Spectral timbre perception in ferrets; discrimination of artificial vowels under different listening conditions

Jennifer K Bizley[*],
Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford, OX1 3PT

Kerry MM Walker,
Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford, OX1 3PT

Andrew J King, and
Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford, OX1 3PT

Jan WH Schnupp
Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford, OX1 3PT

Abstract

Spectral timbre is an acoustic feature that enables human listeners to determine the identity of a spoken vowel. Despite its importance to sound perception, little is known about the neural representation of sound timbre and few psychophysical studies have investigated timbre discrimination in non-human species. In this study, ferrets were positively conditioned to discriminate artificial vowel sounds in a two-alternative-forced-choice paradigm. Animals quickly learned to discriminate the vowel sound /u/ from / /, and were immediately able to generalize across a range of voice pitches. They were further tested in a series of experiments designed to assess how well they could discriminate these vowel sounds under different listening conditions. First, a series of morphed vowels was created by systematically shifting the location of the first and second formant frequencies. Second, the ferrets were tested with single formant stimuli designed to assess which spectral cues they could be using to make their decisions. Finally, vowel discrimination thresholds were derived in the presence of noise maskers presented from either the same or a different spatial location. These data indicate that ferrets show robust vowel discrimination behavior across a range of listening conditions and that this ability shares many similarities with human listeners.

INTRODUCTION

One of the most important functions of the auditory system is to recognize and identify complex sounds, a process which requires that listeners disregard nuisance variables in order to extract informative features and enable classification to some broader category. This is particularly the case in human speech where we must correctly discriminate sounds despite

Corresponding Author: Jennifer Bizley, Address for correspondance, Ear Institute, University College London, 332, Grays Inn Road, London, WC1E 6BT, j.bizley@ucl.ac.uk.
[*]Current Address: Ear Institute, University College London 332, Grays Inn Road, London WC1E 6BT

[3]However, reveberation might influence the extent to which spatial cues are useful for separating a signal and masker. As such, our study may underestimate the benefit provided by spatially separating the vowel from the masking noise.

the natural variation in acoustical structure that occurs across different talkers. Vowel sounds are generated when resonances within the vocal tract filter periodic glottal pulse trains emanating from the vocal cords. The periodic pulse train ensures that vowels possess a harmonic structure, and the resonances create distinct energy peaks in the spectrum, referred to as formants. Individual voices can vary in many ways, including their fundamental frequency and spectral distribution, but by extracting the relative locations of the first and second formant frequencies, it is usually possible to identify the type of vowel being produced regardless of these speaker-dependent differences (Assmann *et al.* 2008, Nearey 1989).

Because the recording of single neuron activity is necessarily invasive we are almost always restricted to using animal models to explore the neural basis for the discrimination and classification of complex sounds, including vowels. This requires that the animal model in question is able to perceive these sounds. A number of observations suggest that the ferret is a suitable species in which to study the neural processes underlying the translation of sound acoustics into perceptual features. Ferrets are highly sensitive to low-frequency pure tones (Kelly *et al.* 1986), there is evidence that they are sensitive to the harmonic fusion of tone complexes (Kalluri *et al.* 2008), they are able to identify the direction of a pitch change (Walker *et al.* 2009) and can detect changes in pitch or spectral timbre of artificial vowels (Walker *et al.* 2011), and, finally, many of their own vocalizations contain low-frequency energy and are strongly periodic. The responses of ferret auditory cortical neurons encode information about the timbre of artificial vowel sounds (Bizley *et al.*, 2009) and support the discrimination of human speech sounds (Mesgarani *et al.* 2008). However, the ability of ferrets to identify sounds according to their spectral envelope has not been studied. Such a behavioral assessment of vowel identification is of interest in the ferret since previous studies have demonstrated that neurons in the auditory cortex of this species are highly sensitive to spectral timbre. In a study that quantified the proportion of the neural response variance that could be attributed to changes in sound location, or pitch, or timbre, the timbre almost always accounted for more variance than the other two features (Bizley *et al.* 2009).

Studies of spectral timbre perception in human listeners typically ask subjects to perform a vowel identification task (Assmann *et al.* 2008, Nearey 1989, Ryalls *et al.* 1982). Since listeners are highly experienced at listening to speech sounds they presumably draw upon their existing knowledge in order to perform this task. Previous studies have demonstrated that non-human primates (Hienz *et al.* 1988, Sinnott *et al.* 1991, Sommers *et al.* 1992), cats (Hienz *et al.* 1996a, Hienz *et al.* 1996b), ferrets (Walker *et al.* 2011), chinchillas (Burdick *et al.* 1975), gerbils (Sinnott *et al.* 2003), birds (Heinz *et al.* 1981) and rats (Eriksson *et al.* 2006) are capable of discriminating vowel-like sounds. In contrast to human studies, the majority of these investigations used change-detection paradigms in order to measure the minimum discriminable change in formant frequency. Such paradigms require an animal to listen to a repeating standard sound, which, after some number of repetitions, may change to a test sound. Animals are rewarded for responding to this change. This approach has been used to measure thresholds for detecting the change in formant location (Hienz *et al.* 1996b, Hienz *et al.* 1988), sensitivity to vowel continuum (Sinnott *et al.* 1991, Sinnott *et al.* 2003) and vowel discrimination thresholds in noise (Hienz *et al.* 1996a). In contrast one early study used instrumental aversive conditioning to train chinchillas to discriminate /a/ from /i/ (Burdick *et al.* 1975). These animals were able to generalize to vowels from different talkers and to artificial vowel sounds. We adopted a similar approach by testing ferrets on a two-alternative forced-choice classification task rather than a change-detection task. This allowed us to measure their ability to recognize and identify the spectral envelope of a sound. While previous studies have investigated the ability of ferrets to detect or discriminate pure tone stimuli (Fritz *et al.* 2010, Hine *et al.* 1994, Kelly *et al.* 1986, Yin *et al.* 2010), localize sounds (Kacelnik *et al.* 2006, Kavanagh *et al.* 1988, Nodal *et al.* 2008,

Parsons *et al.* 1999), discriminate stimulus periodicity (Walker *et al.* 2009) and categorize natural sounds (Schnupp *et al.* 2006), to date no previous study has investigated the ability of ferrets to discriminate artificial vowel sounds, nor has any other study required the identification of stimuli that vary only in their spectral envelope.

We trained ferrets to discriminate the vowel sound /u/ from / / under a range of listening conditions. Our aim was to assess to whether ferrets were able to correctly identify a vowel and to what extent this was dependent on spectral envelope extraction. In the first of four experiments, we examined how changing the fundamental frequency of the vowel on a trial-to-trial basis affected classification ability. We then examined how ferrets classified a continuum of morphed vowel sounds created by systematically shifting the location of the formant frequencies, and used single formant vowels to determine which cues the ferrets may be using in order to make their judgments. Finally, we assessed how well ferrets were able to identify such sounds in the presence of background noise and whether they were able to utilize a spatial cue in order to achieve identification at a lower signal-to-masker ratio.

I. METHODS

A. Subjects

Four adult pigmented female ferrets were used in this study. They were housed in groups of either two or three, with free access to high-protein food pellets and water bottles. On the day before training, water bottles were removed from the home cages and were replaced on the last day of a training run. Training runs lasted for five days or less, with at least two days between each run. On training days, ferrets received drinking water as positive reinforcement while performing a sound discrimination task. Water consumption during training was measured, and was supplemented as wet food in home cages at the end of the day to ensure that each ferret received at least 60 ml of water per kilogram of body weight daily. Regular otoscopic and typanometric examinations were carried out to ensure that the animals' ears were clean and healthy. All experimental procedures were approved by the local ethical review committee and were carried out under licence from the UK Home Office, in accordance with the Animals (Scientific Procedures) Act 1986.

B. Training apparatus

Ferrets were trained to discriminate sounds in custom-built testing chambers constructed from double-glazing units that incorporated a sound-insulating vacuum[1]. The ceiling of the chambers was covered in sound-absorbing foam. The chambers were approximately 45 cm wide, 62 cm long, and 54 cm high (Figure 1A). A plexiglass wall, 12 cm from the front of the box, separated the animal from the electronics and tubing of the apparatus. Three nose-poke holes fitted with infra-red emitters and detectors were mounted on the plexiglass wall, which provided a centrally positioned "start spout" and two "response spouts" positioned to the left and right. Each nose poke hole was also fitted with a water spout. When the ferret

---

[1]Our choice of testing chamber was based around two key requirements; firstly, that the sound attenuation was sufficient to place multiple testing boxes within one laboratory, and secondly, that the surfaces were non-metallic such that the testing boxes was also suitable for performing neural recordings in animals chronically implanted with microelectrodes, whilst they performed sound discrimination tasks. Therefore, the boxes were constructed from double glazing units due to their excellent sound attenuation properties – we had multiple testing chambers placed adjacent to one another so it was obviously crucial that sounds from one chamber were inaudible in the neighbouring ones. A microphone placed inside one testing box while sounds were presented in a next door box confirmed that for the sound levels at which the vowels were presented the sound attenuation produced was effective in that the signal was not detectable above the noise floor in the adjacent testing chamber (measured values for the attenuation of the vowel stimuli between boxes was at least 41 dB, and the attenuation of the brief clicks produced by the solenoids that deliver the water reward was measured to be 37-41 dB. Glass is also an ideal substance to construct a testing arena from as it is non-metallic allowing the animal to remain electrically isolated and preventing artifacts during recording. Moreover, it is easy to keep clean and sterile (essential when animals with chronic implants are to be tested within it).

placed its nose into the nose-poke holes the infra-red beam was broken, enabling the animal's response to be detected. Correct responses were rewarded by water delivery from the spout at the back of the nose poke hole. Sound stimuli as well as acoustic feedback signals were delivered via two loudspeakers (Visaton FRS 8) which were mounted above the central and right hand spouts. In all but Experiment 4 only the speaker above the central response spout was used. These speakers produce a flat response (± 2 dB) from 200 Hz to 20 kHz, with an uncorrected 20 dB drop-off from 200-20 Hz.

The behavioral task, data acquisition, and stimulus generation were all automated using custom software running on personal computers, which communicated with TDT RM1 real-time signal processors (Tucker-Davis Technologies, Alachua, FL).

C. Acoustical stimuli

The stimuli were artificial vowel sounds that were created in MATLAB using custom software, based on an algorithm adapted from Malcolm Slaney's Auditory Toolbox (http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/). The vowel sounds were composed of click trains that were band-pass filtered to add formants centered at 460, 1105, 2735 and 4052 Hz to give the vowel /u/ or 730, 2058, 2979 and 4294 to give the vowel / / (Kewley-Port *et al.* 1996, Kewley-Port *et al.* 1994, Peterson *et al.* 1952). These two vowel sounds were chosen due to their difference in both first and second formant locations. When viewed in F1-F2 space they occupy opposite poles (Peterson *et al.* 1952). A further motivation for these particular vowels was provided by previous extracellular recording studies which demonstrated that a large proportion of auditory cortical neurons in the ferret were modulated by changing vowel timbre (Bizley *et al.* 2009). These two vowels represented the most dissimilar vowels out of the four we tested neurophysiologically and therefore provided an ideal starting place for examining whether ferrets could discriminate such sounds behaviorally. Vowel sounds were ramped on and off with 5 ms wide raised cosine ramps. The repetition rate of the click train determined the vowel fundamental frequency. Unless otherwise stated (e.g. Experiments 1 and 3), vowels were presented at a 200 Hz fundamental frequency. Masking noises (Experiment 4) were either white noise or low-frequency noise. White noise was broadband (1 Hz – 12 kHz) and was 'frozen' for all trials within a single testing session and, so that the results were not influenced by the particular characteristics of any one noise burst, were generated afresh for each testing session. Low-frequency noise was white noise bandpass filtered (2nd order Butterworth filter) between 110-320 Hz. This was was also frozen within a single testing session and varied across sessions. Sound levels were calibrated using a Brüel & Kjær (Norcross, USA) sound level meter and free-field, ½ inch microphone (type 4191). The average sound level of the artificial vowels were approximately 75 dB SPL. For experiments 1-3 the level of the vowels were varied randomly on a trial-to-trial basis over a ± 6 dB range.

D. Training

Prior to psychoacoustic testing, animals were trained on the task by progressing through a four stage behavioral conditioning program using positive reinforcement with water reward. Naive animals took between 28 and 62 days (excluding rest days, mean= 41 days, SD=15 days) of training to progress from entering the testing box for the first time to reaching criterion on the final training stage.

Training Phase 1: Initial shaping—All animals initially spent at least two sessions being shaped to the experimental procedure. In all training and testing stages, a flashing cue light indicated to the animal that the next trial could be initiated. Animals initiated each trial by nose-poking at the centre spout. This caused the cue light to be extinguished until the start of the next trial, and a small water reward was delivered at the centre spout for correct

initiation. The animal then received a larger water reward if it went on to nose poke at either of the two peripheral response spouts. Once the animals were familiar with the routine of initiating trials at the centre "start spout" and then choosing one of the response spouts at either side, they proceeded to the next stage of training. This involved at least 4 sessions in which they were presented with either a target vowel (/u/) or an unfiltered click train presented at the same fundamental frequency and only the spout appropriate to that stimulus was 'active' and able to register responses. The stimuli were presented in a loop, i.e. 250 ms of sound were presented once every second until the animal nose poked the appropriate response spout. Sessions were alternated such that the animal heard only the vowel sound and responded at the left spout for one session, then in the next testing session heard only the click train and was required to respond at the right hand spout. This stage therefore did not require the animals to discriminate the stimuli but was included so that they would learn to associate the /u/ with the left spout and the click train with the right spout. During this and subsequent phases, the time that the animal had to wait at the center start spout before the stimulus was presented was progressively lengthened from 100 ms to several seconds. This delay was important in ensuring that animals remained at the center spout long enough to listen for the sound, and was especially important preparation for experiments 4 and 5 when they would be working at threshold in the presence of a noise masker. Once trained, animals were required to wait for an amount of time randomly chosen to be between 500 and 1500 ms.

Training Phase 2: Basic acoustic discrimination—Animals were next trained to discriminate the target sounds presented in the initial training phase: i.e. the target /u/ vowel sound from an unfiltered click train which had the same 200 Hz fundamental frequency (F0). Animals performed 2-4 training sessions during which the stimuli were presented once a second until a correct response was made. The incorrect spout on each trial was inactivated such that attempts to respond there were not registered and the sound continued to play until the animal triggered the correct spout. Once animals were performing approximately at least 50 trials per session, typically after 2-4 sessions, both spouts were activated so that an incorrect response now terminated the trial with a broad band noise burst, to indicate an incorrect response, and a short time-out interval as a negative reinforcer. During the time-out period, which was initially a few seconds increasing to 14 seconds through the course of training, the animal was unable to initiate a new trial. To prevent ferrets from biasing their responses to one particular response spout, incorrect responses were followed by one or more correction trials in which the previous stimulus was repeated until the ferret made the correct response. Within a few days animals were able to perform this task at >85% correct. Once animals had reached this level of performance on at least 4 consecutive sessions they progressed to phase 3.

Training Phase 3: Vowel discrimination with looped sounds—Once animals could reliably discriminate the unfiltered click train from the vowel sound, the click train was substituted by the second target vowel, again presented in a looped fashion at a constant F0 (200 Hz) with trial-to-trial variability in sound level (pseudorandomly across a 12 dB range). Substituting the click train for the second vowel did not generally affect performance. Animals were tested on this task until they performed at at least 85% correct for at least 4 consecutive sessions. As in previous stages and during all subsequent testing, incorrect responses were marked by a white noise burst and a 14 second 'time-out', and were followed by a correction trial.

Training Phase 4: Vowel discrimination—At this point, we switched from presenting the vowel sounds in a continuous loop until the animals made their response to presenting only two repetitions of the target vowel (each vowel was 250 ms in duration separated by a

750ms inter-stimulus silent interval). This task is represented schematically in Figure 1C. Once animals performed this new task with an accuracy of at least 85% correct across four consecutive sessions, they progressed to the testing stage.

E. Psychoacoustic testing

Experiment 1: The effect of varying fundamental frequency on timbre discrimination—Since the ability to maintain an invariant vowel perception across multiple talkers and voice pitches is a crucial feature of human vowel perception, we were interested to know whether ferrets were able to generalize over a range of voice pitches. The ability to do so indicates that the listener is capable of extracting the underlying spectral envelope and disregarding the changes in fundamental frequency. Once ferrets had reached the criterion level of performance in Training Phase 4, the F0 of the vowel sounds was randomly varied across a range of 150 Hz to 500 Hz, in 25 Hz steps. The mean ($\pm$SD) number of trials per animal in Experiment 1 was 8250$\pm$1580.

Experiment 2: testing the identification of a series of morphed vowels—We generated a series of 8 vowel sounds, which comprised the trained /u/ and / / and six vowels in which the first and second formant frequencies were progressively shifted along a continuum from /u/ to / / in equal, linearly spaced, frequency steps. We tested these morphed vowels to explore whether ferrets perceived vowels categorically, or in a more continuous fashion, like human listeners (Pisoni 1973). Vowels were presented at a fixed F0 of 200 Hz. Ferrets were rewarded for responding at the /u/ spout for the /u/ and for the three vowels that were most /u/ like, and at the / / spout for the / / and for the three vowels whose formant frequencies were closest to those in the / /. All four ferrets were tested on this task and each performed a minimum of 1000 trials (2139 $\pm$ 969, mean $\pm$ SD). Within each testing session, randomly selected morphed vowels were presented as targets on 25% of trials and the remaining 75% of targets were the original complete vowels. Morphed vowels thus represented a minority of test stimuli in each session, as we were primarily interested in how the ferrets naturally classified the sounds, rather than how well they could *learn* to classify the sounds.

Experiment 3: Using single formant vowels to explore which cues ferrets used to discriminate trained vowel sounds—Human listeners rely on the relationship between formant frequencies in order to correctly identify vowels. Since the ferrets in this study were required to discriminate between two vowels only, they could potentially base their judgments on the frequency of either the first or second formant or a combination of the two. In order to probe which spectral cues the ferrets used to inform their vowel discrimination judgments, we generated single-formant versions of the two trained vowels, each with a formant at either the first or second formant location and used these as probe trials to test the ability of two trained ferrets to identify these sounds. F0 was varied from trial to trial across five or six different values (180, 200 (one animal), 255, 330, 405 and 480 Hz). Because we wanted to measure which cues the ferrets were using when discriminating multi-formant vowels, rather than test their ability to learn to discriminate single formant vowels, only a total of 10 trials were collected, across multiple sessions, for each condition. These single formant 'probe' trials were randomly interleaved with the original 2-formant vowels; on the probe trials ferrets were rewarded for correct responses (i.e. identifying the target vowel from which the single formant came) but did not receive any time-out or correction trials for incorrect responses[2]. Testing sessions which included a small number of these probe stimuli (<25%) were included as only one of the twice-daily testing sessions per

---

[2]While this effectively introduces a third class of contingency in this situation, these trials were so rarely inserted that we beleive it highly unlikely that the ferret was able to pick up and form a new rule based upon these limited exposures.

week, and were separated across several weeks, so that the opportunities for ferrets to learn to make the new single-format discriminations were limited.

Experiment 4: Testing the ability of ferrets to identify vowel sounds in noise—
The ability of ferrets to discriminate vowels in the presence of background noise was tested by presenting vowels of varying intensity embedded in a noise masker. Our rational for this experiment was to determine to what extent noise degraded the abilities of trained animals to identify vowel sounds based on spectral timbre. As outlined in the 'Acoustical Stimuli' section, we used two different noise maskers – a broad band white noise designed to mask all frequencies equally, and a low-frequency masker designed to mask the fundamental frequency of the vowels (see stimuli). Masking sounds were maintained at a constant overall amplitude of 65 dB SPL, and the level of the vowel was varied adaptively in a two-up one-down staircase procedure. The step sizes for the first three reversals were 5 dB, and step sizes thereafter were 2 dB. Testing continued until the ferret was satiated. The masking sound was 4 seconds in duration and began 200 ms before the first vowel sound on each trial, and, as in all Experiments, two repetitions of each vowel were presented, each of 250 ms duration, with a 750 ms silent interval. The noise and the vowel sound were either presented from the same central speaker, or the masking noise was presented from a speaker to the animal's right and the target vowel from the central speaker. The targetspeaker was positioned at the front of the testing chamber, offset from the central speaker by 13 cm, giving an azimuthal separation at the ferret's head of approximately 60°. Each animal performed a minimum of 700 trials at each masking (mean number of trials 1526).

F. Data analysis

Correction trials were excluded from the data analysis. Data were analyzed and statistical tests performed using MATLAB (MathWorks, USA). Psychometric functions in Experiment 2 (identification of morphed vowels) were generated using logistic regression, as in previous studies (Bizley *et al.* 2007, Schnupp *et al.* 2005, Walker *et al.* 2009).

In Experiment 4, where signal levels were adjusted adaptively in the presence of noise, the detection threshold value was estimated using critical values of the binomial distribution. Conventional adaptive staircase methods (Levitt 1978) often use threshold estimates based on the average level across a certain number of reversals. However, we found that our ferrets' performance was often too variable to make this a reliable method. Two-down, one-up staircase methods, for example, will normally converge toward the 70.7% performance point, but they do so reliably only if the attention and motivation of the subject remain constant throughout the session. Brief periods of inattention, which seemed not uncommon among our ferret subjects, can cause spurious reversals that would significantly affect threshold estimates based on the reversals only. To obtain robust threshold estimates, we combined data across multiple testing sessions and defined as threshold the lowest stimulus value at which we could say with 99% confidence that the animal's performance was better than just guessing, i.e. > 50% correct. The confidence limits were obtained using critical values of the cumulative binomial distribution. Specifically, if at stimulus level L the ferret performed C out of N trials correctly, then threshold is given by L such that

$$1 - binocdf\left(C\left(L\right), N\left(L\right), 0.5\right) \leq 0.01 \quad \text{Equation 1}$$

Where *binocdf()* is the cumulative binomial distribution. We defined as threshold the smallest L for which the probability that the animal merely guessed is <1%, i.e. 1-*binocdf*(*C*, *N*, 0.5) 0.01. Setting sound levels adaptively using a staircase method ensures that data collection is focused around stimulus values close to threshold, and the use of this 99% confidence criterion allows the threshold to be estimated using all the behavioral data, not

just those responses where reversals occurred. This method is discussed further in the appendix.

II RESULTS

A. Discrimination of artificial vowel sounds which vary in fundamental frequency

Figure 2 A-D plots the ability of four ferrets to discriminate / / and /u/ across voice pitches from 150 Hz to 500 Hz. These data show that ferrets were able to maintain high levels of performance across the pitch range with average scores of 87.5 ± 1.9% (mean ± SD) correct. In two animals, there was a significant negative correlation between F0 and percentage correct scores (Figure 2A, C; slopes of –0.018%/Hz and –0.023%/Hz; R^2 = 0.48 and 0.42; p = 0.002 and 0.005, respectively), but the other two animals showed no significant dependence of performance on F0. In summary, we observed a weak trend for discrimination of the vowels / / and /u/ to be poorer at higher F0. Nevertheless, the animals were clearly able to generalize across F0; indeed this ability was apparent in the very first testing session in which the trial-to-trial F0 was roved, with percentage correct scores on that first session of 84.8 ± 1.5% (mean ± SD).

B. Discrimination of morphed vowel stimuli

Experiment 2 used a continuum of morphed vowels generated by shifting the first and second formants in linear steps from /u/ to / / (Figure 3A). The results are plotted in Figure 3B as a proportion of / / responses. Both the individual data and the fitted psychometric function suggest that ferrets perceived these sounds as a continuum with a smooth change in performance as the formant structure of the stimuli was changed. There is no indication of an abrupt change in performance, as might be expected if the animals perceived these stimuli categorically. The 50% point was at a first-formant frequency of 573 Hz, which is very close to the logarithmic mean of the F1 of the two vowels (579 Hz).

In order to examine whether there was any effect of training on the pattern of responses to the morphed vowels, we compared performance during the first three and last three morphed vowel testing sessions for each ferret (n = 299 ± 90 (mean ± SD) trials in the three sessions combined). No difference was found between the early and late testing blocks in the overall percentage correct scores, the 50% points on the psychometric functions, or the slope estimated from the fitted functions (paired t-tests, p = 0.84, 0.78 and 0.46, respectively).

C. Discrimination of single formant vowels

Experiment 3 used a set of single-formant "vowels" constructed using the first and second formants of the trained vowels, and roved in F0. The amplitude spectra are shown in Figure 4. Figure 5 plots data from two animals, broken down according to the F0 of the vowel sound and the formant that was presented. When the single-formant stimulus comprised the first formant of /u/ (460 Hz) or the second formant of / / (2058 Hz), almost perfect discrimination performance was achieved. However, performance became less reliable when either of the intermediate-frequency formants (i.e. the second formant of /u/ (1105 Hz) or the first formant of / / (730 Hz) was presented; categorization accuracy at these formant values varied with the F0 of the sound.

Because the harmonics of the click train might interact with the position of the formant, we examined the spectra of each of these single formant vowels (Figure 4). It is clear from this figure that the location of the intermediate formants (the second formant of /u/ and the first formant of / /) varied somewhat depending on the F0 of the vowel sound. Importantly, across all fundamental frequencies the location of peak energy (defined as the harmonic with the highest power) for these two formants does not form two discrete distributions; rather it

forms a continuum between the two sets of formants. The across-F0 standard deviations of the location of peak energy for each of the formants was 68Hz (first formant of /u/), 126 Hz (first formant of / /), 95 Hz (second formant of /u/) and 46 Hz (second formant of / /), supporting the observation that the middle-frequency formants were most variable and, in fact, form a continuous distribution. While we cannot be sure what cues the ferrets used when discriminating the full four-formant vowels, these observations suggest that the ferrets can identify these two vowels using just the location of the first formant of /u/ and the second formant of / /, which are the least ambiguous of the formant frequency locations.

D. Vowel discrimination in the presence of noise maskers

In Experiment 4, a variable attenuation was applied to the full four-formant vowel sounds in the presence of a 65 dB SPL noise masker that was presented 200 ms before the target sounds and continued for one second after it. Data were pooled across testing sessions and binned into 3 dB attenuation steps so that the average percentage correct score could be calculated for each target sound intensity. Thresholds were then estimated using the binomial distribution to calculate when performance exceeded chance levels as described in section F of the Methods. Figure 6A shows representative data from one of the ferrets and Figure 6B shows the threshold data for all four animals.

The across- and within-subject data shown in Figure 6 suggest that spatially separating the target, versus collocated conditions (C), and the masking sounds resulted in a release from masking of 5-15 dB. A Friedman test on the thresholds calculated under the four conditions (two noise, two spatial configurations) confirmed that there were significant effects of stimulus condition (p<0.013). To further quantify these effects we ran a three-way ANOVA with animal, space and noise type as factors. Space was the only significant factor ($F_{(1,10)}$ = 12.4, p = 0.006, noise type: $F_{(1,10)}$ = 0.09, p = 0.77; animal $F_{(3,10)}$ = 0.24, p = 0.89), and most (53 %) of the observed variance in the thresholds is attributable to space, with the noise type accounting for only 3 % of the variance, and animal for less than 1%.

Discussion

This dataset represents the first investigation into vowel discrimination in ferrets, and demonstrates that ferrets readily learn to make such discriminations. Importantly, like human listeners, the ferrets were able to generalize over a range of fundamental frequencies and maintain performance in the presence of background noise.

Although human listeners effortlessly generalize vowel identification over the natural range of voice pitches (~100-220 Hz), it has been demonstrated that extending the fundamental frequency above this range impairs performance, with higher pitches leading to reduced vowel discrimination accuracy (Assmann *et al.* 2008, Kewley-Port *et al.* 1996, Miller 1953, Ryalls *et al.* 1982, Slawson 1968). Since the precise location of the formants that characterize vowel sounds varies between speakers, it is thought to be the relationship between the center frequencies of the first and second formants that principally determines vowel identity (Kewley-Port *et al.* 1996). As the fundamental frequency increases, the spacing of the harmonics within the vocalization becomes wider and thus the resolution with which the location of the formant peaks can be estimated declines (de Cheveigne *et al.* 1999). We found that although ferrets were able to discriminate vowels well at the highest fundamental frequencies tested, there was a small overall decline in accuracy with increasing F0. In human listeners it has been noted that perception is particularly impaired once the F0 is higher than the location of the first formant frequency (Smith *et al.* 2005). In the present study, this was the case only at the highest F0 used (500 Hz), which exceeded the first formant frequency of /u/ (460 Hz), but not of / / (730 Hz). This may therefore explain why the ferrets were able to maintain their discrimination between these two vowels. We

may have observed a more marked effect of F0 on discrimination had both target vowels had a first formant frequency below 400 Hz (for example, in discriminating /i/ versus / /). An alternative possibility is that the decrement in performance observed in human listeners relates more to their experience with spoken language; the human voice occupies a restricted range of fundamental frequencies and manipulations outside of this range may impair performance. Indeed, studies of human vowel perception find that listeners are able to deal more readily with changes in fundamental frequency when these changes are correlated with upward shifts (of around 10%) in the formant frequencies (Slawson 1968).

Experiment 2 examined how well ferrets discriminated a series of morphed vowels that were generated by shifting the location of the first and second formant frequencies from /u/ to / / (Figure 3A). Our objective was to examine how ferrets classified a morphed vowel continuum and to explore whether they perceived them categorically or more continuously. Their behavior indicated that they perceived the sounds as existing along a continuum, rather than as two discrete categories. A psychophysical investigation of vowel encoding in gerbils (Sinnott *et al.* 2003) demonstrated a similar pattern of responses to a morphed vowel continuum, again indicating that as in human listeners (Fry *et al.* 1962) a smooth perceptual transition exists, rather than a sharp categorical boundary. Most previous studies with non-human listeners have focused on estimating the smallest change in formant frequency that animals can detect. The threshold for detecting a change in formant location varies with centre frequency and species, with the difference limens in Macaque monkeys being 12.5 Hz at a 500 Hz formant center frequency and 22 Hz at a 1200 Hz center frequency. Cats are capable of detecting second formant changes in the order of 36-87 Hz (Hienz *et al.* 1996b), while the thresholds measured in gerbils for detecting changes in the location of the first and second formants are about 50 Hz and 135 Hz, respectively (Sinnott *et al.* 2003). Our objective was to examine how ferrets classified a morphed vowel continuum, and the smallest changes in first and second formant location within our stimulus set were steps of 39 and 136 Hz respectively, either upwards from /u/ or downwards from / /. Although estimating true thresholds requires a behavioral paradigm designed to explicitly test this, our finding that the ferrets' discrimination accuracy dropped immediately with these step sizes suggests that they can detect differences at least this small.

Experiment 3 used a series of single formant vowels to explore which spectral cues the ferrets may be basing their decision upon. Since stimuli were presented as rarely occurring probe trials, the ferrets had little opportunity to learn to discriminate these sounds. Their responses indicate that formants in the frequency region of the first formant of /u/ and the second formant location of / / were reliably classified. However, when the stimuli included only the intermediary frequencies, the second formant of /u/ and the first formant of / / performance, the animals' responses were much less consistent and, to some degree, dependent upon the F0 of the target sound. Across all values of F0, the first formant of / / (730 Hz) was classified correctly more often than chance, whereas performance for the second formant of /u/ (1105 Hz) was often at chance. Examination of the spectra of these single formant vowels demonstated that the center frequency of the second formant of /u/ and the first formant of / / varied more than the other formants with F0 and that their distribution overlapped. Such observations are consistent with human listeners relying upon the relative locations of the first and second formants since, when the full set of naturally occurring vowels is considered, neither alone uniquely specifies vowel identity (Kewley-Port *et al.* 1996).

In our two-vowel discrimination paradigm, the location of the first formant of /u/ and the second formant of / / uniquely specified the target vowel and it is therefore unsurprising that the ferrets can use these cues to perform the task. Since these vowels were presented as probe trials in a two-alternative forced choice paradigm, we can only assess which cues are

sufficient to support accurate vowel identification. In other words, this does not tell us which cues the trained animals use when discriminating the full four-formant vowel sounds and therefore whether ferrets usually exploit the information available from both formant positions. We can, however, conclude that ferrets are capable of basing their discrimination on the formants that unambiguously distinguish /u/ from / / and are less able to use the more variable cues provided by the second formant of /u/ and the first formant of / /, when presented in isolation.

Experiment 4 estimated signal-to-noise ratios for the discrimination of vowel sounds in the presence of white (1 Hz-12.5 kHz) or low-frequency (110-320 Hz) noise maskers. Discrimination thresholds improved when the spatial location of the target and masking sounds was separated, by between 5 and 15 dB, which is in-line with those observed in ferrets for a tone-in-noise detection task (Hine *et al.* 1994) and budgerigars (Dent *et al.* 1997). Binaural unmasking paradigms most frequently present stimuli over headphones, but a number of studies have also illustrated an advantage of separating the spatial location of a target and its noise masker in the free field in human listeners (Carlile *et al.* 1996, Garadat *et al.* 2007, Saberi *et al.* 1991, Shinn-Cunningham *et al.* 2001). In contrast, our task required that animals identify a target sound rather than simply detect it. A second important difference between this and previous detection studies is that the latter involved presenting a co-localized noise and target sound and compared this to the addition of an in-phase noise source from a separate location Instead, we simply spatially separated the source location of our target and masker and observed that lower discrimination thresholds were obtained. Separating the sources of the target and masking sounds in this fashion will facilitate a 'better ear' effect, whereby the ear closest to the target sound will gain a more favorable signal to noise ratio and, if the ferret is able to attend to this ear, it will gain an advantage in doing so. Since our testing chambers were made from glass, it is possible that the benefits of spatially separating the signal and noise might have been underestimated. For normal hearing and even moderately hearing impaired human listeners, vowel identification abilities are maintained in reveberant listening conditions (Nabelek and Dagenais 1986, Owens et al 1968, Nabelek 1988), but reverberation does negatively impact upon target detection in noise and decrease the benefit gained by spatially separating the target and masker (Koenig *et al.* 1977, Zurek *et al.* 2004).

The ferret provides an excellent animal model for investigating the neural basis of auditory perception, having low frequency hearing (Kelly *et al.* 1986), a well characterized auditory cortex, and being easy to train in sound discrimination tasks. In particular, spectral timbre discrimination seems to be a relatively intuitive task for ferrets – the time taken to train animals in this task (mean ± SD= 41 ±15 sessions, four ferrets), while longer than for a sound-localization task (13 ±7 sessions, 13 ferrets; unpublished observations, though see also (Nodal *et al.* 2008), was less than that required to train other animals in a pitch direction discrimination task (76 ±31, five ferrets, (Walker *et al.* 2009), or a pitch change-detection task (78 ±44, five ferrets) testing sessions (Walker et al., submitted). We have previously employed the same vowel stimuli used in this study to investigate the neural representation of pitch, spectral timbre and sound localization cues across different auditory cortical fields in this species (Bizley *et al.* 2010, Bizley *et al.* 2009). Although the majority of the neurons were found to be sensitive to multiple perceptual features of the sounds that we presented, their responses tended to be most informative about the spectral timbre or vowel identity. Despite the interdependency of neural responses to pitch, timbre and location cues, we also found that single neurons can effectively and unambiguously multiplex information about these sound features within separate temporal windows of their response (Walker *et al.* 2011). We observed that spectral timbre information was present in auditory cortical responses consistently earlier than information about the pitch of vowels, and ferrets trained in a change-detection paradigm could respond to a change in sound timbre more rapidly than

they could detect a change in sound pitch (Walker *et al.* 2011). These results demonstrate that timbre information is encoded by a large number of auditory cortical neurons, and that the timescale of such responses correlates well with behavioral measures. Coupled with our present finding that ferrets are easily trained to discriminate sounds according to their spectral timbre, we conclude from this work that ferrets can provide an excellent animal model for studying the neural encoding of vowel sounds.

In conclusion, we have demonstrated that ferrets are able to accurately classify the vowels / / and /u/, and, importantly, that they are able to do so across a range of listening conditions. Such data add to our knowledge of how different animal species are able to discriminate the acoustical differences in speech-like sounds and potentially provide an animal model for examining the neural basis of such discriminations.

Acknowledgments

Appendix

We derived thresholds for the discrimination of vowels in noise by combining data across multiple testing sessions where the data in each session were collected using an adaptive procedure. This targeted our data collection around the 70.7% performance level but differed from a standard staircase procedure(Levitt 1978) in a number of important ways that are stated and then discussed more fully here.

First, we made no attempt to terminate the staircase after a certain number of reversals, and second, we did not use data from a single session to estimate a threshold. Determining the number of reversals that would be appropriate to estimate a threshold proved tricky as our non-human subjects often took longer to reach a threshold level of performance than a typical human listener and it appeared that by limiting the number of reversals the staircase was effectively being truncated; when left to run longer animals would gradually settle to asymptotic behaviour. Furthermore, since testing provides the ferrets' main access to water, the animals must be allowed to work until they are satiated, and therefore if a staircase is terminated, a new staircase must be initiated, which inevitably did not reach threshold before the animal lost motivation. The disadvantage of not terminating the staircase procedure was that, while performance usually did reach a reasonable threshold, it would subsequently then meander around it with occasional periods of inattention causing rather large increases in the target sound level. Estimating a genuine threshold from these data with standard methods therefore produced variable results. We therefore combined data across testing sessions and used the binomial distribution to estimate when performance was significantly better than chance.

In order to compare the thresholds estimated in this manner with those obtained using the standard procedures we ran a series of simulations allowing us to directly compare the thresolds obtained. Adaptive staircases were simulated from an underlying psychometric function which assumed that d is proportional to signal energy raised to the power k i.e. d = mE^k (McFadden 1966) and threshold (d =1) was set to be 0 dB. Data were simulated using the parameters in the ferret psychoacoustics, i.e. with the first four reversals at 5 dB, and subsequent reversals at 2 dB stepsizes. Staircases were randomly started from between 10 and 16 dB and ran for 8 reversals at the smallest step size with the threshold being an

average of these values. Additionally, 'inattention' was modelled by randomly guessing on a proportion of trials. As the average number of trials contributing to our threshold estimates was 1500, therefore we simulated staircases until the number of trials was 1500 trials (mean number of tracks necessary = 31.8). We were then able to compare the thresholds obtained using our binomial-distribution method, with those obtained from the adaptive staircase, which estimates the 70.7% performance level. We repeated this procedure 200 times simulating levels of inattention of 0, 5, 10 and 20 % of trials. Figure 7 illustrates the thresholds obtained. These methods show that the thresholds obtained using the binomial distribution are always lower than those obtained with the standard methods (by, on average, 4 dB). That the thresholds derived using this method are slightly lower than those estimated using an adaptive staircase is a logical consequence of the analysis method - our use of the binomial distribution asks what performance level is significantly greater than 50%, whereas the adaptive track seeks to find the best approximation of the 70.7% performance level.

To further compare the thresholds derived by pooling across staircases and using our binomial-distribution method with those derived using standard adaptive staircase procedures we repeated our analysis in order to ask at what signal:noise ratio did the ferrets' performance equal or exceed 70.7% correct. In line with the simulated data we found this value to be 4.2 dB ±0.6 (mean ±SEM) higher than the signal-to-noise ratio at which performance was significantly greater than chance (50%) performance.

One caveat of using the binomial- distribution method to estimate thresholds is that the threshold determined will depend on the total number of trials – as trial number increases our ability to detect small but consistent deviations from chance also increases. However, additional simulations confirmed that this effect did not influence the thresholds obtained across the range of trial numbers used in this study, presumably because the number of trials is already large. Two alternative approaches would have been to use the method of constant stimuli to test fixed signal:noise ratios either by randomly selecting values or in a blocked design such that the task got progressively harder. However, such methods yield relatively few trials collected near threshold and might underestimate discrimination ability since, in the first case, the animal is not really penalised for not trying on the rather infrequent difficult trials and, in the second, the animal is often discouraged at the most difficult blocks and, once the task becomes too difficult, often quits. An adaptive procedure by comparison gradually makes the task harder and is perhaps more likely to ensure that the animal gives equal effort as the task gets more difficult, and is less discouraged at very low signal: noise ratios as a series of incorrect trials will yield easier trials. Collecting data in this manner gave us a very large number of trials collected around the animals 70.7% performance level from which we were able to estimate a much more reliable threshold.

REFERENCES

Assmann PF, Nearey TM. Identification of frequency-shifted vowels. J Acoust Soc Am. 2008; 124:3203–12. [PubMed: 19045804]

Bizley JK, Nodal FR, Parsons CH, King AJ. Role of auditory cortex in sound localization in the midsagittal plane. J Neurophysiol. 2007; 98:1763–74. [PubMed: 17596417]

Bizley JK, Walker KM, King AJ, Schnupp JW. Neural ensemble codes for stimulus periodicity in auditory cortex. J Neurosci. 2010; 30:5078–91. [PubMed: 20371828]

Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J Neurosci. 2009; 29:2064–75. [PubMed: 19228960]

Burdick CK, Miller JD. Speech perception by the chinchilla: discrimination of sustained /a/ and /i. J Acoust Soc Am. 1975; 58:415–27. [PubMed: 1184835]

Carlile S, Wardman D. Masking produced by broadband noise presented in virtual auditory space. J Acoust Soc Am. 1996; 100:3761–8. [PubMed: 8969477]

de Cheveigne A, Kawahara H. Missing-data model of vowel identification. J Acoust Soc Am. 1999; 105:3497–508. [PubMed: 10380672]

Dent ML, Larsen ON, Dooling RJ. Free-field binaural unmasking in budgerigars (Melopsittacus undulatus). Behav Neurosci. 1997; 111:590–8. [PubMed: 9189273]

Eriksson JL, Villa AE. Learning of auditory equivalence classes for vowels by rats. Behav Processes. 2006; 73:348–59. [PubMed: 16997507]

Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat Neurosci. 2010; 13:1011–9. [PubMed: 20622871]

Fry DB, Abramson AS, Eimas PD, Liberman AM. The Identification and Discrimination of Synthetic Vowels. Lang Speech. 1962; 5:171–89.

Garadat SN, Litovsky RY. Speech intelligibility in free field: spatial unmasking in preschool children. J Acoust Soc Am. 2007; 121:1047–55. [PubMed: 17348527]

Heinz JM, Sachs MB, Sinnott JM. Discrimination of steady state vowels by blackbirds and pigeons. Journal Acoustical Society of America. 1981; 70:699–706.

Hienz RD, Aleszczyk CM, May BJ. Vowel discrimination in cats: acquisition, effects of stimulus level, and performance in noise. J Acoust Soc Am. 1996a; 99:3656–68. [PubMed: 8655797]

Hienz RD, Aleszczyk CM, May BJ. Vowel discrimination in cats: thresholds for the detection of second formant changes in the vowel /epsilon. J Acoust Soc Am. 1996b; 100:1052–8. [PubMed: 8759958]

Hienz RD, Brady JV. The acquisition of vowel discriminations by nonhuman primates. J Acoust Soc Am. 1988; 84:186–94. [PubMed: 3411047]

Hine JE, Martin RL, Moore DR. Free-field binaural unmasking in ferrets. Behav Neurosci. 1994; 108:196–205. [PubMed: 8192846]

Kacelnik O, Nodal FR, Parsons CH, King AJ. Training-induced plasticity of auditory localization in adult mammals. PLoS Biol. 2006; 4:628–37.

Kalluri S, Depireux DA, Shamma SA. Perception and cortical neural coding of harmonic fusion in ferrets. J Acoust Soc Am. 2008; 123:2701–16. [PubMed: 18529189]

Kavanagh GL, Kelly JB. Hearing in the ferret (Mustela putorius): effects of primary auditory cortical lesions on thresholds for pure tone detection. J Neurophysiol. 1988; 60:879–88. [PubMed: 3171665]

Kelly JB, Kavanagh GL, Dalton JC. Hearing in the ferret (Mustela putorius): thresholds for pure tone detection. Hear Res. 1986; 24:269–75. [PubMed: 3793642]

Kewley-Port D, Li X, Zheng Y, Neel AT. Fundamental frequency effects on thresholds for vowel formant discrimination. J Acoust Soc Am. 1996; 100:2462–70. [PubMed: 8865651]

Kewley-Port D, Watson CS. Formant-frequency discrimination for isolated English vowels. J Acoust Soc Am. 1994; 95:485–96. [PubMed: 8120259]

Koenig AH, Allen JB, Berkley DA, Curtis TH. Determination of masking-level differences in a reverberant environment. J Acoust Soc Am. 1977; 61:1374–6. [PubMed: 881489]

Levitt H. Adaptive testing in audiology. Scand Audiol Suppl. 1978:241–91. [PubMed: 292144]

McFadden D. Masking-level differences with continuous and with burst masking noise. The Journal of the Acoustical Society of America. 1966; 40:1414–9. [PubMed: 5975577]

Mesgarani N, David SV, Fritz JB, Shamma SA. Phoneme representation and classification in primary auditory cortex. J Acoust Soc Am. 2008; 123:899–909. [PubMed: 18247893]

Miller RL. Auditory tests with synthetic vowels. Journal Acoustical Society of America. 1953; 18:114–21.

Nearey TM. Static, dynamic, and relational properties in vowel perception. J Acoust Soc Am. 1989; 85:2088–113. [PubMed: 2659638]

Nodal FR, Bajo VM, Parsons CH, Schnupp JW, King AJ. Sound localization behavior in ferrets: comparison of acoustic orientation and approach-to-target responses. Neuroscience. 2008; 154:397–408. [PubMed: 18281159]

Parsons CH, Lanyon RG, Schnupp JW, King AJ. Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets. J Neurophysiol. 1999; 82:2294–309. [PubMed: 10561407]

Peterson GE, Barney HL. Control methods used in a study of vowels. Journal Acoustical Society of America. 1952; 24:175–84.

Pisoni DB. Auditory and Phonetic Memory Codes in Discrimination of Consonants and Vowels. Percept Psychophys. 1973; 13:253–60. [PubMed: 23226880]

Ryalls JH, Lieberman P. Fundamental frequency and vowel perception. J Acoust Soc Am. 1982; 72:1631–4. [PubMed: 7175033]

Saberi K, Dostal L, Sadralodabai T, Bull V, Perrott DR. Free-field release from masking. J Acoust Soc Am. 1991; 90:1355–70. [PubMed: 1939900]

Schnupp JW, Dawe KL, Pollack GL. The detection of multisensory stimuli in an orthogonal sensory space. Exp Brain Res. 2005; 162:181–90. [PubMed: 15599727]

Schnupp JW, Hall TM, Kokelaar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci. 2006; 26:4785–95. [PubMed: 16672651]

Shinn-Cunningham BG, Schickler J, Kopco N, Litovsky R. Spatial unmasking of nearby speech sources in a simulated anechoic environment. J Acoust Soc Am. 2001; 110:1118–29. [PubMed: 11519579]

Sinnott JM, Kreiter NA. Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans. J Acoust Soc Am. 1991; 89:2421–9. [PubMed: 1861002]

Sinnott JM, Mosqueda SB. Effects of aging on speech sound discrimination in the Mongolian gerbil. Ear Hear. 2003; 24:30–7. [PubMed: 12598811]

Slawson AW. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. J Acoust Soc Am. 1968; 43:87–101. [PubMed: 5636408]

Smith DR, Patterson RD, Turner R, Kawahara H, Irino T. The processing and perception of size information in speech sounds. J Acoust Soc Am. 2005; 117:305–18. [PubMed: 15704423]

Sommers MS, Moody DB, Prosen CA, Stebbins WC. Formant frequency discrimination by Japanese macaques (Macaca fuscata). J Acoust Soc Am. 1992; 91:3499–510. [PubMed: 1619126]

Walker KM, Bizley JK, King AJ, Schnupp JW. Multiplexed and robust representations of sound features in auditory cortex. J Neurosci. 2011; 31:14565–76. [PubMed: 21994373]

Walker KM, Schnupp JW, Hart-Schnupp SM, King AJ, Bizley JK. Pitch discrimination by ferrets for simple and complex sounds. J Acoust Soc Am. 2009; 126:1321–35. [PubMed: 19739746]

Yin P, Fritz JB, Shamma SA. Do ferrets perceive relative pitch? J Acoust Soc Am. 2010; 127:1673–80. [PubMed: 20329865]

Zurek PM, Freyman RL, Balakrishnan U. Auditory target detection in reverberation. J Acoust Soc Am. 2004; 115:1609–20. [PubMed: 15101640]
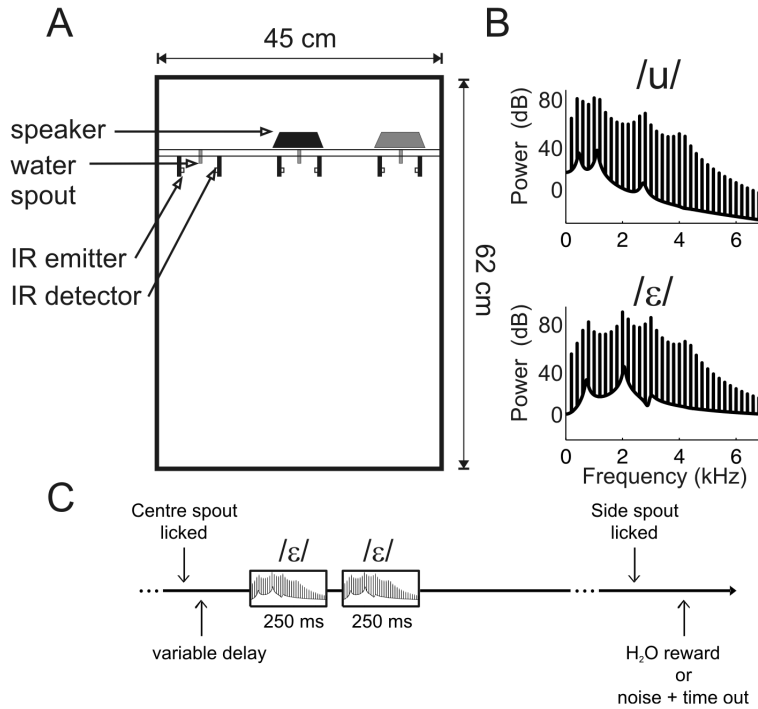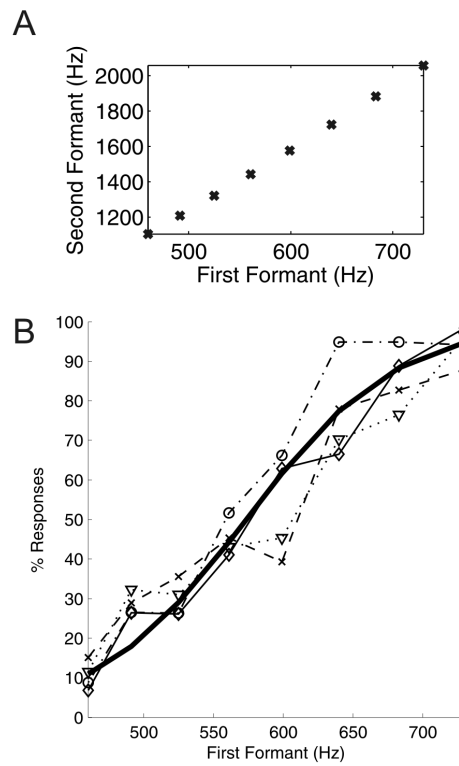
Figure 1.
Testing apparatus, stimuli and behavioral paradigm. (A) Ferrets were trained in a double walled glass testing box with infra-red lick detectors for registering responses, water spouts for delivering rewards, and loud speakers for presenting sounds. In Experiments 1-3 target sounds were presented via the black speaker with the gray loudspeaker only being used in Experiment 4. (B) Spectra of the vowel sounds used in this study (F0 = 200 Hz). (C) Schematic illustration of the time-course of a testing trial. Ferrets initiated a trial by licking a central "go" spout, which, after a variable delay, triggered the presentation of two identical vowel sounds. Ferrets were then rewarded with water or punished with a timeout depending on whether they responded at the correct spout.
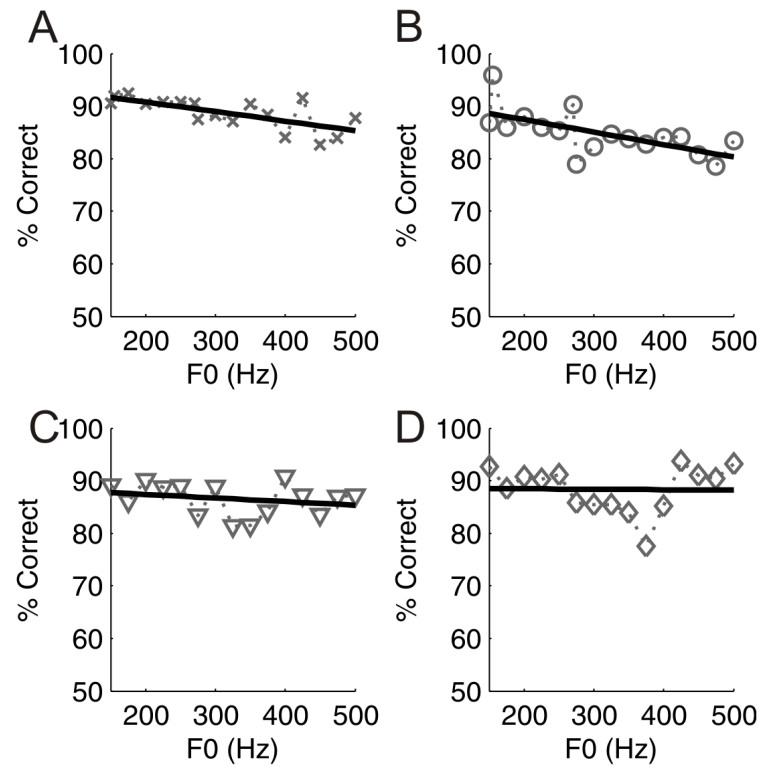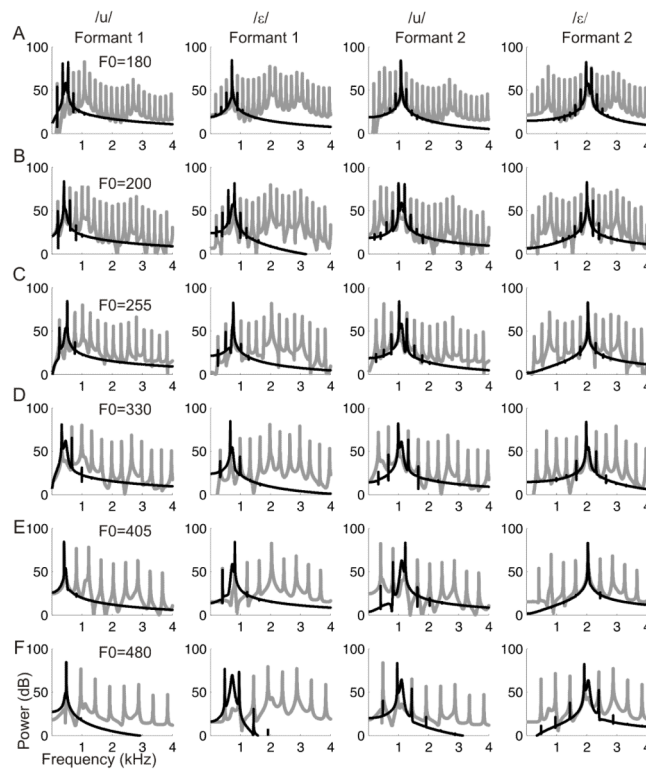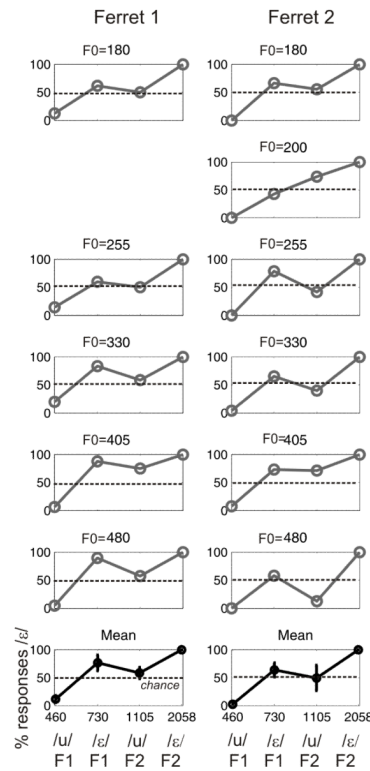
Figure 2.
Spectral timbre discrimination across different fundamental frequencies. (A-D) data from four different ferrets. Percent correct scores at each value of F0 are plotted and overlaid with a regression line.
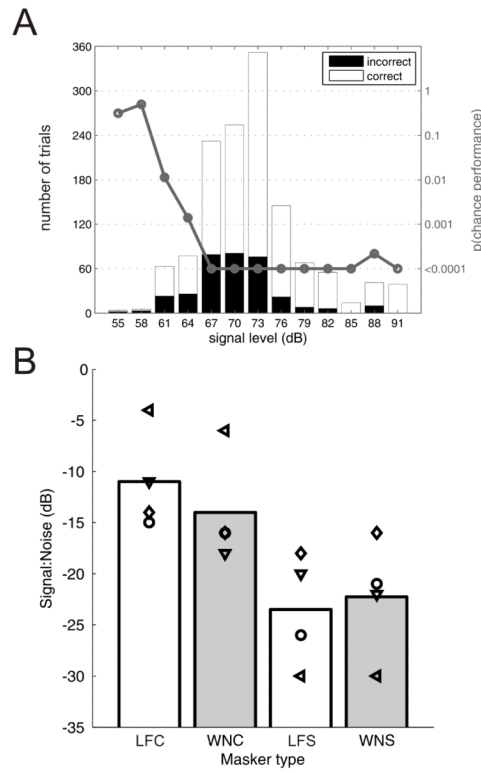
Figure 3.
Discrimination of a morphed vowel continuum. (A) First and second formant frequencies for the morphed vowels which run from /u/ (bottom left) to / / (top right). (B) Behavioral data for four ferrets (individual animals indicated by dashed and stippled lines) showing percentage of responses to the spout corresponding to / / as a function of first formant frequency, together with a fitted logistic psychometric function (solid black line).
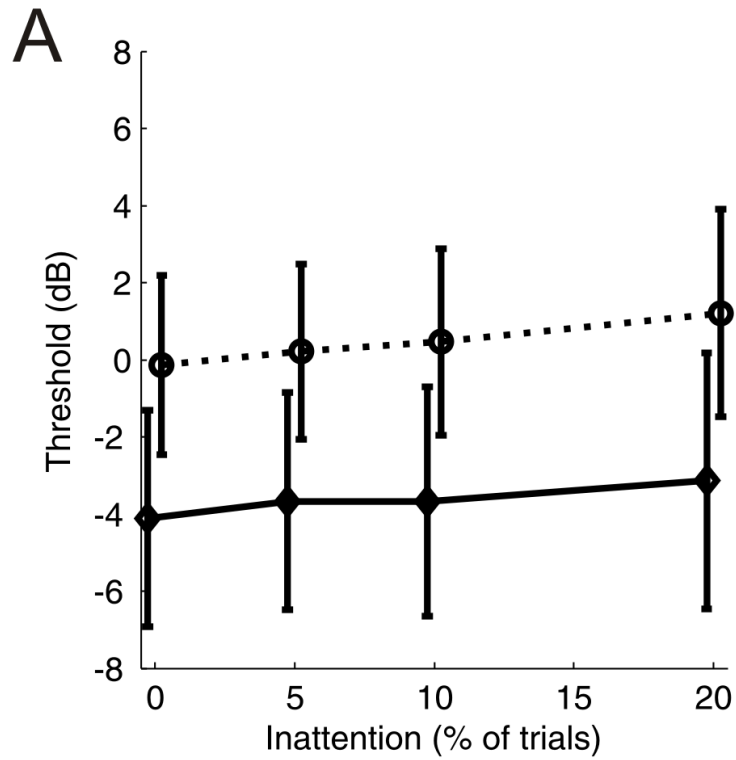
Figure 4.
Power spectra of the "single formant" vowels. Each row corresponds to a different F0, with low F0 stimuli at the top. The columns show the amplitude spectra (in black) from left to right of the first formant of /u/, the first formant of / /, the second formant of /u/ and the second formant of / /. The spectra in gray are the complete, four-formant, vowels.

Figure 5.
Discrimination of the "single formant" vowels illustrated in Figure 4. The plots in each column show the data from two ferrets that were tested with single formant vowels with formant frequencies corresponding to either the first (F1) or second formant (F2) of the /u/ or / / used in the previous experiments. (A-L) The single formant vowels were presented at different fundamental frequencies as shown above each panel (ferret 1 was not tested with an F0 of 200 Hz).

Figure 6.
Vowel discrimination in the presence of masking noise. (A) Data from Ferret #3 discriminating vowels in the presence of white noise, with the location of the vowel sound and the masking noise separated by 60°. The total number of correct responses (white bars) and incorrect responses (black bars) are plotted at each signal level. Overlaid (gray line) is the probability of the observed proportion of correct trials being the same as chance (50% performance) estimated using the binomial distribution. In this case threshold is 64 dBSPL signal level (B) Signal to noise ratio thresholds calculated in each of four listening conditions. Individual animal means are represented by the symbols and the bars show the mean of all animals. Ferrets were tested with either white noise (WN) or speech-shaped noise(Sp), which was either collocated (C) or spatially separated (S) from the target vowel.

Figure 7.
A comparison of thresholds estimated from simulated staricase data using the binomial-distribution method or standard methods. Solid line: mean ±SD threshold estimated using the binomial-distribution method from 200 simulations of repeated two-down one-up staircase procedures (mean number of staircases in each simulation was 31.8 to give at least 1500 trials). Stippled lines show the mean (across all staircases and simulations) threshold estimated using a standard adaptive method, with the error bars denoting the average standard deviation across the ~32 staircases used to simulate each threshold. The underlying psychometric function had a threshold (d =1) of 0 dB. See appendix for further details.