



Research Paper



Selectivity to acoustic features of human speech in the auditory cortex of the mouse

Jennifer L. Mohn^a, Melissa M. Baese-Berk^{b,c}, Santiago Jaramillo^{a,*}

^a Institute of Neuroscience, University of Oregon, Eugene, OR 97403, United States of America

^b Department of Linguistics, University of Oregon, Eugene, OR 97403, United States of America

^c Department of Linguistics, University of Chicago, Chicago, IL 60637, United States of America¹

A B S T R A C T

A better understanding of the neural mechanisms of speech processing can have a major impact in the development of strategies for language learning and in addressing disorders that affect speech comprehension. Technical limitations in research with human subjects hinder a comprehensive exploration of these processes, making animal models essential for advancing the characterization of how neural circuits make speech perception possible. Here, we investigated the mouse as a model organism for studying speech processing and explored whether distinct regions of the mouse auditory cortex are sensitive to specific acoustic features of speech. We found that mice can learn to categorize frequency-shifted human speech sounds based on differences in formant transitions (FT) and voice onset time (VOT). Moreover, neurons across various auditory cortical regions were selective to these speech features, with a higher proportion of speech-selective neurons in the dorso-posterior region. Last, many of these neurons displayed mixed-selectivity for both features, an attribute that was most common in dorsal regions of the auditory cortex. Our results demonstrate that the mouse serves as a valuable model for studying the detailed mechanisms of speech feature encoding and neural plasticity during speech-sound learning.

1. Introduction

Understanding the neural mechanisms underlying speech processing by the auditory system holds paramount importance in our comprehension of how we communicate, of the challenges of learning a new language, and of potential therapies for disorders of speech communication. Despite the wealth of knowledge gained from studies in human subjects (Yi et al., 2021; Oganian et al., 2023), the intricate neural processes involved in speech processing remain challenging to explore fully. Ethical considerations and technical limitations restrict the depth of insight that can be obtained solely through research in humans. A full understanding of how the brain processes and categorizes the acoustic features of speech requires descriptions of how neurons represent these features and how neural connections change when learning novel acoustic categories.

Because this level of investigation is extremely challenging to achieve in human subjects, an appropriate animal model is needed (Kluender, 2000; Lotto et al., 2003). Previous studies have demonstrated that various animal species are capable of learning phonetic categories that share perceptual qualities with humans (Kuhl and Padden, 1983; Kuhl and Miller, 1978; Engineer et al., 2015; Saunders and Wehr, 2019),

suggesting that animal models are appropriate for the study of human speech processing by the auditory system. In particular, the mouse provides an unparalleled level of experimental access to investigate how the activity of individual neurons represents distinct acoustic features and how the circuits these neurons form change during the learning process. However, acoustic communication in mice is markedly different from that in humans. Mouse vocalizations are ultrasonic, the repertoire of “syllables” is more limited, and there is evidence that mouse vocalization is not learned, as vocalizations emitted by deaf mice do not differ in either structure or usage from hearing mice (Portfors, 2007; Portfors and Perkel, 2014; Hammerschmidt et al., 2012; Mahrt et al., 2013). Here, we assessed whether the mouse can serve as a good model organism for studying the neural mechanisms of speech processing, and evaluated whether neurons in distinct regions of the mouse auditory cortex are sensitive to specific features of speech.

We report that mice are able to learn to categorize frequency-shifted human speech sounds that varied according to two features: (1) formant transitions (FT), the spectral change in formant frequencies during a consonant (e.g., the difference between /ba/ and /da/); and (2) voice onset time (VOT), the period of time from the burst of a plosive to the onset of the vocal fold vibration (e.g., the difference between

* Corresponding author.

E-mail address: sjara@uoregon.edu (S. Jaramillo).

¹ Current address.

/ba/ and /pa/). Moreover, we show that neurons across many auditory cortical areas of the mouse (including primary, dorsal, and ventral auditory cortex, as well as temporal association area) are selective to these acoustic features, with the dorso-posterior region of the auditory cortex containing a higher proportion of speech-selective neurons compared to other areas. These combined behavioral and physiological results demonstrate that the mouse can serve as a valid model organism for investigating the encoding of speech features by the auditory system and how these neural representations may change throughout learning.

2. Materials and methods

2.1. Animals

A total of 25 adult C57BL/6J mice (JAX 000664) of both sexes were used in this study. All procedures followed the National Institutes of Health animal care standards and were approved by the University of Oregon Animal Care and Use Committee. Animals were housed in group cages unless implanted for recording, in which case they were individually housed after surgery to prevent damage to implants. Mice used for electrophysiology were between 8 and 18 weeks old at the time of implantation. Mice used for behavioral experiments were between 12 and 13 weeks old at the start of behavioral training. All mice were maintained on a 12 h light/dark cycle.

2.2. Auditory stimuli

Speech sounds were generated using Praat (www.praat.org) and the Python wrapper Parselmouth (Jadoul et al., 2018). Consonant-Vowel (CV) syllables (240 ms long) with different Formant Transitions (FT) and Voice Onset Times (VOT) were generated at frequencies 8 times higher than that of normal human speech to better match the hearing range of mice. The vowel sound had formants at $F_0=800$ Hz, $F_1=5680$ Hz, $F_2=9920$ Hz, and $F_3=20000$ Hz, corresponding to the vowel /a/. VOT sounds were generated by changing the voicing point from 2 ms to 64 ms (logarithmically spaced). The FT sounds were generated by changing the slopes of formants F_2 and F_3 from +9.1 to -9.1 octaves/s for the human-resembling sounds, or from -28.6 to 28.6 octaves/s for the mouse vocalization-resembling sounds. During electrophysiological recordings, only the human-resembling FT slopes were used and only 4 levels of each feature were presented to maintain reasonable recording durations while assuring sufficient number of stimulus presentations for statistical analysis. The average intensity of sounds was calibrated to be around 60 dB-SPL.

In addition to speech sounds, we presented simpler sounds during electrophysiological recordings. One set consisted of 100 ms pure tones (60 and 70 dB-SPL) of 16 different frequencies from 2 kHz to 40 kHz (logarithmically spaced) presented with an interstimulus interval randomized in the range 1–1.4 seconds. We presented 20 repetitions per frequency, randomly sorted. A different set of sounds consisted of sinusoidally amplitude modulated white noise at 11 modulation rates logarithmically spaced between 4 and 128 Hz (100% modulation depth, 500 ms duration, 60 dB-SPL max, 20 repetitions per condition, randomly sorted).

During electrophysiological recordings, auditory stimuli were presented in open-field configuration from a speaker (MF1, Tucker-Davis Technologies) contralateral to the side of recording.

2.3. Behavioral training and assessment

The two-alternative choice sound categorization task was carried out inside single-walled sound-isolation boxes (IAC-Acoustics). Behavioral data was collected using the *taskontrol* platform developed in our laboratory (www.github.com/sjara/taskontrol) using the Python programming language. Mice initiated each trial by poking their noses into the center port of a three-port behavior chamber. After a silent delay of

random duration (150–250 ms, uniformly distributed), a speech sound was presented. Animals were required to choose one of the two side ports to obtain reward (2 μ l of water) according to the identity of sound (e.g., left port for /ba/ and right port for /pa/). If animals chose the incorrect side port, a new trial had to be initiated.

To evaluate psychometric performance for different values of VOT or FT, we synthesized 6 sounds along each feature variation, with 3 sounds indicating reward on the left side and the other 3 sounds indicating reward on the right. The sound on each trial was randomized. During behavioral assessment, a particular cohort of mice was only exposed to variations in either VOT or FT.

Behavioral training consisted of a sequence of stages where animals were first familiarized with the reward ports. They were then required to initiate trials by poking in the center port (obtaining reward after correct choices even if they had made an incorrect choice first), before moving to the stage where they were required to choose the appropriate side port for reward on the first try after each stimulus. During these training stages, animals were presented with only the most extreme values of either VOT and FT, and only after achieving 70% correct for each sound within a session were the sounds with intermediate feature values added to evaluate psychometric performance.

2.4. Surgery

To prepare for head-fixed recordings, mice were implanted with a headbar and bilateral craniotomies were made above auditory cortex. Mice were anesthetized with isoflurane and placed into a stereotactic apparatus (Kopf instruments), a portion of the scalp was then removed and craniotomies were made bilaterally above auditory cortex. The dura mater was removed. A plastic well was fitted and attached to the skull surrounding each craniotomy. Each well was then filled with a silicone elastomer (Sylgard 170, Dow-Corning) to protect the brain. Mice were allowed to recover for at least 3 days before beginning electrophysiological recordings.

2.5. Electrophysiological recordings

Electrophysiological sessions were conducted inside single-walled sound-isolation boxes (IAC Acoustics, Naperville, IL). Electrical signals were collected using Neuropixels 1.0 probes (IMEC) via an NI PXIe-8381 acquisition module and the OpenEphys software (www.open-ephys.org). Probes were coated with a fluorescent dye (DiI: Cat# V22885, or DiD Cat# V22887, Thermo-Fisher Scientific) before penetration of the brain to allow for identification of the recording locations post-mortem. During the experiment, mice were head-fixed and allowed to run atop a wheel. Once the mouse was situated, the silicone elastomer above one of the craniotomies was removed, and the Neuropixels probe was lowered vertically into the auditory cortex with the probe tip reaching approximately 3 mm below the brain surface. The analyzed dataset contains between 1 and 8 recording sessions from each mouse.

2.6. Histology and definition of brain areas

At the conclusion of the experiments, animals were euthanized with euthasol and perfused through the heart with 4% paraformaldehyde. Brains were extracted and left in 4% paraformaldehyde for at least 24 hours before slicing. Brain slices (thickness 100 μ m) were prepared under phosphate-buffered saline using a vibratome (Leica VT1000 S) and imaged using a fluorescence microscope (Axio Imager 2, Carl Zeiss) with a 2.5x objective. To determine the location (brain area and cortical layer) of our recording sites, we manually registered each brain slice to the corresponding coronal section in Allen Mouse Brain Common Coordinate Framework (CCF) (Wang et al., 2020). From this registration, we estimated the location of each recorded cell in CCF coordinates. For our analysis, we combined the Dorsal Auditory Area and the Posterior Auditory Area into a single area we refer to as Dorsal Auditory Area (AudD),

similar to the organization of auditory areas delineated in Franklin and Paxinos (2008).

In addition to parcellating our data according to atlas-defined auditory areas, we also parcellated the recordings according to dorso-ventral (D-V) and anterior-posterior (A-P) coordinates, independent of the boundaries between atlas-defined auditory cortical areas. To do this, we estimated the most anterior, posterior, dorsal, and ventral points of the auditory areas and temporal association area from the Allen CCF Atlas. We used these points to create a rectangle surrounding the auditory areas, and divided this rectangle into equal quadrants: DP, DA, VP and VA.

2.7. Analysis of neural data

Spiking activity of single units was isolated using the spike sorting package Kilosort (Pachitariu et al., 2016), available at <https://github.com/MouseLand/Kilosort>, and manually curated using Phy (Rossant et al., 2016), available at <https://github.com/cortex-lab/phy/>. The resulting data were analyzed using in-house software developed in Python (www.python.org).

Responsiveness to sounds was evaluated by calculating whether there was a statistically significant difference between evoked firing (throughout the duration of each sound) and baseline activity (-200 ms to 0 ms from sound onset), for any of the sounds presented, using a Wilcoxon signed-rank test with Bonferroni correction for multiple-comparisons. To evaluate the pure-tone frequency selectivity of a neuron, a Gaussian function was fit to the spike counts in response to each frequency. The best frequency of a neuron was defined as the peak of this Gaussian function and the tuning width was defined as the width at half-max of this Gaussian fit. For evaluating if cells were significantly responsive to tones within the F2–F3 frequency range (7.1 kHz – 25 kHz), we estimated the cells' responsiveness to each tone (Wilcoxon signed-rank test with $p < \text{Bonferroni corrected } \alpha = 0.05/16 = 0.003$), and a cell was categorized as responsive in the F2–F3 range if it was significantly responsive to any of the tones within that frequency range.

Selectivity to each feature of speech was evaluated for each cell by calculating a selectivity index $SI = \frac{A-B}{A+B}$, where A is the average firing rate from the stimulus that evoked the maximum change from baseline firing rate, and B is the average firing rate from the stimulus that evoked the minimum change from the baseline firing rate. For each SI calculation, we held the other (irrelevant) feature constant (e.g., using the four stimuli represented on the bottom row of Fig. 2C). We chose to hold the irrelevant feature at either the maximum or the minimum value by assessing the condition in which the greatest change was observed between baseline and stimulus-evoked firing rates. To assess if the selectivity of a neuron to a specific feature was statistically significant, we performed a permutation test with 2000 repetitions, where the stimulus associated with each trial was randomized. A cell was categorized as selective for a given feature if the permutation test p-value was less than 0.05. A cell was determined to exhibit mixed-selectivity if it was significantly selective to both VOT and FT.

3. Results

3.1. Mice can discriminate features of speech

To test whether the mouse could be an appropriate animal model for studying the discrimination of speech sound features, we first tested whether mice are capable of learning to categorize human speech sounds. We generated consonant-vowel human speech sounds that were frequency shifted into the mouse hearing range (Fig. 1B). These sounds varied in one of two features: formant transitions (FT), the spectral change in formant features during a consonant (e.g., the difference between /ba/ and /da/); or voice onset time (VOT), the period of time from the burst of a plosive to the onset of the vocal fold vibration (e.g., the difference between /ba/ and /pa/). We trained mice to categorize

speech sounds in a two-alternative choice task (Fig. 1A), based either on changes in VOT or FT, starting with extreme values of these features and then adding sounds with intermediate feature values to characterize the psychometric performance for each feature variation.

Mice were able to successfully categorize sounds according to both VOT (Fig. 1C,D) and FT (Fig. 1E,F), as illustrated by the approximately 80% correct on the extremes of both VOT value and FT slope. As expected, intermediate values of the speech features resulted in performance closer to chance level (50%). While all of the VOT animals successfully categorized sounds by VOT (9 out of 9 total mice), fewer of the FT animals were able to successfully categorize sounds by FT throughout the training period (3 out of 9 total mice, after 106 training sessions). It took on average 49 sessions (min = 30 sessions, max = 73 sessions) for the animals trained to categorize VOT to achieve 70% correct performance categorizing the two extremes of VOT sounds presented (VOT = 2 ms and VOT = 64 ms). The animals that learned to categorize FT took longer to succeed at categorizing the extreme FT sounds presented (+/-9.1 oct/s), on average 64 sessions (min = 56 sessions, max = 76 sessions). Once animals learned the task, they performed 991 trials per session on average.

Since the mice were less successful at learning to categorize FT sounds compared to VOT sounds, we tested whether using sounds with FT slopes that more closely resemble the frequency sweeps that occur in mouse vocalizations would result in better performance. To do this, we generated a new set of FT sounds, such that the extreme slopes were +/-28.6 oct/s, which more closely resemble the features of mouse vocalizations (Portfors, 2007), compared to the +/-9.1 oct/s of the human-resembling FT sounds. We trained a new cohort of mice to discriminate these new FT sounds. A larger fraction of mice in this new cohort was able to learn to categorize FT extremes (8 out of 9 total mice), and they reached 70% correct performance more quickly than the original cohort, in an average of 48 sessions (min = 26 sessions, max = 67 sessions). Overall, these results suggest that while the categorization performance of speech features depends on the specific FT slopes used, mice can learn to discriminate basic features of human speech sounds.

3.2. Neurons across the auditory cortex of mice are responsive to speech sounds

After verifying that mice were able to perceive differences in acoustic features of speech, we wanted to characterize neural responses to these sounds across auditory cortical areas of the mouse. We recorded extracellular electrophysiological responses of neurons from primary and non-primary auditory cortex, as well as the temporal association cortex (TeA) using Neuropixels 1.0 probes (Fig. 2A, 2B). We presented 12 consonant-vowel stimuli that varied in both FT and VOT (Fig. 2C) to awake, naive, head-fixed mice ($n = 7$). Additionally, prior to presenting speech sounds, we presented pure tones of varying frequencies and amplitude modulated noise (AM) of varying rates during each recording session in order to characterize the responses of neurons to simpler sound features. We recorded a total of 1009 single neurons from the auditory cortex and surrounding areas. Of these, 563 (55.8%) were responsive to any of the sounds presented (Fig. 2D, red + blue), and 347 cells were responsive to any of the speech sounds we presented (34.3% of all cells, 61.6% of sound responsive cells; Fig. 2D, red). A cell was classified as responsive to a class of sounds presented if it had a statistically significant difference between the evoked firing rate and baseline firing rate (Wilcoxon signed-rank test, $p < 0.05$, Bonferroni corrected alpha within each class of sounds given the number of stimuli presented). While the resolution of our method for finding the location of each recording may not be precise enough to provide the exact laminar location of each cell, we evaluated whether cells were likely to come from superficial vs. deep cortical layers. We found that most of the recorded cells (92.9%) were located in deep cortical layers (layers 5-6), while only 7.1% were in superficial layers (layers 1-4).

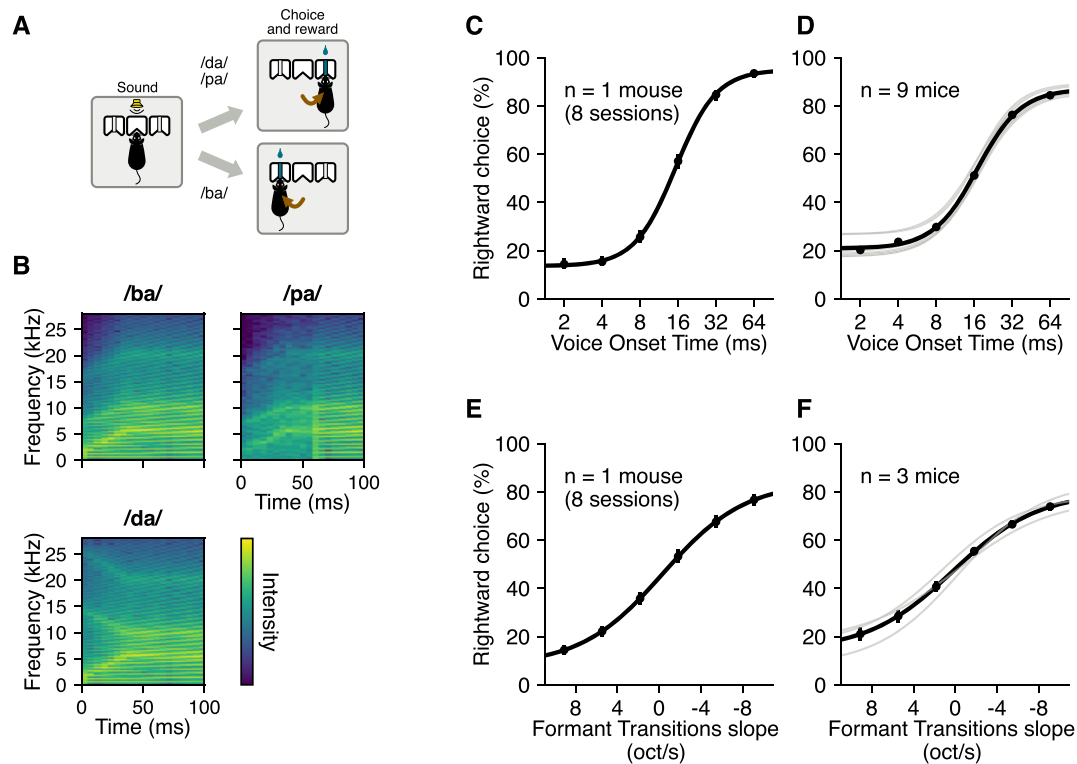


Fig. 1. Mice can categorize speech sounds. (A) Two-alternative choice speech categorization task. Mice initiated a trial by entering the center port, triggering a sound presentation. They then had to choose the left or right port based on which sound was presented: left for short VOT and upwards sloping FT (/ba/), right for long VOT (/pa/), or downwards sloping FT, (/da/). A correct choice resulted in a water reward, an incorrect choice ended the trial without reward. (B) Spectrograms of /ba/, /da/, and /pa/ sounds presented (shifted to the mouse hearing range). Sounds were 240 ms long (only the first 100 ms shown here). Sounds with VOTs intermediate to /ba/ and /pa/ and FTs intermediate to /ba/ and /da/ were also presented. (C) Psychometric curve from a single mouse trained to categorize sounds by VOT. Error bars indicate 95% confidence interval. (D) Average psychometric curve from the VOT cohort (black) and psychometric fit for each mouse (gray). Error bars indicate SEM across mice. (E) As in C, for a mouse trained to categorize sounds by FT. (F) As in D, for the cohort trained to categorize FT.

Once we knew there were cells that were responsive to these speech sounds, we tested whether the proportions of cells that were speech responsive were different across the cortical areas recorded. Using functional and anatomical mapping of the mouse auditory cortex, previous studies have parcellated mouse auditory cortex into anywhere from 3 to 6 sub-areas (Ceballos et al., 2019; Romero et al., 2020; Issa et al., 2017; Tsukano et al., 2016). Because of these discrepancies, we chose to follow two distinct approaches for investigating differences in neural responses across auditory cortical regions. We first evaluated neural activity according to cortical areas defined in the Allen Mouse Brain Atlas: primary auditory cortical area (AudP), dorsal and posterior auditory areas (combined and referred to here as AudD), ventral auditory area (AudV) and TeA, after registering each mouse brain to the atlas. We also used a different approach where we parcellated the whole auditory cortical region according to dorso-ventral (D-V) and anterior-posterior (A-P) coordinates, independent of the boundaries between atlas-defined auditory cortical areas.

When we compared the fraction of speech responsive cells (Fig. 2E, red) as a proportion of total cells (Fig. 2E, all cells) across each atlas-defined area, we found that there were no statistically significant differences in the fraction of speech responsive cells between any of these areas (Table 1). Further, when we restricted our analysis to sound responsive cells (Fig. 2E, red + blue), the proportion of cells that were responsive to the speech sounds (Fig. 2E, red) was also similar across auditory cortical areas (Table 2). Restricting our analyses to cells from deep cortical layers, where most of our recordings were located (Table 3), did not alter the results. We did not have enough cells in other layers to perform analyses for those layers.

In contrast, when we parcellated our data by quadrants in the A-P and D-V directions, instead of atlas-defined areas, we found differences in the proportion of speech responsive cells (Fig. 2F, red) between cortical regions. As a proportion of total cells in each quadrant (Fig. 2F, all cells), we found that the dorso-posterior region had significantly more speech responsive cells (79.5%) compared to each of the other regions (60.0% DA, 55.2% VA, 52.0% VP, Table 4). Further, when we restricted our analysis to the sound responsive cells (Fig. 2F, red + blue) we found that the DP region still had a higher proportion of speech responsive cells compared to VA, but not DP or VP (Table 5). Restricting the analyses to cells located in deep cortical layers, which corresponds to the majority of recorded cells (Table 6), yielded equivalent results. In summary, while we observed no differences between the atlas-defined auditory cortical areas in the proportion of speech responsive cells out of either the total cells recorded or of sound responsive cells, neurons in the dorso-posterior region of the auditory cortex were more likely to be speech responsive compared to the ventral regions.

Having established the presence of speech responsive cells in each of the regions we sampled, we wanted to test whether cells showed selectivity to features of speech or if they were indiscriminately responsive to the speech sounds we presented. We found some cells that changed their firing rate drastically when VOT changed, but had little change in their firing when FT changed (Fig. 3A). Conversely, there were cells that changed their firing rate drastically when FT changed, but their activity had little change when VOT changed (Fig. 3B). There were also cells whose firing rates varied depending on both FT and VOT (Fig. 3C). Finally, some cells were responsive to the speech sounds, but their responses did not change when varying either feature (Fig. 3D).

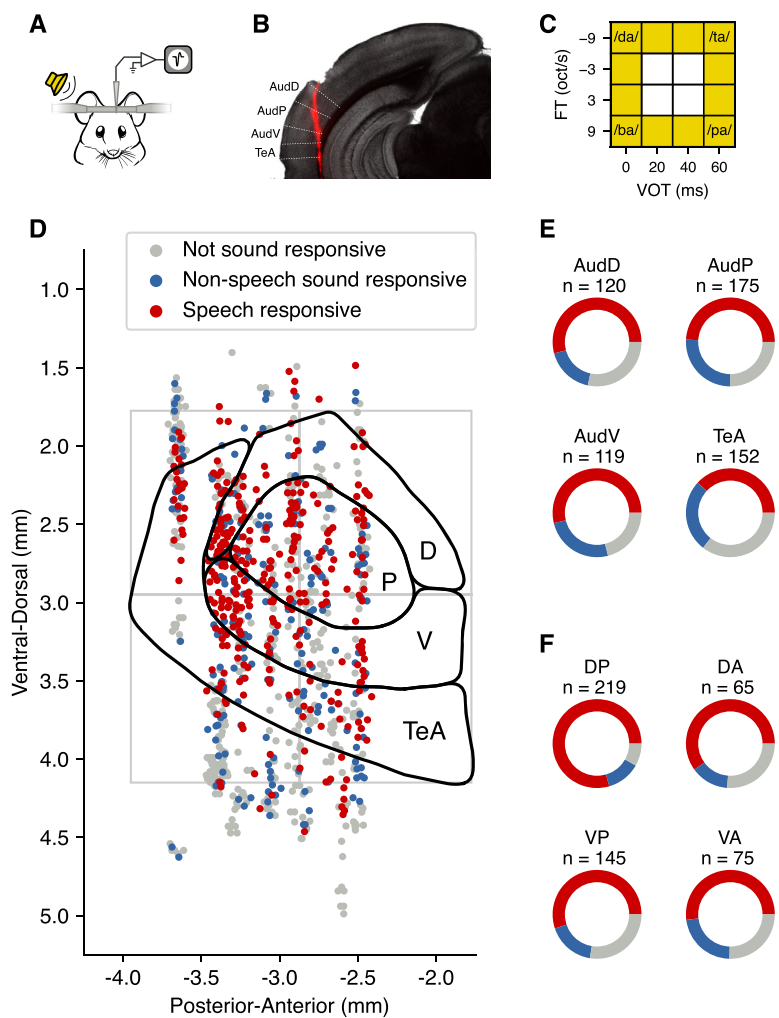


Fig. 2. Electrophysiological recordings from multiple auditory cortical areas. (A) Single-neuron extracellular electrophysiology was recorded from awake, head-fixed mice ($n = 7$) using Neuropixels 1.0 probes. Mice naive to the behavioral task listened passively while sounds were presented contralaterally to the recording site. (B) Example electrode penetration spanning multiple areas of the auditory cortex, shown on a coronal slice. (C) Sounds presented during the recorded sessions varied in FT and VOT. Colored squares indicate presented sounds, empty squares indicate combinations that were not presented. (D) Locations of all cortical cells recorded across all mice, colored according to their responsiveness to sounds. Approximate boundaries of auditory brain areas according to the Allen Mouse Brain Atlas. P: primary, D: dorsal (combining AudPo and AudD), V: ventral, TeA: temporal association area. Gray lines show the borders of each quadrant used for the analysis in F. (E) The proportion of speech responsive cells was similar between each of the atlas-defined areas. (F) Parcellating the auditory cortex into regions along the A-P/D-V axes yielded a higher proportion of speech responsive cells in the dorso-posterior region.

Table 1
The proportion of speech responsive cells with respect to total cells recorded is similar across auditory areas. Right: p-values from comparisons between areas (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Area	Percentage	Responsive/Total		AudP	AudD	AudV	TeA
AudP	49.1%	86/175		—	0.409	0.476	0.057
AudD	54.2%	65/120		—	—	1	0.01
AudV	53.8%	64/119		—	—	—	0.014
TeA	38.2%	58/152		—	—	—	—

Table 2
The proportion of speech responsive cells with respect to sound responsive cells is similar across auditory areas. Right: p-values from comparisons between areas (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Area	Percentage	Speech resp/ Sound resp		AudP	AudD	AudV	TeA
AudP	65.6%	86/131		—	0.133	0.775	0.336
AudD	75.6%	65/86		—	—	0.321	0.02
AudV	68.1%	64/94		—	—	—	0.231
TeA	59.2%	58/98		—	—	—	—

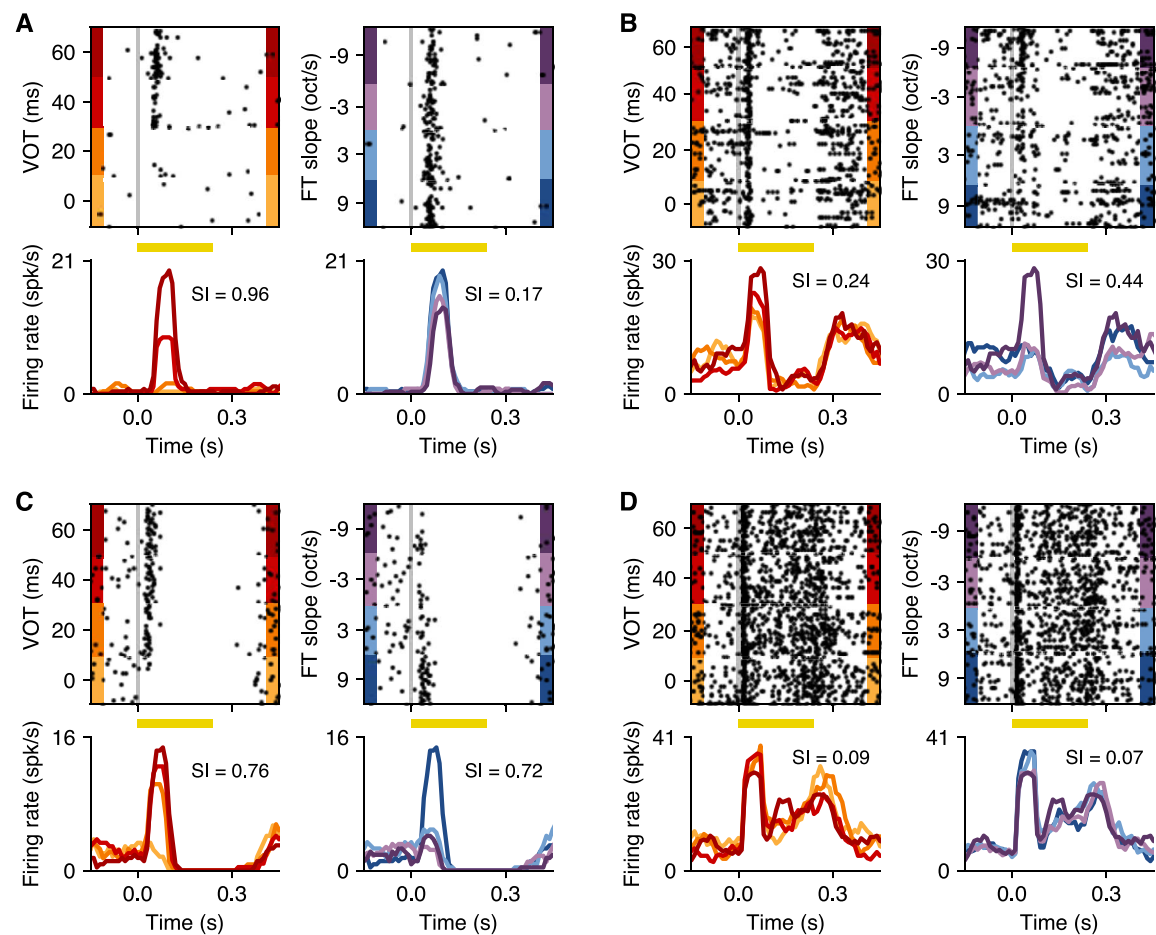


Fig. 3. Neurons in the auditory cortex are selective to speech sounds. (A) Sound evoked responses from an example neuron that is significantly selective to VOT (left panel), but not to FT (right panel). The yellow horizontal bars represent when the auditory stimulus was presented. SI: selectivity index. (B) Example neuron that is significantly selective to FT, but not VOT. (C) Example neuron that is significantly selective to both VOT and FT, which we classify as mixed-selective. (D) Example neuron that is responsive to speech sounds, but not selective for either VOT or FT features.

Table 3
Percentage of cells recorded from superficial vs. deep cortical layers. The large majority of cells recorded in each auditory area came from deep cortical layers (L5 & L6).

	L1–4	L5–6	N total
AudP	4.6%	95.4%	175
AudD	11.7%	88.3%	120
AudV	0%	100%	119
TeA	4.6%	95.4%	152

3.3. Auditory cortical neurons change their firing rate in response to changing VOT and FT

To enable comparing selectivity across neurons and auditory cortical regions, we calculated for each neuron a selectivity index (SI) associated with each feature, while we held the other (irrelevant) feature constant (e.g., using the four stimuli represented on the bottom row of Fig. 2C). We chose to hold the irrelevant feature at either the minimum or the maximum value by assessing the condition in which the greatest change was observed between baseline and stimulus-evoked firing rates. The SIs shown for each neuron in Fig. 3 support the qualitative observations described above illustrating the existence of cells selective to a single feature, cells that had mixed-selectivity (i.e., selectivity to both features), and cells that were responsive to speech sounds but not selective to either feature.

3.4. Selectivity to speech features is greater in the dorso-posterior region of auditory cortex

To compare speech selectivity across areas of the auditory cortex, we first assessed whether the proportion of cells that were selective to each feature was different across the atlas-defined brain areas we sampled. All areas we tested contained cells that were selective to VOT (Fig. 4A). We found that of all the cells recorded, approximately 20-30% of cells in each area were selective to VOT (Fig. 4B), and there were no statistically significant differences in the proportion of cells that were selective to VOT between any of the areas (Table 7). There were also cells selective to FT in each of the areas we tested (Fig. 4E). Similar to VOT, there were no statistically significant differences in the proportion of FT selective cells between any of the areas we sampled (Table 8). Each area had between approximately 5-15% of cells that were FT selective (Fig. 4F). Similarly, when we assessed the proportion of cells selective to VOT or FT in each area with respect to the number of sound responsive cells or the number of speech responsive cells we found no statistically significant differences between any of the areas for either feature (Tables 7 & 8). In addition to calculating the proportion of cells that were selective to speech features in each area, we tested whether the strength of that selectivity was different between areas (Fig. 4A, 4E). When we compared the distribution of selectivity indices between areas among selective cells, we found that there was no difference in the distribution of VOT selectivity indices between areas (Kruskal-Wallis, $p = 0.426$). Likewise, there was no difference in the strength of FT selectivity between any of the areas we sampled (Kruskal-Wallis $p = 0.907$).

Table 4

The proportion of speech responsive cells with respect to all recorded cells is higher in the dorso-posterior region compared to the other regions. Right: p-values from comparisons between regions.

Region	Percentage	Responsive/Total		DP	DA	VP	VA
DP	79.5%	174/219		—	0.003	<0.001	<0.001
DA	60.0%	39/65		—	—	0.549	0.395
VP	55.2%	80/145		—	—	—	0.671
VA	52.0%	39/75		—	—	—	—

Table 5

The proportion of speech responsive cells with respect to sound responsive cells is higher in the dorso-posterior region compared to the ventral-anterior region. Right: p-values from comparisons between regions (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Region	Percentage	Speech resp/ Sound resp		DP	DA	VP	VA
DP	86.6%	174/201		—	0.363	0.025	0.005
DA	81.2%	39/48		—	—	0.536	0.256
VP	76.2%	80/105		—	—	—	0.451
VA	69.6%	39/56		—	—	—	—

Table 6

Percentage of cells recorded from superficial vs. deep cortical layers. The large majority of cells recorded in each region came from deep cortical layers (L5 & L6).

	L1–4	L5–6	N total
DP	5.9%	94.1%	219
DA	4.6%	95.4%	65
VP	2.1%	97.9%	145
VA	0%	100%	75

In contrast to the results from atlas-defined auditory cortical areas, when we parcellated our data into A-P/D-V quadrants we found clear differences in the selectivity to speech features between regions of the auditory cortex (Fig. 4C, 4G). We found that out of all cells, the dorso-posterior quadrant had a significantly higher proportion of cells selective to VOT compared to any of the other quadrants (DP: 40.6%, DA: 18.5%, VP: 19.3%, VA: 18.7%, Table 9). To verify that these results were not being driven by a single animal, we assessed the proportion of VOT selective cells for each region recorded in individual animals and found that this trend was common across most animals (Fig. 4D). Similarly, when we compared the proportion of FT selective cells between these regions, we found that the dorso-posterior region had a significantly higher proportion of FT selective cells compared to all the other regions (DP: 21.9%, DA: 6.2%, VP: 4.8%, VA: 8.0%), and found no statistically significant differences across any of the other regions (Table 10). To verify that this larger proportion of FT selective cells in DP compared to the other regions was not driven by a single animal, we assessed the proportion of FT selective cells in each region for individual animals and found that this trend was consistent across most animals (Fig. 4H). Lastly, the observation that the DP region had a higher proportion of VOT and FT selective cells compared to other regions was apparent also when we restricted our analysis to only sound responsive cells or speech responsive cells (Tables 9, 10), suggesting that this result is not simply the consequence of the cells' responsiveness to the speech sounds presented.

In summary, while we observed no differences in selectivity to speech features between atlas-defined auditory cortical areas, when we parcellated our recorded region into quadrants, we found that neurons in the dorso-posterior region of auditory cortex were more likely to be selective to both VOT and FT compared to the other regions of auditory cortex.

3.5. Selectivity to speech features is not fully explained by frequency selectivity

We next wanted to test whether the differences we found in selectivity to speech features across regions could be explained simply by differences in selectivity to sound frequency. In particular, since FT sounds are composed of varying frequencies in the F2 and F3 formants (7.1 kHz – 25 kHz in the sounds presented, Fig. 1B), it is possible that the differences we observed across regions in selectivity to FT could be explained simply by differences in frequency selectivity given the sounds we presented, rather than selectivity to the speech feature. We focused on FT and not VOT selectivity, because the changes in VOT sounds occur across all of the frequencies, and therefore one does not expect frequency selectivity to explain differences in VOT selectivity.

To test whether selectivity to FT may depend on frequency selectivity, we first calculated the tuning width and best frequency (BF) for each sound responsive cell. Successful estimation of a Gaussian tuning curve was possible in 194 cells. An analysis of these cells yielded a weak negative correlation between the tuning width and the strength of FT selectivity (Spearman $r = -0.18$, $p = 0.005$), suggesting that cells more narrowly tuned to pure tones may have stronger FT selectivity. To test if this might explain the differences in FT selectivity between regions, we compared the distribution of tuning widths between regions. Of the cells where a tuning curve was successfully estimated, we found that the median tuning width (in octaves) in DP was 1.4 ($n = 95$), DA was 1.1 ($n = 28$), VP was 1.7 ($n = 44$) and VA was 2.0 ($n = 27$). When comparing tuning widths between regions, we found that cells in VA had a significantly wider average tuning width compared to those in each of the other regions, while no other significant differences between regions were present (Mann-Whitney U test, DP vs. VA: $p < 0.001$, DA vs. VA: $p < 0.001$, VP vs. VA: $p = 0.006$, $p > 0.05$ all other comparisons). Thus, while differences in tuning width may partially explain the differences in FT selectivity between DP and VA, the stronger FT selectivity in DP compared to DA and VP cannot be explained simply by tuning width differences.

We then compared the distribution of BFs between regions. The distribution of BFs between auditory areas was not significantly different ($p = 0.13$, Kruskal-Wallis), with the median BF of each area falling within the F2–F3 frequency range: 13 kHz for AudP, 12.6 kHz for AudD, 12.6 kHz for AudV and 10.4 kHz for TeA. A comparison between the quadrants, in contrast, yielded a statistically significant difference in the median best frequency across regions ($p < 0.001$, Kruskal-Wallis), yet the median value for each region was still within the F2–F3 range:

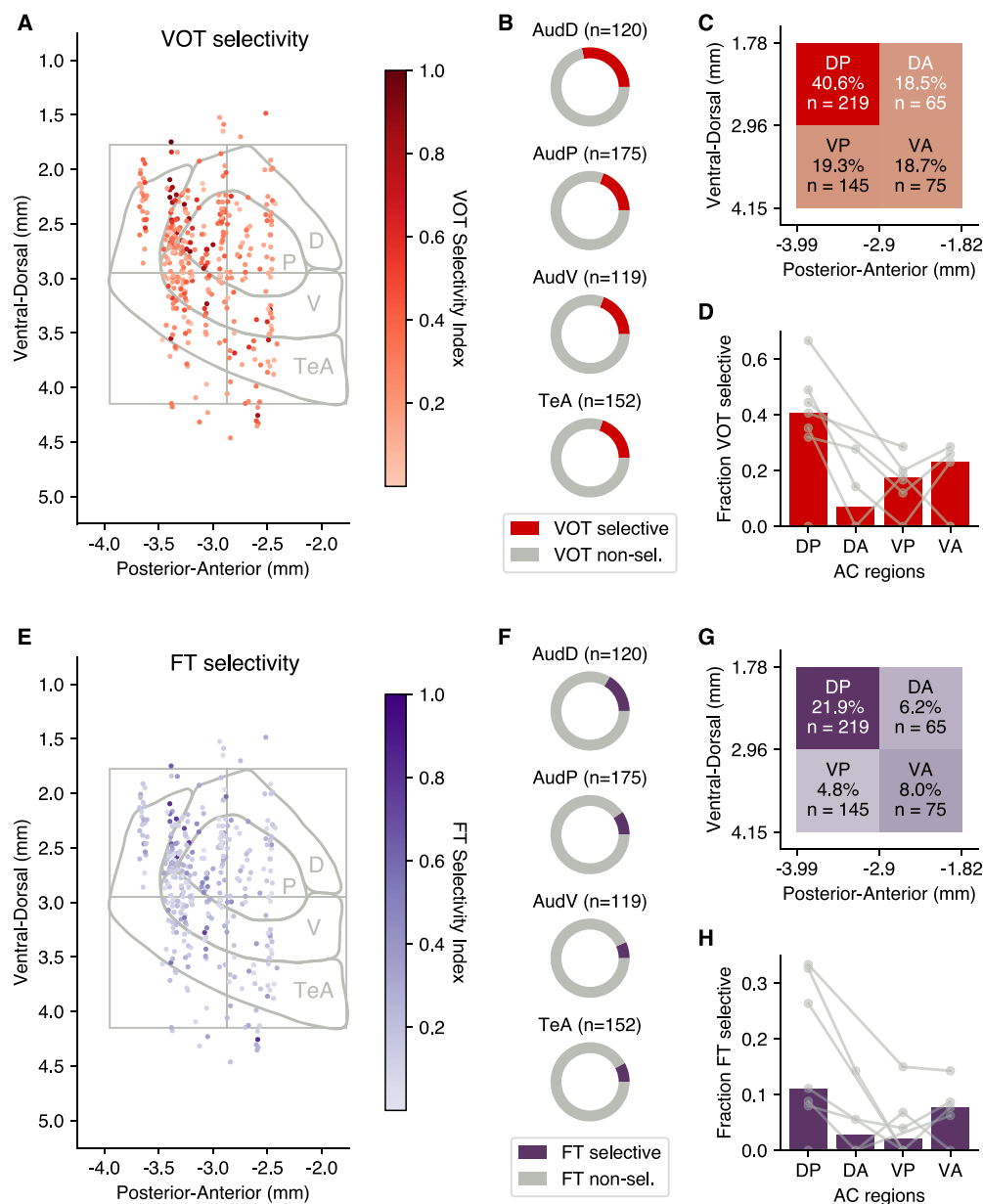


Fig. 4. Selectivity to speech features is greater in the dorso-posterior region of auditory cortex. (A) Locations of speech responsive cells recorded, colored according to VOT selectivity indices. (B) The proportion of VOT selective cells is similar across atlas-defined areas. (C) The proportion of VOT selective cells is greater in the dorso-posterior region of auditory cortex. (D) Proportion of VOT selective cells in each region for each mouse. Dots and connecting lines indicate values for individual animals. Bars indicate medians across animals. Regions with fewer than 3 recorded cells were excluded. (E-H) As in A-D, for FT selectivity.

Table 7
The proportion of cells selective to VOT out of total cells was similar between atlas-defined auditory areas. Right: p-values from comparisons between areas of VOT selective cells out of the total cells (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Area	Of total	VOT Sel/Total	Of sound responsive	Of speech responsive		AudP	AudD	AudV	TeA
AudP	19.4%	34/175	26.0%	39.5%		—	0.091	1	1
AudD	28.3%	34/120	39.5%	52.3%		—	—	0.129	0.114
AudV	19.3%	23/119	24.5%	35.9%		—	—	—	1
TeA	19.7%	30/152	30.6%	51.7%		—	—	—	—

11.1 kHz for DP, 23.1 kHz for DA, 10.3 kHz for VP and 24.3 kHz for VA. However, one would expect that any neuron that responds to tones within the F2–F3 frequency range could be selective to FT, even if their best frequency is not in this range. Therefore, we focused on cells that had a statistically significant response to pure tone frequencies in the

F2–F3 range. When we restrict our analysis of FT selectivity to these subsets of cells, we found that DP had 29.4% of cells selective to FT, while DA only had 9.1%, VP had 10.7%, and VA had 17.2%. Even though in this case only differences between DP and VP reached statistical significance (Table 11), these trends are consistent with the results

Table 8

The proportion of cells selective to FT out of total cells was similar between atlas-defined auditory areas. Right: p-values from comparisons between areas of FT selective cells out of the total cells (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Area	Of total	FT Sel/Total	Of sound responsive	Of speech responsive		AudP	AudD	AudV	TeA
AudP	9.7%	17/175	13.0%	19.8%		—	0.106	0.403	0.697
AudD	16.7%	20/120	23.3%	30.8%		—	—	0.026	0.036
AudV	6.7%	8/119	8.5%	12.5%		—	—	—	0.817
TeA	7.9%	12/152	12.2%	20.7%		—	—	—	—

Table 9

The proportion of cells selective to VOT out of total cells was greater in the dorso-posterior region of auditory cortex compared to the other regions. Right: p-values from comparisons between regions of VOT selective cells out of the total cells (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Region	Of total	VOT Sel/Total	Of sound responsive	Of speech responsive		DP	DA	VP	VA
DP	40.6%	89/219	44.3%	51.1%		—	0.001	<0.001	<0.001
DA	18.5%	12/65	25.0%	30.8%		—	—	1	1
VP	19.3%	28/145	26.7%	35.0%		—	—	—	1
VA	18.7%	14/75	25.0%	35.9%		—	—	—	—

Table 10

The proportion of cells selective to FT out of total cells was greater in the dorso-posterior region of auditory cortex compared to the other regions. Right: p-values from comparisons between regions FT selective cells out of the total cells (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Region	Of total	FT Sel/Total	Of sound responsive	Of speech responsive		DP	DA	VP	VA
DP	21.9%	48/219	23.9%	27.6%		—	0.003	<0.001	0.006
DA	6.2%	4/65	8.3%	10.3%		—	—	0.741	0.751
VP	4.8%	7/145	6.7%	8.8%		—	—	—	0.374
VA	8.0%	6/75	10.7%	15.4%		—	—	—	—

Table 11

The first column shows the proportion of cells responsive to tones in the F2–F3 frequency range. The other columns show that the proportion of cells selective to FT out of cells responsive to tones in the F2–F3 frequency range was significantly higher in the dorso-posterior region of auditory cortex compared to the ventro-posterior region. Right: p-values from comparisons between regions (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Region	F2–F3/Total	Of F2–F3	FT Sel/F2–F3		DP	DA	VP	VA
DP	54.3%	29.4%	35/119		—	0.063	0.007	0.246
DA	33.8%	9.1%	2/22		—	—	1	0.684
VP	38.6%	10.7%	6/56		—	—	—	0.499
VA	38.7%	17.2%	5/29		—	—	—	—

we found using the whole data set. While there was a higher proportion of neurons responsive to pure tone frequencies in the F2–F3 range in DP compared to the other regions, even among these F2–F3 responsive neurons there are more FT selective neurons in DP compared to the other regions. These results suggest that the differences in FT selectivity we found between regions cannot be fully explained by the responsiveness of cells to frequencies of pure tones in the frequency range that varies between FT stimuli.

3.6. Mixed-selectivity is more prevalent in the dorsal portion of auditory cortex

In addition to cells that were selective to just VOT or FT, we found cells that changed their firing rate in response to both features (Fig. 3C). We defined a cell as having mixed-selectivity if it was significantly selective to both VOT and FT. If the cell was significantly selective to only one of the two features, it was classified as single-feature selective. Of the speech-selective cells, we found cells with mixed selectivity in

each of the areas we examined (Fig. 5A). We then wanted to test if the distribution of mixed-selective cells was different across auditory cortical areas. When we parcellated our data by atlas-defined cortical areas, we found that out of the speech responsive cells, there were significantly fewer mixed-selective cells in AudV (6.2%) compared to AudD (24.6%, Table 12, Fig. 5B). While the proportion of mixed-selective cells was trending lower in AudV compared to both AudP and TeA, there was no statistically significant difference in the proportion of mixed-selective cells out of speech responsive cells among these areas (Table 12). Restricting the analysis to only speech-selective cells, yielded similar trends, although there were no statistically significant differences between the areas.

As before, we wanted to test whether there were any differences in the proportion of mixed-selective cells across the A-P/D-V axes, by parcellating the auditory cortex region into quadrants (Fig. 5C). We found that the dorso-posterior quadrant had over twice the proportion of mixed-selective cells compared to each of the other areas (DP: 23.4%, DA: 7.7%, VP: 5.0%, VA: 7.7%), although only the difference between

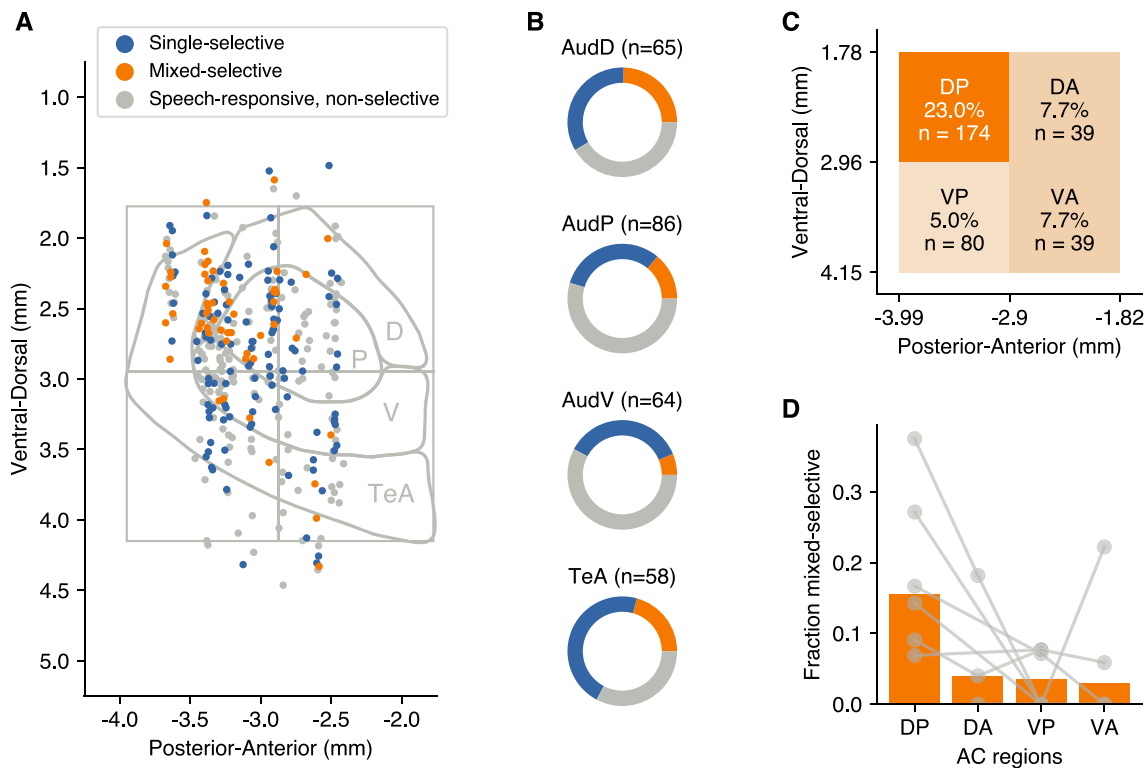


Fig. 5. Mixed selectivity is more prevalent in the dorsal auditory areas. (A) Locations of speech responsive cells recorded, colored according to their selectivity to speech features. (B) The proportion of mixed-selective cells is higher in AudD compared to AudV. (C) The proportion of mixed-selective cells is 3-5 times as high in the DP region compared to the other auditory cortical regions. (D) Proportion of mixed-selective cells out of speech responsive cells for individual animals. Dots and connecting lines indicate values for individual animals. Instances in which an individual had fewer than 3 recorded cells in a region were excluded. Bars indicate the median proportion for each region averaged across animals.

Table 12
The proportion of cells with mixed-selectivity out of speech responsive cells was higher in dorsal auditory area compared to the ventral auditory area. Right: p-values from comparisons between areas of mixed-selective cells out of the speech responsive cells (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Area	Of speech responsive	Mixed/Speech Resp	Of speech selective		AudP	AudD	AudV	TeA
AudP	14.0%	12/86	30.8%		—	0.138	0.182	0.363
AudD	24.6%	16/65	42.1%		—	—	0.006	0.67
AudV	6.2%	4/64	14.8%		—	—	—	0.03
TeA	20.7%	12/58	30.8%		—	—	—	—

Table 13
The proportion of cells with mixed-selectivity out of speech responsive cells was higher in dorso-posterior region compared to the ventral-posterior region. Right: p-values from comparisons between regions of mixed-selective cells out of the speech responsive cells (Fisher Exact Test, Bonferroni corrected $\alpha = 0.05/6 = 0.008$).

Region	Of speech responsive	Mixed/Speech Resp	Of speech selective		DP	DA	VP	VA
DP	23.0%	40/174	41.2%		—	0.045	<0.001	0.045
DA	7.7%	3/39	23.1%		—	—	0.682	1
VP	5.0%	4/80	12.9%		—	—	—	0.682
VA	7.7%	3/39	17.6%		—	—	—	—

regions DP and VP was statistically significant (Table 13). This trend, where region DP had a higher proportion of mixed-selective cells was consistent across individual animals (Fig. 5D). In summary, we found there to be a higher proportion of mixed-selective cells in the atlas-defined dorsal auditory cortex compared to the ventral area, and when parcellating the auditory cortex into quadrants the results followed a similar trend with the dorso-posterior region having a higher proportion of mixed-selective cells than the ventro-posterior region.

4. Discussion

Our behavioral results demonstrate that mice are able to discriminate changes in key acoustic features of human speech, namely, formant transitions and voice onset time. Our finding extend recent results showing that mice can learn to discriminate speech sounds like “sad” vs. “dad” (O’Sullivan et al., 2020) or categorize consonant-vowel pairs for consonants /g/ and /b/ and generalize these categories after changing

the speaker or vowel context (Saunders and Wehr, 2019). Our study also demonstrates that subregions of the auditory cortex of the mouse are specialized in the processing of specific acoustic features present in human speech. These results complement extensive literature on speech processing in other animals, including rat, cat, and primate (Porter et al., 2011; Centanni et al., 2013; Engineer et al., 2015; Eggermont, 1995; Wong and Schreiner, 2003; Steinschneider et al., 2003; Tsunada et al., 2011), and enable the use of advanced tools readily available in the mouse to dissect the neural circuits responsible for speech processing. While acoustic communication in humans involves speech-production anatomical features and complex grammatical structures that are not expected to be well modeled in the mouse, the results above provide evidence that the auditory system of the mouse can effectively process acoustic features of human speech, allowing for a detailed investigation of the neural mechanisms underlying speech processing and speech-sound learning.

A potential limitation of studying speech sound processing using the mouse is the clear difference in frequency hearing range between humans and mice. To address this difference, human speech sounds were shifted up in frequency to the mouse hearing range. The assumption here is that the features of speech under study can be maintained while varying other sound features. While this adjustment can be easily performed for the perception of spectral features, it is more challenging to identify appropriate adjustments for temporal features. For instance, our evaluation of categorization performance using different sets of formant transition slopes yielded varied results, with mice learning the task better when using slopes closer to those naturally occurring in their vocalizations, compared to those common in human speech. It remains an open question if the difference in task performance we observed between FT and VOT categorization may be attributed to having found fewer cells selective to FT compared to VOT for the specific sounds we tested. Overall, these findings further emphasize the importance, when conducting animal studies of human speech processing, of selecting sounds that contain the features of interest within an ethologically relevant range given a particular experimental animal model (Kim and Bao, 2009; Bennur et al., 2013).

In studies with human subjects, functional imaging and neurophysiological investigations have revealed the existence of distinct cortical regions responsible for processing various speech attributes, such as phonemic information, prosody, and voice characteristics (Yi et al., 2019). Therefore, for an animal model to be valid for the study of speech perception, it would be expected that distinct neural circuits specialize in the processing of different features of speech. Consistent with this idea, we found that the dorso-posterior region of the auditory cortex of the mouse contains a higher proportion of neurons that are selective to the speech features tested, compared to other cortical regions.

Differential roles across neural circuits, however, were not readily apparent when we grouped neurons according to atlas-defined cortical areas. Several possibilities could explain these discrepancies. First, a technical limitation of our study is that brain areas were defined by registering each mouse brain to the atlas, yet it is known that the extent of these areas (when functionally defined) can vary significantly across animals (Narayanan et al., 2023). Consequently, the assignment of each neuron to a given brain area may lack precision, potentially affecting the interpretation of comparisons across atlas-defined areas. An additional limitation of this study is that the most anterior regions of the auditory cortex were not fully sampled in our data set, and thus we cannot make conclusions about these regions. Further, since most of our cells came from deep cortical layers, we were unable to draw conclusions about laminar differences in speech feature selectivity in this study. We found, however, that neither frequency tuning nor laminar location fully explained the differences in selectivity to speech features we observed across auditory cortical regions. This suggests that, even without coverage of the whole auditory cortex, organization of selectivity to speech features is apparent. We conclude therefore that, despite the limitations of our study, our observations provide valuable insights

regarding the organization of the auditory cortex and the representation of complex acoustic features. Future studies where brain areas are more densely sampled and are functionally defined for each animal (e.g., via widefield imaging of tonotopic organization) could ameliorate the issues presented above.

A second interpretation of the lack of differences across the cortical areas measured is that only subregions within these brain areas could be specialized in processing each feature. Therefore, a higher level of granularity in how areas are parcellated beyond standard mouse brain atlases we used would be needed to identify specializations within the auditory cortex. In fact, various ways of parcellating the auditory cortex of the mouse have been proposed, ranging from a “lumper” approach yielding only a few areas to a “splitter” approach with many more areas (Ceballo et al., 2019; Issa et al., 2017; Tsukano et al., 2016; Romero et al., 2020). Implementing alternative parcellation strategies may provide a more detailed understanding of functional specialization of speech feature processing within the auditory cortex of the mouse.

In summary, while animal models are not expected to capture all the complexities of human speech perception, our study provides support for the mouse as a useful model for investigating the neural circuits that underlie the processing the basic acoustic features of speech and how these circuits change during learning new sound contrasts, as it occurs during second language acquisition.

CRedit authorship contribution statement

SJ and MMBB conceived the project. JLM conducted the experiments. JLM and SJ analyzed the data. JLM and SJ wrote the paper.

Funding sources

This research was supported by the National Science Foundation (Grant #2024926), the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health (award R56DC015531), and the Office of the Vice President for Research & Innovation at the University of Oregon (through an I3 award).

Declaration of competing interest

The authors declare no competing financial interests.

Data availability

Data will be made available on request.

Acknowledgements

We thank Gabriel Toea and Diyar Dezey for assistance with animal training.

References

- Bennur, S., Tsunada, J., Cohen, Y.E., Liu, R.C., 2013. Understanding the neurophysiological basis of auditory abilities for social communication: a perspective on the value of ethological paradigms. *Hear. Res.* 305, 3–9. <https://linkinghub.elsevier.com/retrieve/pii/S0378595513001998>.
- Ceballo, S., Piwkowska, Z., Bourg, J., Daret, A., Bathellier, B., 2019. Targeted cortical manipulation of auditory perception. *Neuron* 104 (6), 1168–1179.e5. <https://linkinghub.elsevier.com/retrieve/pii/S0896627319308463>.
- Centanni, T.M., Engineer, C.T., Kilgard, M.P., 2013. Cortical speech-evoked response patterns in multiple auditory fields are correlated with behavioral discrimination ability. *J. Neurophysiol.* 110 (1), 177–189. <https://www.physiology.org/doi/10.1152/jn.00092.2013>.
- Eggermont, J.J., 1995. Representation of a voice onset time continuum in primary auditory cortex of the cat. *J. Acoust. Soc. Am.* 98 (2), 911–920. <https://pubs.aip.org/asa/jasa/article/98/2/911-920/932166>.
- Engineer, C.T., Rahebi, K.C., Buell, E.P., Fink, M.K., Kilgard, M.P., 2015. Speech training alters consonant and vowel responses in multiple auditory cortex fields. *Behav. Brain Res.* 287, 256–264. <https://linkinghub.elsevier.com/retrieve/pii/S0166432815002090>.

- Franklin, K.B.J., Paxinos, G., 2008. *The Mouse Brain in Stereotaxic Coordinates*, compact 3. ed edition. Elsevier Academic Press, Amsterdam Heidelberg.
- Hammerschmidt, K., Reisinger, E., Westekemper, K., Ehrenreich, L., Strenke, N., Fischer, J., 2012. Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci.* 13 (1), 40. <https://bmcneurosci.biomedcentral.com/articles/10.1186/1471-2202-13-40>.
- Issa, J.B., Haeffele, B.D., Young, E.D., Yue, D.T., 2017. Multiscale mapping of frequency sweep rate in mouse auditory cortex. *Hear. Res.* 344, 207–222. <https://linkinghub.elsevier.com/retrieve/pii/S0378595516303938>.
- Jadoul, Y., Thompson, B., De Boer, B., 2018. Introducing parselmouth: a python interface to praat. *J. Phon.* 71, 1–15. <https://linkinghub.elsevier.com/retrieve/pii/S0095447017301389>.
- Kim, H., Bao, S., 2009. Selective increase in representations of sounds repeated at an ethological rate. *J. Neurosci.* 29 (16), 5163–5169. <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0365-09.2009>.
- Kluender, K.R., 2000. Contributions of nonhuman animal models to understanding human speech perception. *J. Acoust. Soc. Am.* 107 (5Supplement), 2835. <https://pubs.aip.org/jasa/article/107/5/Supplement/2835/546652/Contributions-of-nonhuman-animal-models-to>.
- Kuhl, P.K., Miller, J.D., 1978. Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *J. Acoust. Soc. Am.* 63 (3), 905–917. <https://pubs.aip.org/jasa/article/63/3/905/681405/Speech-perception-by-the-chinchilla-Identification>.
- Kuhl, P.K., Padden, D.M., 1983. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J. Acoust. Soc. Am.* 73 (3), 1003–1010. <https://pubs.aip.org/jasa/article/73/3/1003/643343/Enhanced-discriminability-at-the-phonetic>.
- Lotto, A., Kluender, K., Holt, L., 2003. *Animal Models of Speech Perception Phenomena*. Chicago Linguist. Soc., p. 33.
- Mahrt, E.J., Perkel, D.J., Tong, L., Rubel, E.W., Portfors, C.V., 2013. Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. *J. Neurosci.* 33 (13), 5573–5583. <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.5054-12.2013>.
- Narayanan, D.P., Tsukano, H., Kline, A.M., Onodera, K., Kato, H.K., 2023. Biological constraints on stereotaxic targeting of functionally-defined cortical areas. *Cereb. Cortex* 33 (6), 3293–3310. <https://academic.oup.com/cercor/article/33/6/3293/6644718>.
- Oganian, Y., Bhaya-Grossman, I., Johnson, K., Chang, E.F., 2023. Vowel and formant representation in the human auditory speech cortex. *Neuron* 111 (13), 2105–2118.e4. <https://linkinghub.elsevier.com/retrieve/pii/S0896627323002660>.
- O'Sullivan, C., Weible, A.P., Wehr, M., 2020. Disruption of early or late epochs of auditory cortical activity impairs speech discrimination in mice. *Front. Neurosci.* 13, 1394. <https://www.frontiersin.org/article/10.3389/fnins.2019.01394/full>.
- Pachitariu, Marius, Steinmetz, Nicholas, Kadir, Shabnam, Carandini, Matteo, Harris Kenneth, D., 2016. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, 061481. <http://biorxiv.org/content/early/2016/06/30/061481.abstract>.
- Porter, B.A., Rosenthal, T.R., Ranasinghe, K.G., Kilgard, M.P., 2011. Discrimination of brief speech sounds is impaired in rats with auditory cortex lesions. *Behav. Brain Res.* 219 (1), 68–74. <https://linkinghub.elsevier.com/retrieve/pii/S0166432810008144>.
- Portfors, C.V., 2007. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J. Am. Assoc. Lab. Anim. Sci.* 46 (1).
- Portfors, C.V., Perkel, D.J., 2014. The role of ultrasonic vocalizations in mouse communication. *Curr. Opin. Neurobiol.* 28, 115–120. <https://linkinghub.elsevier.com/retrieve/pii/S0959438814001287>.
- Romero, S., Hight, A.E., Clayton, K.K., Resnik, J., Williamson, R.S., Hancock, K.E., Polley, D.B., 2020. Cellular and widefield imaging of sound frequency organization in primary and higher order fields of the mouse auditory cortex. *Cereb. Cortex* 30 (3), 1603–1622. <https://academic.oup.com/cercor/article/30/3/1603/5610045>.
- Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., Tolias, A.S., Solomon, S., Buzsáki, G., Carandini, M., Harris, K.D., 2016. Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* 19 (4), 634–641. <https://www.nature.com/articles/nn.4268>.
- Saunders, J.L., Wehr, M., 2019. Mice can learn phonetic categories. *J. Acoust. Soc. Am.* 145 (3), 1168–1177. <https://pubs.aip.org/jasa/article/145/3/1168/827568/Mice-can-learn-phonetic-categories>.
- Steinschneider, M., Fishman, Y.I., Arezzo, J.C., 2003. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J. Acoust. Soc. Am.* 114 (1), 307–321. <https://pubs.aip.org/jasa/article/114/1/307-321/543595>.
- Tsukano, H., Horie, M., Hishida, R., Takahashi, K., Takebayashi, H., Shibuki, K., 2016. Quantitative map of multiple auditory cortical regions with a stereotaxic fine-scale atlas of the mouse brain. *Sci. Rep.* 6 (1), 22315. <https://www.nature.com/articles/srep22315>.
- Tsunada, J., Lee, J.H., Cohen, Y.E., 2011. Representation of speech categories in the primate auditory cortex. *J. Neurophysiol.* 105 (6), 2634–2646. <http://www.physiology.org/doi/10.1152/jn.00037.2011>.
- Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., Dolbeare, T., Blanchard, B., Dee, N., Wakeman, W., Hirokawa, K.E., Szafer, A., Sunkin, S.M., Oh, S.W., Bernard, A., Phillips, J.W., Hawrylycz, M., Koch, C., Zeng, H., Harris, J.A., Ng, L., 2020. The Allen mouse brain common coordinate framework: a 3D reference atlas. *Cell* 181 (4), 936–953.e20. <https://linkinghub.elsevier.com/retrieve/pii/S0092867420304025>.
- Wong, S.W., Schreiner, C.E., 2003. Representation of CV-sounds in cat primary auditory cortex: intensity dependence. *Speech Commun.* 41 (1), 93–106. <https://linkinghub.elsevier.com/retrieve/pii/S0167639302000961>.
- Yi, H.G., Leonard, M.K., Chang, E.F., 2019. The encoding of speech sounds in the superior temporal gyrus. *Neuron* 102 (6), 1096–1110. <https://linkinghub.elsevier.com/retrieve/pii/S0896627319303800>.
- Yi, H.G., Chandrasekaran, B., Nourski, K.V., Rhone, A.E., Schuerman, W.L., Howard, M.A., Chang, E.F., Leonard, M.K., 2021. Learning nonnative speech sounds changes local encoding in the adult human cortex. *Proc. Natl. Acad. Sci.* 118 (36), e2101777118. <https://pnas.org/doi/full/10.1073/pnas.2101777118>.