JARO 13: 527–542 (2012)
DOI: 10.1007/s10162-012-0328-1

JARO
Journal of the Association for Research in Otolaryngology

Research Article

Neural Mechanisms Supporting Robust Discrimination of Spectrally and Temporally Degraded Speech

Kamalini G. Ranasinghe[1], William A. Vrana[1], Chanel J. Matney[1], and Michael P. Kilgard[1]

[1]*School of Behavioral Brain Sciences, The University of Texas at Dallas, 800 West Campbell Road, GR41, Richardson, TX 75080-3021, USA*

Received: 11 October 2011; Accepted: 26 March 2012; Online publication: 2 May 2012

ABSTRACT

Cochlear implants provide good speech discrimination ability despite highly limited amount of information they transmit compared with normal cochlea. Noise vocoded speech, simulating cochlear implants in normal hearing listeners, have demonstrated that spectrally and temporally degraded speech contains sufficient cues to provide accurate speech discrimination. We hypothesized that neural activity patterns generated in the primary auditory cortex by spectrally and temporally degraded speech sounds will account for the robust behavioral discrimination of speech. We examined the behavioral discrimination of noise vocoded consonants and vowels by rats and recorded neural activity patterns from rat primary auditory cortex (A1) for the same sounds. We report the first evidence of behavioral discrimination of degraded speech sounds by an animal model. Our results show that rats are able to accurately discriminate both consonant and vowel sounds even after significant spectral and temporal degradation. The degree of degradation that rats can tolerate is comparable to human listeners. We observed that neural discrimination based on spatiotemporal patterns (spike timing) of A1 neurons is highly correlated with behavioral discrimination of consonants and that neural discrimination based on spatial activity patterns (spike count) of A1 neurons is highly correlated with behavioral discrimination of vowels. The results of the current study indicate that speech discrimination is resistant to degradation as long as the degraded sounds generate distinct patterns of neural activity.

Keywords: speech processing, neural code, noise vocoded speech, primary auditory cortex, cochlear implants, spatiotemporal patterns, spatial patterns

INTRODUCTION

Cochlear implants are one of the most successful neural prostheses ever developed. Bypassing the function of normal cochlea, the cochlear implant signal processor decomposes the input signal into different frequency bands and delivers the temporal envelope modulation of each band to an array of electrodes. Each electrode acts as a channel and stimulates a different set of auditory nerve fibers. Cochlear implants with only eight channels and envelope modulations below 200 Hz provide highly accurate speech discrimination under quiet conditions (Fishman et al. 1997; Hedrick and Carney 1997; Dorman et al. 1998; Kiefer et al. 2000; Loizou et al. 2000; Valimaa et al. 2002; Nie et al. 2006; Won et al. 2007). The spectral and temporal encoding of the speech signal by the cochlear implant speech processor is significantly less detailed than that by the normal cochlea (Loizou 1998, 2006). The neuronal mechanisms involved in processing auditory responses electrically evoked by a cochlear implant are not clearly identified.

Experiments using noise vocoded speech that simulates cochlear implant inputs in normal hearing subjects have suggested that neural mechanisms of speech processing do not depend on fine spectral and temporal details. A noise vocoder decomposes the

Electronic supplementary material The online version of this article (doi:10.1007/s10162-012-0328-1) contains supplementary material, which is available to authorized users.

Correspondence to: Kamalini G. Ranasinghe · School of Behavioral Brain Sciences · The University of Texas at Dallas · 800 West Campbell Road, GR41, Richardson, TX 75080-3021, USA. Telephone: +1-972-8832376; fax: +1-972-883-2491; email: kamalini@utdallas.edu

speech signal into different frequency bands and the temporal envelope of each band is then used to modulate a broadband noise signal (Shannon et al. 1995). Human listeners are able to discriminate noise vocoded speech with high accuracy (Van Tasell et al. 1987; Drullman et al. 1994; Shannon et al. 1995; Dorman and Loizou 1997; Xu et al. 2005). More than 90 % correct word and sentence discrimination is achieved with four to eight spectral bands and envelope modulations low-pass-filtered at 256 Hz (Shannon et al. 1995; Loizou et al. 1999; Xu et al. 2005; Dorman and Loizou 1997; Xu et al. 2005). These results indicate that neural mechanisms underlying speech sound processing can produce similar behavioral performance with few spectral channels and slow envelope modulations.

The neural activity patterns evoked by degraded speech sounds in the central auditory system and the potential ability of these patterns to explain the robust behavioral discrimination are yet to be explored. Responses to noise vocoded speech processed with 1, 2, 3, or 4 spectral bands in chinchilla auditory nerve have shown that distinctions between undegraded consonant sounds remained prominent when noise vocoded with four spectral bands (Loebach and Wickesberg 2006). We propose that neural activity patterns generated in primary auditory cortex (A1) neurons by speech sounds noise vocoded with few channels (8–16) and slow modulations (less than 256 Hz) will generate neural differences that are similar to those generated by undegraded speech sounds. We examined the behavioral performance in rats to discriminate noise vocoded consonants and vowels degraded into ten different levels (nine noise vocoded levels plus undegraded). We recorded spatiotemporal activity patterns from rat A1 for the same stimuli and the differences between activity patterns were used to examine the neural discrimination ability. We predicted that neural differences generated by spectrally and temporally degraded sounds in A1 will be correlated with the behavioral discrimination patterns.

METHODS

Speech stimuli and noise vocoding of speech

We used seven English words including ‘dad,’ ‘bad, ‘sad,’ ‘tad,’ ‘dud,’ ‘deed,’ and ‘dood’ spoken by a native female English speaker in a double-walled sound proof booth (supplementary audio). The fundamental frequency and the spectral envelope of each recorded word were shifted up in frequency by a factor of 2 using the STRAIGHT vocoder (Kawahara 1997) to better match the rat hearing range (Engineer et al. 2008). For example, the fundamental frequency of the octave-shifted ‘Dad’ appeared at 225 Hz. Each octave-shifted word was then degraded through a noise vocoder (Shannon et al. 1995). The first step in the noise vocoder was to pass each word through a bank of bandpass filters. The number of bandpass filters used in this step varied between 2 and 16 (2, 4, 8, 16) and determined the four levels of spectral degradation (Fig. 1). The overall bandwidth of the degraded speech sounds ranged from 300 to 11,000 Hz. Logarithmic filter spacing was used to set the bandwidths of the filters (Loizou et al. 1999). Next, the temporal envelope of each analysis band was extracted by half-wave rectification and low-pass filtering. The low-pass filter cutoff frequency varied between 4 and 256 Hz (4, 16, 64, 256) determining the four levels of temporal degradation (Fig. 1). The extracted envelope was then used to modulate a band of noise that was filtered with the same bandpass filters as the original speech. The modulated noise bands were summed to create the final degraded speech sound. The intensity of all the speech sounds used in the study was adjusted so that the intensity of the most intense 100 ms was at 60 dB SPL. The signal processing was done using MatLab (MathWorks, Natick, MA).

The four spectral levels and the four temporal levels created a combination of 16 degraded levels. Out of this, we selected nine levels for behavior training and neural recordings (Fig. 1). In addition to the degraded speech sounds, we also used the undegraded versions of the seven words for both behavior training and neural recordings. A total of 70 speech stimuli (nine degraded levels × seven words plus seven undegraded words) were used in this study.

Behavior training and analysis

Nine rats underwent behavior training. Each rat was placed in a sound-shielded operant training booth for 1-h session, twice daily for 5 days per week. The booth contained a cage with a lever, a lever light, a house light, and a pellet receptacle. A pellet dispenser was connected to the pellet receptacle and located outside the booth. A calibrated speaker was mounted approximately 20 cm from the mid-point between the lever and the pellet receptacle (which is the most common location of the rat’s head position while inside the booth). The rats were watched via video monitoring during each session. An operant go-no-go paradigm was used to discriminate speech sounds.

The training began with a brief shaping period to teach the rat to press the lever in order to receive a food pellet reward. Each time the rat was in close proximity to the lever, it heard the target sound and received a sugar pellet reward. The target sound during the initial shaping was undegraded ‘dad.’ Eventually, the rat learned to press the lever without assistance. The rat heard the target sound after each lever press and received a sugar pellet. The shaping

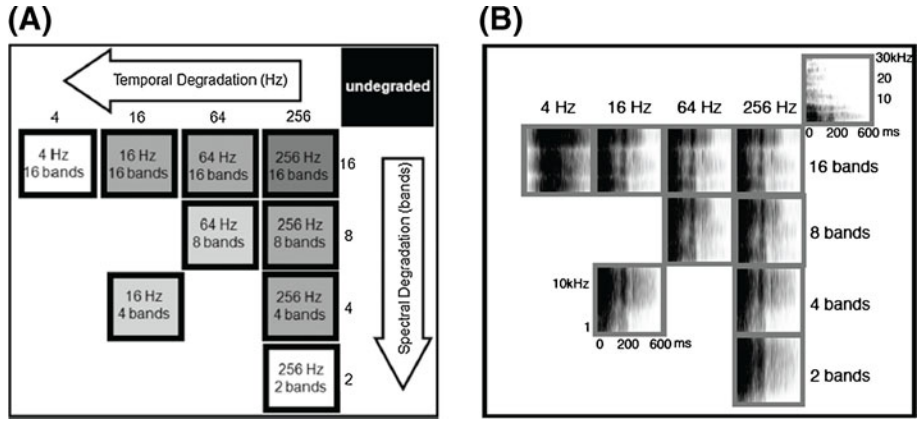


FIG. 1. Levels of degradation. **A** Four spectral levels (along the *vertical axis*) are defined by the number of bandpass filters used in the noise vocoder as 16, 8, 4, and 2. Four temporal levels (along the *horizontal axis*) are defined by the low-pass filter cutoff frequency as 256, 64, 16, and 4 Hz. The nine degraded levels included four spectrally degraded levels (maintaining the temporal LPF at 256 Hz), four temporally degraded levels (maintaining the spectral level at 16 bands), and two levels both spectrally and temporally degraded as eight bands/64 Hz and four bands/16 Hz. Together with the undegraded stimuli this made up to ten levels. **B** Spectrograms of the undegraded 'Dad' and the nine levels of vocoded 'Dad'.

continued until the rat was able to earn a minimum of 120 pellets per session, which lasted on average 3–5 days. The rats were food-deprived to encourage motivation and were weighed daily to make sure that above 85 % of their pre-behavior body weight was maintained. During the next stage of training, the rat began a detection task where it was trained to press the lever when the target sound was presented. Initially, the sound was presented every 10 s, and the rat was given an 8-s window to press the lever. The rat received a sugar pellet reward for pressing the lever within 8 s. The sound interval was gradually reduced to 6 s and the lever press window to 3 s. The target sound was randomly interleaved with silent catch trials during each training session to prevent rats pressing the lever every 6 s habitually. Performance of the rat was monitored using the d' value, which is a measure of discriminability according to signal detection theory (Green and Swets 1989). Training continued until the rats reached a $d' \geq 1.5$ for ten sessions. Following the detection phase, the rats started discrimination training where they learned to discriminate the target sound ('dad') from several non-target sounds. The non-target sounds included 'bad,' 'sad,' 'tad,' 'dud,' 'deed,' and 'dood'. In a given 1-h session, the rat discriminated the target sound against the six non-target sounds that were presented randomly interleaved with silent catch trials. Rats were rewarded with a sugar pellet reward when they pressed the lever within 3 s following the target stimulus. In this training, rats discriminated three consonant tasks as 'dad' vs 'bad,' 'dad' vs 'sad,' 'dad' vs 'tad' (/d/vs/b/, /d/vs/s/and/d/vs/t/) and three vowel tasks as 'dad' vs 'dud,' 'dad' vs 'deed,' and 'dad' vs 'dood' (/æ/vs/ʌ/, /æ/vs/i/,/æ/vs/u/). Once the rat performed the discrimination for 20 sessions (10 days) on undegraded target and undegraded non-target sounds, they were introduced to the degraded speech sounds.

Introduction to the degraded speech sounds began with a brief period where the rats listened to all ten versions of the target sound (undegraded 'dad' plus nine degraded 'dad's) and the six undegraded non-target sounds. The rat received a sugar pellet reward if the lever was pressed within 3 s following the presentation of any version of the target stimulus. Once the rats performed the above task with a $d' \geq 1.5$ (which took ~3.5 days), they started discriminating the degraded target sounds (all versions of 'dad') from degraded non-target sounds (all versions of 'bad, 'sad,' 'tad,' 'dud,' 'deed,' and 'dood'). The sounds were presented in a randomized block design where a given block contained one level of degraded speech sounds. This created a total of ten blocks (one undegraded plus nine degraded levels). For a given 1-h session, a single block was played. Presentation of blocks was pseudo-randomized so that the rats performed a complete cycle of the ten blocks presented in random order and then started on another cycle. Within each cycle, the ten blocks were presented in random order and was carefully counterbalanced. The training continued until the rats completed eight to ten sessions for all the blocks (~10 to 12 weeks). The two measures including the brief training session of vocoder target training of 3.5 days, and the counterbalanced presentations of degradation levels was designed to eliminate any effect of priming caused by the initial training on undegraded speech sounds. Within a block (a single 1-h session), the rats discriminated the target sound (dad) against the six non-target stimuli ('bad,' 'sad,' 'tad,' 'dud,' 'deed,' and 'dood'), all degraded to a certain level. The stimuli were presented randomly and also interleaved

with silent catch trials. Performance of the rat was monitored using the *d'* value as well as the percent correct discriminations (using the correct lever press and correct rejection rates). The latter measure allowed us to compare rat behavior directly with human performance.

Neural recordings and neural data analysis

We recorded neural responses from 120 multi-unit sites of A1 from experimentally naive, anesthetized, female Sprague–Dawley rats (n=5). Rats were anesthetized with pentobarbital (50 mg/kg), and received supplemental dilute pentobarbital (8 mg/ml) every ½ to 1 h, subcutaneously, along with a 1:1 mixture of dextrose and Ringers lactate. The supplemental dosage was adjusted as needed to maintain areflexia. Heart rate and body temperature were monitored throughout the experiment. A1 responses were recorded Parylene-coated tungsten micro-electrodes (1–2 MΩ, FHC). Four electrodes were lowered simultaneously to 600–700 μm below the surface of the right cortex corresponding to layer IV. At each recording site, 25 ms tones were played at 90 logarithmically spaced (1–47 kHz) frequencies at 16 intensities (0–75 dB SPL), to determine the frequency-intensity tuning curves of each site. Speech stimuli consisting of 63 degraded speech sounds and the seven undegraded speech sounds were presented after the tones. Each stimulus was played 20 times per site, in random order. The stimuli were presented every 2,300 ms at 65 dB SPL from a calibrated speaker mounted approximately 10 cm from the base of the left pinna. Stimulus generation, data acquisition, and spike sorting was performed with Tucker–Davis hardware (RP2.1 and RX5) and software (Brainware). Protocols and recording procedures were approved by the University of Texas at Dallas Institutional Animal Care and Use Committee.

Each electrode penetration site was marked on a photograph of the surgically exposed temporal cortex, using cortical blood vessels as landmarks. As in earlier studies (Kilgard and Merzenich 1998), Voroni tessellation was used to transform the discretely sampled surface into a continuous map using the assumption that each point on the map has the response characteristics of the nearest recording site (Fig. 2A). A1 sites were identified based on latency and topography. Tuning curve analysis of each penetration determined the characteristic frequency of each site (Fig. 2B, C).

Neural similarity between speech sounds was computed using Euclidean distance. The Euclidean distance between a given neurogram pair is the square root of the sum of the squared differences between the firing rates for each bin for each recording site. Each neurogram consisted of responses from each of the 120 A1 recording sites on the y axis ordered from low to high characteristic frequency. Neurogram response at each site consisted of the average of the 20 repetitions. Consonant neurograms were constructed from the first 40 ms speech-evoked responses for 'dad,' 'bad,' 'sad,' and 'tad', beginning from the point where neural activity exceeded the spontaneous firing rate by three standard deviations. The two bin sizes used in consonant analysis consisted of 1 ms bins (providing 1 ms precise spike timing information) and 40 ms single bin (providing the total spike count). Vowel neurograms were constructed from the first 400 ms of 'dad,' 'dud,' 'deed,' and 'dood'. The consonant onset of each consonant–vowel–consonant (CVC) syllable was considered as the stimulus onset. The two bin sizes used in vowel analysis consisted of 1 ms bins (providing 1 ms precise spike timing information) and 400 ms single bin (providing the total spike count).

A peristimulus time histogram (PSTH) classifier (Foffani and Moxon 2004; Engineer et al. 2008; Perez et al. 2012) was used to calculate the neural discrimination between speech stimuli in units of percent correct. The classifier randomly picked five sites (each having 20 sweeps) at a time and made a template from 19 sweeps (the sweep being considered was excluded) for each speech sound. The sweep under consideration was then compared with these templates using the Euclidean distance to determine how similar it is to each template. The classifier determined that the response to single sweep was generated by the sound whose template was closest in Euclidean distance. This model assumed that the average templates are analogous to the neural memory of the different sounds. For consonant analysis, the classifier used the first 40 ms of the speech evoked responses of 'dad,' 'bad,' 'sad,' and 'tad', beginning from the point when neural activity exceeded the spontaneous firing rate by three standard deviations. For vowel analysis, the classifier used the first 400 ms of the speech-evoked activity of 'dad,' 'dud,' 'deed,' and 'dood', beginning from the point where neural activity of the initial consonant exceeded the spontaneous firing rate by three standard deviations. Pearson's correlation coefficient was used to examine the relationship between neural and behavioral discrimination.

RESULTS

Rats, like humans, can reliably discriminate noise vocoded speech

We trained nine rats to discriminate noise vocoded speech sounds (CVC syllables) including three conso-

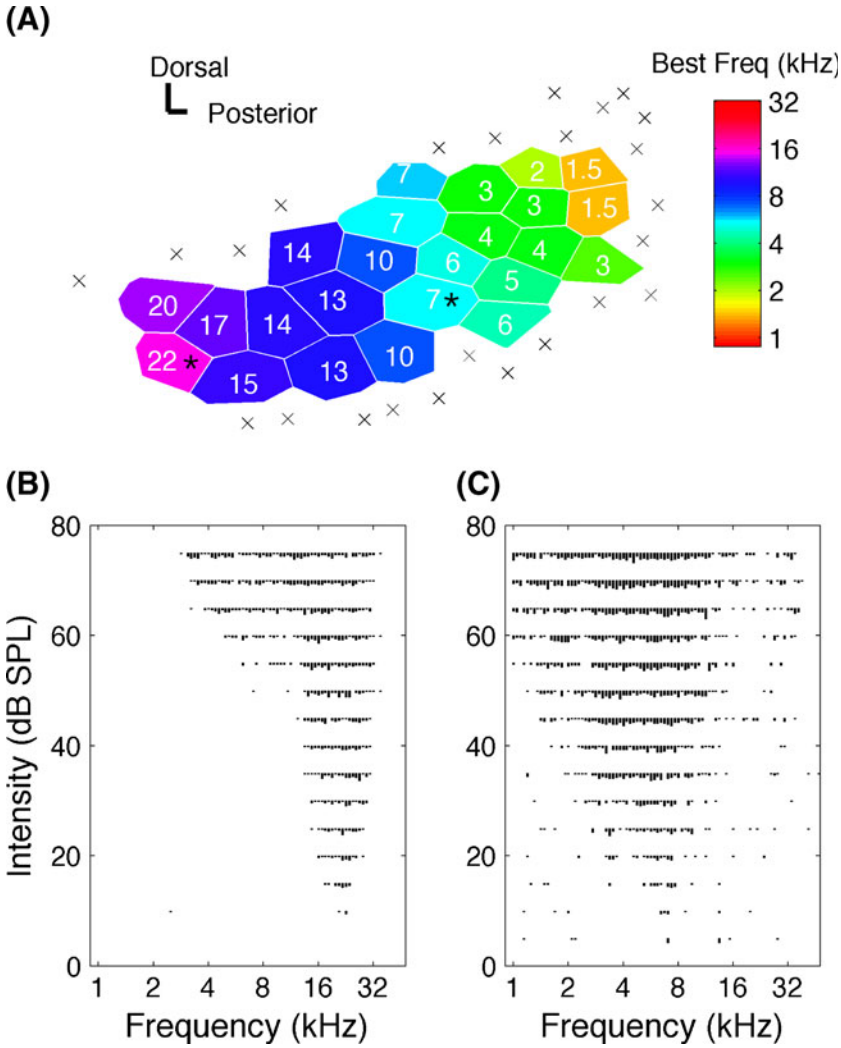
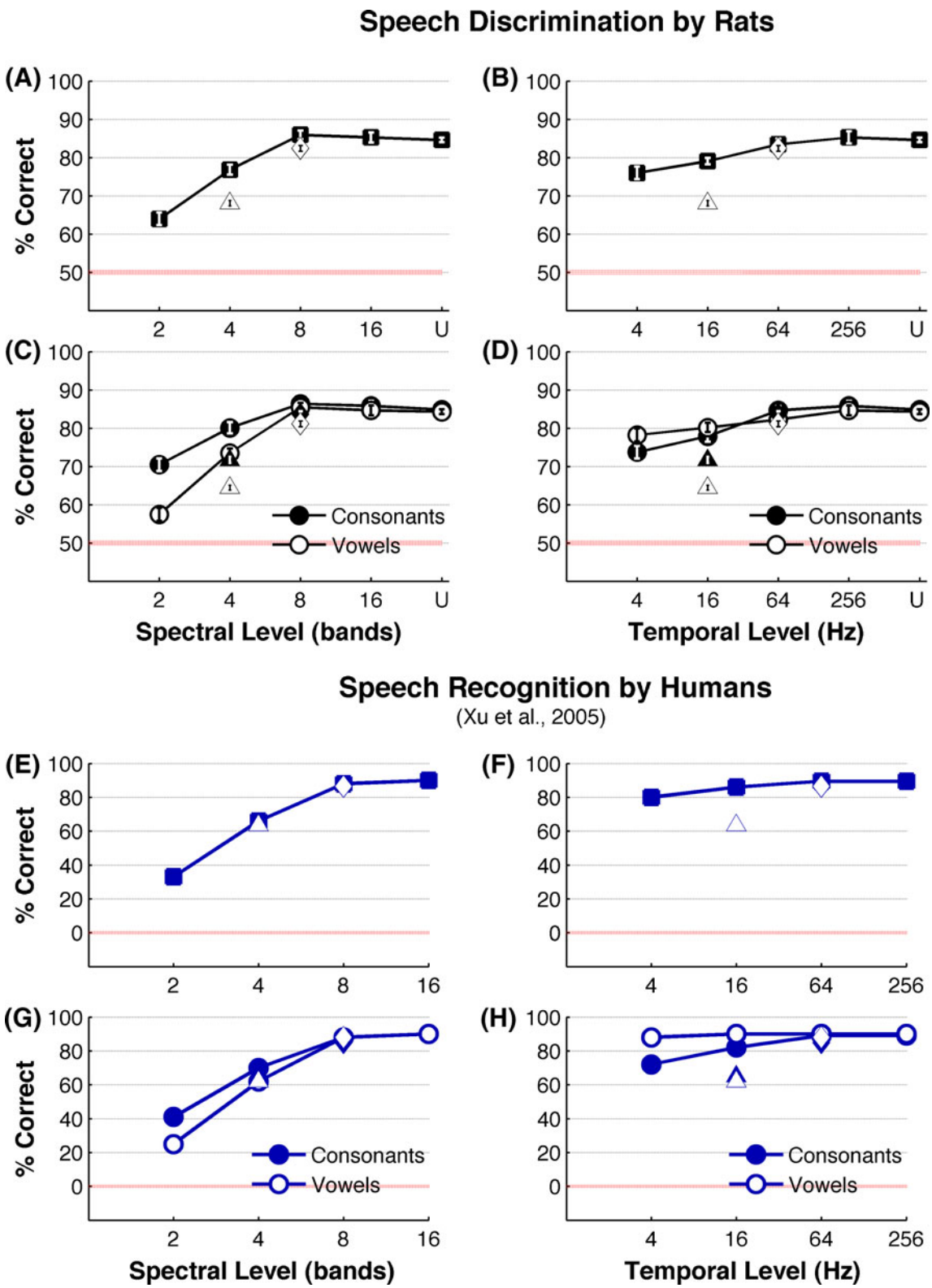


FIG. 2. An example A1 map and tuning curve responses. **A** An example map of auditory cortex in a naïve anesthetized recording experiment. Each *polygon* (Voronoi tessellation) represents a single electrode penetration. *Color* of each polygon represents the value of the characteristic frequency. Sites that did not meet the definition of A1 are indicated by a *multiplication sign*. Scale *bars* indicate a distance of 0.125 mm. The responses of the sites marked by *stars* are used to illustrate tuning curve responses below. **B** Representative tuning curves illustrating the frequency–intensity response of a site tuned to 22 and **C** 7 KHz.

nant pairs (/d/vs/b/, /d/vs/s/, /d/vs/t/) and three vowel pairs (/dad/vs/dud/, /dad/vs/deed/, /dad/vs/dood/). We recorded neural activity patterns for the same stimuli from 120 A1 sites in five naïve rats under anesthesia. The noise vocoder degraded the spectral and temporal content of the stimuli defining nine levels of degradations (Fig. 1, and see "Methods").

Rats were able to accurately discriminate speech sounds that were spectrally or temporally degraded. Rats correctly discriminated undegraded speech sounds on 85±1 % of trials (Fig. 3A) (corresponding d'=2.5±0.04) and were able to achieve the same accuracy with eight spectral bands (86±3 %, P>0.05, Tukey's HSD; corresponding d'=2.5±0.14) (Fig. 3A). Even with the spectral degradation as low as four bands, rats discriminated speech sounds at 77±1 % correct, which is well above the chance performance (50 %) (Fig. 3A). More severe degradation of spectral content down to two spectral bands reduced the rats' accuracy to 64±3 % (Fig. 3A). Across all levels of temporal degradation, rats continued to discriminate speech with high accuracy (Fig. 3B). Rats were able to achieve the same degree of performance with 64 Hz low-pass filtered speech (83±1 %, P>0.05, Tukey's HSD) compared with undegraded speech (Fig. 3B). Even at the highest level of temporal degradation (4 Hz low-pass), performance remained well above chance behavior (76±4 %, Fig. 3B) (corresponding d'=1.7±0.14). Discrimination of noise vocoded speech by rats in our results are remarkably similar to noise vocoded speech discrimination by human listeners who also reach their asymptotic performance at eight spectral bands and continue to show robust behavioral perception with the temporal envelope low-pass filtered at or above 4 Hz (Shannon et al. 1995; Xu et al. 2005) (Fig. 3E, F). These findings show that rats, like humans, can reliably discriminate speech sounds without the fine spectral and temporal details that are normally contained in the signal.

Human psychophysical studies have shown that spectral degradation and temporal degradation differently affect vowel recognition and consonant



recognition. Vowel recognition is more sensitive to spectral degradation, while consonant recognition is more sensitive to temporal degradation (Drullman et al. 1994; Shannon et al. 1995; Xu et al. 2005; Nie et al. 2006). Consistent with these patterns, vowel discrimination by rats was significantly affected by spectral degradation, compared with consonant discrimination (Fig. 3C; $F(1,8)=77.74$, $MSE=8.52$, $P<0.0001$).

FIG. 3. Behavioral discrimination of degraded speech sounds. **A** Percent correct discrimination of speech (average across all six tasks which includes /d/ vs /b/, /d/ vs /s/, /d/ vs /t/, /dad/ vs /dud/, /dad/ vs /deed/, /dad/ vs dood/) by rats (n=9) with spectral degradation. The *squares* indicate stimuli that are low-pass-filtered at 256 Hz, *diamond* at 64 Hz, and the *triangle* at 16 Hz. **B** Percent correct discrimination by rats (average across all six tasks which includes /d/ vs /b/, /d/ vs /s/, /d/ vs /t/, /dad/ vs /dud/, /dad/ vs /deed/, /dad/ vs /dood/) with temporal degradation. The *squares* indicate stimuli that are bandpass-filtered with 16 channels, *diamond* with eight channels, and the *triangle* with four channels. **C** Percent correct discrimination of consonant tasks (average of thee consonant tasks which includes /d/ vs /b/, /d/ vs /s/, /d/ vs /t/) and vowel tasks (average of three vowel tasks which includes /dad/ vs /dud/, /dad/ vs /deed/, /dad/ vs dood/) across levels of spectral degradation. The *circles* indicate stimuli that are low-pass-filtered at 256 Hz, *diamonds* at 64 Hz, and the *triangles* at 16 Hz. **D** Percent correct discrimination for consonants and vowels across levels of temporal degradation. The *circles* indicate stimuli that are bandpass-filtered with 16 channels, *diamonds* with eight channels, and the *triangles* with four channels. For **A** through **D**, error *bars* indicate SE. Note that most error *bars* of **A** through **D** appear smaller than the *symbols*. In all six tasks, 'dad' was the target stimulus. The rats correctly pressed the lever for a food reward for 'dad' and correctly rejected the lever for non-target stimuli. **E** Speech recognition performance by humans under spectral degradation. The *squares* indicate stimuli that are low-pass-filtered at 256 Hz, *diamond* at 64 Hz, and the *triangle* at 16 Hz. **F** Speech discrimination performance by humans under temporal degradation. The *squares* indicate stimuli that are bandpass-filtered with 16 channels, *diamond* with eight channels, and the *triangle* with four channels. **G** Recognition of vowels and consonants by humans under spectral degradation. The *circles* indicate stimuli that are low-pass-filtered at 256 Hz, *diamonds* at 64 Hz, and the *triangles* at 16 Hz. Recognition of vowels and consonants by humans under temporal degradation. The *circles* indicate stimuli that are bandpass-filtered with 16 channels, *diamonds* with eight channels, and the *triangles* with four channels. The *red line* marks the chance performance. Note that human study has the chance performance at 0 %, and the current study has the chance performance at 50 %. The data for **E** through **H** are from Xu et al. 2005. *U* = undegraded.

For example, reducing the spectral frequency bands from 16 to 2 reduced the vowel recognition by 32 % (from 85±1 % to 57±2 %) but only reduced the consonant recognition by 18 % (from 86±1 % to 70± 2 %). Finer resolution of F1 and F2 representation of vowels as a result of increased spectral number of bands has been suggested as a likely explanation for the dominant effect of spectral degradation on vowel recognition (Peterson and Barney 1952; Xu et al. 2005). Temporal degradations below 64 Hz affected consonant recognition by rats significantly more than it affected vowel recognition (Fig. 3D). For example, when the temporal envelope is low-pass-filtered at 16 or at 4 Hz, while maintaining the spectral number of bands at 16, the performance accuracies of consonant recognition (78±4 % and 74±5 %, respectively, $F(1,32)=92.7$, $P<0.01$) were significantly below that of vowels(80±3 %, 78±5 %, respectively, $F(1,32)=23.39$, $P<0.01$) (Fig. 3D). These patterns of behavioral recognition of spectrally and temporally degraded consonants and vowels by rats are remarkably similar to those reported in human psychophysical studies (Xu et al. 2005) (Fig. 2G, H). Our results indicate that, like in humans, discrimination of vowels by rats is more sensitive to spectral degradation while discrimination of consonants is more sensitive to temporal degradation.

Neural differences between consonants remain robust with spectral and temporal degradation

Multiunit activity was recorded from 120 A1 sites in response to each of the behaviorally tested speech sounds. The spatiotemporal activity pattern evoked by each sound was illustrated using neurograms, based on PSTH of each A1 site ordered by its characteristic frequency. The neurograms contained precise spike timing information of 1-ms bins (Fig. 4A–D). As described in previous experiments, each consonant evoked a different firing pattern in A1 neurons (Engineer et al. 2008; Shetake et al. 2011). For example, /d/ evoked high-frequency tuned neurons early in their response followed by low-frequency tuned neurons. In contrast, /b/ evoked an early low-frequency tuned neuronal response followed by a late high-frequency response (Fig. 4A). The response patterns for the undegraded /d/, /b/, /s/, and /t/ (Fig. 4A) were consistent with the observations reported in earlier studies using normal speech (Steinschneider et al. 1999, 2003, 2005; Engineer et al. 2008). For example, the stop consonants (/d/, /b/, /t/) generated sharper onset activity compared with the fricative /s/ (Fig. 4A). These results confirm earlier observations that each consonant generate a unique spatiotemporal activity pattern in A1.

The unique patterns of activity generated by different consonants remained distinguishable, with spectral and temporal degradation of the sounds. For example, when the signal was degraded down to 16 bands and low-pass-filtered at 256 Hz (Fig. 4B), or even down to eight spectral bands and low-pass-filtered at 64 Hz (Fig. 4C), the early firing of high-frequency neurons followed by late firing of low-frequency neurons evoked by /d/ still remained clearly identifiable. Similarly, the early low-frequency firing followed by late high-frequency firing pattern of /b/ stayed clearly identifiable at 16 bands/256 Hz stage (Fig. 4B) as well as eight bands/64 Hz level (Fig. 4C). With further degradation of spectral and temporal content of the signal, the neural activity patterns became less distinct. For example, when the signals were degraded down to four spectral bands and low-pass-filtered at 16 Hz (4bands/16 Hz), the A1 neurograms of /d/ and /b/ looked very similar (Fig. 4D). These observations suggest that, even after speech signals are bandpass-filtered with few channels (i.e., eight bands) and low-pass-filtered with slow

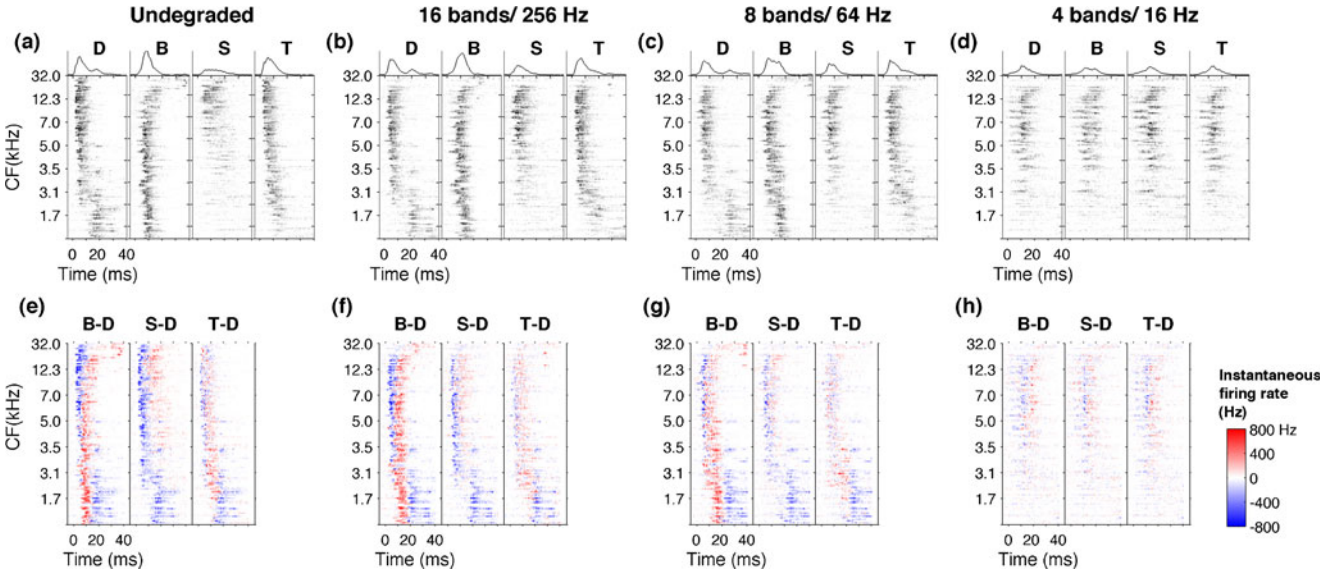


FIG. 4. Neural responses to consonants neurograms depicting the evoked activity over the first 40 ms for the consonants /d/, /b/, /s/, and /t/ of undegraded stimuli (**A**). The unique patterns evoked by each consonant are clearly visible when noise vocoded with 16 bands and 256 Hz (**B**) or eight bands and 64 Hz (**C**). The activity patterns are degraded and become less distinct with more severe degradations like four bands and 16 Hz (**D**). Difference neurograms plotting the neural activity of /d/ subtracted from /b/, /s/, and /t/ for undegraded stimuli (**E**). The neural differences generated in A1 remain clearly visible when sounds are noise vocoded with 16 bands and 256 Hz (**F**) or eight bands and 64 Hz (**G**). With more severe degrees of spectral and temporal degradation neural differences become diminished as shown for four bands and 16 Hz (**H**).

modulations (i.e., 64 Hz), A1 neurons generate significant neural activity patterns.

We created difference neurograms by subtracting the spatiotemporal patterns generated by /d/ from those of /b/, /s/, and/t/(Fig. 3E, F, G, H). At undegraded condition, owing to their own unique firing patterns, the difference plots produced robust residual patterns. For example, the different neurograms for /b/ vs /d/ clearly showed the excess of early low-frequency firing of /b/ compared with /d/ (Fig. 4E, for B-D, the lower red region) and the excess of late high-frequency firing (Fig. 4E, for B-D, the upper red region). When the signals are degraded down to 16 spectral bands and low-pass-filtered at 256 Hz, or even further down to eight bands and 64 Hz, these residual patterns remained clearly identifiable (Fig. 4F, G). With further spectral and temporal degradation of the signal, where the spatiotemporal patterns became less distinct, the difference neurograms showed minimal activity (Fig. 4H). These results show that spatiotemporal activity patterns generated in A1 neurons continue to encode stimulus differences in the absence of fine spectral and temporal details of speech until it fails at severe degrees of signal degradation. We predicted that A1 neural differences generated by degraded consonants could account for the robust behavioral discrimination patterns.

As a measure of neural dissimilarity, we calculated the Euclidean distance between the onset neurograms for /d/ vs /b/, /d/ vs /s/, and /d/ vs /t/ at each signal degradation level. As we predicted, the pattern of consonant discrimination across different levels of degradation was well explained by the neural differences (r^2=0.48, P<0.001; Fig. 5A). We then further quantified the neural discrimination ability in units of percent correct. To accomplish this, we used a neural classifier (Engineer et al. 2008), which uses each individual sweep (of the 20 sweeps) of stimulus presentation (see, "Methods"). The neural classifier performance when using the precise temporal spike timing information (1 ms resolution) was strongly and significantly correlated with the behavioral discrimination of consonants by rats (r^2=0.58, P<0.001, Fig. 5B). As reported in earlier studies (Engineer et al. 2008), we found that spike timing information using analysis bins from 1–20 ms are significantly correlated with behavioral discrimination of consonants. The neural discrimination between consonant pairs remained high across the degradation levels that rats found it easier to discriminate compared with the levels that rats found difficult to discriminate. For example, /d/ vs /b/ was the easiest task for rats across all levels of degradation, except the two lowest temporal stages of 16 and 4 Hz (Fig. 4C, D). The same pattern was seen with the neural classifier discrimination where /d/ vs /b/ achieved the highest percent correct values across all levels of degradation except at 16 and 4 Hz stages (Fig. 5E, F). Similarly, /d/ vs /t/ remained the most difficult task for rats across all levels of degradation, except at 4 Hz

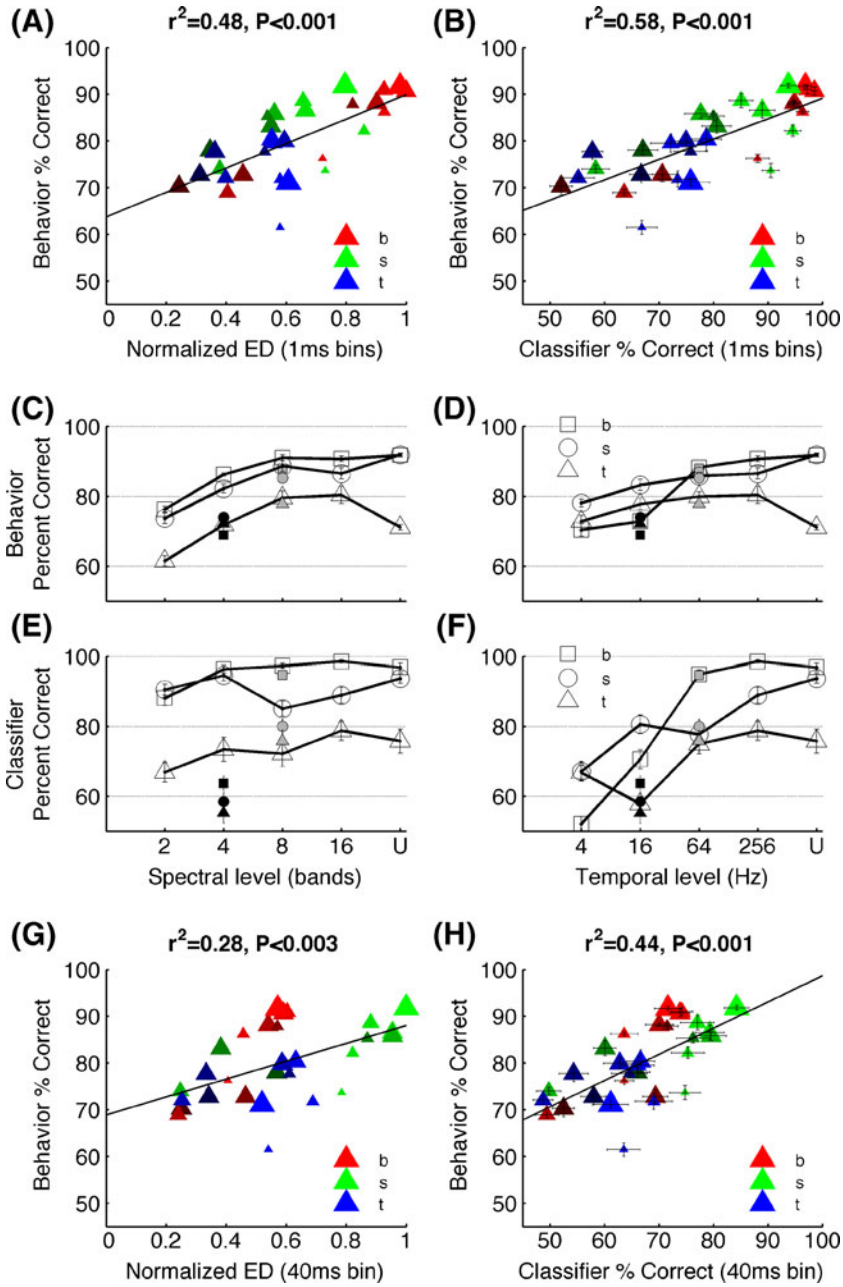


FIG. 5. Neural and behavioral correlates of consonant discrimination behavioral discrimination of consonants is significantly correlated with the neural differences estimated by the Euclidean distances between the onset neurograms which considers the precise spike timing information of 1 ms (**A**) and the neural classifier based using the spike timing information (**B**). The behavioral discrimination patterns observed for the three consonants under different levels of spectral (**C**) and temporal (**D**) degradation matched the neural classifier discrimination patterns (**E**) and (**F**). The *open symbols* in (**C**) and (**E**) indicate stimuli low-pass-filtered at 256 Hz, the *gray symbols* at 64 Hz, and the *black symbols* at 16 Hz. The *open symbols* of (**D**) and (**F**) indicate stimuli noise vocoded with 16 bands, the *gray* with 8 bands, and the *black* with 4 bands. For **C** through **F**, error *bars* indicate SE. Note that most error *bars* of **C** and **D** appear smaller than the *symbols*. The behavioral patterns of consonant discrimination was also correlated with the Euclidean distances between neurograms considering the average spike count of the first 40 ms onset response (disregarding the spike timing information) (**G**), and the neural classifier based on average spike count of the first 40 ms response (**H**). Note that inclusion of spike timing (**A** and **B**) explained more variability of behavioral performance than elimination of spike timing (**G** and **H**). For **A**, **B**, **G**, and **H**, the *sizes* of the symbols indicate spectral resolution where the largest shows the undegraded consonants and the *brightness* of the symbols indicates temporal resolution where the brightest shows undegraded consonants. *U* = undegraded.

(Fig. 5C, D) and for the neural classifier (Fig. 4E, F). These findings indicate that the spatiotemporal activity patterns generated in A1 neurons sufficiently encode neural dissimilarities under spectral and temporal degradation of speech and suggest that behavioral discrimination of speech remains robust as long as A1 neural activity patterns remains distinct.

Absence of precise spike timing information decreased the strength of neural and behavioral correlates of noise vocoded consonant discrimination. Neural differences calculated from Euclidean distance between the spike count during the first 40 ms discarding the precise spike timing information only explained 28 % variance of the behavioral discrimination (r^2=0.28, P<0.003; Fig. 5G). The neural classifier based on spike count over the 40 ms window explained 44 % of the behavioral variance of consonant discrimination (r^2=0.44, P<0.001; Fig. 5H). These results highlight the importance of temporal patterns of neural activity when encoding stimuli with rapidly varying spectral details such as consonants. This finding is consistent with previous neurophysiological and behavioral studies showing that spike timing information with temporal precision strongly contributes to consonant representation in the auditory cortex (Engineer et al. 2008).

Neural differences between vowels remain robust with spectral and temporal degradation

In contrast to rapid spectral changes of consonants, vowels represent relatively stable spectral patterns. Previous neurophysiological studies have reported that different vowel sounds generate distinct spatial activity patterns in the central auditory system (Sachs and Young 1979; Delgutte and Kiang 1984; Ohl and Scheich 1997; Versnel and Shamma 1998). These observations suggest that the number of spikes (spike count) averaged over stimulus duration, as a function of characteristic frequency could be used to discriminate vowel sounds. We examined the neural activity patterns generated in A1 by each syllable with a different vowel (/dad/, /dud/, /deed/, and/dood/) over a 400-ms window, which contains most of the vowel component (Fig. 6A, B, C, D). The spike count profiles for each vowel were created, depicting the number of spikes evoked by A1 neurons arranged according to their characteristic frequency (Fig. 6E, F, G, H). Syllables with different vowels generated a distinct spatial activity pattern in A1. For example, in the undegraded condition, /dad/ evoked a greater number of spikes in low-frequency tuned neurons and a lesser number of spikes in high-frequency tuned neurons (Fig. 6E). In contrast, /dood/ evoked fewer spikes in low-frequency tuned neurons and more spikes in high-frequency tuned neurons (Fig. 6E). These patterns remained clearly identifiable when the stimuli were degraded with 16 spectral bands and low-pass-filtered at 256 Hz and even further down to eight bands and 64 Hz (Fig. 6F, G). With more severe spectral and temporal degradation of the signal at four bands and 16 Hz, the spatial patterns between /dad/, /dud/, /deed/, and /dood/became more similar (Fig. 6H). These results suggest that spike count profiles of A1 neurons produce significant neural activity under spectral and temporal degradation of the signal.

The difference plots created by subtracting the /dad/ spike count profile, from those of /dud/, /deed/, and /dood/illustrated the relative differences between vowel pairs (Fig. 6I, J, K, L). For example, subtraction of /dad/ spike count from that of /dood/ produced a response which lies clearly away from the zero line (Fig. 5I, /dood/ minus /dad / plotted in red) (the dashed line corresponds to /dad/ minus /dad/ or zero difference, Fig. 6I). In contrast, subtraction of /dad/ from /dud/ produced a response which lies closer to the zero difference line (Fig. 6I, /dud/ minus /dad/ plotted in blue). The distinctiveness of spike count patterns produced by vowels remained resistant to spectral and temporal

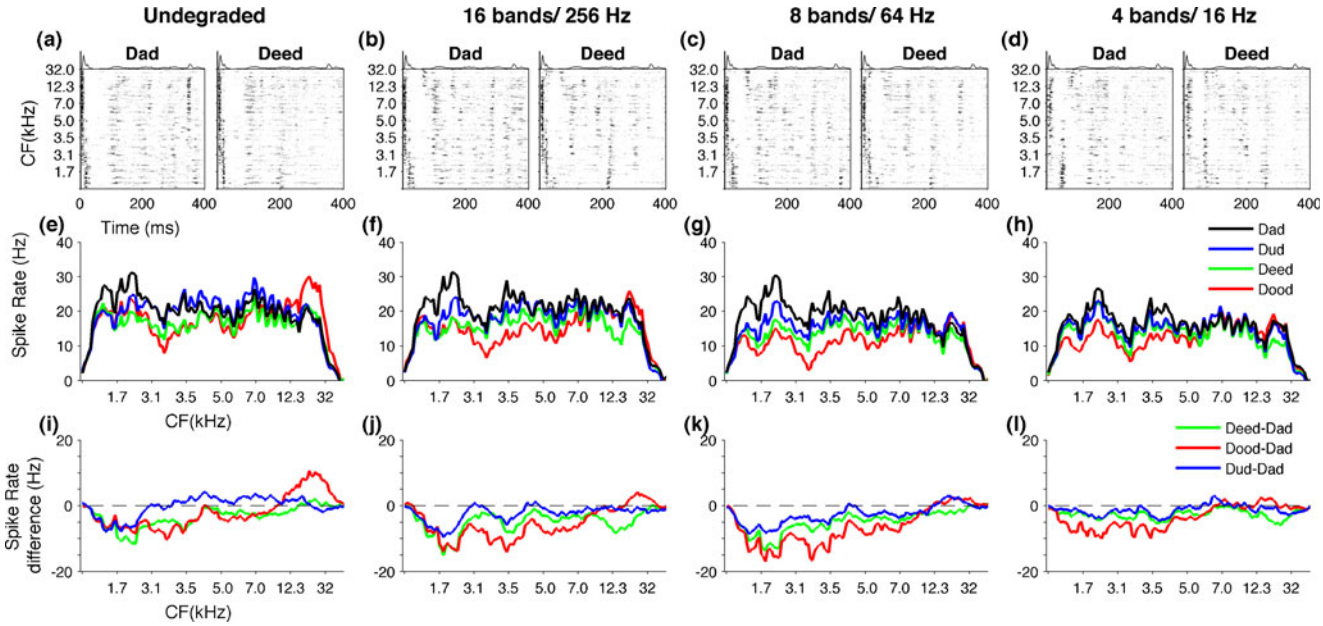


FIG. 6. Neural responses to vowels neurograms depicting the evoked activity over the first 400 ms for /dad/ and /deed/ of undegraded stimuli (**A**), 16 bands and 256 Hz (**B**), eight bands and 64 Hz (**C**), and four bands and 16 Hz (**D**). The spike sound profiles (spatial patterns), plotting the number of spikes fired for the 400 ms duration by A1 neurons arranged according to their characteristic frequency. Each syllable with a different vowel (/dad/, /dud/, /deed/, /dood/) generates a unique spatial pattern in A1 at undegraded condition (**E**). The spatial patterns remain distinct when stimuli are noise vocoded with 16 bands and 256 Hz (**F**) or eight bands and 64 Hz. The patterns become less distinct with more severe degradations like four bands and 16 Hz. The neural differences generated in A1 are plotted by subtracting the spatial pattern of /dad/ from /dud/, /deed/, and /dood/ at undegraded condition (**I**).The zero line indicates /dad/ minus /dad/. The neural differences between spatial patterns remain clearly visible when sounds are noise vocoded with 16 bands and 256 Hz (**E**) or eight bands and 64 Hz (**F**). With more severe degrees of spectral and temporal degradation neural differences become closer to zero line indicating reduced neural dissimilarity, as shown for four bands and 16 Hz (**G**).

degradation of the signal. For example, the residual activity patterns plotted by subtracting the spike count of /dad/ from other three vowels remained clearly away from the zero line at 16 bands/256 Hz stage as well as eight bands/64 Hz stage (Fig. 6J, K). With more severe spectral and temporal degradation of the signal, the spatial patterns became less distinct. For example, when the signal was degraded down to four spectral bands and low-pass-filtered at 16 Hz, the residual activity patterns created by /dud/ minus /dad/, /deed/ minus /dad/, and /dood/ minus /dad/ got closer to zero line (Fig. 6L). These results show that neural dissimilarities generated by spatial patterns of A1 for vowels continue to remain robust when spectral information is limited to few bands (i.e., eight bands) and temporal envelope low-pass-filtered at slow modulations (i.e., 64 Hz).

The neural dissimilarity quantified by spike count profile for each vowel accurately predicted the behavioral discrimination pattern of noise vocoded vowels. The Euclidean distance between the number of spikes evoked in each site for /dad/ vs /dud/, /dad/ vs /deed/, and /dad/ vs /dud/ across different levels of degradation was significantly correlated with the behavioral discrimination of vowels by rats (r^2=0.52, P<0.001, Fig. 7A). Neural discrimina-

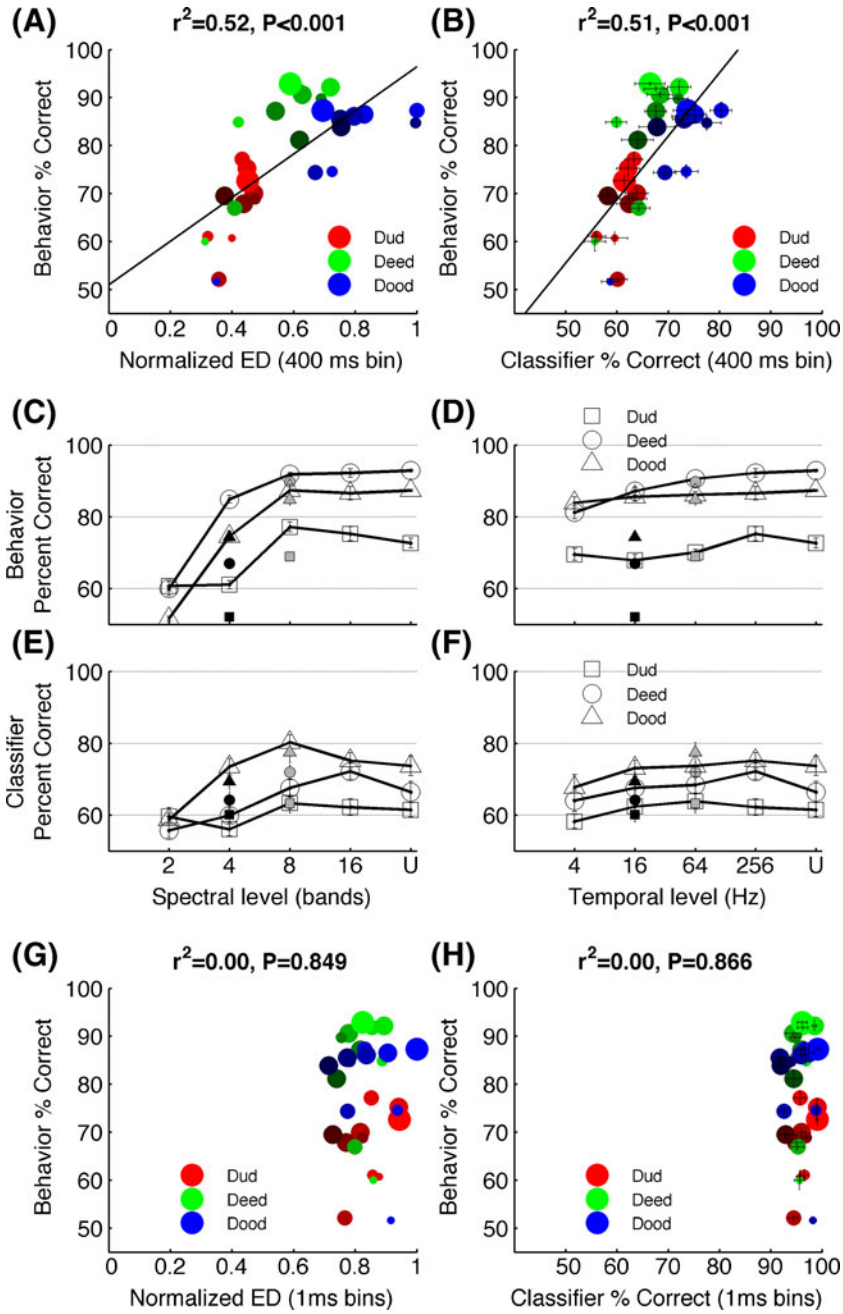


FIG. 7. Neural and behavioral correlates of vowel discrimination behavioral discrimination of vowels is significantly correlated with the neural differences estimated by the Euclidean distances between the spatial patterns generated in A1 (**A**) and the neural classifier based using the spike rate of the 400 ms duration (**B**). The behavioral discrimination patterns observed for the three vowel tasks under different levels of spectral (**C**) and temporal (**D**) degradation closely matched the neural classifier discrimination patterns (**E**) and (**F**). The *open symbols* in (**C**) and (**E**) indicate stimuli low-pass-filtered at 256 Hz, the *gray symbols* at 64 Hz, and the *black symbols* at 16 Hz. The *open symbols* of (**D**) and (**F**) indicate stimuli noise vocoded with 16 bands, the *gray* with 8 bands, and the *black* with 4 bands. For **C** through **F**, error *bars* indicate SE. Note that most error *bars* of **C** and **D** appear smaller than the symbols. The behavioral patterns of vowel discrimination was not correlated with the Euclidean distances between neurograms considering the precise spike timing information (**G**), or the neural classifier based on spike timing (**H**). For **A**, **B**, **G**, and **H**, the *sizes* of the *symbols* indicate spectral resolution where the *largest* shows the undegraded vowels and the *brightness* of the symbols indicates temporal resolution where the brightest shows undegraded vowels. *U* = undegraded.

tion in units of percent correct was measured by using the neural classifier. The classifier used the spike counts of each individual sweeps of stimulus presentation (see "Methods"). The neural classifier discrimination based on average spike count was strongly correlated with behavioral discrimination of noise vocoded vowels bay rats (r^2=0.51, P<0.001; Fig. 7B). The neural discrimination between vowel pairs remained high across the degradation levels that rats found it easier to discriminate compared with the levels that rats found difficult to discriminate. For example, the rats found /dad/ vs /dud/ the most difficult task across all levels of degradation except the lowest spectral stage (two bands) (Fig. 7C, D). Similarly, the neural classifier achieved the lowest percent correct values for /dad/ vs /dud/ across all the degradation levels except two bands stage. Rats found /dad/ vs /dood/ and /dad/ vs /deed/ equally easy across most of the degradation levels. The neural classifier also discriminated /dad/ vs /dood/ and /dad/ vs /deed/ equally better across most temporal stages, although across some of the spectral stages (four, eight bands), it performed better on /dad/ vs /dood/ (Fig. 7E, F). Collectively, these findings indicate that neural dissimilarities encoded in spatial patterns generated by vowels in A1 could account for the behavioral discrimination of vowels under spectral and temporal degradation.

Spatiotemporal neural activity patterns containing precise timing of action potentials failed to explain behavioral discrimination of vowels. Inclusion of precise spike timing information over the 400-ms window created a ceiling effect placing the neural discriminability at a level far exceeding the behavioral ability of vowel discrimination. The neural dissimilarity computed by Euclidean distance of 1-ms precise spike timing patterns between /dad/ vs /dud/, /dad/ vs /deed/ and /dad/ vs /dood/ was not correlated with behavioral discrimination vowels (r^2=0.0001, P= 0.849; Fig. 7G). The neural classifier, when provided with spike timing information in a 400-ms window, also failed to produce significant neural correlates with behavioral discrimination of vowels (r^2=0.0001, P= 0.866; Fig. 7H). The neural classifier discrimination based on the first 40 ms spike timing information was also not correlated with behavioral discrimination of vowels (r^2=0.01, P=0.521). These results suggest that the potential neural mechanisms underlying vowel perception operate on spatial patterns as opposed to the spatiotemporal patterns of consonants.

DISCUSSION

Summary

In this study, we examined the behavioral discrimination of noise vocoded consonants and vowels by rats and recorded neural activity patterns generated in rat A1 for the same stimuli. Rats, like humans, discriminated noise vocoded speech processed with 8 to 16 spectral bands and normal speech with similar accuracy. Spatiotemporal activity patterns generated in rat A1 by speech sounds noise vocoded with 8 to 16 spectral bands and envelope modulations below 256 Hz produced neurograms similar to undegraded speech. Behavioral discrimination ability was correlated with the neural discrimination based on neural activity patterns generated in A1. The behavioral discrimination of speech remained accurate as long as the degraded stimuli generated robust neural differences in A1 neurons.

The rats' ability to discriminate noise vocoded speech reported in this study is comparable to human psychophysical findings in several aspects. First, the degree of tolerance rats showed in their behavioral discrimination of noise vocoded sounds was very similar to human data. Both humans and rats performed speech discrimination above 75 % correct when the spectral information was limited to four bands or above and when the temporal envelope information was low-pass-filtered at 4 Hz or above (Shannon et al. 1995; Loizou et al. 1999; Xu et al. 2005). Second, the level of degradation at which behavior reached a performance maximum was the same in rats and humans. Speech discrimination in quiet by human subjects either listening through a cochlear implant or to noise vocoded speech, reaches a peak at eight spectral bands and increasing the spectral bands more than eight does not improve the performance (Fishman et al. 1997; Loizou et al. 1999; Xu et al. 2005). These observations are essentially replicated in our results, where rats' performance reached an asymptote with eight spectral bands. Third, spectral degradation affected vowel discrimination by rats, significantly more than consonants. Recognition of noise vocoded vowels by normal hearing people also reported the same pattern with spectral degradation (Shannon et al. 1995; Xu et al. 2005). Fourth, temporal degradation affected consonant discrimination by rats, significantly more than vowels, and the human behavior on noise vocoded speech also showed similar patterns (Drullman et al. 1994; Xu et al. 2005). These findings suggest that general auditory processing mechanisms underlying speech discrimination are likely shared by humans and animals.

A noise vocoder only transmits limited spectral and temporal details of original speech. For example, a noise vocoder with eight bandpass filters provides a coarser frequency analysis of speech, compared with normal cochlea. The formant structure created by vocal tract resonances is significantly altered in vocoded speech. Formant transitions and pitch are removed (Shannon et al. 1995). Our results show that the neural differences generated in A1 by noise

vocoded stimuli processed with eight spectral bands and modulations below 256 Hz are indistinguishable from differences generated by undegraded stimuli. This finding suggests that central auditory representation of speech may not be highly dependent on the detailed frequency assessment provided by physiological cochlea. Our result is consistent with recent human imaging evidence that showed superior and middle temporal gyri exhibiting equally strong activations for normal speech and intelligible noise vocoded speech (Scott et al. 2000; Davis and Johnsrude 2003; Scott et al. 2006). Based on these results, it is reasonable to conclude that a cochlear implant with eight spectral channels and envelope modulations up to 256 Hz would successfully recreate the degree of neural differences produced by normal speech in A1 (Fishman et al. 1997; Hedrick and Carney 1997; Kiefer et al. 2000; Loizou et al. 2000; Loebach and Wickesberg 2006). Successful generation of distinct auditory neural patterns similar to those generated by normal speech provides a potential neural substrate for processing of auditory neural responses generated by cochlear implants.

In post-linguistically deaf cochlear implant recipients the speech perception performance gradually improves over the first few months after implantation (Hughes et al. 2001; Krueger et al. 2008). This adaptation is generally attributed to the slow process of learning how to interpret inputs from the cochlear implant (Tyler and Summerfield 1996). This process appears to involve substantial plasticity in the central auditory system (Kral and Tillein 2006; Sharma et al. 2007; Beitel et al. 2011). It is not clear whether this plasticity is required either to learn the patterns generated by cochlear implant, or simply because the auditory system has been deprived of inputs for such a long period of time (Moore and Shannon 2009). Our results indicate that spectral and temporal degradation of auditory inputs has a surprisingly small impact on the activity patterns evoked by speech. The fact that both human and rat listeners can rapidly learn to accurately discriminate noise vocoded speech demonstrates that large-scale plasticity is not required to process degraded speech (Shannon et al. 1995; Xu et al. 2005). Thus, a most likely explanation for the long period of adaptation following cochlear implantation is that a slow restoration of the normal operations of an auditory system long deprived of auditory inputs. Adaptation to the pronounced differences between the neural activation evoked by electrical stimulation and by physiological firing of inner hair cells consists of other significant processes that may take place during this period (Kral et al. 2006). Deafness is known to cause increased spontaneous activity, tonotopic distortion, homeostatic changes, and strengthening of non-auditory inputs in auditory cortex (Kral 2007; Rao et al. 2010). It is likely that the time a cochlear implant recipient spends adapting to the device is a period of recovery from sensory deprivation rather than a protracted period required for learning to compensate for degraded speech. These observations suggest that the process of adaptation might be accelerated to achieve higher levels of asymptotic performance by focusing on strategies designed to restore normal operation to the auditory system. For example, patients might be better served by listening to scales, rather than speech sounds which simultaneously activate many channels (Gatehouse 1992).

Sensory systems encode stimulus characteristics in multiple time scales distributed from milliseconds to few hundred seconds. Precise spike timing of action potentials as well as spike count averaged over long integration windows have been shown to transmit information about perceptually important stimulus characteristics in different sensory modalities including vision, audition, and somatosensation. Psychophysical and neurophysiological studies of vibrotactile and visual motion processing have revealed that the average firing rate of cortical neurons are well correlated with perception (Britten et al. 1992; Romo and Salinas 2003; Liu and Newsome 2005). The additional information encoded in precise timing of action potentials was reportedly not useful in these tasks. On the other hand, the relative timing of neural activity was required to represent complex stimuli in the auditory system (Engineer et al. 2008). The behavioral perception of consonant recognition tasks were only correlated when the precise spike timing information was included in the neural discrimination and were not correlated with the average spike count. Recent work reviewing these opposite arguments have suggested that sensory information is likely encoded in multiple time scales including both millisecond precision and averaging of spikes (Kumar et al. 2010; Panzeri et al. 2010). Our data support this argument by showing that, within the auditory system, dynamic acoustic properties of consonants are better represented by spike timing of A1 neurons, while the steady state properties of vowels are better represented in spike count. These results are in agreement with the neural representation of consonants and vowels observed in previous experiments (Sachs and Young 1979; Delgutte and Kiang 1984; Ohl and Scheich 1997; Versnel and Shamma 1998; Engineer et al. 2008). The choice of temporal resolution used by the neural decoding mechanisms appears to be determined by the properties of stimulus characteristics. These conclusions are also consistent with previous theoretical explanations of the ability of auditory system to encode stimulus properties in multiple timescales in bilateral primary auditory areas, later lateralizing into right and left hemispheres for subsequent stages of processing (Poeppel 2003; Boemio et al. 2005). It has been suggested that such

multiple coding inputs provide complementary rather than competing information in the representation of temporally varying spectral components of complex signals (Panzeri et al. 2010).

Our previous experiments reported that differences created by precise spike timing patterns of A1 neurons by consonants are highly correlated with the behavioral ability to discriminate consonants (Engineer et al. 2008; Shetake et al. 2011). This does not imply that only responses in A1 would be expected to be correlated with behavior. Recent experiments on rat inferior colliculus responses have extended this observation along the central auditory axis by reporting significant neural and behavioral correlates of consonant discrimination (Perez et al. 2012). These findings support the view that temporal coding of central auditory neurons represents acoustic transients of consonants in a continuous neural encoding of 1–10 ms resolution of action potential firing. Since the consonants itself carry millisecond precise acoustic transients, it is possible that the spike timing precision of A1 activity patterns to be totally driven by stimulus properties. In the current study, we removed the fast temporal modulations of the stimuli by low-pass-filtering. Our results demonstrated that noise vocoded consonant recognition was not significantly different from that of normal consonants even after removing envelope modulations above 64 Hz, and that millisecond precise spike timing of A1 neural activity is still highly correlated with this behavior. These results indicate that it is possible to represent the differences between consonants in precise timing of action potential firing, independent of the timing of acoustic transformations. Our results are also in agreement with several theoretical studies showing that temporal coding of neurons have the ability to encode not only the time varying temporal patterns of stimulus characteristics but also static spatial patterns (Buonomano and Merzenich 1995; Buonomano and Merzenich 1999; Mauk and Buonomano 2004). For example, the temporal patterns generated in a model of visual cortex neurons by handwritten numerals potentially represented distinct neural firing patterns for each numeral (Buonomano and Merzenich 1999). Collectively, these findings substantiate the fact that temporal codes are not simply faithful representations of time varying stimulus features but an abstract representation of input signals distributed in action potential timing.

Future studies are needed to evaluate the potential effect of anesthesia and attention on speech sound responses. Earlier studies have reported qualitatively similar response properties in awake and anesthetized preparations of auditory recordings including A1 (Watanabe 1978; Steinschneider et al. 1982; Sinex and Chen 2000; Cunningham et al. 2002; Engineer et al. 2008). Engineer et al. reported that A1 responses recorded in both under anesthesia and in awake rats are well correlated with behavioral discrimination of speech contrasts (Engineer et al. 2008). However, anesthesia has been shown to reduce the maximum rate of click trains that A1 neurons can respond to, so it is likely that anesthesia reduces the A1 response to speech sounds, especially the sustained response to vowel sounds (Anderson et al. 2006; Rennaker et al. 2007). Although human imaging studies have shown little to no effect of attention on primary auditory cortex responses to speech sounds, single unit studies are needed to better understand how attention might shape the cortical representation of speech (Grady et al. 1997; Hugdahl et al. 2003; Christensen et al. 2008). Given the recent reports of A1 coding strategy being modified in noisy environments (Shetake, et al. 2011), it is likely that other forms of stimulus degradations may also modify the analysis windows used to represent speech sounds.

The current study reports the first evidence of noise vocoded speech discrimination in an animal model. Our results demonstrate that speech signals separated into a small number of spectral channels and limited to slow modulations generate similar degree of neural discrimination ability in A1 compared with normal speech. Apart from providing a potential neural mechanism to explain the remarkable success of cochlear implants, our results clarify the level of detail the neural differences of phonemic contrasts are represented in auditory cortex. Future neural and behavioral studies of speech sound processing in animals with cochlear implants are needed to confirm that these devices are capable of fully restoring speech evoked activity patterns in auditory cortex.

ACKNOWLEDGMENTS

The authors would like to thank T. R. Rosenthal, E. M. Renfroe, T. K. Jasti, E. Tran, K. Ram, R. Cheung, H. Shepard, Z. Ghneim, C. Xie, D. Vuppula, T. Nguyen, and R. Joseph for help with behavioral training. We would also like to thank, A. Moller, P. Assmann, E. Tobey, F. G. Zeng, C. Engineer, B. Porter, J. Shetake, and A. Reed for their comments and suggestions on earlier versions of the manuscript. We would also like to thank P.C. Loizou in his assistance in signal processing and for his comments and suggestions on the manuscript. We would also like to thank H. Abdi for his guidance in statistical analyses. This work was supported by Award Numbers R01DC010433 and R15DC006624 from the National Institute on Deafness and Other Communication Disorders.

REFERENCES

ANDERSON SE, KILGARD MP, SLOAN AM, RENNAKER RL (2006) Response to broadband repetitive stimuli in auditory cortex of the unanesthetized rat. Hearing Res 213:107–117

BEITEL RE, VOLLMER M, RAGGIO MW, SCHREINER CE (2011) Behavioral training enhances cortical temporal processing in neonatally deafened juvenile cats. J Neurophysiol 106:944–959

BOEMIO A, FROMM S, BRAUN A, POEPPEL D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395

BRITTEN KH, SHADLEN MN, NEWSOME WT, MOVSHON JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. The J of neurosci: the Off J of the Soc for Neurosci 12:4745–4765

BUONOMANO DV, MERZENICH MM (1995) Temporal information transformed into a spatial code by a neural-network with realistic properties. Science 267:1028–1030

BUONOMANO DV, MERZENICH M (1999) A neural network model of temporal code generation and position-invariant pattern recognition. Neural Comput 11:103–116

CHRISTENSEN TA, ANTONUCCI SM, LOCKWOOD JL, KITTLESON M, PLANTE E (2008) Cortical and subcortical contributions to the attentive processing of speech. Neuroreport 19:1101–1105

CUNNINGHAM J, NICOL T, KING C, ZECKER SG, KRAUS N (2002) Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. Hearing Res 169:97–111

DAVIS MH, JOHNSRUDE IS (2003) Hierarchical processing in spoken language comprehension. J of Neurosci J Soc Neurosci 23:3423–3431

DELGUTTE B, KIANG NY (1984) Speech coding in the auditory nerve: I. Vowel-like sounds. J Acoust Soc Am 75:866–878

DORMAN MF, LOIZOU PC (1997) Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. Am J Otol 18: S113–S114

DORMAN MF, LOIZOU PC, FITZKE J, TU Z (1998) The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. J Acoust Soc Am 104:3583–3585

DRULLMAN R, FESTEN JM, PLOMP R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064

ENGINEER CT, PEREZ CA, CHEN YTH, CARRAWAY RS, REED AC, SHETAKE JA, JAKKAMSETTI V, CHANG KQ, KILGARD MP (2008) Cortical activity patterns predict speech discrimination ability. Nat Neurosci 11:603–608

FISHMAN KE, SHANNON RV, SLATTERY WH (1997) Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. J Speech Lang Hear R 40:1201–1215

FOFFANI G, MOXON KA (2004) PSTH-based classification of sensory stimuli using ensembles of single neurons. J Neurosci Meth 135:107–120

GATEHOUSE S (1992) The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids. J Acoust Soc Am 92:1258–1268

GRADY CL, VAN METER JW, MAISOG JM, PIETRINI P, KRASUSKI J, RAUSCHECKER JP (1997) Attention-related modulation of activity in primary and secondary auditory cortex. Neuroreport 8:2511–2516

GREEN DM, SWETS JA (1989) Signal detection theory and psychophysics. Los Altos, CA: Peninsula Publishing

HEDRICK MS, CARNEY AE (1997) Effect of relative amplitude rind formant transitions on perception of place of articulation by adult listeners with cochlear implants. J Speech Lang Hear R 40:1445–1457

HUGDAHL K, THOMSEN T, ERSLAND L, RIMOL LM, NIEMI J (2003) The effects of attention on speech perception: an fMRI study. Brain Lang 85:37–48

HUGHES ML, VANDER WERFF KR, BROWN CJ, ABBAS PJ, KELSAY DM, TEAGLE HF, LOWDER MW (2001) A longitudinal study of electrode impedance, the electrically evoked compound action potential, and behavioral measures in nucleus 24 cochlear implant users. Ear Hear 22:471–486

KAWAHARA H (1997) Speech representation and transformation using adaptive interpolation of weighted spectrum. IEEE Trans Acoust Speech Sig Process 2:1303–1306

KIEFER J, VON ILBERG C, RUPPRECHT V, HUBNER-EGNER J, KNECHT R (2000) Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: effect of variations in stimulation rate and number of channels. Ann Otol Rhinol Laryngol 109:1009–1020

KILGARD MP, MERZENICH MM (1998) Cortical map reorganization enabled by nucleus basalis activity. Science 279:1714–1718

KRAL A (2007) Unimodal and cross-modal plasticity in the 'deaf' auditory cortex. Int J Audiol 46:479–493

KRAL A, TILLEIN J (2006) Brain plasticity under cochlear implant stimulation. Adv Otorhinolaryngol 64:89–108

KRAL A, TILLEIN J, HEID S, KLINKE R, HARTMANN R (2006) Cochlear implants: cortical plasticity in congenital deprivation. Prog Brain Res 157:283–313

KRUEGER B, JOSEPH G, ROST U, STRAUSS-SCHIER A, LENARZ T, BUECHNER A (2008) Performance groups in adult cochlear implant users: speech perception results from 1984 until today. Otol Neurotol 29:509–512

KUMAR A, ROTTER S, AERTSEN A (2010) Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. Nat Rev Neurosci 11:615–627

LIU J, NEWSOME WT (2005) Correlation between speed perception and neural activity in the middle temporal visual area. The Journal of neuroscience: the official journal of the Society for Neuroscience 25:711–722

LOEBACH JL, WICKESBERG RE (2006) The representation of noise vocoded speech in the auditory nerve of the chinchilla: physiological correlates of the perception of spectrally reduced speech. Hearing Res 213:130–144

LOIZOU PC (1998) Mimicking the human ear. Ieee Signal Proc Mag 15:101–130

LOIZOU PC (2006) Speech processing in vocoder-centric cochlear implants. Adv Otorhinolaryngol 64:109–143

LOIZOU PC, DORMAN M, TU ZM (1999) On the number of channels needed to understand speech. J Acoust Soc Am 106:2097–2103

LOIZOU PC, POROY O, DORMAN M (2000) The effect of parametric variations of cochlear implant processors on speech understanding. J Acoust Soc Am 108:790–802

MAUK MD, BUONOMANO DV (2004) The neural basis of temporal processing. Annu Rev Neurosci 27:307–340

MOORE DR, SHANNON RV (2009) Beyond cochlear implants: awakening the deafened brain. Nat Neurosci 12:686–691

NIE K, BARCO A, ZENG FG (2006) Spectral and temporal cues in cochlear implant speech perception. Ear Hear 27:208–217

OHL FW, SCHEICH H (1997) Orderly cortical representation of vowels based on formant interaction. P Natl Acad Sci U S A 94:9440–9444

PANZERI S, BRUNEL N, LOGOTHETIS NK, KAYSER C (2010) Sensory neural codes using multiplexed temporal scales. Trends Neurosci 33:111–120

PEREZ CA, ENGINEER CT, JAKKAMSETTI V, CARRAWAY RS, PERRY MS, KILGARD MP (2012) Different time scales for the neural coding of consonant and vowel sounds. New York, NY: Cereb Cortex

PETERSON G, BARNEY H (1952) Control methods used in a study of the vowels. J Acoust Soc Am 24:175–183

POEPPEL D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun 41:245–255

RAO D, BASURA GJ, ROCHE J, DANIELS S, MANCILLA JG, MANIS PB (2010) Hearing loss alters serotonergic modulation of intrinsic excitability in auditory cortex. J Neurophysiol 104:2693–2703

RENNAKER RL, CAREY HL, ANDERSON SE, SLOAN AM, KILGARD MP (2007) Anesthesia suppresses nonsynchronous responses to repetitive broadband stimuli. Neuroscience 145:357–369

ROMO R, SALINAS E (2003) Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4:203–218

SACHS MB, YOUNG ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. J Acoust Soc Am 66:470–479

SCOTT SK, BLANK CC, ROSEN S, WISE RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123(Pt 12):2400–2406

SCOTT SK, ROSEN S, LANG H, WISE RJ (2006) Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. J Acoust Soc Am 120:1075–1083

SHANNON RV, ZENG FG, KAMATH V, WYGONSKI J, EKELID M (1995) Speech recognition with primarily temporal cues. Science 270:303–304

SHARMA A, GILLEY PM, DORMAN MF, BALDWIN R (2007) Deprivation-induced cortical reorganization in children with cochlear implants. Int J Audiol 46:494–499

SHETAKE JA, WOLF JT, CHEUNG RJ, ENGINEER CT, RAM SK, KILGARD MP (2011) Cortical activity patterns predict robust speech discrimination ability in noise. Eur J Neurosci 34:1823–1838

SINEX DG, CHEN GD (2000) Neural responses to the onset of voicing are unrelated to other measures of temporal resolution. J Acoust Soc Am 107:486–495

STEINSCHNEIDER M, AREZZO J, VAUGHAN HG JR (1982) Speech evoked activity in the auditory radiations and cortex of the awake monkey. Brain Res 252:353–365

STEINSCHNEIDER M, VOLKOV IO, NOH MD, GARELL PC, HOWARD MA 3RD (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol 82:2346–2357

STEINSCHNEIDER M, FISHMAN YI, AREZZO JC (2003) Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. J Acoust Soc Am 114:307–321

STEINSCHNEIDER M, VOLKOV IO, FISHMAN YI, OYA H, AREZZO JC, HOWARD MA 3RD (2005) 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 15:170–186

TYLER RS, SUMMERFIELD AQ (1996) Cochlear implantation: relationships with research on auditory deprivation and acclimatization. Ear Hear 17:38S–50S

VALIMAA TT, MAATTA TK, LOPPONEN HJ, SORRI MJ (2002) Phoneme recognition and confusions with multichannel cochlear implants: consonants. J Speech Lang Hear Res 45:1055–1069

VAN TASELL DJ, SOLI SD, KIRBY VM, WIDIN GP (1987) Speech waveform envelope cues for consonant recognition. J Acoust Soc Am 82:1152–1161

VERSNEL H, SHAMMA SA (1998) Spectral-ripple representation of steady-state vowels in primary auditory cortex. J Acoust Soc Am 103:2502–2514

WATANABE T (1978) Responses of the cat's collicular auditory neuron to human speech. J Acoust Soc Am 64:333–337

WON JH, DRENNAN WR, RUBINSTEIN JT (2007) Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. J Assoc Res Otolaryngol 8:384–392

XU L, THOMPSON CS, PFINGST BE (2005) Relative contributions of spectral and temporal cues for phoneme recognition. J Acoust Soc Am 117:3255–3267