Published in final edited form as:
Eur J Neurosci. 2011 December ; 34(11): 1823–1838. doi:10.1111/j.1460-9568.2011.07887.x.

Cortical activity patterns predict robust speech discrimination ability in noise

Jai A. Shetake, **Jordan T. Wolf**, **Ryan J. Cheung**, **Crystal T. Engineer**, **Satyananda K. Ram**, and **Michael P. Kilgard**
The University of Texas at Dallas, School of Behavioral Brain Sciences, 800 West Campbell Road, GR41 Richardson, TX 75080-3021, USA

Abstract

The neural mechanisms that support speech discrimination in noisy conditions are poorly understood. In quiet conditions, spike timing information appears to be used in the discrimination of speech sounds. In this study, we evaluated the hypothesis that spike timing is also used to distinguish between speech sounds in noisy conditions that significantly degrade neural responses to speech sounds. We tested speech sound discrimination in rats and recorded primary auditory cortex (A1) responses to speech sounds in background noise of different intensities and spectral compositions. Our behavioral results indicate that rats, like humans, are able to accurately discriminate consonant sounds even in the presence of background noise that is as loud as the speech signal. Our neural recordings confirm that speech sounds evoke degraded but detectable responses in noise. Finally, we developed a novel neural classifier that mimics behavioral discrimination. The classifier discriminates between speech sounds by comparing the A1 spatiotemporal activity patterns evoked on single trials with the average spatiotemporal patterns evoked by known sounds. Unlike classifiers in most previous studies, this classifier is not provided with the stimulus onset time. Neural activity analyzed with the use of relative spike timing was well correlated with behavioral speech discrimination in quiet and in noise. Spike timing information integrated over longer intervals was required to accurately predict rat behavioral speech discrimination in noisy conditions. The similarity of neural and behavioral discrimination of speech in noise suggests that humans and rats may employ similar brain mechanisms to solve this problem.

Keywords

neural basis of speech; rodent; similarities between human and animal; speech in noise; temporal integration

Introduction

Communication sounds often occur in the presence of environmental noise. Humans and many species of animal can reliably identify behaviorally relevant communication sounds in the presence of other interfering sounds (Ehret & Gerhardt, 1980; Gerhardt & Klump, 1988; Hulse *et al.*, 1997; Lohr *et al.*, 2003; Narayan *et al.*, 2007). Normal-hearing humans can discriminate between many speech sounds above chance level, even when the speech sounds and background noise are of equal intensity (Miller & Nicely, 1955; House *et al.*, 1965;

Correspondence: M. P. Kilgard, as above. kilgard@utdallas.edu.

Wang & Bilger, 1973; Dubno & Levitt, 1981; Phatak & Allen, 2007). The normal auditory system is remarkably adept at processing acoustic information in noisy situations.

The neural mechanisms that allow speech discrimination to be robust to background noise are not well understood (Fitch *et al.*, 1997; Bishop & Miller, 2009; Song *et al.*, 2010). It is not clear whether humans have specialized mechanisms to process speech in noise or whether standard acoustic processing mechanisms are sufficient to explain the robustness of speech discrimination in noise. A number of previous studies suggest that the underlying processing of speech sounds is similar in humans and animals (Kuhl & Miller, 1975; Tallal *et al.*, 1993; Fitch *et al.*, 1997; Cunningham *et al.*, 2002; Reed *et al.*, 2003; Mesgarani *et al.*, 2008). Behavioral discrimination of speech sounds in quiet by many species of animal is similar to that in humans (Kuhl & Miller, 1975; Kluender *et al.*, 1987; Ramus *et al.*, 2000; Reed *et al.*, 2003; Engineer *et al.*, 2008). Neural responses to speech sounds presented in quiet are also similar in humans and animals (Steinschneider *et al.*, 1999; Wong & Schreiner, 2003; Engineer *et al.*, 2008; Mesgarani *et al.*, 2008). Few studies have directly compared neural and behavioral discrimination of speech sounds to determine the neural mechanisms for encoding of speech sounds (Engineer *et al.*, 2008). Neural activity in the rat primary auditory cortex (A1) is capable of accurately predicting rat speech discrimination (Engineer *et al.*, 2008). Speech sounds that evoke similar spatiotemporal neural activity patterns are more difficult to discriminate than sounds that evoke distinct patterns. Spike timing information precise to 1–10 ms appears to be necessary to account for rat behavioral discrimination between different consonant sounds in quiet situations. In this study, we documented neural and behavioral discrimination of speech sounds in background noise.

The experiments in this study were designed to test the hypotheses that: (i) the neural mechanisms used to distinguish between similar speech sounds are different in noisy conditions; (ii) neural responses can explain why some speech contrasts are more easily masked than others; and (iii) animals, like humans, are able to discriminate speech sounds in high levels of background noise. Our results provide new insights into the neural mechanisms that contribute to robust speech sound processing in noisy environments, and suggest that the rat may prove to be a useful model for studying the neural basis of human speech sound processing deficits (Merzenich *et al.*, 1993; Tallal *et al.*, 1993; Cunningham *et al.*, 2002; Threlkeld *et al.*, 2007).

Materials and methods

Stimuli

Speech stimuli—We used 11 of the 20 English consonant–vowel–consonant words, ending in `ad' (as in `tad') used in the Engineer *et al.* (2008) study. These sounds were `bad', `dad', `gad', `pad', `tad', `sad', `yad', `rad', `lad', `shad', and `chad'. A detailed description of the recording and processing of these sounds can be found in Engineer *et al.* (2008). In brief, we recorded these sounds in a double-walled, sound-proof booth. All speech sounds were produced by a female speaker. The fundamental frequency and spectrum envelope of each word was shifted up in frequency by a factor of two with the STRAIGHT vocoder (Kawahara, 1997), to better match the rat hearing range (Sally & Kelly, 1988). The intensity of the speech sounds was adjusted so that the intensity during the most intense 100 ms was 60 dB SPL. Speech sounds were approximately 500 ms long.

Noise stimuli—White noise was generated with a random number generator in MATLAB, and covered a frequency range of 0.2–50 kHz. The speech-shaped noise was generated by passing white noise through first-order Butterworth filters. Butterworth filters available in the FDATOOL function in MATLAB were used. As the speech stimuli in our experiment were produced by a female speaker, we used the female long-term average speech spectrum. The

long-term average speech spectrum for females is flat from ~200 to ~700 Hz, and falls off by approximately 7 dB SPL per octave on either side (Byrne *et al.*, 1994). Like those of speech stimuli, the frequency limits of speech-shaped noise were shifted up by a factor of two in order to adjust for the rat frequency hearing range. Therefore, the speech-shaped noise used in this study had a flat spectrum between ~400 and ~1400 Hz, and sloped by 7 dB SPL per octave on either side. Continuous background noise was used because it has a greater impact on speech sound processing than discontinuous noise (Miller & Licklider, 1950; Buus, 1985; Moore, 1985; Hall & Grose, 1991). Both white noise and speech-shaped noises were calibrated to 48, 60 and 72 dB SPL.

Behavioral training procedure

We trained 12 female Sprague–Dawley rats to discriminate speech sounds in quiet and noise. An operant go / no-go training procedure was used for speech discrimination training (Fig. 1). Half of the rats ($n = 6$) were trained to press a lever in response to the target sound `dad' and ignore the non-target sounds `bad', `gad', `tad', and `sad'. The other half ($n = 6$) were trained to press the lever in response to the target sound `shad' and ignore the non-target sounds `bad', `dad', `yad', `sad', `pad', `tad', and `chad'. The rats weighed an average of 275 ± 15 g, and were housed under a 12 : 12-h reversed light cycle at constant humidity and temperature. Rats were food-deprived to provide motivation for food reward during the behavioral training, but were maintained at no < 85% of their normal body weight. Access to water was free at all times except during the behavioral training session, which was 1 h long. Rats were trained for two training sessions each day for 5 days a week in a sound-shielded operant-training booth. The booth contained a video camera for monitoring, a house light, a cage, and a speaker. The cage contained a lever, a lever light, and a pellet receptacle. The pellet dispenser was mounted outside the training booth to minimize noise. The speaker was mounted approximately 10 cm away from the midpoint between the lever and the pellet receptacle, as rats generally stayed in this area during behavioral training sessions.

Rats progressed through the following behavioral training stages: 1, shaping; 2, detection of sounds; 3, discrimination in quiet; and 4, discrimination in noise. The training protocol used up to stage 2, namely detection of sounds, was similar to that in our previous study (Engineer *et al.*, 2008). In brief, during the shaping stage, the rats were trained to press the lever for food reward. During the initial detection of sounds training stage, rats were given the food reward only if they pressed the lever after listening to a target speech sound. The same target speech sound was used every time, so that the rats associated it with the food reward. We calculated d prime (d′), to compare the hit rates to the false alarm rates (Engineer *et al.*, 2008). A d′ of 0 indicates the rat is pressing the lever equally often to both target and distracter, while a positive d′ indicates that the rat has a higher hit rate than false alarm rate. A d′ of 1.5 or higher indicates that the rats were reliably detecting the target sound. Once rats reached the performance level d ′≥ 1.5 for 10 sessions, they were advanced to discriminate between consonant sounds in quiet.

During each consonant discrimination task, rats learned to discriminate the target sound from the non-target sounds. Initially, the rats learned to discriminate sounds in quiet. Trials began every 6 s, and rats were only rewarded for lever presses in response to the target (conditioned) stimuli, which were randomly interleaved for 50% of the time. Silent catch stimuli and non-target sounds were randomly interleaved for 50% of the time, and pressing the lever in these trials resulted in a time-out of 6 s. After 20 sessions of discrimination training in quiet, the rats were trained to discriminate between speech sounds in background noise. During each session, rats were trained to discriminate speech sounds in both speech-shaped noise and white noise of 48, 60 and 72 dB SPL. We also continued to test rat behavioral speech discrimination in quiet, to allow direct comparison between discrimination in quiet and in noise on the same days. Each session started with 13 trials of

discrimination in quiet. For future reference, we will refer to 13 trials with a particular noise intensity and noise type as a noise block. Sounds presented in each block were chosen randomly to maintain a 1 : 1 presentation ratio of target sound and non-target sound. After discrimination of trials in quiet, either white noise or speech-shaped noise of 48 dB SPL was randomly chosen to be played in the background. Noise intensity was gradually increased every noise block to the maximum level, so that the rat could habituate to the noise, and then gradually decreased. After this sequence, the noise was changed to a random noise intensity and noise type for every block. Noise had a 3-s ramp on to avoid startling the rat with abrupt noise bursts. After this sequence, noise blocks were played randomly. Training performance was quantified as percentage correct and d′ measure. Percentage correct was used to correlate neural discrimination which was obtained from a neural classifier; this is explained later in the section. The percentage of trials in which the rats pressed the lever in response to the target sound and the percentage of trials in which the rats refrained from pressing the lever in response to the non-target sounds were averaged together to calculate behavior percentage correct. The behavioral discrimination performance of rats in noise improved consistently up to the 50th session, after which their performance reached a plateau. Behavioral data collected on the last 20 training sessions were averaged together to determine the speech discrimination of rats. In order to compare the discrimination performance of our neural classifier with a wide variety of behavioral tasks, we included behavioral data of five additional tasks from our earlier study in quiet, and averaged behavioral discrimination of six tasks common to both studies. The discrimination tasks common to both studies were `dad' vs. `bad', `dad' vs. `gad', `dad' vs. `tad', `dad' vs. `sad', `shad' vs. `sad', and `shad' vs. `chad'. The discrimination tasks used only in our previous study were `shad' vs. `fad', `shad' vs. `jad', `shad' vs. `had', `rad' vs. `lad', and `mad' vs. `nad'. sas statistical software (SAS Institute, Cary, NC, USA) was used for all statistical analyses. We used a two-factor repeated-measures ANOVA to test the behavioral discrimination in the different noise conditions and noise types, and Tukey *post hoc* comparisons with significance set to $P < 0.05$ to test the effects of individual contrasts under the different noise conditions. The Z-statistic was used to test whether Pearson correlation coefficients were significantly different from each other. Handling, housing and testing of the animals were approved by the University of Texas Institutional Animal Care and Use Committee.

Neural data: electrophysiology recordings

We recorded multiunit responses to speech in quiet and in six noise conditions from 133 sites in the right A1 of eight anesthetized naïve rats. Neural responses from our earlier study ($n = 445$ A1 sites from 11 rats) (Engineer *et al*., 2008) recorded in quiet were used to evaluate neural discrimination for the five additional behavioral tasks that were only tested in quiet. Neural responses were also recorded from six ($n = 4$ rats with target sound as `dad', and n = 2 rats with target sound as `shad') of the trained rats. Rats were anesthetized with pentobarbital for surgery (50 mg/kg), and a state of areflexia was maintained throughout the experiment with supplemental doses of dilute pentobarbital (0.2–0.5 mL; 8 mg/mL). Anesthesia depth was monitored by heart rate, breathing rate, corneal reflexes, and response to toe pinch. Nourishment was provided with a 1 : 1 mixture of dextrose (5%) and standard Ringer's lactate solution, and body temperature was maintained at 37 °C. A tracheotomy was performed to minimize breathing problems and breathing sounds, and a cisternal drain was made to minimize cerebral edema. A part of the skull over the temporal ridge was removed to expose the right A1. The dura was removed, and the cortex was maintained under a thin film of silicone oil to prevent desiccation. Four parylene-coated tungsten microelectrodes [Fred Haer Company (FHC), Bowdoin, ME; 1–2 MΩ] were lowered simultaneously to a depth of 600 μm so that they were in layer IV/V of A1. To determine the characteristic frequency at each site, we played 90 logarithmically spaced tones ranging from 1 to 47 kHz at 16 intensities ranging from 0 to 75 dB SPL. The tones were 25 ms in length, and their

presentation was randomly interleaved. We placed the speaker 10 cm away from the left ear. Neural response characteristics such as start latency, end latency and characteristic frequency at each recording site were used to determine whether the electrodes were placed in A1. After all of the tones, speech sounds were played. The speech stimulus set was composed of the same 11 monosyllabic consonant–vowel–consonant words used in behavioral training, a silence stimulus, and two noise types: white noise and speech-shaped noise. Each speech sound was separated by 2300 ms and presented 20 times in quiet and in three noise intensity levels: 48, 60 and 72 dB SPL of both white noise and speech-shaped noise. The different noise conditions, that is, noise levels and noise types, were interleaved to avoid the state of anesthesia from affecting neural responses in any particular noise condition. Within each noise type, the noise intensities were interleaved as follows. All of the sounds (11 speech sounds and silent stimulus) were repeated five times at one noise intensity. From here on, we will refer to this set of sounds, that is, 12 sounds repeated five times played in a particular noise intensity and noise type, as one block. Within each block, the speech sounds were randomly interleaved. The sequence in which the blocks were played is as follows. First, one block of sounds was played in quiet. Either white noise or speech-shaped noise was added in a linearly increasing order of intensity and then in a linearly decreasing order, to avoid sudden changes in neural responses that could result from random presentation of noise. We also added a 500-ms ramp with change in noise intensity to ensure that there were no clicks or other disturbances in the noise and avoid neural response to sudden onset of noise. The delay between the onset of noise and the speech sounds was never < 10 s, to eliminate any transients related to noise onset. Because the noise conditions were varied in blocks, the background noise was typically on for minutes before changing in intensity or type. This sequence resulted in each of the sounds being played 10 times in one noise type. This sequence was repeated with the other noise type, and then again with the first noise type, and finally with the second noise type. Stimulus generation and data acquisition were performed with Tucker–Davis hardware (RP2.1 and RX5) and software (BRAINWARE). Surgery protocols and recording procedures were approved by the University of Texas at Dallas Institutional Animal Care and Use Committee.

Neural data: electrophysiology data analysis

Neural response characteristics—The total number of spikes was calculated over the first 100 ms of the neural response onset above spontaneous. Onset start latency (milliseconds) was defined as the time from stimulus onset to the earliest reliable neural response, and was determined as the time when average neural responses from all recording sites were at least 3 SD above spontaneous activity. The end of peak latency (milliseconds) was the time when the average neural responses returned to baseline, and was determined as the time when neural responses after the peak-driven response were not significantly different from spontaneous activity. Latency was based on the average population data instead of latency data from individual recording sites, because it was difficult to establish the latency for individual recording sites in high background noise, and the evoked response was small. Distinctiveness between neural response patterns was calculated by the use of city-block and Euclidean distance, to test the correlation between neural distinctiveness and behavioral discrimination. For statistical analysis of neural responses, we used a two-factor repeated-measures ANOVA and Tukey *post hoc* comparisons with significance set to $P < 0.01$.

Classifier—To quantify neural discrimination, we modified a well-studied nearest neighbor classifier (Foffani *et al*., 2004; Schnupp *et al*., 2006; Engineer et al., 2008). Unlike our earlier classifier, which was explicitly given the stimulus start time (Engineer *et al*., 2008), the version of classifier used in the current study was not given information about the stimulus start time.

The neural classifier identifies sounds on the basis of action potential activity produced by a single presentation. The classifier attempts to identify which of two possible sounds was presented by looking for the spatiotemporal activity patterns generated by each sound. Pairwise analysis was performed to ensure simplicity and to maintain historical continuity with our earlier study. The two activity patterns were derived from the average spatiotemporal response from a set of recording sites distributed across A1. The performance of the classifier was tested by systematically varying the durations and bin sizes of the average spatiotemporal patterns. Specifically, the durations were varied between 10 and 700 ms in steps of 10 ms, and the bin sizes were varied between 1 and 700 ms in steps of 5 ms, to find the classifier parameters that could accurately predict behavior. Analysis was based on activity from 1, 5, 10, 25, 40, 60 or 133 recording sites. The sites were randomly chosen recording sites from the complete list of 133 sites. As an example of the parameters used, Fig. 8A and B shows the average response of 60 A1 sites to the word `dad' and `bad'. In this case, the duration of the pattern was 100 ms, and the bin size was 10 ms.

The classifier determines which of the two patterns was more likely to have occurred during the single trial response period, which is 750 ms in length (Fig. 8C). The classifier computes the city-block distance (Fig. 8C, top) between the average patterns and the single trial activity, assuming all possible start times. The average spatio-temporal activity patterns were generated from 19 trials, and the current trial was not included. City-block distance is simply the average difference between the firing rate at each bin of the single trial activity and the average pattern. All results were also tested by the use of Euclidian distance, and were similar to the results obtained with city-block distance. The classifier guesses that the sound that was presented on each trial was the sound whose average spatiotemporal pattern was closest (Fig. 8C, asterisk) to the single trial activity. Neural discrimination for a group of sites was determined by calculating the percentage of trials on which the classifier correctly guessed the presented sound.

$Classifier_{100/10\ ms}$ refers to the classifier when 100 ms of activity from 60 sites binned every 10 ms was used for the neural discrimination. $Classifier_{400/60\ ms}$ refers to the classifier when 400 ms of activity from 60 sites binned every 60 ms was used. The hybrid classifier uses 100 ms of activity binned every 10 ms in quiet conditions and 400 ms of activity binned every 60 ms in noise. We also analyzed neural responses to sounds from our previous study in quiet, with the classifiers used in this study, and determined whether our new neural classifier could predict the behavioral discrimination of tasks from our previous study (Engineer *et al*., 2008).

Results

Behavioral results

Our results provide the first evidence that animals, like humans, can discriminate between numerous speech sounds even when the speech signals and background noise are of equal intensity (Fig. 2). We trained 12 rats to discriminate between different consonant sounds in noise. Half of the rats ($n = 6$) were first trained to discriminate the target word `dad' from the non-target words `bad', `gad', `tad' and `sad' in quiet. The other half were first trained to discriminate the target word `shad' from the non-target words `bad', `dad', `pad', `tad', `yad', `sad' and `chad' in quiet. All of the speech sounds were presented such that the loudest 100 ms was at 60 dB SPL. After the rats had learned to accurately perform the tasks, both groups were required to discriminate the same stimuli in white noise and speech-shaped noise of 48, 60 and 72 dB SPL.

Consistent with our earlier report, the rats were able to discriminate the target from the non-target sounds in more than 85% of trials when tested in quiet (Fig. 2A) (Engineer *et al*., 2008). As expected, discrimination could be impaired by adding background noise, and the extent of impairment depended primarily on the intensity of the noise. Moderate levels of background noise (i.e. 48 dB SPL) did not significantly impair speech discrimination as compared with quiet for most speech sound contrasts (Fig. 2; $F_{1,66} = 0.54$, $P = 0.5$, two-way repeated-measures ANOVA). Only the discrimination of `shad' from `chad' was significantly impaired by 48 dB SPL noise (Fig. 2B and C; $P = 0.005$, Tukey *post hoc*). In 60 dB SPL noise, rats were able to perform each of the discrimination tasks at well above chance levels, but performance was significantly impaired as compared with quiet (Fig. 2; $P = 0.005$, Tukey *post hoc*). In 72 dB SPL noise, discrimination on all tasks, except `dad' vs. `tad', was significantly impaired as compared with quiet (Fig. 2; $P = 0.005$, Tukey *post hoc*). These results demonstrate that rats can discriminate speech sounds at a minimum signal-to-noise ratio that is similar to that for humans (Miller & Nicely, 1955; House *et al*., 1965; Horii *et al*., 1971; Wang & Bilger, 1973; Dubno & Levitt, 1981; Phatak *et al*., 2008).

Moderate background noise does not impair discrimination by preventing detection of the target sounds. In 60 dB SPL noise, rats pressed the lever in response to both target sounds significantly more than in response to silent catch trials (60 ± 5% more responses to target, $P = 0.001$, Tukey *post hoc*). Even in 72 dB SPL white noise, rats were able to detect both targets (37 ± 7% more responses to target, $P < 0.01$). In 72 dB SPL speech-shaped noise, rats were able to detect `dad' (13 ± 7% more responses to target, $P < 0.05$) but not `shad' (5 ± 8% more responses to target, $P = 0.61$). These results are consistent with earlier observations that discrimination is more sensitive to background noise than detection (Hawkins & Stevens, 1950; Whiting *et al*., 1998).

Behavioral discrimination in most tasks was impaired to similar extents by white noise and speech-shaped noise when the noise intensities were 48 or 60 dB SPL (Fig. 2; by 2 ± 1.9% in 48 dB SPL noise, and by 9.5 ± 2% in 60 dB SPL noise, respectively). However, in 72 dB SPL noise, speech discrimination was significantly more impaired by speech-shaped noise (average percentage correct, 51 ± 2%) than by white noise (average percentage correct, 60 ± 3%) (Fig. 2A; $F_{1,66} = 10.20$, $P = 0.001$, two-way repeated-measures ANOVA). Rats could discriminate six of 11 tasks significantly above chance level in 72 dB SPL white noise and only three of 11 tasks in 72 dB SPL speech-shaped noise (Fig. 2B and C; $P = 0.005$, Tukey *post hoc*). The observation that speech-shaped noise generally causes greater impairment in speech discrimination than white noise is consistent with psychophysical studies in humans (Busch & Eldredge, 1967; Dubno & Levitt, 1981).

There were significant differences in the degree to which different speech contrasts were impaired by the different noise conditions ($P = 0.01$, Tukey *post hoc*). In quiet, `dad' vs. `bad' and `dad' vs. `gad' were the easiest discrimination tasks, and `dad' vs. `tad' and `shad' vs. `chad' were the hardest discrimination tasks, consistent with our previous report (Fig. 2B and C) (Engineer *et al*., 2008). In the presence of background noise, `dad' vs. `bad' and `dad' vs. `gad' continued to be among the easiest discrimination tasks, and `shad' vs. `chad' continued to be the hardest discrimination task (Fig. 2B and C). The ability of rats to discriminate `dad' vs. `tad' in white noise, however, was quite robust, and showed the least impairment of all of the discrimination tasks (Fig. 2B and C; $P = 0.005$, Tukey *post hoc*). For example, the behavioral discrimination for `dad' vs. `gad' fell by 23 ± 3% from quiet to 72 dB SPL white noise, whereas that of `dad' vs. `tad' fell by 6 ± 2%. This result is consistent with results from psychophysical studies in humans, which show that voicing tasks are the most robust to white noise (Miller & Nicely, 1955; Wang & Bilger, 1973). Discrimination for `dad' vs. `tad' was not as robust in speech-shaped noise, and fell by 27 ± 3% when 72 dB SPL speech-shaped noise was added. This result is consistent with previous

results in humans, which show that voicing discrimination is not as robust in speech-shaped noise as in white noise (Dubno & Levitt, 1981).

Discrimination tasks with `dad' as the target stimulus were significantly more robust to 72 dB SPL white noise than discrimination tasks with the target stimulus `shad' (Fig. 2B and C; $P = 0.01$, Tukey *post hoc*), but were not more robust in 72 dB SPL speech-shaped noise ($P = 0.08$, Tukey *post hoc*). The greater masking of `shad' by white noise is consistent with previous results in humans, which show that fricatives and affricates are more sensitive than other sounds to white noise (Miller & Nicely, 1955; Busch & Eldredge, 1967; Horii *et al*., 1971; Phatak *et al*., 2008). The only tasks that were more impaired in white noise than in speech-shaped noise were `dad' vs. `sad' and `shad' vs. `chad' in 48 dB SPL noise, and `dad' vs. `gad', `shad' vs. `tad' and `shad' vs. `chad' in 60 dB SPL noise ($P = 0.005$, Tukey *post hoc*). We expected that A1 responses to speech sounds in each of the noise types and intensities would clarify the auditory mechanisms that support robust speech processing in noisy environments.

Neural responses to speech sounds in background noisy situations

Neural responses were recorded from 133 multiunit clusters of A1 neurons in barbiturate-anesthetized rats. Neural responses were obtained in response to the same 11 consonant sounds that were behaviorally tested in quiet and in 48, 60 and 72 dB SPL speech-shaped noise and white noise. The responses recorded in silence were consistent with results from previous studies in humans and animals (Steinschneider *et al*., 1999, 2005; Wong & Schreiner, 2003; Engineer *et al*., 2008; Skoe & Kraus, 2010). For example, voiced stop consonants (`b', `d', and `g') evoked a single burst of activity, whereas unvoiced stop consonants (`p' and `t') resulted in a second peak of activity corresponding to the voicing onset. Stop consonants (`b', `d', `g', `p', and `t') evoked a greater neural onset response than fricatives and affricates (`s', `sh', and `ch') (average of 3.0 ± 0.1 spikes vs. 2.3 ± 0.4 spikes; $P = 0.0001$; Tukey *post hoc*). Neural responses to consonant sounds were degraded in the presence of background noise. Background noise reduced the number of action potentials evoked by speech sounds and delayed the latency of the response (Fig. 2). Increasing the level of background noise transiently increased the average firing rate from 25 to 95 Hz (~500-ms half-life). Background noise did not significantly increase the spontaneous firing activity during the time period when speech sounds were presented (> 10 s after each change in noise level; $P > 0.05$). The average baseline spontaneous firing rates were 25, 26, 26 and 27 Hz in quiet, and 48, 60 and 72 dB, respectively (two-way repeated measures ANOVA; $P > 0.05$).

The amount of degradation of the neural onset response depended on the intensity of the noise, the type of speech sound, the spectral composition of the noise, and the characteristic frequency at each recording site. Increasing the intensity of background noise caused significant degradation of the neural onset response to all consonant sounds. Specifically, increasing the intensity of background noise reduced the total number of spikes and increased both start and end latency of the neural onset response to consonant sounds (Fig. 3; $F_{3,396} = 141.79$, $P = 0.0001$, $F_{3,396} = 29.14$, $P = 0.0001$, and $F_{3,396} = 15.72$, $P = 0.0001$, respectively, two-factor repeated-measures ANOVA). For example, the average number of spikes evoked by consonant sounds in the first 100 ms was 1.7 ± 0.2 less in 60 dB SPL noise than in quiet ($P = 0.0001$, Tukey *post hoc*). The average start and end latency of sounds increased by 35 ± 14 ms and 62.5 ± 14.5 ms, respectively, in 60 dB SPL noise as compared with quiet ($P = 0.0001$, Tukey *post hoc*). The neural onset response to most sounds was prominent in 48 dB SPL noise and still present when the background noise and speech signal were of the same intensity (i.e. 60 dB SPL; Fig. 4). The onset response to most speech sounds was eliminated by the presence of 72 dB SPL background noise. The severe

reduction in neural activity by 72 dB SPL noise is consistent with our behavioral results showing that speech discrimination is severely impaired in this noise intensity.

Neurophysiology studies in humans show that the cortical responses to some sounds are more degraded by noise than those to other sounds (Whiting *et al.*, 1998; Martin *et al.*, 1999; Billings *et al.*, 2010). Our recordings in rats also showed a significant effect of stimulus on cortical responses in noise (Fig. 4; $F_{30,3960} = 68.67$, $P = 0.0001$, two-factor repeated-measures ANOVA). Voiced stop consonants (`b', `d', and `g') were least affected by the presence of background noise (Fig. 4A–C; $P = 0.0001$, Tukey *post hoc*). The number of spikes in the onset response evoked by voiced stop consonants was significantly greater than that evoked by all of the other consonant groups (i.e. unvoiced stop consonants, fricatives, affricates, and glides; $P = 0.0001$, Tukey *post hoc*). For example, sounds `b', `d' and `g' evoked an average of 1.67 ± 0.10 spikes, significantly above spontaneous activity, even in 72 dB SPL white noise (Fig. 4; $P = 0.0001$, Tukey *post hoc*). On the other hand, the neural onset response to the fricative sound `s' was almost completely eliminated, even in the presence of 48 dB SPL white noise (Fig. 4F; $P = 0.0001$, Tukey *post hoc*). The onset responses to unvoiced sounds were eliminated in 72 dB SPL white noise (Fig. 4D–K). The prominence of the neural response for `d' and the absence of a neural response for `sh' in 72 dB SPL white noise supports our behavioral finding of robust speech discrimination of tasks with target sound `d'. As in previous studies, in both humans and animals, neural responses to vowel sounds were more robust than those to consonant sounds (Cunningham *et al.*, 2002; Russo *et al.*, 2004; Song *et al.*, 2010). For example, the neural response to vowel onset was present even in 72 dB SPL white noise for most sounds, whereas the response to the unvoiced consonant was eliminated (Fig. 4D–K). The differential effect of noise on neural responses to different sounds may clarify the greater behavioral sensitivity of certain tasks to noise.

The spectral composition of background noise significantly altered the degree of degradation of the neural response. Speech-shaped noise had a greater impact on most sounds than white noise (Fig. 4; $F_{1,132} = 391.14$, $P = 0.0001$). The average number of spikes evoked by sounds `b', `d', `g', `p', `y', `r' and `l' was reduced by 0.94 ± 0.14 in the presence of speech-shaped noise as compared with white noise of equal intensity (60 dB SPL, $P = 0.0001$, Tukey *post hoc*). Three speech sounds, `b', `d', and `g', caused significantly driven responses in 72 dB SPL white noise, whereas neural responses to no speech sounds were significantly different from spontaneous activity in 72 dB SPL speech-shaped noise ($P = 0.0001$, Tukey *post hoc*). These neural results are consistent with our behavioral results in rats and previous results in humans showing that speech-shaped noise is more impairing than white noise for most speech sounds (Busch & Eldredge, 1967; Dubno & Levitt, 1981).

The neural responses evoked by the sounds `s', `sh', `ch', and `t', which primarily contain high-frequency energy (> 10 kHz), were more degraded in white noise than in speech-shaped noise. The average number of spikes evoked by `s', `sh', `ch' and `t' was reduced by 0.31 ± 0.12 in white noise as compared with speech-shaped noise (60 dB SPL, $P = 0.0001$, Tukey *post hoc*). Our neural results show that white noise impairs responses to high-frequency speech sounds more than those to low-frequency sounds, as seen in previous human psychophysical studies (Miller & Nicely, 1955; Busch & Eldredge, 1967; Wang & Bilger, 1973; Phatak *et al.*, 2008).

The high spatial resolution of neural responses obtained from our study allowed us to explain the differential effect of noise type by comparing responses of neurons tuned to different frequencies. Although both noise types were presented at the same overall intensities, white noise contains approximately 9 dB more high-frequency (> 10 kHz) energy than speech-shaped noise, which contains approximately 9 dB more low-frequency (< 4 kHz) energy. As a result, speech-shaped noise degraded responses of low-frequency neurons

(characteristic frequency < 4 kHz) more than white noise (Figs 5 and 6), whereas white noise degraded neural responses of high-frequency neurons (characteristic frequency > 10 kHz) more than speech-shaped noise (60 dB SPL, $P = 0.0001$, Tukey *post hoc*; Fig. 6). The number of spikes evoked in low-frequency neurons by most speech sounds was 1.84 ± 0.26 less in speech-shaped noise than in white noise (60 dB SPL, $P = 0.0001$, Tukey *post hoc*; Fig. 5). The number of spikes evoked in high-frequency neurons by high-frequency speech sounds (i.e. `s', `sh', `ch', and `t') was 0.80 ± 0.09 less in white noise than in speech-shaped noise (60 dB SPL, $P = 0.0001$, Tukey *post hoc*). Most speech sounds in our study contained mostly low-frequency energy and evoked greater responses in low-frequency neurons. As speech-shaped noise degraded neural responses in the low-frequency region more than white noise, it is not surprising that most speech sounds were degraded to a greater extent by speech-shaped noise than by white noise (Fig. 5). The consonant sounds, which primarily contain high-frequency energy, that is, sounds `s', `sh', `ch', and `t', evoked the most activity in the high-frequency group of neurons (Fig. 6). As white noise degraded neural activity in the high-frequency neurons more than speech-shaped noise, it is not surprising that white noise degraded neural responses to high-frequency sounds to a greater extent than speech-shaped noise. Our results are consistent with earlier reports in humans that the spectral contents of both the speech signal and the background noise influence neural responses to speech sounds (Whiting *et al.*, 1998; Martin *et al.*, 1999; Kozou *et al.*, 2005; Martin & Stapells, 2005; Billings *et al.*, 2010).

The different noise conditions had similar effects on behavioral and neural responses. We found a high correlation between the average number of spikes evoked by sounds in each discrimination task and the behavioral discrimination on that task ($R^2 = 0.63$, $P = 10^{-18}$, Fig. 7A). These results are consistent with electroencephalography and imaging (blood oxygen level-dependent) studies in humans, which show degradation of behavioral and of neural responses that are proportional to the noise intensity (Whiting *et al.*, 1998; Muller-Gass *et al.*, 2001; Binder *et al.*, 2004).

Our observation that the average number of spikes evoked by speech sounds in noise was correlated with behavioral discrimination should not be taken as evidence that spike count is sufficient to discriminate between speech sounds. To discriminate between stimuli, there must be a difference in neural activity. When we quantified the difference in the number of spikes evoked by pairs of sounds in each speech contrast, we found that the difference was poorly correlated with behavioral discrimination ($R^2 = 0.05$, $P = 0.05$; Fig. 7B). This observation is consistent with our earlier report that speech discrimination in quiet is not correlated with the difference in the total number of spikes, that is, firing rate, evoked by each sound at each multiunit recording site (Engineer *et al.*, 2008). In our previous study, speech discrimination in quiet was only correlated with A1 activity when spike timing was used. In this study, we also found that distinctiveness of neural activity patterns was well correlated with behavioral speech discrimination when neural responses were analyzed by spike timing. Neural activity was best correlated with behavioral discrimination when 100 ms of neural activity was binned with 10-ms spike timing precision ($R^2 = 0.61$, $P = 10^{-17}$; Fig. 7C). Neural activity was also significantly well correlated when neural responses were analyzed over 70–180 ms with 1–30-ms spike timing precision ($R^2 > 0.55$; $P < 10^{-12}$). The correlation between behavioral and neural responses was high ($R^2 > 0.55$) when either city-block distance or Euclidean distance was used to quantify distinctiveness of neural responses. These results suggest that the brain can use differences in spike timing to discriminate between speech sounds in noise.

Although this result extents those of our earlier report to noisy conditions, this method requires the stimulus onset time to be known. In the following section, we will explain the

disadvantage of specifying the stimulus onset time, and describe a neural analysis method that does not require the stimulus onset to be specified.

Quantifying neural discrimination without reference to sound onset

Previous studies of neural coding using spike timing strategies generally assumed that the decoding mechanism knew the precise stimulus start time (Gawne *et al*., 1996; Furukawa *et al*., 2000; Ahissar *et al*., 2001; Panzeri *et al*., 2001; Foffani *et al*., 2004, 2008; Schnupp *et al*., 2006; Wang *et al*., 2007; Engineer *et al*., 2008). The stimulus start time is often calculated from the average response of a large group of neurons (Chase & Young, 2007; Engineer *et al*., 2008). Although it is reasonable to expect that stimulus onset time information is available in quiet situations, loud background noise significantly degrades neural responses, making it difficult to determine the stimulus onset time, even based on the average activity of many neurons. Therefore in this study, we developed a new form of neural classifier that is able to determine which speech sound was presented by analyzing neural activity without precise knowledge of stimulus onset time. We compared classifier discrimination with behavioral speech discrimination to determine the neural analysis methods that are correlated with behavior.

In brief, the neural classifier examines activity from a single trial collected from a set of A1 neurons and attempts to identify which of two possible sounds was presented by looking for the spatiotemporal activity patterns generated by each sound (Fig. 7; see Materials and methods). Neural discrimination is determined by calculating the percentage of trials in which the classifier correctly guessed the speech sounds.

We first tested whether the new classifier could mimic the behavioral discrimination in the 11 consonant tasks reported in our earlier study (Engineer *et al*., 2008). Using single trial data from groups of 60 A1 sites, the classifier was able to discriminate `dad' from `bad' in quiet for 99 ± 1% of the time, which is comparable to behavioral discrimination. Classifier discrimination on the 11 tasks was highly correlated with behavioral discrimination in quiet when the average spatiotemporal patterns were composed of 100 ms of neural activity binned with 10-ms precision ($R^2 = 0.75$, $P = 0.0005$). This correlation was similar to the correlation observed with the classifier from our previous study, which was given the exact start time of each stimulus ($R^2 = 0.66$, $P = 0.005$) (Engineer *et al*., 2008).

The new classifier required neural activity from at least 25 recording sites to achieve a performance that was comparable to that of the old classifier using one site. The performance of the new classifier was almost at chance level when only one A1 site was provided (51 ± 0%). This poor performance when neural data from only one recording site are given is attributable to the fact that the classifier could not distinguish spontaneous activity from driven activity without knowledge of the stimulus onset time. When the spatiotemporal activity patterns included many sites, the new classifier was able to reliably discriminate between speech sounds by using data from a single trial. Using a large number of sites provided the classifier with additional information about the spatial pattern of activity which was not available when only one recording site was used. When neural data from a large number of sites were analyzed together, classifier performance improved, because the spatiotemporal pattern of driven activity across sites was distinct from spontaneous activity patterns. The sound `dad', for example, caused high-frequency neurons to fire 5–10 ms before low-frequency neurons, whereas the sound `bad' caused the neurons to fire in the opposite order (Fig. 8A and B). When activity from many sites was analyzed, this pattern could be easily distinguished from spontaneous activity occurring during the 750-ms analysis window (Fig. 8C). This analysis [tested with current and previously published behavioral and neural data in Engineer *et al*. (2008)] suggests that it is not necessary to know the precise stimulus onset time to decode A1 patterns.

In the current study, we tested five additional consonant discrimination tasks that were not included in our earlier study (Engineer *et al*., 2008), resulting in a total of 16 consonant discrimination tasks tested in quiet. Behavioral and neural discrimination from each dataset were averaged together for the six tasks that were examined in both studies. We observed a high correlation between behavioral and classifier discrimination ($R^2 = 0.72$, $P = 10^{-5}$; Fig. 9A) when the classifier analyzed neural activity using 100 ms of activity binned with 10-ms precision. The correlation between neural and behavioral discrimination was also high when the average patterns had a duration of 70–150 ms and were binned with 5–20-ms precision ($R^2 > 0.55$, $P < 0.0001$). For future reference, we will refer to the classifier that analyzes activity from 60 sites using 100-ms durations binned with 10-ms precision as $\text{classifier}_{100/10\ \text{ms}}$. In our previous study in quiet, spike timing information was well correlated with behavioral discrimination, and average spike count was not correlated with behavioral consonant discrimination. Neural responses analyzed with our new classifier also were very poorly correlated with behavioral consonant discrimination on the 16 tasks when an average spike count over a 100-ms duration was used ($R^2 = 0.11$, $P = 0.3$). These results support the hypothesis from our earlier study that spike timing information is required to discriminate between consonant sounds in quiet conditions.

After confirming that the new classifier was correlated with behavioral discrimination in quiet, we tested whether the classifier discrimination was correlated with speech discrimination in noise ($n = 11$ tasks). $\text{Classifier}_{100/10\ \text{ms}}$ was unable to discriminate between speech sounds presented in even moderate noise. For example, $\text{classifier}_{100/10\ \text{ms}}$ discrimination of `dad' vs. `bad' in 48 dB SPL white noise was at chance levels (Fig. 9B). Behavioral discrimination in this condition was 91 ± 1%. This discrepancy between neural and behavioral discrimination was seen across all noise conditions and discrimination tasks, leading to a poor correlation between behavior and $\text{classifier}_{100/10\ \text{ms}}$ discrimination in noise ($R^2 = 0.2$; Fig. 9B). None of the spike timing analysis ranges that could predict behavioral discrimination in quiet could predict behavioral discrimination in noise. This result led us to hypothesize that neural responses in noise are analyzed differently from those in quiet.

We tested classifier discrimination in noise by using multiple combinations of durations and temporal precisions, that is, bin sizes. The range of tested durations was from 10 to 700 ms, and the range of tested bin sizes was from 1 to 700 ms. As neural responses were delayed in noisy situations, we suspected that a classifier that used longer durations of activity would result in a better correlation with behavior. Classifier discrimination was best correlated with behavioral discrimination on all tasks in the different noise conditions when the average spatiotemporal patterns from 60 A1 sites were analyzed with a duration of 400 ms and binned with a temporal precision of 60 ms ($R^2 = 0.68$, $P = 10^{-17}$; Fig. 10B). For example, classifier discrimination on `dad' vs. `bad' in 48 dB SPL white noise was 96 ± 1%. For future reference, we refer to the classifier that analyzes activity from 60 sites using 400-ms durations binned with 60-ms precision as $\text{classifier}_{400/60\ \text{ms}}$. In 48–60 dB SPL noise, $\text{classifier}_{400/60\ \text{ms}}$ discrimination was significantly better on all 11 tasks than $\text{classifier}_{100/10\ \text{ms}}$ discrimination (Figs 9B and 10B; $F_{1,10} = 12.95$, $P = 0.001$, two-way repeated-measures ANOVA). $\text{Classifier}_{400/60\ \text{ms}}$ discrimination remained close to chance for most tasks in 72 dB SPL noise ($P = 0.005$ for 16 of 22 tasks), which is similar to the behavioral discrimination. $\text{Classifier}_{400/60\ \text{ms}}$ discrimination was well correlated with behavioral discrimination in noise when neural responses were analyzed with 25 or more sites together ($R^2 > 0.55$). Classifier discrimination was well correlated with behavioral discrimination in noise, provided that neural activity was analyzed over 300–600-ms durations and binned with 50–100-ms temporal precision ($R^2 > 0.55$, $P < 10^{-10}$). None of the analysis ranges that were well correlated with behavioral discrimination in noise were well correlated with behavioral discrimination in quiet. For example, $\text{classifier}_{400/60\ \text{ms}}$ was poorly correlated with behavioral discrimination in quiet, because of a ceiling effect

observed for many tasks ($R^2 = 0.11$, $P = 0.2$; Fig. 10A). Our observations that (i) neural responses analyzed over small integration windows can predict behavioral discrimination in quiet but not in noise and (ii) responses analyzed with longer integration windows can predict behavioral discrimination in noise but not in quiet support our hypothesis that neural responses are analyzed differently in quiet and noisy situations. These results show that neural responses need to be integrated over longer time scales to accurately predict behavioral discrimination in noisy conditions.

A hybrid classifier that uses the 100/10-ms parameters in quiet and the 400/60-ms parameters in noise was highly correlated with behavioral discrimination under all conditions tested ($R^2 = 0.66$, $P = 10^{-19}$; Fig. 11). Hybrid classifier performance was well correlated with behavioral discrimination when neural responses of more than 25 sites were grouped together ($R^2 > 0.55$, $P < 10^{-11}$; Fig. 12A). Classifier accuracy was closest to behavioral discrimination (Fig. 2A) when it was provided with neural activity recorded from 60 sites (Fig. 12B). Classifier discrimination was well correlated with behavioral discrimination when the average patterns in quiet were analyzed with 70–150-ms durations and binned with 5–20-ms precision, and when the average patterns in noise were analyzed with 300–600-ms durations and binned with 50–100-ms precision ($R^2 > 0.55$, $P < 10^{-11}$). These results show that neural responses can explain behavioral discrimination provided that they are analyzed over longer integration periods and with less temporal precision in noise.

We also recorded neural activity from six of the 12 rats trained to discriminate speech sounds in noise ($n = 4$ rats with target sound as `dad', and $n = 2$ rats with target sound as `shad'). Neural discrimination found with activity from trained rats was well correlated with behavioral discrimination ($R^2 > 0.60$, $P = 10^{-18}$), and was not significantly different from that in naïve rats ($F_{6,10} = 2.36$, $P > 0.05$, two-way repeated-measures ANOVA). Although the correlation between neural and behavioral discrimination was slightly higher when data from naïve rats were used ($R^2 > 0.66$, $P < 10^{-19}$), the difference between the correlations was not statistically significant ($P = 0.42$, chi-square statistic with confidence intervals of 0.72–0.84). These results suggest that 2 months of training on a difficult speech in noise discrimination task did not significantly alter A1 responses to speech sounds. This lack of plasticity could possibly be explained by recent theories proposing that cortical plasticity might be necessary only in the first few days, when learning occurs, and is not necessary for sustained performance, as seen in our animals after 2 months of training (Reed *et al.*, 2011).

Although the prediction of the hybrid classifier could be off by as much as 30% for a specific task in a specific noise condition (i.e. `shad' vs. `chad' in 48 dB SPL speech-shaped noise), the hybrid classifier was highly accurate in predicting the average discrimination of rats across 11 tasks in different noise conditions ($R^2 = 0.89$, $P = 0.002$; Fig. 13A). If relative spike timing information was not used (i.e. spike count only), the hybrid classifier was not able to match this level of correlation, regardless of the number of sites or duration of activity examined ($R^2 < 0.35$). For example, classifier discrimination on most tasks was poorly correlated with behavioral discrimination even when noise-dependent integration windows (i.e. 100 ms in quiet and 400 ms in noise) for spike count were used ($R^2 = 0.03$). Classifier discrimination in both quiet and noise was poorly correlated with behavioral discrimination when spike count over the entire duration of the sound was used ($R^2 = 0.1$). Collectively, our results suggest that spike timing information is required to explain behavioral speech discrimination and that the precision of the spike timing information depends on whether significant background noise is present.

Discussion

The human auditory system is highly effective at extracting speech information from background noise. Speech discrimination is well above chance even when background noise is 10 dB louder than the speech signal (Miller & Nicely, 1955; Wang & Bilger, 1973; Phatak *et al.*, 2008). Our study provides the first evidence that rats can accurately discriminate between speech signals when the background noise is 10 dB louder than the speech signal. The similarity of speech in noise performance supports the hypothesis that the neural mechanisms that support speech sound processing could be similar in humans and other mammals (Kuhl & Miller, 1975; Tallal *et al.*, 1993; Fitch *et al.*, 1997; Cunningham *et al.*, 2002; Reed *et al.*, 2003; Mesgarani *et al.*, 2008). In both humans and rats, certain speech contrasts are more sensitive to noise than others. These differences in noise sensitivity appear to result from an inability to distinguish between the spatiotemporal patterns evoked by similar speech sounds, and not simply from an inability to detect the sounds. We have developed a novel neural classifier that mimics behavioral discrimination. Unlike previous classifiers, this new classifier does not assume that the stimulus onset time is known. The classifier was only able to mimic behavioral discrimination when neural responses to speech in noise were integrated over longer time windows than were needed to mimic discrimination of speech sounds in quiet. Our results indicate that the rat may prove to be a useful model of human speech sound processing, and could be used to develop models of speech sound processing disorders.

Comparison with human psychophysical literature

In addition to showing that rats and humans have similar thresholds for speech in noise, our study reveals a number of other similarities between these species. Human psychophysical studies have observed that certain speech sounds are more easily masked by noise than others, and that some types of noise mask certain sounds more than others. These results from previous human studies provided us with another opportunity to evaluate the similarity in behavioral speech discrimination in rats and humans. Our behavioral results in rats confirm several key findings from human psychophysical studies. In both rats and humans, most speech sounds are more impaired in speech-shaped noise than in white noise (Busch & Eldredge, 1967; Dubno & Levitt, 1981), but high-frequency sounds, such as fricatives, are more impaired in white noise than in speech-shaped noise (Miller & Nicely, 1955; Wang & Bilger, 1973). Voicing contrasts are much more robust than other speech contrasts in white noise in both rats and humans (Miller & Nicely, 1955; Wang & Bilger, 1973; Dubno & Levitt, 1981). These additional similarities between rat and human speech performance in noise corroborate the hypothesis that at least the primary stages of speech sound processing employ the basic auditory processing mechanisms common to all mammals (Kuhl & Miller, 1975; Merzenich *et al.*, 1993; Tallal *et al.*, 1993; Cunningham *et al.*, 2002; Reed *et al.*, 2003; Mesgarani *et al.*, 2008). Further studies are needed to test the conditions under which the basic speech processing would be similar between animals and humans. For example, it would be interesting to know whether rats, like humans, can use glimpsing strategies to improve speech discrimination in comodulated noise (Buus, 1985; Moore, 1985; Hall & Grose, 1991).

Predictors of speech intelligibility, such as the Articulation Index, the Speech Intelligibility Index, and the Speech Transmission Index, have been shown to predict human speech intelligibility in noise (Dorman *et al.*, 1997; Steeneken & Houtgast, 2002; Cox & Vinagre, 2004; Chen & Loizou, 2010). These systems are good at predicting the average intelligibility of the speech signal in noise, but cannot predict the intelligibility of individual speech contrasts or explain why some speech contrasts are more confusable than others in noise. Moreover, these systems require adjustments such as frequency weighting functions, which differ according to the speech signal and the acoustic conditions (Steeneken & Houtgast,

1980, 2002; Studebaker *et al.*, 1987; Chen & Loizou, 2010). Our study could lead to a generic neural-inspired model of speech sound discrimination in noise. The neural mechanisms found in our study may provide insights that are useful for designing systems to discriminate between speech sounds in noisy conditions.

Comparison with human neurophysiological studies

Our neurophysiological results reveal a number of important similarities between neural responses to speech in noise recorded in rats and humans. The addition of background noise increases the latency and decreases the amplitude of consonant responses in both rats and humans (Whiting *et al.*, 1998; Martin *et al.*, 1999; Martin & Stapells, 2005). Our recordings mimic human imaging results, which show that the neural representation of vowel sounds is more robust to noise than that of consonant sounds (Russo *et al.*, 2004; Song *et al.*, 2010). This robustness is probably attributable to the fact that vowel sounds are generally louder than consonants (Brinton, 2000). Previous psychophysical studies have explained the effect of noise in terms of articulatory features (e.g. voicing and place of articulation), but have provided no biological hypothesis for these differences (Miller & Nicely, 1955; Busch & Eldredge, 1967; Dubno & Levitt, 1981; Phatak *et al.*, 2008). Our results provide the first evidence that discrimination of different speech contrasts in noise can be explained by reduced differences in the evoked spatiotemporal neural activity patterns.

Additional studies are needed to confirm that our recordings from anesthetized rats are comparable to recordings from awake rats who are actively attending the speech sounds. Earlier neural recordings in awake and anesthetized monkeys, cats, ferrets and rats suggest that the basic pattern of speech-evoked responses is only modestly altered by anesthesia (Steinschneider *et al.*, 1999, 2005; Wong & Schreiner, 2003; Engineer *et al.*, 2008; Mesgarani *et al.*, 2008). Speech-evoked responses in awake humans were also similar to responses recorded in awake rats and monkeys (Steinschneider *et al.*, 1999, 2005; Engineer *et al.*, 2008). The similarity of behavioral and neural responses between rats and humans suggests that the rat is a reasonable animal model of speech sound processing and that the rat model could be useful in understanding the greater noise sensitivity of certain clinical populations, including individuals with hearing loss and learning impairments (Cunningham *et al.*, 2002; Ziegler *et al.*, 2005, 2009; Threlkeld *et al.*, 2007; Anderson *et al.*, 2010; Engineer *et al.*, 2011).

Clinical relevance

People with hearing and learning impairments have difficulty in understanding speech in noisy conditions (Hawkins & Stevens, 1950; Snowling *et al.*, 1986; Hygge *et al.*, 1992; Stuart & Phillips, 1996; Cunningham *et al.*, 2001; Appeltants *et al.*, 2005; Sperling *et al.*, 2005; Harris *et al.*, 2009). For example, learning-impaired children are three times worse than normal children in understanding conversational speech in noise (Cunningham *et al.*, 2001). An understanding of how auditory neurons represent speech sounds may help in determining the nature of neural encoding deficits in these populations. Although animals do not interpret speech sounds linguistically, our study indicates that the basic underlying processing of speech sounds may be similar in humans and animals, even under highly noisy conditions. Our study shows that temporal analysis of neural activity is required to accurately discriminate between speech sounds in noise. Previous studies have shown that improvements in temporal processing increase speech processing capabilities in learning-impaired children (Merzenich *et al.*, 1996; Tallal *et al.*, 1996; Russo *et al.*, 2010). Further studies in rat models of dyslexia could provide insights into abnormal encoding of speech sounds in these populations (Threlkeld *et al.*, 2007) and have implications for clinical management of speech processing disorders.

Comparison with neural decoding literature

An extensive literature on neural decoding has shown that spatiotemporal patterns in the cortex provide significant information about the sensory world (Abeles *et al.*, 1993; Villa *et al.*, 1999; Butts *et al.*, 2007; Kayser *et al.*, 2009; Huetz *et al.*, 2010). Although the amount of information is almost invariably greater when spike timing is included in the analysis, there is relatively little behavioral evidence that spike timing information can predict behavior (Reinagel & Reid, 2000; Panzeri *et al.*, 2001; Foffani *et al.*, 2004; Butts *et al.*, 2007; Huetz *et al.*, 2010; Panzeri & Diamond, 2010). A recent study showed that there is a high correlation between behavioral consonant discrimination in quiet and the distinctiveness of cortical activity patterns, but only when spike timing information is included (Engineer *et al.*, 2008). This result differed from those of earlier studies in primates showing that cortical decoding based on spike count was best correlated with behavior (Romo & Salinas, 2003; Liu & Newsome, 2005; Lemus *et al.*, 2009). The most likely explanation for the apparent contradiction is that the primate studies used continuous or periodic stimuli that lack the kind of temporal transitions that are present in consonant stimuli. In this study, we tested consonant sounds embedded in background noise to determine the effect of degrading stimulus onset timing on behavioral and neural discrimination. Because background noise makes it harder to determine when the sounds occur, we developed a method of decoding neural responses without providing the classifier with the stimulus onset time. Our results provide the first behavioral evidence that a code based on relative spike timing can account for behavior over a wide range of conditions without the need for precise knowledge of stimulus start time. The observation that neural decoding is well correlated with behavior supports the long-held hypothesis that sensory scenes are encoded in highly distributed spatiotemporal activity patterns (Abeles *et al.*, 1993; Villa *et al.*, 1999; Butts *et al.*, 2007; Kayser *et al.*, 2009; Huetz *et al.*, 2010).

The precision of spike timing needed to accurately explain consonant discrimination behavior seems to depend on whether the sounds are presented in quiet or noise. Neural activity analyzed with longer temporal integration than for quiet is necessary to account for consonant discrimination in noise. We know of no study that has directly related neural discrimination in cortex based on different levels of temporal integration with behavioral discrimination in high-noise and low-noise situations. Our combined neural classifier model uses spike timing activity integrated over a longer window in noise, and results in neural discrimination that is comparable to observed behavioral discrimination.

Limitations

An important limitation of our study was that the classifier was provided with neural responses over a fixed time period. We tested classifier discrimination with neural activity from 100 ms before and 150 ms after stimulus presentation (total duration of 750 ms). This time period given to the classifier can be viewed as an expectation period during which the stimulus can appear. In real-life situations, the brain has to process information without any expectation period; that is, there is an infinite stimulus time uncertainty. A previous study showed that an increase in this window of expectation decreased classifier discrimination if a spike count strategy was used, but not when a spike timing strategy was used (Panzeri & Diamond, 2010). These results suggest that neural responses analyzed with spike timing measures could be robust to larger stimulus time uncertainties than tested in our study. Neural network models that work in real time are needed to shed further light on how the brain handles these challenges.

Another limitation of our classifier model is that it does not explain how computations in the classifier are performed by neurons. An important part of the classifier is comparison between the average post-stimulus time histogram (PSTH) template of sounds and single

trial activity of neurons. Our classifier does not explain the neural mechanisms underlying these computations; that is, it is not mechanistically realistic, and serves as an estimate of neural distinctiveness. Another limitation is that although the brain can clearly detect background noise (Salvi *et al.*, 2002; Binder *et al.*, 2004; Wong *et al.*, 2008, 2009), our classifier does not specify these mechanisms or how they alter neural decoding. Furthermore, our classifier divides neural activity into multiple intervals and bins to test which analysis strategies are correlated with behavior. The neural mechanisms capable of such temporal processing abilities are not clear. The biophysical properties underlying such computations are investigated by studies that focus on developing realistic neural network models showing how spike timing strategies are learnt and used by these networks to discriminate between complex stimuli (Buonomano & Merzenich, 1995; Buonomano & Maass, 2009; Gutig & Sompolinsky, 2006;). Further work in these areas will help us better understand how the brain accomplishes such computations.

Acknowledgments

The authors would like to thank K. Ranasinghe, W. Vrana and J. Riley for their assistance with microelectrode mappings. We would like to thank B. Porter and A. Reed for their help with the behavioral experiment set-up, and T. Rosenthal, W. Vrana and K. Fitch for their help with behavioral training. We would also like to thank P. Assmann, H. Abdi, M. Atzori, C. McIntyre, B. Porter, K. Ranasinghe, A. Reed, T. Rosen and M. Borland for their suggestions about earlier versions of the manuscript. This work was supported by grants from the US National Institute for Deafness and Other Communicative Disorders (R15DC006624 and R01DC010433).

Abbreviations

A1	primary auditory cortex
PSTH	post-stimulus time histogram

References

Abeles M, Bergman H, Margalit E, Vaadia E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol. 1993; 70:1629–1638. [PubMed: 8283219]

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. USA. 2001; 98:13367–13372. [PubMed: 11698688]

Anderson S, Skoe E, Chandrasekaran B, Kraus N. Neural timing is linked to speech perception in noise. J. Neurosci. 2010; 30:4922–4926. [PubMed: 20371812]

Appeltants D, Gentner TQ, Hulse SH, Balthazart J, Ball GF. The effect of auditory distractors on song discrimination in male canaries (Serinus canaria). Behav. Processes. 2005; 69:331–341. [PubMed: 15896531]

Billings CJ, Bennett KO, Molis MR, Leek MR. Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. Ear Hear. 2010; 32:53–60. [PubMed: 20890206]

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 2004; 7:295–301. [PubMed: 14966525]

Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. J. Cogn. Neurosci. 2009; 21:1790–1805. [PubMed: 18823249]

Brinton, L. The Structure of Modern English: a Linguistic Introduction. Vol. 1. John Benjamin Publishing Co.; Amsterdam: 2000. p. 42

Buonomano DV, Maass W. State-dependent computations: spatiotemporal processing in cortical networks. Nat. Rev. Neurosci. 2009; 10:113–125. [PubMed: 19145235]

Buonomano DV, Merzenich MM. Temporal information transformed into a spatial code by a neural network with realistic properties. Science (New York, N.Y.). 1995; 267:1028–1030.

Busch AC, Eldredge D. The effect of differing noise spectra on the consistency of identification of consonants. Lang. Speech. 1967; 10:194–202. [PubMed: 5583885]

Butts DA, Weng C, Jin J, Yeh CI, Lesica NA, Alonso JM, Stanley GB. Temporal precision in the neural code and the timescales of natural vision. Nature. 2007; 449:92–95. [PubMed: 17805296]

Buus S. Release from masking caused by envelope fluctuations. J. Acoust. Soc. Am. 1985; 78:1958–1965. [PubMed: 4078172]

Byrne D, Harvey D, Khanh T, Stig A, Keith W, Robyn C, Bjorn H, Raymond H, Joseph K, Lui C, Jurgen K, Kotby MN, Nasser HAN, Wafaa AHEK, Yasuko N, Herbert O, Richard P, Dafydd S, Rhys M, Tony S, George T, Gregory IF, Soren W, Carl L. An international comparison of long-term average speech spectra. J. Acoust. Soc. Am. 1994; 90:2108–2120.

Chase SM, Young ED. First-spike latency information in single neurons increases when referenced to population onset. Proc. Natl. Acad. Sci. USA. 2007; 104:5175–5180. [PubMed: 17360369]

Chen F, Loizou PC. Analysis of a simplified normalized covariance measure based on binary weighting functions for predicting the intelligibility of noise-suppressed speech. J. Acoust. Soc. Am. 2010; 128:3715–3723. [PubMed: 21218903]

Cox S, Vinagre L. Modelling of confusions in aircraft call-signs. Speech Commun. 2004; 42:289–312.

Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin. Neurophysiol. 2001; 112:758–767. [PubMed: 11336890]

Cunningham J, Nicol T, King C, Zecker SG, Kraus N. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. Hear. Res. 2002; 169:97–111. [PubMed: 12121743]

Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Am. 1997; 102:2403–2411. [PubMed: 9348698]

Dubno JR, Levitt H. Predicting consonant confusions from acoustic analysis. J. Acoust. Soc. Am. 1981; 69:249–261. [PubMed: 7217523]

Ehret G, Gerhardt HC. Auditory masking and effects of noise on responses of the green treefrog (Hyla cinerea) to synthetic mating calls. J. Comp. Physiol. 1980; 141:13–18.

Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP. Cortical activity patterns predict speech discrimination ability. Nat. Neurosci. 2008; 11:603–608. [PubMed: 18425123]

Engineer ND, Riley JR, Seale JD, Vrana WA, Shetake JA, Sudanagunta SP, Borland MS, Kilgard MP. Reversing pathological neural activity using targeted plasticity. Nature. 2011; 470:101–104. [PubMed: 21228773]

Fitch RH, Miller S, Tallal P. Neurobiology of speech perception. Annu. Rev. Neurosci. 1997; 20:331–353. [PubMed: 9056717]

Foffani G, Tutunculer B, Moxon KA. Role of spike timing in the forelimb somatosensory cortex of the rat. J. Neurosci. 2004; 24:7266–7271. [PubMed: 15317852]

Foffani G, Chapin JK, Moxon KA. Computational role of large receptive fields in the primary somatosensory cortex. J. Neurophysiol. 2008; 100:268–280. [PubMed: 18400959]

Furukawa S, Xu L, Middlebrooks JC. Coding of sound-source location by ensembles of cortical neurons. J. Neurosci. 2000; 20:1216–1228. [PubMed: 10648726]

Gawne TJ, Kjaer TW, Richmond BJ. Latency: another potential code for feature binding in striate cortex. J. Neurophysiol. 1996; 76:1356–1360. [PubMed: 8871243]

Gerhardt HC, Klump GM. Masking of acoustic signals by the chorus background noise in the green tree frog: a limitation on mate choice. Anim. Behav. 1988; 36:1247–1249.

Gutig R, Sompolinsky H. The tempotron: a neuron that learns spike timing-based decisions. Nat. Neurosci. 2006; 9:420–428. [PubMed: 16474393]

Hall JW III, Grose JH. Relative contributions of envelope maxima and minima to comodulation masking release. Q. J. Exp. Psychol. 1991; 43:349–372.

Harris KC, Dubno JR, Keren NI, Ahlstrom JB, Eckert MA. Speech recognition in younger and older adults: a dependency on low-level auditory cortex. J. Neurosci. 2009; 29:6078–6087. [PubMed: 19439585]

Hawkins JE Jr, Stevens SS. The masking of pure tones and of speech by white noise. J. Acoust. Soc. Am. 1950; 22:6–13.

Horii Y, House AS, Hughes GW. A masking noise with speech-envelope characteristics for studying intelligibility. J. Acoust. Soc. Am. 1971; 49:1849–1856. [PubMed: 5125732]

House AS, Williams CE, Heker MH, Kryter KD. Articulation-testing methods: consonantal differentiation with a closed-response set. J. Acoust. Soc. Am. 1965; 37:158–166. [PubMed: 14265103]

Huetz C, Gourevitch B, Edeline JM. Neural codes in the thalamocortical auditory system: from artificial stimuli to communication sounds. Hear. Res. 2010; 241:147–158. [PubMed: 20116422]

Hulse SH, MacDougall-Shackleton SA, Wisniewski AB. Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (Sturnus vulgaris). J. Comp. Psychol. 1997; 111:3–13. [PubMed: 9090135]

Hygge S, Ronnberg J, Larsby B, Arlinger S. Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. J. Speech Hear. Res. 1992; 35:208–215. [PubMed: 1370969]

Kawahara H. Speech Representation and Transformation using Adaptive Interpolation of Weighted Spectrum: Vocoder Revisited. Proceedings of IEEE Int. Conf. Acoust., Speech, Signal Process. 1997; 1997:1303–1306.

Kayser C, Montemurro MA, Logothetis NK, Panzeri S. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron. 2009; 61:597–608. [PubMed: 19249279]

Kluender KR, Diehl RL, Killeen PR. Japanese quail can learn phonetic categories. Science (New York, N.Y.). 1987; 237:1195–1197.

Kozou H, Kujala T, Shtyrov Y, Toppila E, Starck J, Alku P, Naatanen R. The effect of different noise types on the speech and non-speech elicited mismatch negativity. Hear. Res. 2005; 199:31–39. [PubMed: 15574298]

Kuhl PK, Miller JD. Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants. Science (New York, N.Y.). 1975; 190:69–72.

Lemus L, Hernandez A, Romo R. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc. Natl. Acad. Sci. USA. 2009; 106:9471–9476. [PubMed: 19458263]

Liu J, Newsome WT. Correlation between speech perception and neural activity in the middle temporal visual area. J. Neurosci. 2005; 25:711–722. [PubMed: 15659609]

Lohr B, Wright TF, Dooling RJ. Detection and discrimination of natural calls in masking noise by birds: estimating the active space of a signal. Anim. Behav. 2003; 65:763–777.

Martin BA, Stapells DR. Effects of low-pass noise masking on auditory event-related potentials to speech. Ear Hear. 2005; 26:195–213. [PubMed: 15809545]

Martin BA, Kurtzberg D, Stapells DR. The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds/ba/and/da. J. Speech Lang. Hear. Res. 1999; 42:271–286. [PubMed: 10229446]

Merzenich MM, Schreiner C, Jenkins W, Wang X. Neural mechanisms underlying temporal integration, segmentation, and input sequence representation: some implications for the origin of learning disabilities. Ann. NY Acad. Sci. 1993; 682:1–22. [PubMed: 8323106]

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P. Temporal processing deficits of language-learning impaired children ameliorated by training. Science (New York, N.Y.). 1996; 271:77–81.

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 GA, Licklider JCR. The intelligibility of interrupted speech. J. Acoust. Soc. Am. 1950; 22:167–173.

Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 1955; 27:338–352.

Moore BC. Additivity of simultaneous masking, revisited. J. Acoust. Soc. Am. 1985; 78:488–494. [PubMed: 4031248]

Muller-Gass A, Marcoux A, Logan J, Campbell KB. The intensity of masking noise affects the mismatch negativity to speech sounds in human subjects. Neurosci. Lett. 2001; 299:197–200. [PubMed: 11165769]

Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K. Cortical interference effects in the cocktail party problem. Nat. Neurosci. 2007; 10:1601–1607. [PubMed: 17994016]

Panzeri S, Diamond ME. Information carried by population spike times in the whisker sensory cortex can be decoded without knowldege of stimulus time. Front. Syn. Neurosci. 2010; 2:17.

Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron. 2001; 29:769–777. [PubMed: 11301035]

Phatak SA, Allen JB. Consonant and vowel confusions in speech-weighted noise. J. Acoust. Soc. Am. 2007; 121:2312–2326. [PubMed: 17471744]

Phatak SA, Lovitt A, Allen JB. Consonant confusions in white noise. J. Acoust. Soc. Am. 2008; 124:1220–1233. [PubMed: 18681609]

Ramus F, Hauser MD, Miller C, Morris D, Mehler J. Language discrimination by human newborns and by cotton-top tamarin monkeys. Science (New York, N.Y.). 2000; 288:349–351.

Reed P, Howell P, Sackin S, Pizzimenti L, Rosen S. Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. J. Exp. Anal. Behav. 2003; 80:205–215. [PubMed: 14674729]

Reed A, Riley J, Carraway R, Carrasco A, Perez C, Jakkamsetti V, Kilgard MP. Cortical map plasticity improves learning but is not necessary for improved performance. Neuron. 2011; 70:121–131. [PubMed: 21482361]

Reinagel P, Reid RC. Temporal coding of visual information in the thalamus. J. Neurosci. 2000; 20:5392–5400. [PubMed: 10884324]

Romo R, Salinas E. Flutter discrimination: neural codes, perception, memory and decision making. Nat. Rev. Neurosci. 2003; 4:203–218. [PubMed: 12612633]

Russo N, Nicol T, Musacchia G, Kraus N. Brainstem responses to speech syllables. Clin. Neurophysiol. 2004; 115:2021–2030. [PubMed: 15294204]

Russo N, Hornickel J, Nicol T, Zecker S, Kraus N. Biological changes in auditory function following training in children with autism spectrum disorders. Behav. Brain. Funct. 2010; 6:60–68. [PubMed: 20950487]

Sally SL, Kelly JB. Organization of auditory cortex in the albino rat: sound frequency. J. Neurophysiol. 1988; 59:1627–1638. [PubMed: 3385476]

Salvi RJ, Lockwood AH, Frisina RD, Coad ML, Wack DS, Frisina DR. PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. Hear. Res. 2002; 170:96–106. [PubMed: 12208544]

Schnupp JW, Hall TM, Kokelaar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J. Neurosci. 2006; 26:4785–4795. [PubMed: 16672651]

Skoe E, Kraus N. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 2010; 31:302–324. [PubMed: 20084007]

Snowling M, Goulandris N, Bowlby M, Howell P. Segmentation and speech perception in relation to reading skill: a developmental analysis. J. Exp. Child Psychol. 1986; 41:489–507. [PubMed: 3734692]

Song J, Skoe E, Banai K, Kraus N. Perception of speech in noise: neural correlates. J. Cogn. Neurosci. 2010; 23:2268–2279. [PubMed: 20681749]

Sperling AJ, Lu Z-L, Manis FR, Seidenberg MS. Deficits in perceptual noise exclusion in developmental dyslexia. Nat. Neurosci. 2005; 8:862–863. [PubMed: 15924138]

Steeneken HJ, Houtgast T. A physical method for measuring speech-transmission quality. J. Acoust. Soc. Am. 1980; 67:318–326. [PubMed: 7354199]

Steeneken HJM, Houtgast T. Phoneme-group specific octave-band weights in predicting speech intelligibility. Speech Commun. 2002; 38:399–411.

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA III. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J. Neurophysiol. 1999; 82:2346–2357. [PubMed: 10561410]

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA III. Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb. Cortex. 2005; 15:170–186. [PubMed: 15238437]

Stuart A, Phillips DP. Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. Ear Hear. 1996; 17:478–489. [PubMed: 8979036]

Studebaker GA, Pavlovic CV, Sherbecoe RL. A frequency importance function for continuous discourse. J. Acoust. Soc. Am. 1987; 81:1130–1138. [PubMed: 3571730]

Tallal P, Miller S, Fitch RH. Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann. NY Acad. Sci. 1993; 682:27–47. [PubMed: 7686725]

Tallal P, Miller SL, Bedi G, Byma G, Wang X, Nagarajan SS, Schreiner C, Jenkins WM, Merzenich MM. Language comprehension in language-learning impaired children improved with acoustically modified speech. Science (New York, N.Y.). 1996; 271:81–84.

Threlkeld SW, McClure MM, Bai J, Wang Y, LoTurco JJ, Rosen GD, Fitch RH. Developmental disruptions and behavioral impairments in rats following in utero RNAi of Dyx1c1. Brain Res. Bull. 2007; 71:508–514. [PubMed: 17259020]

Villa AE, Tetko IV, Hyland B, Najem A. Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. Proc. Natl. Acad. Sci. USA. 1999; 96:1106–1111. [PubMed: 9927701]

Wang MD, Bilger RC. Consonant confusions in noise: a study of perceptual features. J. Acoust. Soc. Am. 1973; 54:1248–1266. [PubMed: 4765809]

Wang L, Narayan R, Grana G, Shamir M, Sen K. Cortical discrimination of complex natural stimuli: can single neurons match behavior? J. Neurosci. 2007; 27:582–589. [PubMed: 17234590]

Whiting KA, Martin BA, Stapells DR. The effects of broadband noise masking on cortical event-related potentials to speech sounds / ba / and / da. Ear Hear. 1998; 19:218–231. [PubMed: 9657596]

Wong SW, Schreiner CE. Representation of CV-sounds in cat primary auditory cortex: intensity dependence. Speech Commun. 2003; 41:93–106.

Wong PC, Uppunda AK, Parrish TB, Dhar S. Cortical mechanisms of speech perception in noise. J. Speech Lang. Hear. Res. 2008; 51:1026–1041. [PubMed: 18658069]

Wong PC, Jin JX, Gunasekera GM, Abel R, Lee ER, Dhar S. Aging and cortical mechanisms of speech perception in noise. Neuropsychologia. 2009; 47:693–703. [PubMed: 19124032]

Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C. Deficits in speech perception predict language learning impairment. Proc. Natl. Acad. Sci. USA. 2005; 102:14110–14115. [PubMed: 16162673]

Ziegler JC, Pech-Georgel C, George F, Lorenzi C. Speech-perception-in-noise deficits in dyslexia. Dev. Sci. 2009; 12:732–745. [PubMed: 19702766]

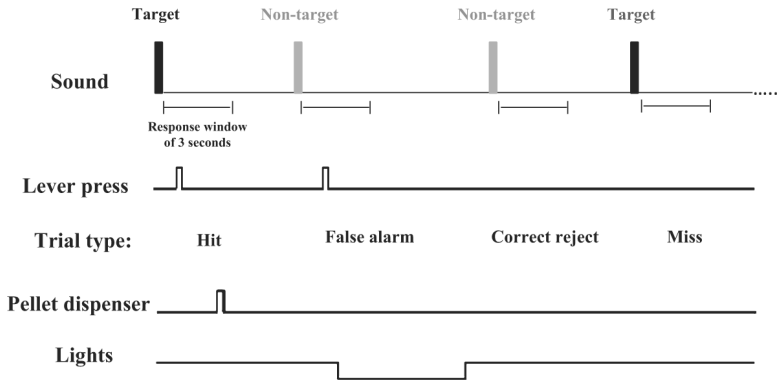

Fig. 1.

Schematic diagram of the go / no-go speech discrimination task. The target sound was a word spoken by a female talker and shifted up by one octave with the STRAIGHT vocoder to better match the rat hearing range (Engineer *et al.*, 2008). The non-targets differed in the initial consonant sound. Continuous background noise was added to make the task more difficult.

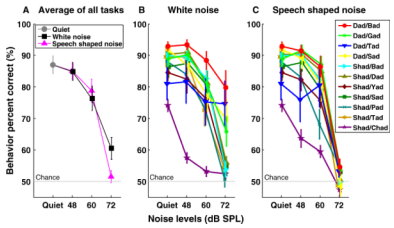

Fig. 2.

Behavioral discrimination of consonant sounds in different intensities of white noise and speech-shaped noise. (A) Average speech discrimination on all tasks. Rats could discriminate between consonant sounds well above chance level, even when speech and noise were of equal intensity, that is, 60 dB SPL ($P = 0.005$, Tukey *post hoc*). (B) Discrimination of 11 consonant discrimination tasks in quiet and different intensities of white noise. (C) Discrimination of 11 consonant discrimination tasks in quiet and different intensities of speech-shaped noise. Chance level performance is shown as light gray lines. Speech-shaped noise of 72 dB SPL was more impairing than white noise of 72 dB SPL ($F_{1,66} = 10.20$, Mean Square Error (MSE) = 244.02, $P = 0.001$).

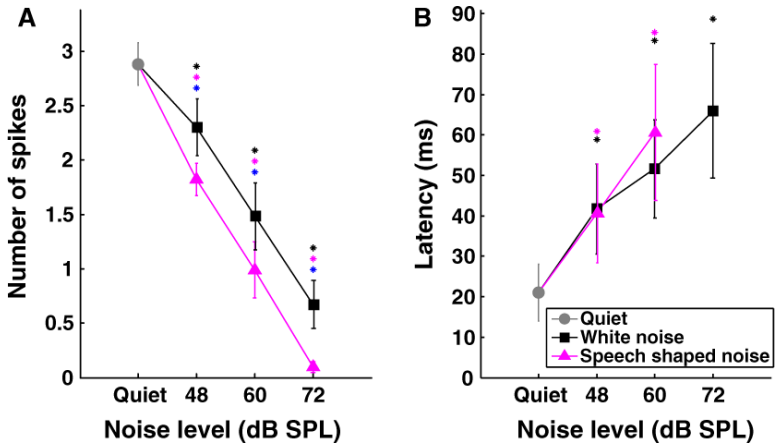

Fig. 3.

Degradation of neural responses in different intensities of white noise and speech-shaped noise. (A) Average number of spikes evoked within 100 ms of consonant sound onset. None of the sounds had significantly driven activity in 72 dB SPL speech-shaped noise ($P > 0.05$, $n = 133$ A1 sites). (B) Average start latency of the population response evoked by all consonant sounds. As there were no significantly driven spikes in 72 dB SPL speech-shaped noise, there was no latency in this condition. Black and magenta asterisks indicate neural responses in white noise and speech-shaped noise, respectively; these were significantly degraded as compared with quiet ($P = 0.0001$, Tukey *post hoc*). Blue asterisks indicate that neural responses in white noise were significantly less degraded than responses in speech-shaped noise at that intensity ($P = 0.0001$, Tukey *post hoc*). Error bars indicate standard errors of the mean across the 11 sounds tested. For interpretation of color references in figure legend, please refer to the Web version of this article.

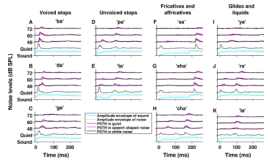

Fig. 4.

Average PSTH responses evoked by each speech sound in three different intensities of white noise and speech-shaped noise. Sounds are grouped according to manner of articulation. Cyan and light gray lines show the amplitude envelopes of speech and noise signals. Speech signals were calibrated so that loudest 100 ms was at 60 dB SPL and noises were at 48, 60 and 72 dB SPL. In quiet, most sounds, except for voiced stop consonants, evoked a two-peaked response; the first peak was evoked by the consonant part of the sound, and the second peak was evoked by the vowel part of the sound. Stop consonants evoked the strongest onset response in quiet. For reference, sound `t' evoked the strongest neural onset response at 377 Hz. Voiced stop consonants (A–C) were the most robust in all noise conditions ($P = 0.0001$, Tukey *post hoc*). Fricatives and affricatives evoked the weakest response in quiet and noise (F and H). For most sounds, neural responses in white noise were significantly more robust than those in speech-shaped noise of equal intensity ($P = 0.0001$, Tukey *post hoc*). For interpretation of color references in figure legend, please refer to the Web version of this article.

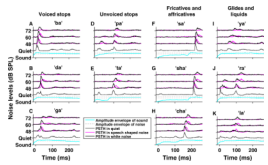

Fig. 5.

Average PSTH responses evoked by different speech sounds in the 45 low-frequency A1 sites. Neural responses in low-frequency neurons (characteristic frequency of neurons, < 4 kHz) were more robust in white noise than in speech-shaped noise ($P = 0.0001$, Tukey *post hoc*). High-frequency sounds (e.g. `t', `s', `sh', and `ch') evoked weak responses in low-frequency neurons even in quiet, and these were completely eliminated in 48 dB SPL noise (E, F, G, and H).

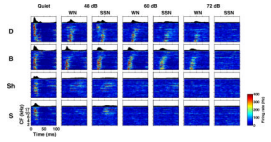

Fig. 6.

Neurograms depicting onset response to speech sounds `dad', `bad', `shad' and `sad' in quiet and six different types of continuous background noise. These neurograms represent the spatiotemporal activity patterns of 133 multiunit A1 recording sites arranged according to the characteristic frequency (CF) of each site. Each horizontal line represents the average PSTH from one recording site in response to 20 stimulus repeats. Time is represented on the *x*-axis (0–100 ms). The firing rate of each site is represented in color, whereby dark blue indicates the lowest firing rate, that is, spontaneous activity, and dark red represents the highest firing rate. For comparison, the mean population PSTH evoked by each sound is plotted above the corresponding neurogram. SSN, speech-shaped noise; WN, white noise. For interpretation of color references in figure legend, please refer to the Web version of this article.

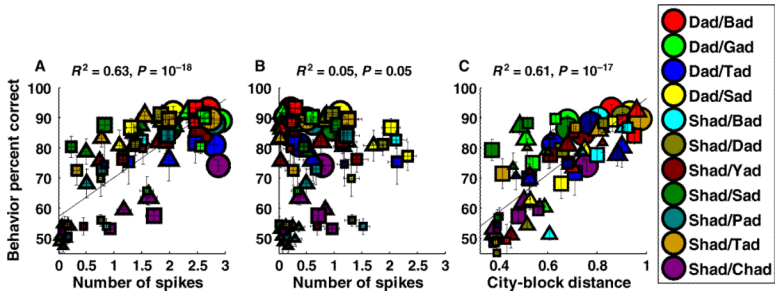

Fig. 7.

Correlation between neural and behavioral responses in all tasks. (A) The average number of spikes evoked by two sounds in the discrimination task is well correlated with behavioral discrimination. (B) The difference in the average number of spikes evoked by two sounds in the discrimination task is poorly correlated with behavioral discrimination. In a and b, the numbers of spikes were averaged together over the first 100 ms. (C) Normalized city-block distances between the average spatiotemporal patterns evoked by each sound involved in the discrimination tasks were well correlated with behavioral discrimination when the neural responses were analyzed over 100 ms and binned with 10-ms spike timing precision. Circles, triangles and squares represent performance in quiet, speech-shaped noise, and white noise, respectively. The sizes of the symbols indicate noise level, smaller symbols indicating greater amount of noise (i.e. lower signal-to-noise ratio).

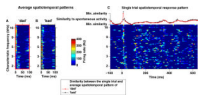

Fig. 8.

Consonant discrimination using relative spike timing. This figure illustrates how the classifier evaluates which of two speech sounds (`dad' or `bad' in this case) was presented from a single trial of neural activity recorded at 60 A1 sites. (A and B) In this example, the average spatiotemporal patterns of sounds `dad' and `bad' were analyzed over 100 ms and binned with 10-ms spike timing precision. This classifier is referred to as classifier$_{100 / 10 ms}$. Neural activity from different A1 sites is arranged according to the characteristic frequency of each site. (C) Single trial spatiotemporal activity from the same 60 sites is shown from 100 ms before the stimulus onset to 150 ms after the stimulus end. The similarity of the single trial to each of the average spatiotemporal activity patterns is shown as red and black lines at the top of the figure, each point indicating the different possible stimulus start times. The point of greatest similarity occurred immediately after the word `dad' was presented (asterisk). The classifier correctly identified that `dad' had been presented, because the similarity of the single trial was highest to the average pattern generated by `dad'. Neural discrimination was determined by calculating the percentage of trials in which the classifier correctly guessed the presented sound (see Materials and methods). The set of A1 sites shown in this example was able to correctly identify whether `dad' or `bad' was presented on 39 of 40 presentations in quiet (97.5% correct). For interpretation of color references in figure legend, please refer to the Web version of this article.

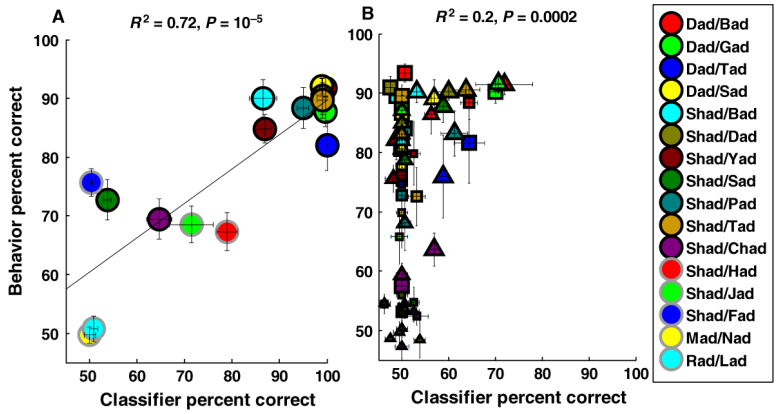

Fig. 9.

Correlation between classifier$_{100\ /\ 10\ ms}$ and behavioral discrimination in quiet and noise. (A) Classifier discrimination was significantly correlated with behavioral discrimination in quiet, when average spatiotemporal patterns were analyzed with 100-ms duration and binned with 10-ms precision. (B) In noise, classifier discrimination analyzed with the same parameters was poorly correlated with behavioral discrimination. Symbols with black borders are the tasks used in this study. Symbols with gray borders are extra tasks used from our previous study (Engineer *et al.*, 2008) (see Materials and methods). Circles, triangles and squares represent discrimination in quiet, speech-shaped noise, and white noise, respectively. The sizes of the symbols indicate noise level, smaller symbols indicating greater amount of noise (i.e. lower signal-to-noise ratio).

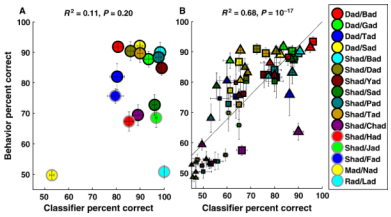

Fig. 10.

Correlation between classifier$_{400\ /\ 60\ ms}$ discrimination and behavioral discrimination in quiet and noise. (A) In quiet, classifier discrimination analyzed with 400-ms duration and binned with 60-ms precision was poorly correlated with behavioral discrimination. (B) In noise, classifier discrimination was significantly correlated with behavioral discrimination, when the average spatiotemporal patterns were analyzed with 400-ms duration and binned with 60-ms precision. Symbols with black borders are the tasks used in this study. Symbols with gray borders are extra tasks used from our previous study (Engineer *et al.*, 2008) (see Materials and methods). Circles, triangles and squares represent discrimination in quiet, speech-shaped noise, and white noise, respectively. The sizes of the symbols indicate noise level, smaller symbols indicating greater amount of noise (i.e. lower signal-to-noise ratio).

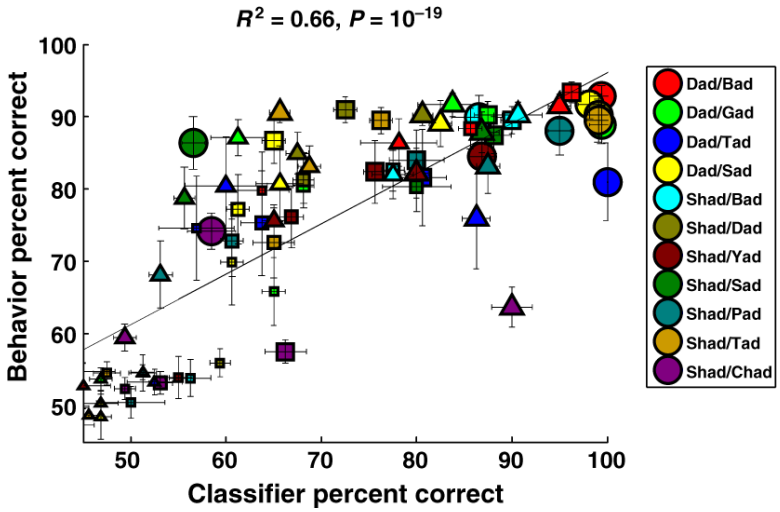

Fig. 11.

Single trial responses analyzed with longer integration time scales than in quiet were well correlated with behavioral discrimination in all conditions. Hybrid classifier performance when the average spatiotemporal patterns in quiet were analyzed with 100-ms duration and binned with 10-ms precision and the average spatiotemporal patterns in noise were analyzed with 400-ms duration and binned with 60-ms precision was correlated with behavioral discrimination in quiet all conditions. Circles, triangles and squares represent discrimination in quiet, speech-shaped noise, and white noise, respectively. The sizes of the symbols indicate noise level, smaller symbols indicating greater amount of noise (i.e. lower signal-to-noise ratio).

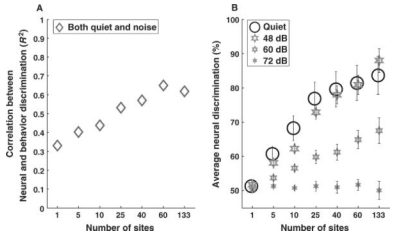

Fig. 12.

Average neural discrimination and correlation between the hybrid classifier and behavioral discrimination as a function of site. (A) The correlation (R^2) between neural and behavioral discrimination improved significantly ($F_{1,10}$ = 12.95, Mean Square Error (MSE) = 0.11, P = 0.001) as a function of site. Neural discrimination in quiet and noise was significantly correlated with behavior when neural responses from more than 25 sites were analyzed together ($R^2 > 0.55$; $P < 10^{-11}$). Neural discrimination from 60 multiunit clusters was best correlated with behavior ($R^2 = 0.68$, $P = 10^{-17}$). (B) Average neural discrimination in all tasks in quiet, 48 dB SPL noise and 60 dB SPL noise increased as a function of site, and was comparable to behavioral discrimination. Increasing the number of sites did not affect neural discrimination in 72 dB SPL noise, which is also comparable with what was found for behavioral discrimination at this noise level.

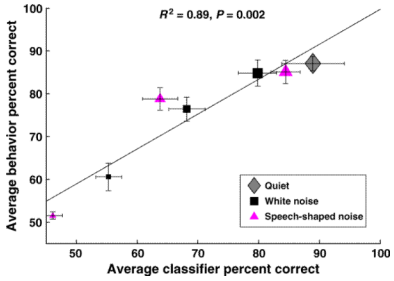

Fig. 13.

Hybrid classifier percentage correct can explain the variance in behavioral discrimination caused by either noise conditions or discrimination tasks alone. The hybrid classifier can explain 89% of the variance (P = 0.002) caused by the seven noise conditions when the classifier discrimination is averaged across all 11 tasks. The sizes of the symbols indicate noise level, smaller symbols indicating greater amount of noise (i.e. lower signal-to-noise ratio).