Neural correlates of perceptual constancy in Auditory Cortex

Stephen M Town, Katherine C Wood & Jennifer K Bizley

Ear Institute, University College London, 332 Gray's Inn Road, London, WC1X 8EE, UK

Corresponding author: Stephen Town (s.town@ucl.ac.uk)

Author contributions:

SMT and JKB designed the experiments and wrote the paper; SMT, KCW and JKB collected the data; SMT analysed the data.

Acknowledgements

This work was funded by grants to JKB from the BBSRC (BB/H016813/1), the Royal Society (DH0900058) and Wellcome Trust / Royal Society (WT098418MA)

Summary

Perceptual constancy describes the ability to represent objects in the world across variation in sensory input such as recognizing a person from different angles or a spoken word across talkers. This ability requires neural representations that are sensitive to some aspects of a stimulus (such as the spectral envelope of a sound) while tolerant to other variations in stimuli (such periodicity). In hearing, such representations have been observed in auditory cortex but never in combination with behavioural testing, which is essential in order to link neural codes to perceptual constancy. By testing ferrets in a vowel discrimination task which they perform across multiple stimulus dimensions and recording neuronal activity in auditory cortex we directly correlate neural tolerance with perceptual constancy. Subjects reported vowel identity across variations in fundamental frequency, sound location, and sound level, but failed to consistently generalize across voicing from voiced to whispered sounds. We decoded the responses of simultaneously recorded units in auditory cortex to identity units informative about vowel identity across each of these task-orthogonal variations in acoustic input. Significant proportions of units were vowel informative across each of these conditions, although fewer units were informative about vowel identity across voicing. For about half of vowel informative units, information about vowel identity was conserved across multiple orthogonal variables. The time of best decoding was also used to identify the relative timing and temporal multiplexing of sound features. Our results show that neural tolerance can be observed within single units in auditory cortex in animals demonstrating perceptual constancy.

Introduction

Sensory systems must represent stimulus identity with tolerance for variation in physical input; in the vision an object must be recognized from multiple angles, lighting conditions and other identity preserving transformations (DiCarlo and Cox 2007, DiCarlo, Zoccolan et al. 2012) while in hearing, sounds such as individual words or phonemes can be recognized across talkers, voice pitch, background noise and other acoustic transformations (Sharpee, Atencio et al. 2011).

The ability to recognize objects across variations in sensory input is known as perceptual constancy (or perceptual invariance) and is observed for both visual and auditory stimuli (Logothetis and Sheinberg 1996, Griffiths and Warren 2004, Bizley and Cohen 2013). For sounds, abstraction of tolerant representations is a key step in auditory object formation (Griffiths and Warren 2004) and it is possible for humans and other animals to recognize features of sounds such as loudness across variation in location (Zahorik and Wightman 2001). Similarly for vocalizations including speech sounds, sounds can be recognized across talkers (Kojima and Kiritani 1989, Ohms, Gill et al. 2010), vocal tract length (Ghazanfar, Turesson et al. 2007, Schebesch, Lingner et al. 2010) and fundamental frequency (a determinant of perceived pitch)(Bizley, Walker et al. 2013, Town, Atilgan et al. 2015). Perceptual constancy is thus a general property of hearing across species and key in the analysis of auditory scenes (Bregman 1990, Bizley and Cohen 2013).

The importance of perceptual constancy has led to the study of its underlying neural mechanisms. Tolerant (or invariant) representations of a range of sound types including animal vocalizations (Billimoria, Kraus et al. 2008, Meliza and Margoliash 2012, Carruthers, Laplagne et al. 2015) and synthetic stimuli such as pure tones (Sadagopan and Wang 2008) and pulse trains (Bendor and Wang 2007) emerge in and across auditory cortex. Likewise, tolerance to background noise increases from midbrain to cortex (Rabinowitz, Willmore et al. 2013) where auditory cortical neurons can represent complex sounds including speech across noisy environments (Schneider and Woolley 2013, Mesgarani, David et al. 2014). However while there are many examples of neural

tolerance or invariance to specific sound features, it is unclear to what extent these are directly relevant for behavior as few examples of such phenomena have been shown in subjects actively demonstrating perceptual constancy.

To study the tolerance of neural representations during perceptual constancy, we recorded the activity of auditory cortical neurons in ferrets performing a vowel discrimination task in which synthesized vowels were varied along a number of identity-preserving, task-orthogonal transformations. Auditory cortical neurons represent multiple features of vowels including identity and dimensions such as fundamental frequency and virtual location that can be varied independently (Bizley, Walker et al. 2009, Walker, Bizley et al. 2011). Here we asked if, and how, neurons could represent vowel identity across fundamental frequency, real-world location, sound level and voicing, and how auditory encoding related to perceptual constancy during behavior.

Results

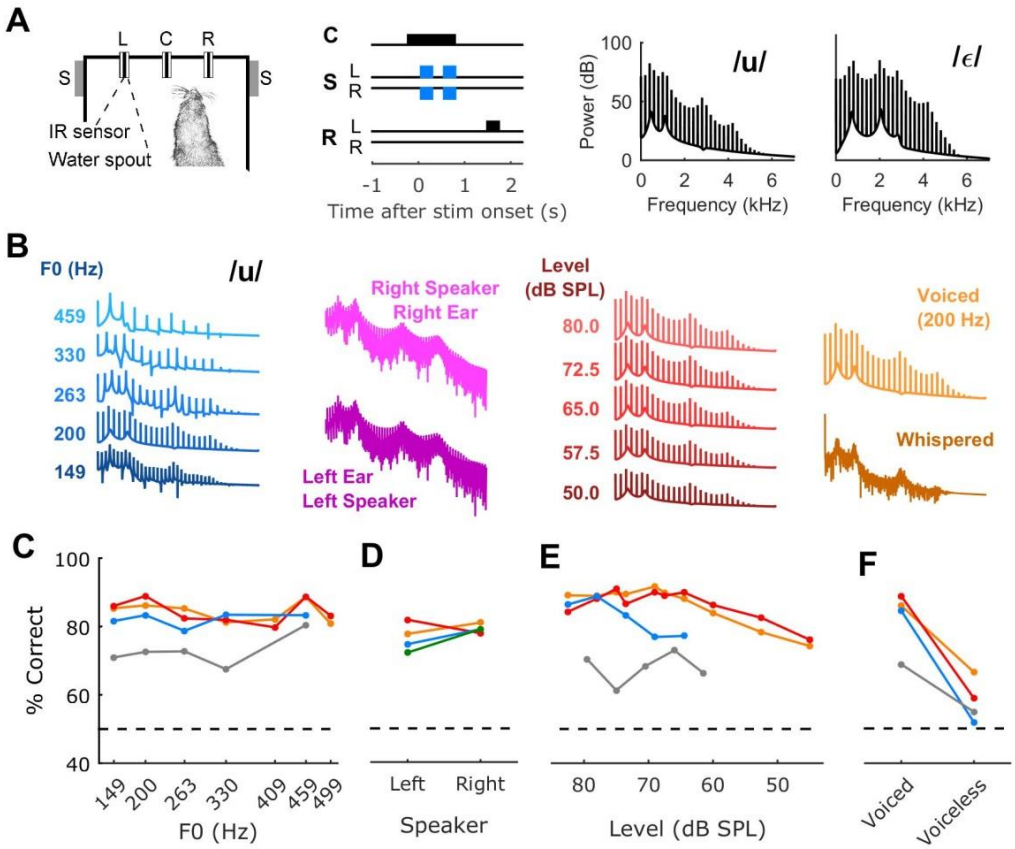
Perceptual constancy during vowel discrimination

Ferrets discriminated vowel identity in a two-alternative forced choice task (Fig. 1A) in which vowels were synthesized with varying fundamental frequency (F0) or voicing, or presented at varying levels or from different locations. Variation in these task-orthogonal variables produced a variety of different spectra while preserving formants in the spectral envelope critical for vowel identification (Fig. 1B)(Peterson and Barney 1952, Town and Bizley 2013). On each trial, the animal nose poked at a central port to trigger stimulus presentation with a variable delay. The stimuli was *t*wo tokens of *the same* vowel, each lasting 250 ms with an interval of 250 ms. Animals responded at either left or right ports based on vowel identity and were rewarded with water for correct responses while errors resulted in brief timeout (1-5 s).

Ferrets were able to discriminate vowels accurately across task-relevant dimensions: Across F0 and sound source location, we found consistent performance across variation in F0 (Fig. 1C) or

Figure 1 Perceptual constancy during vowel discrimination

(A) Task design in which ferrets discriminated vowel sounds. **(B)** Spectra for 13 examples of a single vowel [u] with varying F0, location, sound level and voicing. Spectra for sounds across location generated in virtual acoustic space (Schnupp, Booth et al. 2003). **(C-F)** Behavioral performance when discriminating vowels across F0 (C), location (D), level (E) and location (F). Colors indicate individual subjects.



location (Fig. 1D) (no effect of either factor logistic regression, $p > 0.05$, Table 1) and performance across all F0s and all sound locations was significantly better than chance for each subject (binomial test vs. 50%, $p < 0.001$, Table 2). For sound level, performance increased significantly with sound level in three of four ferrets ($p < 0.01$; Table 1) with discrimination saturating at approximately 90% correct (Fig. 1E). For all sound levels, performance was better than chance (Table 2). We also presented whispered vowels on 20% of trials as probe stimuli, however discrimination of whispered

vowels was significantly worse than for voiced vowels (Fig. 1F) and only two ferrets were able to discriminate whispered stimuli better than chance (F1201: $p < 0.001$, and F1203: $p < 0.001$, Table 2). This suggests either that ferrets are poor at generalizing across voicing or that feedback absent on probe trials was required to maintain high levels of discrimination. Together these results show that ferrets can perceive a constant vowel identity across variations in acoustic input related to fundamental frequency, sound location and sound level.

Decoding vowel identity

We implanted moveable microelectrodes in left and right auditory cortex to recorded multi-unit (n = 469 units) and LFP activity during task performance from sound-responsive sites in left and right auditory cortex (Fig. S1). For each unit, we measured responses to vowels across F0, sound location, level and voicing (Fig. 2A) and measured the information available about vowel identity across each task orthogonal variable. To measure information, we decoded stimulus parameters (e.g. vowel identity, F0 etc.) from responses of individual units using spike distance based classification (Fig. S2A). The time window of the response decoded was variable and we searched for those parameters (start time and window duration) that gave best performance (Fig. S2B). We initially only decoded responses from correct trials as we reasoned these would provide the best demonstration of auditory cortical encoding. For each unit, we could assess whether decoding performance was significantly better than chance (permutation test, $p < 0.05$) and characterize both the number of significantly informative units and overall decoder performance across auditory cortex (Fig. 2B and 2C).

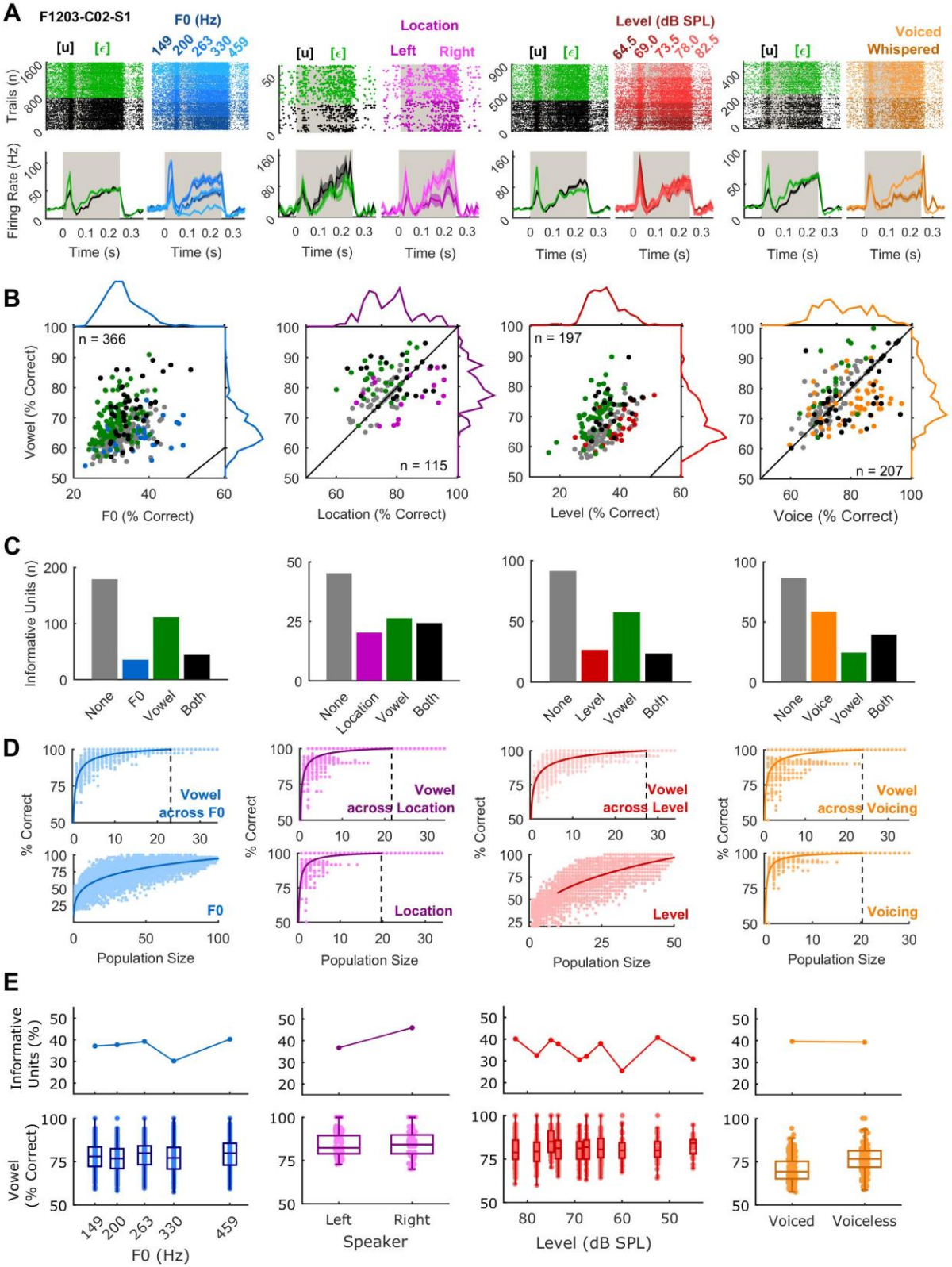
We first tested the hypothesis that neural units should be capable of encoding vowel identity across the same variation in stimulus dimensions that the animals could behaviourally. Consistent with this, when classified by statistical significance, the proportion of vowel informative units was highest across dimensions over which behavioral performance was most constant: 154 / 366 (42.1%) were informative about vowel identity across fundamental frequency, 50 / 122 (41.0%)

across sound location and 80 / 197 (40.6%) across sound level, whereas only 63/207 (30.4%) of units were informative about vowel identity across voicing (Fig. 2C). Across all sound-responsive units (regardless of informative classification), we correctly decoded vowel identity across five fundamental frequencies on 66.6 ± 0.33% of trials (mean ± s.e.m., 366 units), across two sound locations on 79.5 ± 0.62% of trials (115 units), across five sound levels on 67.5 ± 0.46% of trials (197 units) and across two voicing conditions on 76.9 ± 0.55% of trials (207 units)(Fig. 2B). When we considered vowel decoding over the most extreme two fundamental frequencies or sound levels tested (Fig S3A and S3B), decoder performance was comparable to testing across location or voicing (mean ± s.e.m. performance across F0 = 73.0 ± 0.40%; across sound level = 75.7 ± 0.86%). The pattern of decoder performance therefore reflected the sensitivity of raw decoder performance to the number of orthogonal levels (i.e. number of F0s or sound locations) over which vowels were varied and once this was taken into account decoder performance was equivalent across all the stimulus dimensions tested.

Figure 2 Neural responses and decoding stimulus parameters

(A) Raster and peri-stimulus time histograms (PSTHs) from an example unit in l response to vowels with different fundamental frequency, sound location, sound level and voicing. Data shown for the first vowel presented during the trial; PSTHs show mean ± s.e.m. firing rate. **(B)** Decoding performance across all units when reconstructing vowel identity and task orthogonal values from single trial responses of individual units. (**C**) Number of units informative about vowel identity and / or task orthogonal values when considering responses across all stimuli. **(D)** Population decoding of vowel identity and task-orthogonal values. Individual data points show populations with different member units; full line shows curve fit for a two-term power series model; dotted lines indicate number of units required to reach 100% decoding performance. The top row indicates vowel decoding across (left-right) F0, space, level and voicing, and the bottom row indicates decoding

performance in the orthogonal dimension. **(E)** Number of vowel informative units and decoder performance at each task-orthogonal value.



While we are able to generalise across particular stimulus dimensions (e.g. pitch) this does not mean that we are unable to detect changes in other stimulus dimensions. Accordingly, in addition to vowel identity, auditory cortical units were also informative about task orthogonal variables across vowels. We found 78 / 366 units (21.3%) were informative about fundamental frequency (mean ± s.e.m. decoding performance = 32.9 ± 0.25%), 44 / 115 units (38.3%) about sound location (77.4 ± 0.74%), 49 / 197 units (24.9%) about sound level (34.3 ± 0.38%) and 97 / 207 units (46.9%) about voicing (77.3 ± 0.58%)(Fig. 2D). The proportion of units informative about F0 and sound level were smaller than sound location or voicing in part because of the increased number of decoded classes (five vs. two respectively). However when we decoded F0 or sound level across the most extreme values tested (Fig. S3A and S3B), the proportion of orthogonal informative units increased (F0: 100 / 350 or 28.6% of units informative; Level: 33 / 65 or 50.8%). We also noted that the proportion of units informative about sound level but not F0 was associated with the interval between sound levels (Fig. S3C-D). For sound level we tested units across two stimulus sets ranging from 64.5 to 82.5 dB SPL in 4.5 dB steps, or from 45 to 75 dB SPL in 7.5 dB steps. While we found units informative about vowel identity and / or sound level for both stimulus sets, more units were informative about sound level (alone or together with vowel identity) when stimuli were varied from 45 to 75 dB SPL (Fig. S4A).

In order to attempt to more directly link neural encoding to perceptual constancy, we next asked how decoding performance compared to the animal's behavioural ability. Since our decoder considered only correct trials, 100% decoding performance would indicate that a unit's information content matched the subject's behavior. While some units were close to this level (>90%) we noted that for no individual unit did decoding performance reach this criterion. Nevertheless, by decoding responses from small populations of units (Fig. 2D) it was possible to decode vowel identity with 100% performance across F0 (23 units), sound location (22 units), sound level (28 units) and voicing (24 units). Similarly, we could also decode task-orthogonal variables with high accuracy using comparable population sizes for sound location (20 units) and voicing (21 units), although more

neurons were required to decode F0 (n = 129) and sound level (56 units). Thus across small populations of auditory cortical units, neural responses could be decoded with the same performance as animal's behavioral performance and represent multiple stimulus dimensions near-perfectly.

Finally we asked how units encoded vowel identity across individual task orthogonal values: if the responses of neural units underlie the animals' perceptual constancy we would expect similar levels of decoder performance across the variations in acoustic structure of vowels that animals' displayed invariance. We found that, as with behavior (Fig. 1C-F), the proportion of vowel informative units did not differ significantly with fundamental frequency (logistic regression, χ^2 = 0.035, p = 0.852) or sound location (χ^2 = 1.85, p = 0.174). Furthermore we also found no effect of either sound level (χ^2 = 0.416, p = 0.519) or voicing (χ^2 = 0.002, p = 0.964) on proportions of vowel informative units, suggesting that for neural encoding of vowel identity was robust across all dimensions tested.

Conserved information content

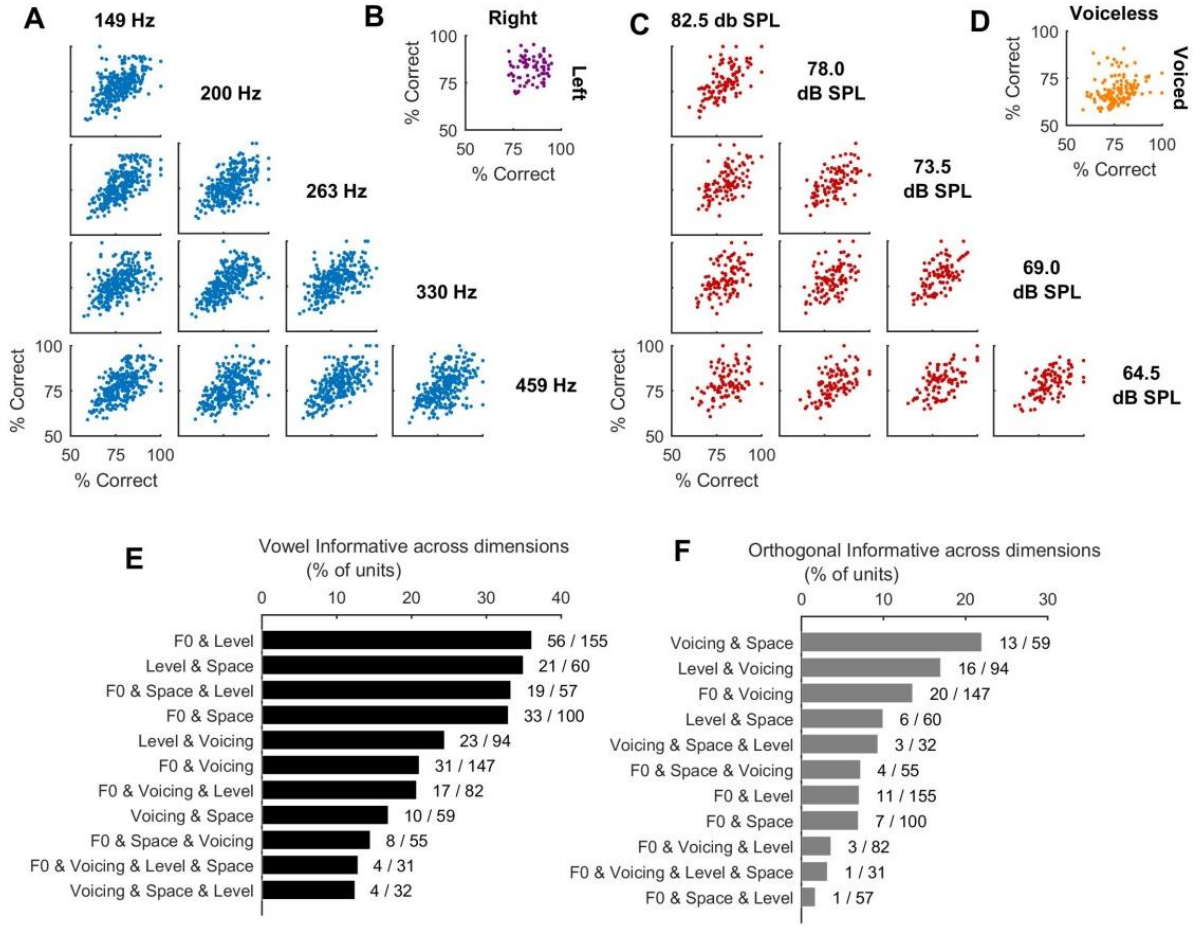
We recorded units with vowel informative responses across task-orthogonal dimensions and found that across the proportion of recorded units, the proportion of vowel informative units was constant at different task-orthogonal values. This could emerge due to the presence of "invariant" neural units that are tolerant to changes within and across task-orthogonal dimensions, or, alternatively, it could be that tolerance to task-orthogonal values only emerges across populations of units. To test these hypotheses, we asked whether *the same units* were informative about vowel identity across all orthogonal dimensions. We considered this question in two stages: firstly, within a single task-orthogonal dimension (e.g. F0), did the same units encode vowel identity across the different values tested. Secondly, we asked whether the same units provided tolerant information across different task-irrelevant dimensions. We addressed the first question in two ways by (a) comparing vowel decoding performance across all task orthogonal values, and (b) predicting a unit's

classification as informative about vowel identity at a specific task orthogonal value (e.g. F0 = 200 Hz) from that unit's classification at all other values tested (e.g. F0 = 149, 263, 330 and 459 Hz) using multivariate logistic regression modelling. For logistic regressions, we reasoned that units with conserved function would be informative about vowel identity at multiple orthogonal values and so a unit's classification for a subset of values should allow us to predict classification on a held-out value. In contrast if function was not conserved, the classification at each orthogonal value would vary randomly and we should not be able to predict classification better than chance for held-out data.

Vowel decoding at different F0s was positively correlated (Fig 3A, mean ± s.e.m. $R^2 = 0.359 \pm 0.0179$) indicating that units informative about vowel identity at one F0 were also did informative at other F0s. In support of this, unit classification (vowel informative / uninformative) at each F0 could be predicted from unit classification at all other F0s (Table 3, $p \leq 0.01$) though the weighting of each predictor F0 varied (Fig S4). Similarly, vowel decoding was also correlated at different sound levels (Fig 3C and Fig. S5, mean ± s.e.m. $R^2 = 0.290 \pm 0.02$) although unit classification at individual sound levels was not predicted from classification at other sound levels tested ($p > 0.05$). Across space, vowel decoding performance was not correlated for stimuli presented at left and right speakers (Fig. 3B; $R^2 = 0.0353$) and unit classification at one location was not predictive of classification at the other location ($\chi^2 = 0.072$, $p = 0.788$). Vowel decoding performance was weakly correlated for voiced and voiceless stimuli (Fig. 3D, $R^2 = 0.063$) and unit classification on voiced trials was not predictive of classification on voiceless trials ($\chi^2 = 1.28$, $p = 0.258$). Together this suggests that information about vowel identity at different F0s and, to a lesser extent, sound level was conserved across units (i.e. the same units could signal vowel identity under different conditions within a stimulus dimension) whereas information about vowel identity at different locations or in different voicing conditions was provided by different units.

Figure 3 Conserved Information content

(A-D) Paired comparisons of decoding performance for vowel identity at different fundamental frequencies (**A**), sound locations (**B**), sound levels (**C**) or voicing conditions (**D**). **(E)** Number of units classified as vowel informative *across* multiple task-orthogonal dimensions. **(F)** As E but for units classified as being informative *about* multiple task-orthogonal dimensions.



We next asked if vowel decoding was conserved across stimulus dimensions by counting the number of units that were vowel informative across each of dimension sound location, level, voicing and fundamental frequency. While not every unit was tested across every task-orthogonal dimension, we observed that the dimensions over which behavioral performance was constant (F0, sound location and level) were also the dimensions over which the highest proportions of units remained informative across dimensions (Fig. 3E). Across fundamental frequency and sound level,

we recorded 56/155 units (36.1%) that were vowel informative across both dimensions. Similar proportions of conserved units were also identified across fundamental frequency and location (33 /100; 33%), sound level and location (21 / 60 units; 35%) and the combination of F0, sound level and location (19 / 57; 33.3%). Notably fewer conserved units were observed for other combinations of task orthogonal factors (<25%) – all of which included voicing (across which animals generalized poorly). This suggests that the same auditory cortical units represent vowel identity across (as well as within) task-orthogonal variations that the animals also perceive as having a constant identity.

If the same neurons are informative about vowel identity across different task-orthogonal variables, it should be possible to predict a unit's classification (vowel informative / uninformative) in one condition (e.g. vowel discrimination across F0) based on its classification in other conditions (e.g. vowel discrimination across level). Using logistic regression models, we confirmed this was true for fundamental frequency and sound level ($\chi^2 = 29.3$, $p = 6.34 \times 10^{-8}$), F0 and sound location ($\chi^2 = 8.21$, $p = 4.16 \times 10^{-3}$) as well as F0 and voicing ($\chi^2 = 4.68$, $p = 0.0305$), and sound level and voicing ($\chi^2 = 2.77$, $p = 6.34 \times 10^{-4}$). Models in which unit classification across the target dimension (e.g. F0) was predicted by classification in more than one other dimension (e.g. sound level and location) were also significant (Table 4). Thus our statistical analysis confirmed that information about vowel identity was conserved across units when considering multiple task-orthogonal variables consistent with the idea that tolerant information was a property of single units and not only neural ensembles.

We also asked whether units that were informative about one task-orthogonal variable (e.g. sound location) were also informative about other task-orthogonal variables (e.g. voicing). While we did observe small populations of units that were informative about multiple dimensions (Fig. 3F) such populations were smaller than the populations that were vowel informative across dimensions (sign-rank test, $p = 0.0029$). Furthermore, we found no combination of task-orthogonal dimensions in which unit classification was predictive (Table 4). This suggests that while information about vowel

identity is conserved across task orthogonal dimensions, at least during the performance of a vowel identification task, few units are informative about multiple task-orthogonal dimensions.

Temporal multiplexing of sound features

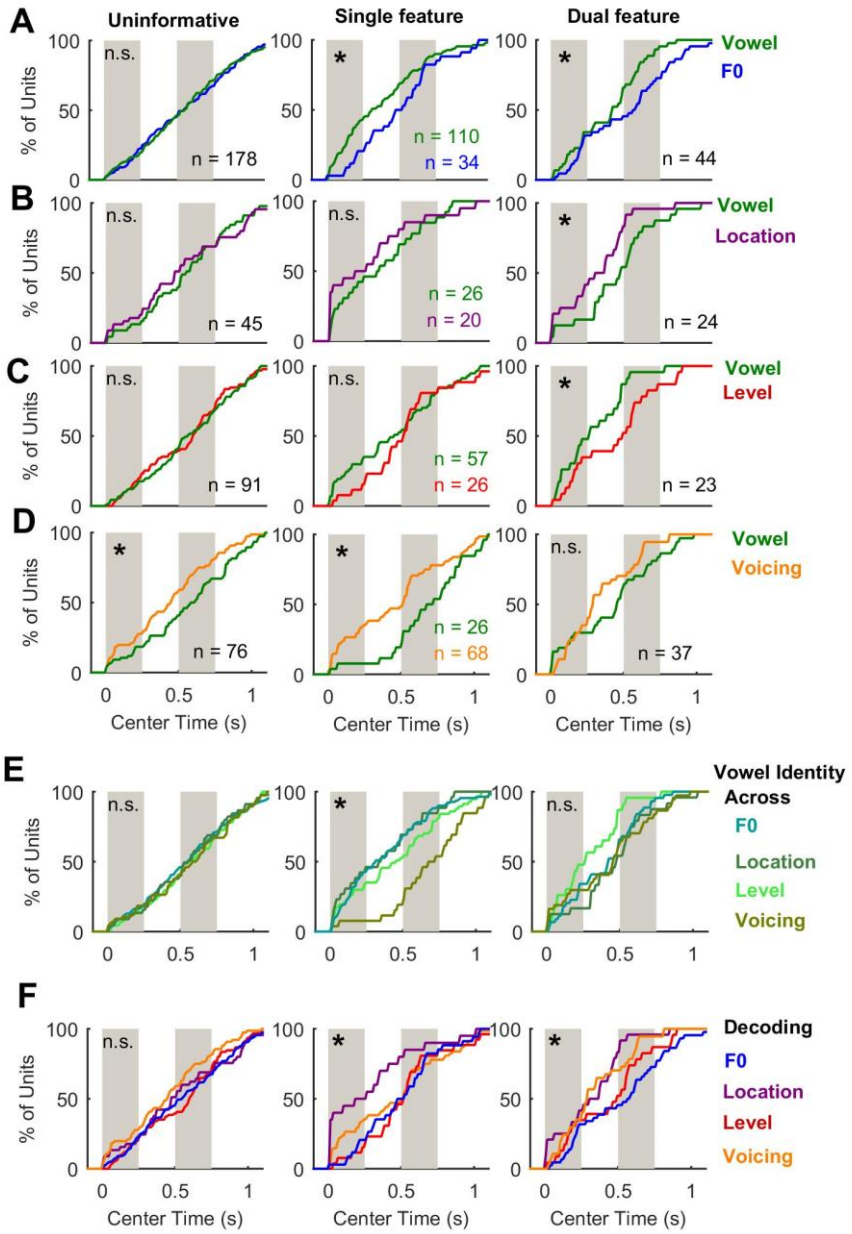
We next turned our attention to the timing of information about auditory features. Responses of auditory cortical neurons in anesthetized ferret contain mutual information about vowel identity, F0 and virtual location in distinct time windows (temporal multiplexing)(Walker, Bizley et al. 2011). Here we asked whether the same was true of neural responses recorded in animals actively discriminating sounds using decoding rather than information theoretic measures (which together provide complementary approaches to understanding neural encoding (Quian Quiroga and Panzeri 2009)). We also extended the study of temporal multiplexing to sound level and voicing.

To investigate temporal multiplexing, we compared the time windows after trial onset that gave best performance using our decoder (Fig. S2B). A time window was defined by the start time and window duration, and we summarized these two parameters by calculating the midpoint of the window (i.e. start time + duration / 2). For each unit, we measured the center time for decoding of vowel identity across each task-orthogonal dimension, and for decoding task-orthogonal values across vowels. The cumulative distribution functions (CDFs) of center times across units was then compared when decoding vowel identity and task-orthogonal values (Fig 4).

We found that information about vowel identity arose earlier than about fundamental frequency (Fig. 4A) in units who were informative about both F0 and vowel identity (Sign-rank test: $z = -2.43$, $p = 0.0150$) and units that were informative about F0 or vowel identity (Rank-sum test: $z = -2.31$, $p = 0.0206$) but not uninformative units (Sign-rank test: $z = -0.0734$, $p = 0.942$). Similarly, information about sound location was also best decoded earlier than vowel identity in units informative about both dimensions (Fig. 4B, $z = 2.26$, $p = 0.0240$) but not either dimension alone ($z = 1.29$, $p = 0.198$) or uninformative units ($z = 0.722$, $p = 0.470$). Information about vowel identity was

Figure 4 Temporal multiplexing in auditory cortex

(A-D) Cumulative distributions showing center times (start time + duration / 2) for best performance when decoding vowel identity or task-orthogonal variables (**A:** F0, **B**: location, **C**: level and **D**: voicing). Units shown separately by classification as informative about vowel identity and task-orthogonal values (Dual feature units), either vowel identity or task task-orthogonal (Single feature units) or neither feature (uninformative). **(E)** CDFs for decoding vowel identity across each task-orthogonal variable. **(F)** CDFs for decoding task-orthogonal values across vowels.



best decoded before sound level (Fig. 4C) in units who were informative about both features ($z = -2.13$, $p = 0.0333$) but not units informative about individual features ($z = -0.933$, $p = 0.351$) or uninformative units ($z = 0.471$, $p = 0.638$). Information about voicing was best decoded earlier than vowel identity (Fig. 4D) but in contrast to other comparisons, this was only significant in uninformative units ($z = 2.40$, $p = 0.0164$) and units informative about vowel or voicing ($z = 2.79$, $p = 0.005$) but not units informative about both features ($z = 1.33$, $p = 0.184$). Together these results suggest that units encoding multiple sound features use temporal multiplexing to represent these features at different times. Notably, multiplexing in units sensitive to several stimulus features was only observed when we varied F0, sound level or location - features that animals also perceive with a constant identity.

Our comparison between vowel and task-orthogonal decoding revealed differences in timing of best decoding. However we also asked if vowel identity was always best decoded with a fixed temporal profile or if best decoding of vowel identity differed between task-orthogonal variables. For units that only encoded vowel identity, timing of vowel information significant differed across task orthogonal variables (Fig. 4E)(Kruskall-Wallis one-way ANOVA: $\chi^2 = 19.95$, $p = 1.74 \times 10^{-4}$). Post-hoc comparisons confirmed that vowel identity was best decoded across voicing significantly later than across F0 (tukay-kramer correction for pairwise comparisons, $p = 0.001$), sound location ($p = 0.0015$) and sound level ($p = 0.013$). Thus for these units, information about vowel identity was not fixed and could vary depending on the orthogonal features of sounds. We found no differences in the timing of vowel identity information between sound level, location or F0 and no significant differences between task-orthogonal variables for units informative about vowel identity and task-orthogonal values ($\chi^2 = 5.76$, $p = 0.124$) or uninformative units ($\chi^2 = 0.677$, $p = 0.879$).

We also compared the timing of information about task-orthogonal variables decoded across vowels. Here we found significant differences in timing of task-orthogonal information for units that were informative only about task-orthogonal values ($\chi^2 = 11.04$, $p = 0.0206$) and units

informative about both task-orthogonal values and vowel identity (χ^2 = 9.77, p = 0.0115). Post-hoc comparisons for units informative about only task-orthogonal values showed that sound location was best decoded significantly before F0 (p = 0.0255) and sound level (p = 0.0283). For units informative about both task-orthogonal values and vowel identity, post-hoc comparisons showed that F0 was best decoded significantly later than information about space (p = 0.0199) and voicing (p = 0.0468). No further post-hoc comparisons were significant and timing across uninformative units was not significantly different (χ^2 = 7.033, p = 0.0708). These results indicate that task-orthogonal variables are encoded at different times in neural responses to sound and support the suggestion that temporal multiplexing is a general phenomenon that maybe observed beyond the vowels studied here such as stimuli varying only in level or location (REFERENCES).

Discussion

Here we demonstrated that auditory cortical neurons can reliably represent vowel sounds across a range of acoustic transformations that preserve sound identity in perception. By recording neurons in animals engaged in a task that demonstrates perceptual constancy, our results provide a direct comparison of neural encoding and behavior.

Brain and behavior

A central aim of this study was to directly test the tolerance (or invariance) of neural coding during perceptual constancy. We confirmed that vowel discrimination performance was indeed consistent across fundamental frequency, location and a range of sound levels during recording, although animals were unable to generalize performance to whispered stimuli (but this may be due to their presentation as probe trials). Across all task-orthogonal dimensions, we could decode vowel identity better than chance for many units and match behavioral performance by decoding over populations of units. When comparing behavior and neural decoding, it was notable that the proportion of units informative about vowel identity across voicing was lower (30.4%) than any other task orthogonal

dimension (40 – 42% of units). This correlation suggest that processing of auditory cortical neurons reflects behaviourally relevant information and that such neurons are less able to represent vowels across voicing conditions, leading to impaired generalization of vowel identity in the task. This is further supported by the timing at which we best decoded vowel identity across voicing conditions, which was significantly later than any other orthogonal variable for units that were informative about vowel identity (Fig. 4E).

Why did auditory cortical units fail to generalize across voicing as well as other dimensions? One explanation may relate to the energy content of whispered and voiced stimuli; in contrast to voiced vowels that concentrate energy at harmonic frequencies, whispered sounds contain energy distributed across frequency (Fig. 1b). Versnel and Shamma (1998) reported similar responses in primary auditory cortex (A1) to voiced and unvoiced (whispered) vowels presented with durations of 90 – 140 ms. We observed that for units such as that shown in Fig. 2A, the onset response to voiced and whispered vowels was indeed similar. However during the sustained period of our stimuli (100 – 250 ms) responses differed, suggesting that the broadband versus harmonic spectra of unvoiced versus voiced vowels may be reflected in the sustained firing of the unit. This suggestion is consistent with the late optimization of voicing information when compared to other orthogonal features (Fig. 4F).

Temporal multiplexing

We found that for units informative about both vowel identity and task-orthogonal values were best decoded using response activity in consistently distinct time windows. Given we decoded vowel identity and task-orthogonal variables independently and without *a priori* bias to particular times in the neural response after stimulus onset, it is most likely the case multiplexing represents the properties of the units rather than of the decoder. Our results showing that best decoding of vowel identity and sound location was faster than voicing or F0 are consistent with findings that perception of sound location and vowel identity are associated with the onset of sounds (Litovsky,

Colburn et al. 1999, Stecker and Hafter 2002) whereas fundamental frequency requires time for listeners to estimate (Gray 1942, Mckeown and Patterson 1995, Walker, Bizley et al. 2011). These differences in timing may be intuitive given multiple cycles of a vowel are required to F0 from the amplitude waveform arriving at the ear. However we also showed several previously undemonstrated differences in the timing of auditory information including rapid rise in decoded information about voicing and that sound level took longer to decode best than vowel identity, sound location or voicing. The time-course of sound level decoding may reflect the temporal integration of the auditory system (Buus, Florentine et al. 1997, Glasberg and Moore 2002). Thus neural organization of sound features in time mirrored perception that are thought to be driven by stimulus acoustics (e.g. vowel identity vs. F0) and the dynamics of auditory scene analysis (e.g. the precedence effect and its importance in localizing sound sources, particularly in reverberant environments).

Multi-feature representations

The representation of multiple sound features, sometimes within the same units, supports previous reports of distributed coding in auditory cortex (Bizley, Walker et al. 2009) and mirrors to co-encoding of multiple stimulus features in the ventral visual stream (Hong, Yamins et al. 2016). As in vision, the function of the auditory system is to recognize objects across identity-preserving transformations. A key function of visual cortex may be to extract invariant or tolerant representations of objects through hierarchical networks (Yamins and DiCarlo 2016) and the emergence of tolerant representations in primary and non-primary auditory cortex suggests similar principles may apply to the auditory system. However primary auditory and visual cortices may not be equivalent given the extensive subcortical processing of sounds in the auditory system (King and Nelken 2009) but rather early auditory cortex may be higher in the auditory invariance hierarchy than early visual cortices (V1 / V2) in the visual invariance hierarchy. Our results and other studies (Bizley, Walker et al. 2009) support this view as we sampled neurons that represented vowel identity

across orthogonal variables from both primary and non-primary cortices. However we did not map the precise boundaries between cortical subfields in this study and our results only show tolerant representations exist early in auditory cortex rather than determining whether tolerance increases from primary to non-primary areas. As the emergence of tolerant representations has been mapped progressively through the visual system (Hong, Yamins et al. 2016), it will also be important to study the tolerance of auditory representations at stages in the ascending auditory system before cortex; in the medial geniculate nucleus and inferior colliculus of animals demonstrating perceptual constancy.

References

Bendor, D. and X. Q. Wang (2007). "Differential neural coding of acoustic flutter within primate auditory cortex." Nature Neuroscience **10**(6): 763-771.

Billimoria, C. P., B. J. Kraus, R. Narayan, R. K. Maddox and K. Sen (2008). "Invariance and sensitivity to intensity in neural discrimination of natural sounds." Journal of Neuroscience **28**(25): 6304-6308.

Bizley, J. K. and Y. E. Cohen (2013). "The what, where and how of auditory-object perception." Nat Rev Neurosci **14**(10): 693-707.

Bizley, J. K., K. M. Walker, A. J. King and J. W. Schnupp (2013). "Spectral timbre perception in ferrets: discrimination of artificial vowels under different listening conditions." J Acoust Soc Am **133**(1): 365-376.

Bizley, J. K., K. M. Walker, F. R. Nodal, A. J. King and J. W. Schnupp (2013). "Auditory cortex represents both pitch judgments and the corresponding acoustic cues." Curr Biol **23**(7): 620-625.

Bizley, J. K., K. M. Walker, B. W. Silverman, A. J. King and J. W. Schnupp (2009). "Interdependent encoding of pitch, timbre, and spatial location in auditory cortex." J Neurosci **29**(7): 2064-2075.

Bregman, A. S. (1990). Auditory scene analysis. Cambridge, MA, MIT Press.

Buus, S., M. Florentine and T. Poulsen (1997). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function." J Acoust Soc Am **101**(2): 669-680.

Carruthers, I. M., D. A. Laplagne, A. Jaegle, J. J. Briguglio, L. Mwilambwe-Tshilobo, R. G. Natan and M. N. Geffen (2015). "Emergence of invariant representation of vocalizations in the auditory cortex." J Neurophysiol **114**(5): 2726-2740.

DiCarlo, J. J. and D. D. Cox (2007). "Untangling invariant object recognition." Trends Cogn Sci **11**(8): 333-341.

DiCarlo, J. J., D. Zoccolan and N. C. Rust (2012). "How does the brain solve visual object recognition?" Neuron **73**(3): 415-434.

Ghazanfar, A. A., H. K. Turesson, J. X. Maier, R. van Dinther, R. D. Patterson and N. K. Logothetis (2007). "Vocal-tract resonances as indexical cues in rhesus monkeys." Current Biology **17**(5): 425-430.

Glasberg, B. R. and B. C. J. Moore (2002). "A model of loudness applicable to time-varying sounds." Journal of the Audio Engineering Society **50**(5): 331-342.

Gray, G. W. (1942). "Phonemic Microtomy: The Minimum Duration of Perceptible Speech Sounds." Speech Monographs **9**: 75-90.

Griffiths, T. D. and J. D. Warren (2004). "What is an auditory object?" Nature Reviews Neuroscience **5**(11): 887-892.

Hong, H., D. L. Yamins, N. J. Majaj and J. J. DiCarlo (2016). "Explicit information for category-orthogonal object properties increases along the ventral stream." Nat Neurosci **19**(4): 613-622.

King, A. J. and I. Nelken (2009). "Unraveling the principles of auditory cortical processing: can we learn from the visual system?" Nat Neurosci **12**(6): 698-701.

Kojima, S. and S. Kiritani (1989). "Vocal-Auditory Functions in the Chimpanzee - Vowel Perception." International Journal of Primatology **10**(3): 199-213.

Litovsky, R. Y., H. S. Colburn, W. A. Yost and S. J. Guzman (1999). "The precedence effect." Journal of the Acoustical Society of America **106**(4): 1633-1654.

Logothetis, N. K. and D. L. Sheinberg (1996). "Visual object recognition." Annual Review of Neuroscience **19**: 577-621.

Mckeown, J. D. and R. D. Patterson (1995). "The Time-Course of Auditory Segregation - Concurrent Vowels That Vary in Duration." Journal of the Acoustical Society of America **98**(4): 1866-1877.

Meliza, C. D. and D. Margoliash (2012). "Emergence of Selectivity and Tolerance in the Avian Auditory Cortex." Journal of Neuroscience **32**(43): 15158-15168.

Mesgarani, N., S. V. David, J. B. Fritz and S. A. Shamma (2014). "Mechanisms of noise robust representation of speech in primary auditory cortex." Proc Natl Acad Sci U S A **111**(18): 6792-6797.

Musial, P. G., S. N. Baker, G. L. Gerstein, E. A. King and J. G. Keating (2002). "Signal-to-noise ratio improvement in multiple electrode recording." J Neurosci Methods **115**(1): 29-43.

Ohms, V. R., A. Gill, C. A. A. Van Heijningen, G. J. L. Beckers and C. ten Cate (2010). "Zebra finches exhibit speaker-independent phonetic perception of human speech." Proceedings of the Royal Society B-Biological Sciences **277**(1684): 1003-1009.

Peterson, G. E. and H. L. Barney (1952). "Control methods used in a study of vowels." J Acoust Soc Am **24**: 175-184.

Quian Quiroga, R. and S. Panzeri (2009). "Extracting information from neuronal populations: information theory and decoding approaches." Nat Rev Neurosci **10**(3): 173-185.

Rabinowitz, N. C., B. D. B. Willmore, A. J. King and J. W. H. Schnupp (2013). "Constructing Noise-Invariant Representations of Sound in the Auditory Pathway." Plos Biology **11**(11).

Sadagopan, S. and X. Wang (2008). "Level invariant representation of sounds by populations of neurons in primary auditory cortex." Journal of Neuroscience **28**(13): 3415-3426.

Schebesch, G., A. Lingner, U. Firzlaff, L. Wiegrebe and B. Grothe (2010). "Perception and neural representation of size-variant human vowels in the Mongolian gerbil (Meriones unguiculatus)." Hearing Research **261**(1-2): 1-8.

Schneider, D. M. and S. M. N. Woolley (2013). "Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes." Neuron **79**(1): 141-152.

Schnupp, J. W., J. Booth and A. J. King (2003). "Modeling individual differences in ferret external ear transfer functions." J Acoust Soc Am **113**(4 Pt 1): 2021-2030.

Sharpee, T. O., C. A. Atencio and C. E. Schreiner (2011). "Hierarchical representations in the auditory cortex." Curr Opin Neurobiol **21**(5): 761-767.

Stecker, G. C. and E. R. Hafter (2002). "Temporal weighting in sound localization." Journal of the Acoustical Society of America **112**(3): 1046-1057.

Town, S. M., H. Atilgan, K. C. Wood and J. K. Bizley (2015). "The role of spectral cues in timbre discrimination by ferrets and humans." J Acoust Soc Am **137**(5): 2870-2883.

Town, S. M. and J. K. Bizley (2013). "Neural and behavioral investigations into timbre perception." Front Syst Neurosci **7**: 88.

Versnel, H. and S. A. Shamma (1998). "Spectral-ripple representation of steady-state vowels in primary auditory cortex." J Acoust Soc Am **103**(5 Pt 1): 2502-2514.

Walker, K. M., J. K. Bizley, A. J. King and J. W. Schnupp (2011). "Multiplexed and robust representations of sound features in auditory cortex." J Neurosci **31**(41): 14565-14576.

Yamins, D. L. and J. J. DiCarlo (2016). "Using goal-driven deep learning models to understand sensory cortex." Nat Neurosci **19**(3): 356-365.

Zahorik, P. and F. L. Wightman (2001). "Loudness constancy with varying sound source distance." Nature Neuroscience **4**(1): 78-83.

Tables

Table 1

Results of logistic regressions comparing performance across task-orthogonal variables.

	Task-orthogonal Dimension			
Ferret	**F0**	**Voicing**	**Sound Level**	**Location**
F1201	*df* = 12034, *p* = 0.7731	*df* = 5802, *p* < 0.001	*df* = 3490, *p* < 0.001	*df* = 879, *p* = 0.219
F1203	*df* = 9945, *p* = 0.764	*df* = 4212, *p* < 0.001	*df* = 3063, *p* = 0.005	*df* = 701, *p* = 0.185
F1217	*df* = 4790, *p* = 0.368	*df* = 3784, *p* < 0.001	*df* = 2214, *p* < 0.001	*df* = 145, *p* = 0.523
F1304	*df* = 1485, *p* = 0.388	*df* = 617, *p* = 0.002	*df* = 455, *p* = 0.882	*Not tested*

Table 2

Comparison of observed vowel discrimination against chance performance (50%); data shown as fraction of trials correct and probability of observed performance (binomial test).

		Ferret			
Dimension	**Data Set**	**F1201**	**F1203**	**F1217**	**F1304**
F0	All	10214 / 12036 p < 0.001	8502 / 9947 p < 0.001	3946 / 4792 p < 0.001	1051 / 1487 p < 0.001
Voicing	Voiced	3830 / 4472 *p* < 0.001	2833 / 3207 *p* < 0.001	2585 / 3038 *p* < 0.001	333 / 473 *p* < 0.001
	Whispered	870 / 1332 *p* < 0.001	596 / 1007 *p* < 0.001	386 / 748 *p* = 0.400	83 / 146 *p* = 0.116
Sound level (dB SPL)	45	138 / 186 p < 0.001	149 / 196 p < 0.001	*Not tested*	*Not tested*
	52.5	151 / 193 p < 0.001	169 / 205 p < 0.001	*Not tested*	*Not tested*
	60	161 / 192 p < 0.001	162 / 188 p < 0.001	*Not tested*	*Not tested*
	64.5	438 / 498 p < 0.001	372 / 414 p < 0.001	341 / 442 p < 0.001	*Not tested*
	67.5	175 / 195 p < 0.001	176 / 198 p < 0.001	*Not tested*	*Not tested*
	69	479 / 523 p < 0.001	357 / 397 p < 0.001	345 / 449 p < 0.001	*Not tested*
	73.5	460 / 515 p < 0.001	352 / 407 p < 0.001	380 / 457 p < 0.001	*Not tested*
	75	168 / 187 p < 0.001	179 / 197 p < 0.001	*Not tested*	*Not tested*

	78	445 / 501 p < 0.001	382 / 434 p < 0.001	392 / 442 p < 0.001	*Not tested*
	82.5	445 / 500 p < 0.001	357 / 424 p < 0.001	366 / 424 p < 0.001	*Not tested*
	All	*Not tested*	*Not tested*	*Not tested*	309 / 455 p < 0.001
Location	All	701 / 879 *p* < 0.001	564 / 703 *p* < 0.001	112 / 145 *p* < 0.001	*Not tested*

Table 3

Logistic regression results for determining if a unit was vowel informative at a particular F0 value, predicted from unit classification at all other fundamental frequencies.

Orthogonal Dimension	**Orthogonal value**	**χ^2**	**P**
Fundamental Frequency (Hz)	149	24.3	6.91×10^{-5}
	200	22.6	1.52×10^{-4}
	263	13.1	0.0107
	330	16.9	2.07×10^{-3}
	459	26.7	2.27×10^{-5}

Table 4

Logistic regression results for predicting (1) if a unit would be vowel informative across one task-orthogonal dimension (Response) based on unit classification across other task-orthogonal variables (Predictors), or (2) if a unit was informative about one task-orthogonal dimension based on its classification as informative about other task-orthogonal dimensions. For single predictor models, the association between response and predictor was bidirectional – i.e. if classification across space predicted classification across F0, then classification across F0 also predicted classification across space.

			Vowel Informative		Orthogonal Informative	
Response	**Predictors**	**DFE**	**χ^2**	**p**	**χ^2**	**p**
F0	Space	98	8.21	**4.16×10^{-3}**	0.278	0.598
F0	Level	153	29.3	**6.34×10^{-8}**	0.264	0.607
F0	Voicing	145	4.68	**0.0305**	2.590	0.107
Space	Level	58	0.426	0.514	0.544	0.460
Space	Voicing	57	0.583	0.445	0.198	0.656
Level	Voicing	92	13.2	**2.77×10^{-4}**	1.50	0.221
F0	Space & Voicing	52	11.32	**3.48×10^{-3}**	1.09	0.580
F0	Space & Level	54	17.77	**1.38×10^{-4}**	0.109	0.947
F0	Level & Voice	79	15.39	**4.55×10^{-4}**	0.868	0.648
Space	F0 & Voicing	52	6.067	**0.0481**	0.411	0.814
Space	F0 & Level	54	5.06	0.0798	0.453	0.797
Space	Level & Voicing	29	3.10	0.110	1.49	0.474
Level	F0 & Space	54	15.28	**4.80×10^{-4}**	0.344	0.842
Level	F0 & Voice	79	22.89	**1.07×10^{-5}**	1.19	0.551
Level	Space & Voicing	29	4.41	0.212	1.29	0.524
Voicing	F0 & Space	52	8.752	**0.0126**	0.801	0.670
Voicing	F0 & Level	79	11.97	**2.51×10^{-3}**	2.05	0.359
Voicing	Space & Level	29	2.21	0.331	0.903	0.637
F0	Space, Level & Voicing	27	11.8	**8.09×10^{-3}**	5.67	0.129
Space	F0, Level & Voicing	27	5.587	0.134	4.90	0.179
Level	F0, Space & Voicing	27	11.86	**7.87×10^{-3}**	0.961	0.811
Voicing	F0, Space & Level	27	6.582	0.0865	2.99	0.393

Methods

Animals

Subjects were four pigmented female ferrets (1-5 years old) trained to discriminate vowels across fundamental frequency, sound level, voicing and location (Bizley, Walker et al. 2013, Town, Atilgan et al. 2015). Each ferret was chronically implanted with Warp-16 microdrives (Neuralynx, MT) housing sixteen independently moveable tungsten microelectrodes (WPI Inc., FL) positioned over primary and posterior fields of left or right auditory cortex (Fig. S1). Details of the surgical implantation procedures and histological confirmation of electrode position are described elsewhere (Bizley, Walker et al. 2013). A further six ferrets (also pigmented females) implanted with the same microdrives were used as naïve animals for passive recording. These animals were trained in a variety of psychophysical tasks that did not involve the vowel sounds presented here.

Subjects were water restricted prior to testing; on each day of testing, subjects received a minimum of 60ml/kg of water either during testing or supplemented as a wet mash made from water and ground high-protein pellets. Subjects were tested in morning and afternoon sessions on each day for up to five days in a week. Test sessions lasted between 10 and 50 minutes and ended when the animal lost interest in performing the task.

The weight and water consumption of all animals was measured throughout the experiment. Regular otoscopic examinations were made to ensure the cleanliness and health of ferrets' ears. All experimental procedures were approved by a local ethical review committee and performed under license from the UK Home Office and in accordance with the Animals (Scientific Procedures) Act 1986.

Apparatus

Ferrets were trained to discriminate sounds in a customized pet cage (80 cm x 48 cm x 60 cm, length x width x height) within a sound-attenuating chamber (IAC) lined with sound-attenuating

foam. The floor of the cage was made from plastic, with an additional plastic skirting into which three spouts (center, left and right) were inserted. Each spout contained an infra-red sensor (OB710, TT electronics, UK) that detected nose-pokes and an open-ended tube through which water could be delivered.

Sound stimuli were presented through two loud speakers (Visaton FRS 8) positioned on the left and right sides of the head at equal distance and approximate head height. These speakers produce a flat response (±2 dB) from 200Hz to 20 kHz, with an uncorrected 20 dB drop-off from 200 to 20 Hz when measured in an anechoic environment using a microphone positioned at a height and distance equivalent to that of the ferrets in the testing chamber. An LED was also mounted above the center spout and flashed (flash rate: 3 Hz) to indicate the availability of a trial. The LED was continually illuminated whenever the animal successfully made contact with the IR sensor within the center spout until a trial was initiated. The LED remained inactive during the trial to indicate the expectation of a peripheral response and was also inactive during a time-out following an incorrect response.

The behavioral task, data acquisition, and stimulus generation were all automated using custom software running on personal computers, which communicated with TDT real-time signal processors (RZ2 and RZ6, Tucker-Davis Technologies, Alachua, FL).

Task Design, Stimuli and Behavioral Testing

Ferrets discriminated vowel identity in a two-alternative forced choice task described elsewhere (Town, Atilgan et al. 2015). Briefly, on each trial the animal was required to approach the center spout and hold head position for a variable period (0 – 500 ms) before stimulus presentation. Each stimulus consisted of a 250 ms artificial vowel sound repeated once with an interval of 250 ms. Animals were required to maintain contact with the center spout until the end of the interval between repeats (i.e. 500 – 1000 ms after initial nose-poke) and could then respond at either left or right spout. Correct responses were rewarded with water delivery whereas incorrect responses led

to a variable length time-out (3 - 8 s). Incorrect responses were also followed by a correction trial on which animals were presented with the same stimuli with the same time. Correction trials and trials on which the animal failed to respond within the trial window (60 s) were not analyzed.

Stimuli were artificial vowel sounds synthesized in MATLAB (Mathworks, USA) based on an algorithm adapted from Malcolm Slaney's Auditory Toolbox (https://engineering.purdue.edu/~malcolm/interval/1998-010/). The adapted algorithm simulates vowels by passing a sound source (either a click train, to mimic a glottal pulse train for voiced stimuli, or broadband noise for voiceless stimuli) through a biquad filter with appropriate numerators such that formants are introduced in parallel. Four formants (F1-4) were modelled: three subjects were trained to discriminate [u] (F1-4: 460, 1105, 2857, 4205 Hz) from [ɛ] (730, 2058, 2857, 4205 Hz) while one subject was trained to discriminate [a] (936, 1551, 2975, 4263 Hz) from [i] (437, 2761, 2975, 4263 Hz). Selection of formant frequencies was based on previously published data (Peterson and Barney 1952, Town, Atilgan et al. 2015) and synthesis produced sounds consistent with the intended phonetic identity. Formant bandwidths were kept constant at 80, 70, 160 and 300 Hz (F1-4 respectively) and all sounds were ramped on and off with 5 ms cosine ramps.

To test perceptual constancy, we varied the rate of the pulse train to generate different fundamental frequencies and used broadband noise rather than pulse train to generate voiceless vowel. For sound level we simply attenuated signals in software prior to stimulus generation. For sound location, we presented vowels only from the left or right speaker whereas for all other tests sounds were presented from both speakers. Across variations in F0, voicing and space, we fixed sound level at 70 dB SPL. For tests across sound level and location, voiced vowels were generated with 200 Hz fundamental frequency. Sound levels were calibrated using a Brüel & Kjær (Norcross, USA) sound level meter and free-field [1/2] inch microphone (4191) placed at the position of the animal's head during trial initiation.

Neural Recording

Neural activity in auditory cortex was recorded continuously throughout task performance. On each electrode, voltage traces were recorded using TDT System III hardware (RX8 and RZ2) and OpenEx software (Tucker-Davis Technologies, Alachua, FL) with a sample rate of 25 kHz. For extraction of action potentials, data were bandpass filtered between 300 and 5000 Hz and motion artefacts were removed using a decorrelation procedure applied to all voltage traces recorded from the same microdrive in a given session (Musial, Baker et al. 2002). For each channel within the array, we identified candidate events as those with amplitudes between -2.5 and -6 times the RMS value of the voltage trace and defined waveforms of events using a 32-sample window centered on threshold crossings.

In the current study, waveform shapes were not sorted and data from multiple test sessions combined across days. The activity for each unit thus represents the unsorted multi-unit activity of a small population of cells at the recording site. We identified sound responsive units as those whose stimulus evoked response within the 300 ms after onset of first token differed significantly from spontaneous activity in the 300 ms before making contact with the spout (t-test, $p < 0.05$).

Decoding procedure

We decoded vowel identity across trials using a simple spike-distance decoder with leave-one-out cross-validation (LOCV). For every trial in a given condition (see below), we calculated template responses as the mean PSTH response to each vowel from all but one test trial and used the Euclidean distance between test trial and template responses to recover vowel identity (Fig. S2A). Where equal distances were observed between templates, we randomly assigned vowel identity. This procedure was repeated for all trials and decoding performance measured as the percentage of trials on which vowel identity was correctly recovered.

Auditory cortical units showed a wide variety of response profiles that made it difficult to select a single fixed time window over which to generate PSTHs. To accommodate the heterogeneity of auditory cortical neurons and identify the time at which stimulus information arose, we repeated our decoding procedure using a series of time windows with varying start time (-0.5 to 1 s after stimulus onset, varied at 0.1 s intervals) and duration (10 to 500 ms, 10 ms intervals) (Fig S2B). Within this parameter space we then reported the parameters that gave best decoding performance. Where several parameters gave best performance we reported the time window with earliest start time and shortest duration.

To assess the significance of decoding performance, we conducted a permutation test in which the decoding procedure (including temporal optimization) was repeated 100 times but with vowel identity randomly shuffled between trials to give a null distribution of decoder performance. The null distribution was then parameterized by fitting a Gaussian probability density function, for which we then calculated the probability of observing the real decoding performance. Here we only considered units for which there was a minimum of five trials from which to generate template responses for each stimulus condition. As such, trial number was the limiting factor in our choice of decoder as more complex approaches (e.g. linear discriminant analysis, mutual information etc.) required more trials for each stimulus condition than our study allowed.

Population Decoding

To decode vowel identity from the responses of populations of units on a given trial (y), we used a maximum likelihood estimate in which the likelihood of each stimulus class (e.g. vowel identity, F0, level etc.) was simply the weighted sum of individual unit estimates:

$$likelihood(stim) = \sum_{i=1}^{N} w_i$$

Where N is the number of units in the population, and w_i is a confidence weighting based on spike distances:

$$w = \frac{d_{stim}}{\sum_{j=1}^{n} d_j}$$

Where d_1 to d_n are the spike distances between the test trial response and template responses (for n stimulus classes – e.g. vowel identity or F0), and d_1 represents the minimum distance.

We tested populations of up to 50 units, by which point decoder performance had typically saturated at 100% (with the exception of decoding F0 and sound level). Populations were constructed by random sampling without replacement from the possible cohort of units under investigation. For each population size for which the number of possible combinations of units exceeded the number of units available, we repeated sampling 100 times.