somesthetic stimuli: Psychophysical measurements correlated with neuronal events in primate medial premotor cortex. Cereb Cortex. 1997;7: 317–326. doi:10.1093/cercor/7.4.317
26. Feng G, Gan Z, Wang S, Wong PCM, Chandrasekaran B. Task-General and Acoustic-Invariant Neural Representation of Speech Categories in the Human Brain. Cereb Cortex. 2018;28: 3241–3254. doi:10.1093/cercor/bhx195
27. Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, et al. Demixed principal component analysis of neural population data. Elife. 2016. doi:10.7554/eLife.10989
28. Melchor J, Morán I, Figueroa T, Lemus L. Perceptual Invariance of Words and Other Learned Sounds in Non-human Primates. bioRxiv. bioRxiv; 2019. p. 805218. doi:10.1101/805218
29. Brockelman WY, Schilling D. Inheritance of stereotyped gibbon calls. Nature. 1984;312: 634–636. doi:10.1038/312634a0
30. Hammerschmidt K, Fischer J. Baboon vocal repertoires and the evolution of primate vocal diversity. J Hum Evol. 2019;126: 1–13. doi:10.1016/j.jhevol.2018.10.010
31. Nieder A, Mooney R. The neurobiology of innate, volitional and learned vocalizations in mammals and birds. Philosophical Transactions of the Royal Society B: Biological Sciences. 2020. doi:10.1098/rstb.2019.0054
32. Owren MJ, Dieter JA, Seyfarth RM, Cheney DL. '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. 1992;120: 218–231. doi:10.1163/156853992X00615
33. Zador AM. A critique of pure learning and what artificial neural networks can learn from

animal brains. Nat Commun. 2019;10: 3770. doi:10.1038/s41467-019-11786-6
34. Fritz J, Mishkin M, Saunders RC. In search of an auditory engram. Proc Natl Acad Sci. 2005;102: 9359–9364. doi:10.1073/pnas.0503998102
35. Scott BH, Mishkin M, Yin P. Monkeys have a limited form of short-term memory in audition. Proc Natl Acad Sci U S A. 2012;109: 12237–12241. doi:10.1073/pnas.1209685109
36. Wright AA. Auditory list memory and interference processes in monkeys. J Exp Psychol Anim Behav Process. 1999;25: 284–296.
37. Takahashi DY, Liao DA, Ghazanfar AA. Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. Curr Biol. 2017;27: 1844-1852.e6. doi:10.1016/j.cub.2017.05.004
38. Zhao L, Rad BB, Wang X. Long-lasting vocal plasticity in adult marmoset monkeys. Proc R Soc B Biol Sci. 2019;286. doi:10.1098/rspb.2019.0817
39. Ng CW, Plakke B, Poremba A. Primate auditory recognition memory performance varies with sound type. Hear Res. 2009;256: 64–74. doi:10.1016/j.heares.2009.06.014
40. Fritz JB, Malloy M, Mishkin M, Saunders RC. Monkey's short-term auditory memory nearly abolished by combined removal of the rostral superior temporal gyrus and rhinal cortices. Brain Research. Elsevier B.V.; 2016. pp. 289–298. doi:10.1016/j.brainres.2015.12.012
41. Muñoz-López M, Insausti R, Mohedano-Moriano A, Mishkin M, Saunders RC. Anatomical pathways for auditory memory II: Information from rostral superior temporal gyrus to dorsolateral temporal pole and medial temporal cortex. Front Neurosci. 2015;9: 1–21. doi:10.3389/fnins.2015.00158
42. Munoz-Lopez MM, Mohedano-Moriano A, Insausti R. Anatomical pathways for auditory memory in primates. Frontiers in Neuroanatomy. 2010. doi:10.3389/fnana.2010.00129
43. Vergara J, Rivera N, Rossi-Pool R, Romo R. A Neural Parametric Code for Storing Information of More than One Sensory Modality in Working Memory. Neuron. 2016;89: 54–62. doi:10.1016/j.neuron.2015.11.026
44. Freiwald WA, Tsao DY. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science (80-). 2010;330: 845–851. doi:10.1126/science.1194908
45. Hesse JK, Tsao DY. The macaque face patch system: a turtle's underbelly for the brain. Nat Rev Neurosci. 2020;21: 695–716. doi:10.1038/s41583-020-00393-w
46. Seger CA, Miller EK. Category Learning in the Brain. Annu Rev Neurosci. 2010;33: 203–219. doi:10.1146/annurev.neuro.051508.135546
47. Cromer JA, Roy JE, Miller EK. Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. Neuron. 2010;66: 796–807. doi:10.1016/j.neuron.2010.05.005
48. Roy JE, Buschman TJ, Miller EK. PFC neurons reflect categorical decisions about ambiguous stimuli. J Cogn Neurosci. 2014;26: 1283–1291. doi:10.1162/jocn_a_00568
49. Russ BE, Orr LE, Cohen YE. Prefrontal Neurons Predict Choices during an Auditory Same-Different Task. Curr Biol. 2008;18: 1483–1488. doi:10.1016/j.cub.2008.08.054
50. Lee JH, Russ BE, Orr LE, Cohen YE. Prefrontal activity predicts monkeys' decisions during an auditory category task. Front Integr Neurosci. 2009;3: 1–12. doi:10.3389/neuro.07.016.2009
51. Huang Y, Brosch M. Associations between sounds and actions in primate prefrontal cortex. Brain Res. 2020;1738: 1–22. doi:10.1016/j.brainres.2020.146775

405 # 10 Extended Data and Table

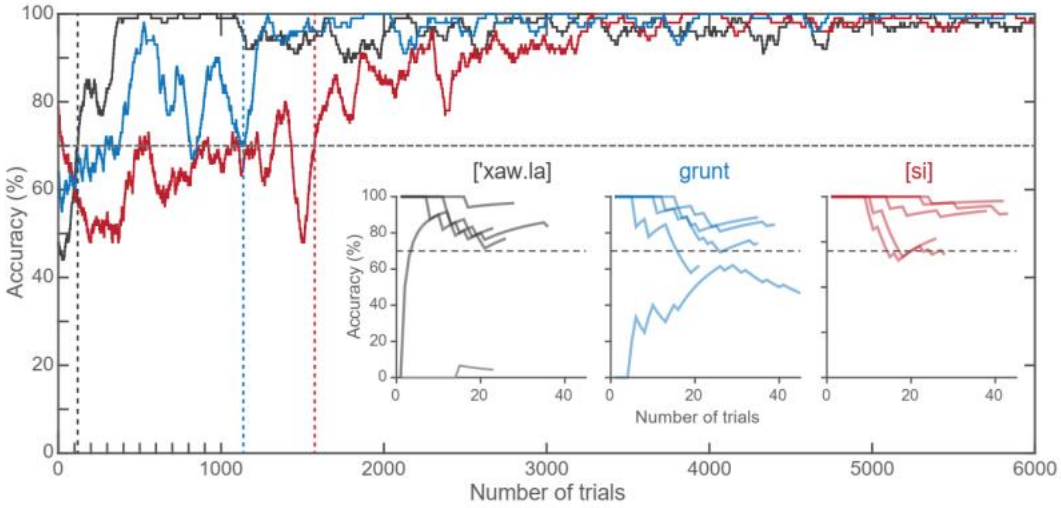


Figure 1-1. Trials for learning sounds as compared to the trials for recognition of novel acoustic versions. The recognition of three sounds became stable above 70% accuracy in hundreds of training trials throughout several weeks. Insets: the recognition of 80% of the exemplar learned sounds' versions reached 70% accuracy in no more than 40 trials. Dashed horizontal lines indicate performance at 70%. Dashed vertical lines indicate trials from the last record of the training that was preceded by several months of previous trainings.

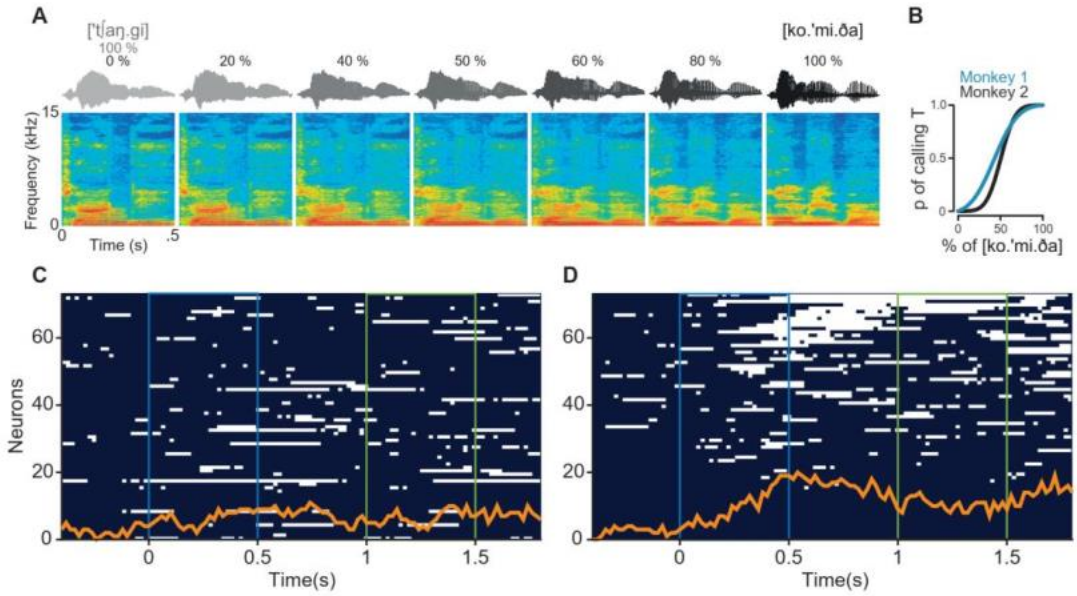


Figure 4-1. SMA neurons with significant linear regression slopes between firing rate and acoustic information. Besides their categorical responses to T and nT, the neurons also modulated their activity as a function of sounds morphed in T and nT proportions. **(A)** Spectrograms of a morphing set. **(B)** Monkeys' behavior during the recognition of a morphing set. **(C, D)** SMA neurons tested with T and nT, respectively, during morphing sets of learned sounds. In white are the periods of significant linear regression slopes for each neuron throughout the task's components (permutation test: $p < 0.05$). The orange line indicates the total number of neurons with acoustic information at each time bin. Each neuron in A corresponds to each neuron in B. All 65 neurons coded the versions of the learned sounds as well (see Fig. 3*A*).

406

Table 1-1. Description of acoustic stimuli

Target	Description	non-Target	Description
coo	Monkey vocalization	grunt	Monkey vocalization
warble	Monkey vocalization	grunt-2	Monkey vocalization
[ko.'mi.ða]	Spanish word for food	shrill bark	Monkey vocalization
['pwɛr.ta]	Spanish word for door	pulsed	Monkey vocalization
[a.ni.'ma.les]	Spanish word for animals	['lo.ka]	Spanish word for crazy
['ři.o]	Spanish word for river	[kimi]	Invented word
[no]	Spanish word for not	['tʃaŋ.gi]	Invented word
[la.'βa.βo]	Spanish word for sink	[si]	Spanish word for yes
moo	Cow vocalization	['xaw.la]	Spanish word for cage
bounce	Bouncing tone	[mo.ni.'tor]	Spanish word for screen
PT500	500 Hz Pure tone	['po.sa]	Invented word
		['pa.si]	Invented word
		[i.'ɣlu]	Spanish word for igloo
		[pa.lo.'mi.tas]	Spanish word for popcorn
		meow	Cat meowing
		chirp	Bird vocalization
		screech	Parrot screech
		caw	Crow squawk
		woof	Dog bark
		hoot	Owl hooting
		AM-tone	Modulated tone (1 kHz)
		buzz	Mosquito whine
		ring	ringtone
		ring-2	bell ring
		noise	Band pass noise (1-4 kHz)

7. DISCUSIÓN

7.1 Los monos rhesus discriminan sonidos mediante sus formantes

A diferencia del lenguaje hablado que constituye un sistema de comunicación simbólico aprendido, la estructura acústica de las vocalizaciones de los PNH se considera que está determinada genéticamente y es poco o nada dependiente de la experiencia (Brockelman y Schilling, 1984; Owren et al., 1992; Hammerschmidt y Fischer, 2019; Nieder y Mooney, 2020). No obstante, recientemente se han presentado algunas evidencias de aprendizaje de producción vocal durante el desarrollo y en monos adultos (Takahashi et al., 2017; Zhao et al., 2019). Adicionalmente, se han identificado áreas corticales que responden de forma preferente a las vocalizaciones conespecíficas en: el cinturón lateral (Rauschecker et al., 1995), el giro temporal superior (Leaver y Rauschecker, 2010; Ortiz-Rios et al., 2015; Belin et al., 2018) y la corteza prefrontal (Romanski et al., 1999; Rauschecker y Romanski, 2011). Sin embargo, aún no se ha establecido si la capacidad de aprender o reconocer sonidos también está determinada de forma innata, si depende de la experiencia, o es el resultado de ambas.

Para estudiar las bases psicofísicas y los mecanismos neurobiológicos de la percepción auditiva desarrollamos un modelo experimental de discriminación de sonidos en monos rhesus. El óptimo desempeño psicofísico mostrado por los macacos aporta evidencia contundente de su capacidad para aprender y discriminar sonidos heteroespecíficos, con modulaciones espectrotemporales variadas. Previo a nuestro estudio, solo existían datos etológicos que sugerían memoria a largo plazo de vocalizaciones conespecíficas (Seyfarth et al., 1980).

Gran parte de nuestros conocimientos del procesamiento auditivo en PNH se obtuvo estudiando propiedades acústicas específicas (p. ej. frecuencia, amplitud) de sonidos simples (tonos puros, ruidos pasa-banda o tonos de amplitud modulada) con animales anestesiados o durante escucha pasiva (Rauschecker et al., 1995; Tian et al., 2001; Romanski y Goldman-Rakic, 2002; Recanzone, 2008; Fukushima et al., 2014). Los escasos estudios con monos reconociendo estímulos complejos reportaron un desempeño considerablemente limitado (Fritz et al., 2005; Ng et al., 2009; Scott et al., 2012). Además, con nuestro paradigma todas las categorías acústicas presentadas

(vocalizaciones conespecíficas, heteroespecíficas y sonidos artificiales) tuvieron un desempeño similar; a diferencia de lo previamente reportado, donde los sonidos conespecíficos fueron identificados mejor (Ng et al., 2009). Estudios futuros podrían determinar si tales diferencias son debidas a diferencias entre la memoria auditiva a corto y largo plazo, que es una las principales diferencias entre nuestra tarea y las precedentes (Muñoz-López et al., 2010; Muñoz-López et al., 2015; Fritz et al., 2016).

La categorización auditiva es un proceso cognitivo de clasificación de sonidos a partir de propiedades espectrotemporales y asociaciones semánticas (Tsunada y Cohen, 2014). Nuestros resultados de la categorización de las mezclas de sonidos ("*morphs*") fue consistente con estudios similares en humanos (Chakladar et al., 2008; Furuyama et al., 2017; Jiang et al., 2018), y con el único reporte existente en monos, donde aprendieron a categorizar mezclas entre dos palabras en inglés que diferían en la primera consonante: /bad/ y /dad/ (Russ et al., 2008; Tsunada et al., 2011).

Los formantes constituyen las modulaciones frecuenciales de mayor energía en los sonidos de comunicación como las palabras. Los análisis de correlación entre la psicofísica y los formantes contenidos en los *morphs*, sumado a cómo los humanos usan los formantes para reconocer vocales, nos llevaron a estudiar de forma más específica la contribución de los formantes en la discriminación de los sonidos. Dado que no podíamos descartar que otras propiedades acústicas afectaran la percepción durante la presentación de los *morphs*, decidimos presentar estímulos que solo tuvieran las frecuencias formantes (F1, F2 y F1F2) de los sonidos aprendidos.

Encontramos que la información acústica contenida en los dos primeros fue suficiente para que los monos discriminaran los sonidos. Resultados semejantes ya habían sido reportados en humanos (Peterson y Barney, 1952; Remez et al., 1981), mientras que en los monos solo se había propuesto que ayudaban a extraer información de las características del emisor (Ghazanfar et al., 2007; Furuyama et al., 2016, 2017) y posiblemente de sus propias vocalizaciones (Fitch y Fritz, 2006).

7.2 Percepción invariante en el AMS

La habilidad de reconocer un sonido de forma invariante, requiere tolerancia perceptual para diferentes modificaciones en su estructura espectrotemporal (variantes acústicas).

Los mecanismos cerebrales subyacentes de esta habilidad solo se comprenden de forma parcial y no encontramos reportes previos de su estudio en PNH hasta el presente trabajo (Melchor et al 2020; 2021).

El reconocimiento de variantes de vocalizaciones conespecíficas, vocales y consonantes ha sido descrito en pinzones, hurones y roedores, respectivamente (Elie y Theunissen, 2015; Town et al., 2018; Saunders y Wehr, 2019). En humanos, usando resonancia magnética se encontró que el AMS se activa durante el reconocimiento invariante de palabras para diferentes emisores (Feng et al., 2018). Nuestros datos nos permiten aseverar que los monos también identifican variantes acústicas fonéticamente similares, pero de emisores a los que no habían sido expuestos con anterioridad.

La corteza premotora ventral (homóloga al área de broca en humanos) de los macacos participa en discriminaciones acústicas (Lemus et al., 2009), mientras que la actividad del AMS correlaciona con las decisiones basadas en la comparación de sonidos (Vergara et al., 2016). Por lo tanto, el AMS no solo estaría participando en el control del movimiento voluntario y la memoria de trabajo (Lemus et al., 2007; Nachev et al., 2008; Lara et al., 2018), sino también en la integración de señales sensoriales y de memoria para la toma de decisiones basadas en estímulos acústicos (Morán et al., 2021).

En nuestro estudio encontramos que las respuestas neuronales del AMS correlacionan con la percepción invariante de los monos. Las neuronas no distinguen entre los sonidos presentados, pero sí entre las categorías (target y no-target), de manera que se forman dos señales que permiten resolver la tarea. Tanto la actividad neuronal a nivel individual y poblacional concuerdan con la decisión perceptual de los monos. Es decir, las variantes acústicas presentadas se asociaron a las mismas respuestas categóricas de los sonidos aprendidos. En consecuencia, investigar la codificación invariante de las vocalizaciones en cada área cortical de la vía ventral auditiva es uno de los próximos objetivos del laboratorio.

8. CONCLUSIÓN

Los monos rhesus tiene la capacidad de aprender, categorizar y discriminar sonidos complejos heteroespecíficos, así como estímulos con diversas modificaciones espectrotemporales, o que sólo contengan los dos principales formantes. Además, descubrimos que las neuronas del área motora suplementaria representan las categorías conductuales durante la percepción invariante de sonidos.

9. PERSPECTIVAS

Investigar cómo se codifican las vocalizaciones conespecíficas y heteroespecíficas a nivel cortical puede potencialmente contribuir a nuestra comprensión de los mecanismos de comunicación y el lenguaje. Por tanto, el paradigma experimental aquí presentado, podría ayudar a entender cómo la actividad cortical a lo largo de la vía ventral durante la identificación activa de estímulos acústicos es procesada en áreas cerebrales de integración sensorio-motora para producir conductas.

10. REFERENCIAS

Ahveninen, J., Huang, S., Nummenmaa, A., Belliveau, J. W., Hung, A. Y., Jääskeläinen, I. P., Rauschecker, J. P., Rossi, S., Tiitinen, H., & Raij, T. (2013). Evidence for distinct human auditory cortex regions for sound location versus identity processing. *Nature Communications*, *4*(May), 1–8. https://doi.org/10.1038/ncomms3585

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). *‘“ What ”’ and ‘“ where ”’ in the human auditory system*. *98*(21).

Barton, B., Venezia, J. H., Saberi, K., Hickok, G., & Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(50), 20738–20743. https://doi.org/10.1073/pnas.1213381109

Baumann, S., Joly, O., Rees, A., Petkov, C. I., Sun, L., Thiele, A., & Griffiths, T. D. (2015). The topography of frequency and time representation in primate auditory cortices. *ELife*, *2015*(4), 1–15. https://doi.org/10.7554/eLife.03256

Belin, P., Bodin, C., & Aglieri, V. (2018). A “voice patch” system in the primate brain for processing vocal information? *Hearing Research*, *366*, 65–74. https://doi.org/10.1016/j.heares.2018.04.010

Belin, P., Zatorre, R., & Pike, B. G. (2000). *Voice-selectiveareas in humanauditorycortex Modulating top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music View project Neural Mechanisms of Voice Processing in Marmosets. View project*. https://www.researchgate.net/publication/235653597

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165. https://doi.org/10.1038/nature03867

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, *16*(4), 391–399. https://doi.org/10.1016/j.conb.2006.07.001

Billimoria, C. P., Kraus, B. J., Narayan, R., Maddox, R. K., & Sen, K. (2008). Invariance and Sensitivity to Intensity in Neural Discrimination of Natural Sounds. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0961-08.2008

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693–707. https://doi.org/10.1038/nrn3565

Bizley, J. K., Walker, K. M. M., King, A. J., & Schnupp, J. W. H. (2013). Spectral timbre perception in ferrets: Discrimination of artificial vowels under different listening conditions. *The Journal of the Acoustical Society of America*, *133*(1), 365–376. https://doi.org/10.1121/1.4768798

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K. H., & Fitch, W. T. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, *7*, 1–11. https://doi.org/10.1038/srep41070

Brewer, A. A., & Barton, B. (2016). Maps of the Auditory Cortex. *Annual Review of Neuroscience*, *39*(1), 385–407. https://doi.org/10.1146/annurev-neuro-070815-014045

Brockelman, W. Y., & Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature*, *312*(5995), 634–636. https://doi.org/10.1038/312634a0

Camalier, C. R., D’Angelo, W. R., Sterbing-D’Angelo, S. J., De La Mothe, L. A., & Hackett, T. A.

(2012). Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 18168–18173. https://doi.org/10.1073/pnas.1206387109

Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J. J., Mwilambwe-Tshilobo, L., Natan, R. G., & Geffen, M. N. (2015). Emergence of invariant representation of vocalizations in the auditory cortex. *Journal of Neurophysiology*, *114*(5), 2726–2740. https://doi.org/10.1152/jn.00095.2015

Chakladar, S., Logothetis, N. K., & Petkov, C. I. (2008). Morphing rhesus monkey vocalizations. *Journal of Neuroscience Methods*, *170*(1), 45–55. https://doi.org/10.1016/j.jneumeth.2007.12.023

Cohen, Y. E., Bennur, S., Christison-Lagay, K., Gifford, A. M., & Tsunada, J. (2016). Functional Organization of the Ventral Auditory Pathway. *Advances in Experimental Medicine and Biology*, *894*, 381–388. https://doi.org/10.1007/978-3-319-25474-6_40

Cohen, Y. E., Theunissen, F., Russ, B. E., & Gill, P. (2007). Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *Journal of Neurophysiology*, *97*(2), 1470–1484. https://doi.org/10.1152/jn.00769.2006

Daniel Meliza, C., & Margoliash, D. (2012). Emergence of selectivity and tolerance in the avian auditory cortex. *Journal of Neuroscience*, *32*(43), 15158–15168. https://doi.org/10.1523/JNEUROSCI.0845-12.2012

Elie, J. E., & Theunissen, F. E. (2015). Meaning in the avian auditory cortex: Neural representation of communication calls. *European Journal of Neuroscience*, *41*(5), 546–567. https://doi.org/10.1111/ejn.12812

Ey, E., Pfefferle, D., & Fischer, J. (2007). Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates*, *48*(4), 253–267. https://doi.org/10.1007/s10329-006-0033-y

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage*, *23*(3), 840–848. https://doi.org/10.1016/j.neuroimage.2004.09.019

Feng, G., Gan, Z., Wang, S., Wong, P. C. M., & Chandrasekaran, B. (2018). Task-General and Acoustic-Invariant Neural Representation of Speech Categories in the Human Brain. *Cerebral Cortex*, *28*, 3241–3254. https://doi.org/10.1093/cercor/bhx195

Fettiplace, R. (2020). Diverse Mechanisms of Sound Frequency Discrimination in the Vertebrate Cochlea. *Trends in Neurosciences*, *43*(2), 88–102. https://doi.org/10.1016/j.tins.2019.12.003

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222. https://doi.org/10.1121/1.421048

Fitch, W. T., & Fritz, J. B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *The Journal of the Acoustical Society of America*, *120*(4), 2132–2141. https://doi.org/10.1121/1.2258499

Frank, M., Muhlack, B., Zebe, F., & Scharinger, M. (2020). Contributions of pitch and spectral information to cortical vowel categorization. *Journal of Phonetics*, *79*, 100963. https://doi.org/10.1016/j.wocn.2020.100963

Fritz, J. B., Malloy, M., Mishkin, M., & 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 Research*, *1640*, 289–298. https://doi.org/10.1016/j.brainres.2015.12.012

Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research*, *206*(1–2), 159–176. https://doi.org/10.1016/j.heares.2005.01.015

Fukushima, M., Saunders, R. C., Leopold, D. A., Mishkin, M., & Averbeck, B. B. (2014). Differential coding of conspecific vocalizations in the ventral auditory cortical stream. *Journal of Neuroscience*, *34*(13), 4665–4676. https://doi.org/10.1523/JNEUROSCI.3969-13.2014

Furuyama, T., Kobayasi, K. I., & Riquimaroux, H. (2016). Role of vocal tract characteristics in individual discrimination by Japanese macaques (Macaca fuscata). *Scientific Reports*, *6*. https://doi.org/10.1038/srep32042

Furuyama, T., Kobayasi, K. I., & Riquimaroux, H. (2017). Acoustic characteristics used by Japanese macaques for individual discrimination. *Journal of Experimental Biology*, *220*(19), 3571–3578. https://doi.org/10.1242/jeb.154765

Ghazanfar, A. A., & Rendall, D. (2008). Evolution of human vocal production. In *Current Biology* (Vol. 18, Issue 11). Curr Biol. https://doi.org/10.1016/j.cub.2008.03.030

Ghazanfar, A. A., Smith-Rohrberg, D., & Hauser, M. D. (2001). The role of temporal cues in rhesus monkey vocal recognition: Orienting asymmetries to reversed calls. *Brain, Behavior and Evolution*, *58*(3), 163–172. https://doi.org/10.1159/000047270

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-Tract Resonances as Indexical Cues in Rhesus Monkeys. *Current Biology*, *17*(5), 425–430. https://doi.org/10.1016/j.cub.2007.01.029

Gifford, G. W., MacLean, K. A., Hauser, M. D., & Cohen, Y. E. (2005). The neurophysiology of functionally meaningful categories: Macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *Journal of Cognitive Neuroscience*, *17*(9), 1471–1482. https://doi.org/10.1162/0898929054985464

Goldstein, E. B., & Brockmole, J. (2016). *Sensation and perception* (Cengage Learning (ed.)).

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892. https://doi.org/10.1038/nrn1538

Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hearing Research*, *271*(1–2), 133–146. https://doi.org/10.1016/j.heares.2010.01.011

Hammerschmidt, K., & Fischer, J. (2019). Baboon vocal repertoires and the evolution of primate vocal diversity. *Journal of Human Evolution*, *126*, 1–13. https://doi.org/10.1016/j.jhevol.2018.10.010

Hauser, M D, & Marler, P. (1993). Food-associated calls in rhesus macaques (Macaca mulatta). 1. Socioecological factors influencing call production. *Behavioral Ecology*, *4*(3), 194–205.

Hauser, Marc David. (1998). Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. *Animal Behaviour*, *55*(6), 1647–1658. https://doi.org/10.1006/anbe.1997.0712

Heelan, C., Lee, J., O'Shea, R., Lynch, L., Brandman, D. M., Truccolo, W., & Nurmikko, A. V. (2019). Decoding speech from spike-based neural population recordings in secondary auditory

cortex of non-human primates. *Communications Biology*, *2*(1). https://doi.org/10.1038/s42003-019-0707-9

Herdener, M., Esposito, F., Scheffler, K., Schneider, P., Logothetis, N. K., Uludag, K., & Kayser, C. (2013). Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex*, *49*(10), 2822–2833. https://doi.org/10.1016/j.cortex.2013.04.003

Hienz, R. D., & Brady, J. V. (1988). The acquisition of vowel discriminations by nonhuman primates. *Journal of the Acoustical Society of America*, *84*(1), 186–194. https://doi.org/10.1121/1.396963

Hienz, R. D., Sachs, M. B., & Sinnott, J. M. (1981). Discrimination of steady-state vowels by blackbirds and pigeons. *Journal of the Acoustical Society of America*, *70*(3), 699–706. https://doi.org/10.1121/1.386933

Hienz, Robert D., Jones, A. M., & Weerts, E. M. (2004). The discrimination of baboon grunt calls and human vowel sounds by baboons. *The Journal of the Acoustical Society of America*, *116*(3), 1692–1697. https://doi.org/10.1121/1.1778902

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Honorof, D. N., & Whalen, D. H. (2010). Identification of speaker sex from one vowel across a range of fundamental frequencies. *The Journal of the Acoustical Society of America*, *128*(5), 3095–3104. https://doi.org/10.1121/1.3488347

Javitt, D. C., & Sweet, R. A. (2015). Auditory dysfunction in schizophrenia: Integrating clinical and basic features. *Nature Reviews Neuroscience*, *16*(9), 535–550. https://doi.org/10.1038/nrn4002

Jiang, X., Chevillet, M. A., Rauschecker, J. P., & Riesenhuber, M. (2018). Training Humans to Categorize Monkey Calls: Auditory Feature- and Category-Selective Neural Tuning Changes. *Neuron*, *98*(2), 405-416.e4. https://doi.org/10.1016/j.neuron.2018.03.014

Johnson, K., & Sjerps, M. J. (2021). Speaker Normalization in Speech Perception. *The Handbook of Speech Perception*, 145–176. https://doi.org/10.1002/9781119184096.CH6

Kaas, J. H., & Hackett, T. A. (1999). "What" and "where" processing in auditory cortex. *Nature Neuroscience*, *2*(12), 1045–1047. https://doi.org/10.1038/15967

Kajikawa, Y., Frey, S., Ross, D., Falchier, A., Hackett, T. A., & Schroeder, C. E. (2015). Auditory properties in the parabelt regions of the superior temporal gyrus in the awake macaque monkey: An initial survey. *Journal of Neuroscience*, *35*(10), 4140–4150. https://doi.org/10.1523/JNEUROSCI.3556-14.2015

Kell, A. J. E., & McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-11710-y

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders*, *74*(November 2017), 74–97. https://doi.org/10.1016/j.jcomdis.2018.05.004

Kojima, S., & Kiritani, S. (1989). Vocal-auditory functions in the chimpanzee: Vowel perception. *International Journal of Primatology*, *10*(3), 199–213. https://doi.org/10.1007/BF02735200

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948–955. https://doi.org/10.1016/j.neuroimage.2004.02.020

Langner, G, Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex. *Journal of Comparative Physiology A*, *181*, 665–676. papers2://publication/uuid/20ED004C-8CEF-45A4-83BA-DA0696324B51

Langner, Gerald, Dinse, H. R., & Godde, B. (2009). A map of periodicity orthogonal to frequency representation in the cat auditory cortex. *Frontiers in Integrative Neuroscience*, *0*(NOV), 27. https://doi.org/10.3389/NEURO.07.027.2009

Lara, A. H., Cunningham, J. P., & Churchland, M. M. (2018). Different population dynamics in the supplementary motor area and motor cortex during reaching. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05146-z

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience*, *30*(22), 7604–7612. https://doi.org/10.1523/JNEUROSCI.0296-10.2010

Leaver, A. M., & Rauschecker, J. P. (2016). Functional topography of human auditory cortex. *Journal of Neuroscience*, *36*(4), 1416–1428. https://doi.org/10.1523/JNEUROSCI.0226-15.2016

Lemus, L., Hernández, A., Luna, R., Zainos, A., Nácher, V., & Romo, R. (2007). Neural correlates of a postponed decision report. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17174–17179. https://doi.org/10.1073/pnas.0707961104

Lemus, L., Hernández, A., & Romo, R. (2009). Neural encoding of auditory discrimination in ventral premotor cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(34), 14640–14645. https://doi.org/10.1073/pnas.0907505106

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–575. https://doi.org/10.1038/nature04951

Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., & Brefczynski-Lewis, J. A. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *Journal of Neuroscience*, *29*(7), 2283–2296. https://doi.org/10.1523/JNEUROSCI.4145-08.2009

Lieberman, P., & Blumstein, S. E. (1988). Speech Physiology, Speech Perception, and Acoustic Phonetics. In *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge University Press. https://doi.org/10.1017/cbo9781139165952

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, *5*(3), 270–288. https://doi.org/10.1093/cercor/5.3.270

Melchor, J., Morán, I., Vergara, J., Figueroa, T., Perez-Orive, J., & Lemus, L. (2020). Neuronal Correlates of the Perceptual Invariance of Words and Other Sounds in the Supplementary Motor Area of Macaques. *BioRxiv*, 2020.12.22.424045. https://doi.org/10.1101/2020.12.22.424045

Melchor, J., Vergara, J., Figueroa, T., Morán, I., & Lemus, L. (2021). Formant-Based Recognition of Words and Other Naturalistic Sounds in Rhesus Monkeys. *Frontiers in neuroscience*,

1452.
https://doi.org/10.3389/fnins.2021.728686

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(18), 6792–6797. https://doi.org/10.1073/pnas.1318017111

Middlebrooks, J. C. (2015). Auditory System: Central Pathways☆. In *Reference Module in Biomedical Sciences*. Elsevier. https://doi.org/10.1016/b978-0-12-801238-3.04506-2

Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, *8*(8 JUL), 1–14. https://doi.org/10.3389/fnins.2014.00225

Moore, R. C., Lee, T., & Theunissen, F. E. (2013). Noise-invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise. *PLoS Computational Biology*, *9*(3). https://doi.org/10.1371/journal.pcbi.1002942

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

Muñoz-López, M., Insausti, R., Mohedano-Moriano, A., Mishkin, M., & Saunders, R. C. (2015). Anatomical pathways for auditory memory II: Information from rostral superior temporal gyrus to dorsolateral temporal pole and medial temporal cortex. *Frontiers in Neuroscience*, *9*(APR), 1–21. https://doi.org/10.3389/fnins.2015.00158

Muñoz-López, M., Mohedano-Moriano, A., & Insausti, R. (2010). Anatomical pathways for auditory memory in primates. *Frontiers in Neuroanatomy*, *4*(OCT), 1–13. https://doi.org/10.3389/fnana.2010.00129

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856–869. https://doi.org/10.1038/nrn2478

Ng, C. W., Plakke, B., & Poremba, A. (2009). Primate auditory recognition memory performance varies with sound type. *Hearing Research*, *256*(1–2), 64–74. https://doi.org/10.1016/j.heares.2009.06.014

Nieder, A., & Mooney, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Issue 1789). Royal Society Publishing. https://doi.org/10.1098/rstb.2019.0054

Nusbaum, H., & Magnuson, J. S. (1997). Talker Normalization : Phonetic Constancy as a Cognitive Process. *Talker Variability and Speech Processing*, *June 2016*, 109–132. https://doi.org/10.1121/1.2028337

Ortiz-Rios, M., Kuśmierek, P., DeWitt, I., Archakov, D., Azevedo, F. A. C., Sams, M., Jääskeläinen, I. P., Keliris, G. A., & Rauschecker, J. P. (2015). Functional MRI of the vocalization-processing network in the macaque brain. *Frontiers in Neuroscience*, *9*(APR). https://doi.org/10.3389/fnins.2015.00113

Owren, M. J., Dieter, J. A., Seyfarth, R. M., & 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*(3–4),

218–231. https://doi.org/10.1163/156853992X00615

Oxenham, A. J. (2018). How We Hear: The Perception and Neural Coding of Sound. *Annual Review of Psychology*, *69*(1), 27–50. https://doi.org/10.1146/annurev-psych-122216-011635

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The Processing of Temporal Pitch and Melody Information in Auditory Cortex et al This is consistent with the hierarchy of processing pro-posed for auditory cortex on the basis of recent anatomi- Centre for the Neural Basis of Hearing cal studies in the mac. *Physiology Department Kaas and Hackett*, *36*, 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, *24*(30), 6810–6815. https://doi.org/10.1523/JNEUROSCI.0383-04.2004

Perrodin, C., Kayser, C., Abel, T. J., Logothetis, N. K., & Petkov, C. I. (2015). Who is That? Brain Networks and Mechanisms for Identifying Individuals. *Trends in Cognitive Sciences*, *19*(12), 783–796. https://doi.org/10.1016/j.tics.2015.09.002

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Current Biology*, *21*(16), 1408–1415. https://doi.org/10.1016/j.cub.2011.07.028

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*(3), 367–374. https://doi.org/10.1038/nn2043

Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*, *98*, 262–268. https://doi.org/10.1016/j.cortex.2017.10.020

Rauschecker, J. P., & Romanski, L. M. (2011). Auditory Cortical Organization: Evidence for Functional Streams. *The Auditory Cortex*, 99–116. https://doi.org/10.1007/978-1-4419-0074-6_4

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 97, Issue 22, pp. 11800–11806). National Academy of Sciences. https://doi.org/10.1073/pnas.97.22.11800

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, *268*(5207), 111–114. https://doi.org/10.1126/science.7701330

Recanzone, G. H. (2008). Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey. *Journal of Neuroscience*, *28*(49), 13184–13193. https://doi.org/10.1523/JNEUROSCI.3619-08.2008

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–950. https://doi.org/10.1126/science.7233191

Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations . *The Journal of the Acoustical Society of America*, *103*(1), 602–614. https://doi.org/10.1121/1.421104

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour*, *51*(5), 1007–1015. https://doi.org/10.1006/anbe.1996.0103

Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research 1986 65:1*, *65*(1), 38–48. https://doi.org/10.1007/BF00243828

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

Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, *5*(1), 15–16. https://doi.org/10.1038/nn781

Romanski, Lizabeth M., & Averbeck, B. B. (2009). The Primate Cortical Auditory System and Neural Representation of Conspecific Vocalizations. *Annual Review of Neuroscience*, *32*(1), 315–346. https://doi.org/10.1146/annurev.neuro.051508.135431

Romanski, Lizabeth M., Averbeck, B. B., & Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *Journal of Neurophysiology*, *93*(2), 734–747. https://doi.org/10.1152/jn.00675.2004

Romo, R., Lemus, L., & de Lafuente, V. (2012). Sense, memory, and decision-making in the somatosensory cortical network. *Current Opinion in Neurobiology*, *22*(6), 914–919. https://doi.org/10.1016/J.CONB.2012.08.002

Roy, J. E., Buschman, T. J., & Miller, E. K. (2014). PFC neurons reflect categorical decisions about ambiguous stimuli. *Journal of Cognitive Neuroscience*, *26*(6), 1283–1291. https://doi.org/10.1162/jocn_a_00568

Russ, B. E., Orr, L. E., & Cohen, Y. E. (2008). Prefrontal Neurons Predict Choices during an Auditory Same-Different Task. *Current Biology*, *18*(19), 1483–1488. https://doi.org/10.1016/j.cub.2008.08.054

Sadagopan, S., & Wang, X. (2008). Level invariant representation of sounds by populations of neurons in primary auditory cortex. *Journal of Neuroscience*, *28*(13), 3415–3426. https://doi.org/10.1523/JNEUROSCI.2743-07.2008

Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. In *Hearing Research* (Vol. 307, pp. 42–52). Elsevier. https://doi.org/10.1016/j.heares.2013.07.016

Saunders, J. L., & Wehr, M. (2019). Mice can learn phonetic categories. *The Journal of the Acoustical Society of America*, *145*(3), 1168–1177. https://doi.org/10.1121/1.5091776

Scott, B. H., Leccese, P. A., Saleem, K. S., Kikuchi, Y., Mullarkey, M. P., Fukushima, M., Mishkin, M., & Saunders, R. C. (2017). Intrinsic Connections of the Core Auditory Cortical Regions and Rostral Supratemporal Plane in the Macaque Monkey. *Cerebral Cortex (New York,*

N.Y. : 1991), *27*(1), 809–840. https://doi.org/10.1093/cercor/bhv277

Scott, B. H., Mishkin, M., & Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(30), 12237–12241. https://doi.org/10.1073/pnas.1209685109

Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803. https://doi.org/10.1126/science.7433999

Sinnott, J. M. (1989). Detection and discrimination of synthetic English vowels by Old World monkeys (Cercopithecus, Macaca) and humans. *Journal of the Acoustical Society of America*, *86*(2), 557–565. https://doi.org/10.1121/1.398235

Sinnott, J. M., Brown, C. H., Malik, W. T., & Kressley, R. A. (1997). A multidimensional scaling analysis of vowel discrimination in humans and monkeys. *Perception and Psychophysics*, *59*(8), 1214–1224. https://doi.org/10.3758/BF03214209

Sinnott, J. M., & Kreiter, N. A. (1991). Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans. *Journal of the Acoustical Society of America*, *89*(5), 2421–2429. https://doi.org/10.1121/1.400974

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America*, *117*(1), 305–318. https://doi.org/10.1121/1.1828637

Song, X., Osmanski, M. S., Guo, Y., & Wang, X. (2016). Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(3), 781–786. https://doi.org/10.1073/pnas.1516120113

Takahashi, D. Y., Liao, D. A., & Ghazanfar, A. A. (2017). Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Current Biology*, *27*(12), 1844-1852.e6. https://doi.org/10.1016/j.cub.2017.05.004

Tani, T., Abe, H., Hayami, T., Banno, T., Miyakawa, N., Kitamura, N., Mashiko, H., Ichinohe, N., & Suzuki, W. (2018). Sound frequency representation in the auditory cortex of the common marmoset visualized using optical intrinsic signal imaging. *ENeuro*, *5*(2). https://doi.org/10.1523/ENEURO.0078-18.2018

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, *292*(5515), 290–293. https://doi.org/10.1126/science.1058911

Tippens, P. E. (2005). *Physics*. 782.

Tomlinson, R. W., & Schwarz, D. W. F. (1988). Perception of the missing fundamental in nonhuman primates. *Journal of the Acoustical Society of America*, *84*(2), 560–565. https://doi.org/10.1121/1.396833

Town, S. M., & Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Frontiers in Systems Neuroscience*, *7*(NOV), 1–14. https://doi.org/10.3389/fnsys.2013.00088

Town, S. M., Wood, K. C., & Bizley, J. K. (2018). Sound identity is represented robustly in auditory cortex during perceptual constancy. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07237-3

Tsunada, J., Lee, J. H., & Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *Journal of Neurophysiology*, *105*(6), 2634–2646. https://doi.org/10.1152/jn.00037.2011

Tsunada, Joji, & Cohen, Y. E. (2014). Neural mechanisms of auditory categorization: From across brain areas to within local microcircuits. *Frontiers in Neuroscience*, *8*(8 JUN), 1–10. https://doi.org/10.3389/fnins.2014.00161

Vergara, J., Rivera, N., Rossi-Pool, R., & Romo, R. (2016). A Neural Parametric Code for Storing Information of More than One Sensory Modality in Working Memory. *Neuron*, *89*(1), 54–62. https://doi.org/10.1016/j.neuron.2015.11.026

Wang, X., Merzenich, M. M., Beitel, R., & Schreiner, C. E. (1995). Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. *Journal of Neurophysiology*, *74*(6), 2685–2706. https://doi.org/10.1152/jn.1995.74.6.2685

Wang, Xiaoqin. (2018). Cortical Coding of Auditory Features. *Annual Review of Neuroscience*, *41*(1), 527–552. https://doi.org/10.1146/annurev-neuro-072116-031302

Warren, J. D., Jennings, a. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, *24*(4), 1052–1057. https://doi.org/10.1016/j.neuroimage.2004.10.031

Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., & Recanzone, G. H. (2006). Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *Journal of Neurophysiology*, *96*(6), 3323–3337. https://doi.org/10.1152/jn.00392.2006

Zhao, L., Rad, B. B., & Wang, X. (2019). Long-lasting vocal plasticity in adult marmoset monkeys. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1905). https://doi.org/10.1098/rspb.2019.0817

Zuberbühler, K. (2000). Interspecies semantic communication in two forest primates. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1444), 713–718. https://doi.org/10.1098/rspb.2000.1061