Perception of degraded speech by chinchillas (*Chinchilla laniger*): Word-level stimulus generalization

Authors: William P. Shofner, Nicole Yacko and Kristina Bowdrie

Department of Speech and Hearing Sciences
Indiana University
200 S. Jordan Ave.
Bloomington, IN 47405

Keywords: chinchilla, speech perception, noise-vocoder, stimulus generalization

Running Head: Speech perception by chinchillas

Acknowledgments

The authors gratefully acknowledge Dr. Rachael Holt for providing the speech sounds. The authors would like to thank Gifty Easow and Angie Nguyen for assistance in testing the chinchillas.

Correspondence regarding the article should be addressed to:

William P. Shofner
Department of Speech and Hearing Sciences
Indiana University
200 S. Jordan Ave.
Bloomington, IN 47405
Email: wshofner@indiana.edu

Abstract

One characteristic of human speech perception is a remarkable ability to recognize speech when the speech signal is highly degraded. It has been argued that this ability to perceive highly degraded speech reflects speech-specific mechanisms. The present study tested this hypothesis by measuring the ability of chinchillas to recognize noise-vocoded versions of naturally-spoken monosyllabic words using operant conditioning in a stimulus generalization paradigm. Chinchillas do not generalize the vocoded words to be perceptually equivalent to the naturally-spoken words. The responses from chinchillas to the vocoded words fall well below their responses to the naturally-spoken words. In this case, pitch cues rather than speech cues may be controlling the behavioral responses. In order to reduce pitch cues, chinchillas were re-trained using 64-channel noise-vocoded words. The responses from chinchillas to the vocoded test words were now similar to those of the 64-channel versions and were similar to those obtained from human listeners. However, responses obtained from chinchillas to time-reversed versions were high and similar to responses obtained to time-normal versions suggesting that the cue controlling behavioral responses was the phonetic structure of the words. These results show that chinchillas used different acoustic cues than human listeners. The ability of chinchillas to recognize noise-vocoded words as being perceptually equivalent to the naturally-spoken versions is inferior compared to that of human listeners. The findings suggest that the ability of human listeners to recognize highly degraded speech is unlikely to be based solely on the general auditory and perceptual mechanisms that are common among mammals.

Introduction

Humans possess a remarkable ability to recognize and understand speech when the speech signal is highly degraded, and the existence of specialized speech mechanisms could account for this ability (e.g. Remez, Rubin, Pisoni & Carrell, 1981; Remez, Rubin, Berns, Pardo & Lang, 1994). However, the ability to recognize degraded speech alone provides insufficient evidence for the existence of specialized speech mechanisms. In order to argue for the existence of speech-specific mechanisms, it is necessary to demonstrate an inability to perceive degraded speech based solely on general auditory processing mechanisms (Fitch, 2011). A comparison of speech perception in humans to speech perception based solely on general auditory processing mechanisms in non-human mammals provides constructive insights into understanding the adaptations that may be enhanced or specialized for speech processing in humans. Only a few studies have examined the perception of degraded speech by non-human mammals (Heimbauer, Beran & Owren, 2011; Ranasinghe, Vrana, Matney & Kilgard, 2012; Shofner, 2014). Of particular interest here, is the study of Heimbauer et al. (2011), in which they report on noise-vocoded (NV) word recognition in Panzee, a linguistically trained chimpanzee. Noise-vocoding is a common method for degrading speech sounds used in human perceptual studies, in part, because of the relationship between vocoding and the processing in cochlear implants.

Heimbauer et al. (2011) showed a parallel in NV word recognition performance for the same set of words between Panzee and a group of human listeners, suggesting that the mechanisms underlying degraded speech perception may have been present in the common ancestor of humans and chimpanzees. One conclusion of their paper is that linguistic experience plays a critical role in speech perception in humans and Panzee. Although the role of linguistic experience may seem important intuitively, its importance for degraded speech perception is challenging to test directly. Specifically, it is critical to measure degraded word recognition performance in the absence of linguistic experience, but the absence of linguistic experience is difficult to achieve. For example, one approach might be to study vocoded word recognition in human listeners using nonsense words. However, Remez et al. (1981) demonstrated that when listeners were presented with sentences based on sine-wave speech, they did not recognize the sounds as speech, but when listeners were simply told the sound they were hearing was computer-generated speech, listeners could then recognize and understand a substantial number of words in the highly degraded, sine wave sentences. Thus, if listeners are presented with natural and vocoded nonsense words, there is no guarantee that they would not be tapping into existing linguistic mechanisms. Another approach might be to compare speech perception by adult cochlear implant users with pre-lingual deafness to those with post-lingual deafness (Teoh et al., 2004a). However, this approach may be confounded by deafness-induced degeneration of central auditory structures along the auditory pathway (Teoh et al., 2004b). An obvious approach would be to repeat the experiments of Heimbauer et al. (2011) with chimpanzees that are not linguistically trained as Panzee was. However, this is challenging given the present limitations of using chimpanzees in biomedical and behavioral research (Institute of Medicine, 2011). We argue that the chinchilla offers a good alternative animal model to this latter approach.

Chinchillas are rodents, but unlike many other rodents, they have a range of hearing similar to human listeners (Heffner and Heffner, 1991). Although psychophysically measured thresholds are generally higher in chinchillas than humans, functional relationships are often similar between chinchillas and humans as with frequency discrimination (Nelson & Kiester, 1978) and noise intensity discrimination (Shofner, Yost & Sheft, 1993, Shofner & Sheft, 1994).

Chinchilla auditory filters derived from simultaneous masking using notched-noise are similar to those of humans (Niemiec, Yost & Shofner, 1992), and chinchillas appear to possess a spectral dominance region and missing fundamental percept for pitch that are similar to those of humans (Shofner & Yost, 1997; Shofner, 2011). Chinchillas show phonetic boundaries consistent with categorical perception of voice onset time that are similar to the boundaries of humans (Kuhl & Miller, 1975; 1978; Ohlemiller, Jones, Heidbreder, Clark & Miller, 1999). However, unlike humans and Panzee, chinchillas lack linguistic experience. Thus, we argue that behavioral responses obtained from chinchillas to degraded words will reflect speech perception based solely on the general auditory and perceptual mechanisms that are common between chinchillas and humans, and are presumably common among mammals. The present study reports on the recognition of NV monosyllabic words by chinchillas as measured in a stimulus generalization paradigm.

In stimulus generalization paradigms, an animal is trained to respond to a specific stimulus and then responses are measured to test stimuli that vary systematically along one or more stimulus dimensions (Malott & Malott, 1970). The signal is presented frequently and is thus expected whereas test stimuli are presented infrequently and are unexpected. A systematic change in behavioral response along the physical dimension of the stimulus is known as a generalization gradient and is consistent with the hypothesis that the animal possesses a perceptual or psychological dimension related to the physical dimension (Guttman, 1963). Thus, data from stimulus generalization paradigms can indicate what acoustic features control the behavioral response of the animal. Generalization data are often interpreted to indicate similarities in an animal's perception between test and signal stimuli (Guttman, 1963). Test stimuli that evoke similar behavioral responses as the signal stimulus indicate a perceptual equivalence (Hulse, 1995) among these stimuli. Rock, Lasker & Simon (1969) argue that generalization by animals occurs through recognition processes. Thus, we interpret generalization to NV words to reflect the recognition of NV words.

How then does an animal like the chinchilla, which has the same basic auditory system as a human but lacks any speech-specific mechanisms or linguistic experience, perceive a word like "cut", for example? To be more specific, what are the acoustic cues available in the word that control the behavioral response of the animal in the generalization paradigm? Words will have no meaning to the chinchillas and will simply be a type of complex sound. If there is a perceptual equivalence among vocoded test words and the naturally-spoken words *and* the acoustic cues controlling the behavioral responses in the generalization paradigm are the phonetic structures of the words, then it would suggest that the mechanisms underlying word recognition are similar between humans and chinchillas. The results of the present study indicate that the acoustic cues controlling the behavioral responses of the chinchillas differ from those used by human listeners suggesting that general auditory and perceptual mechanisms alone are inadequate to account for word recognition.

General Methods

The procedures used were approved by the Institutional Animal Care and Use Committee and the Institutional Review Board for Indiana University. All human participants provided informed consent.

Subjects

Five adult chinchillas (*Chinchilla laniger*) and 16 human listeners served as subjects in these experiments. All 5 chinchillas had experience with the stimulus generalization task (Shofner, 2014). Chinchillas received food pellet rewards during behavioral testing, and their body weights were maintained between 85-90% of their normal weight. American English was the first language for all 16 human participants, and audiometric thresholds for 125-8000 Hz were ≤ 25 dB HL. Listeners were paid an hourly stipend for participation.

Acoustic Stimuli

Naturally-spoken monosyllabic words spoken by a female voice were obtained from the Lexical Neighborhood Test (Kirk, Pisoni & Osberger, 1995). Table 1 summarizes the formant and fundamental frequencies (F0s) of the naturally-spoken words used in the present study; the spectral analysis of the words was carried out using Praat (http://www.fon.hum.uva.nl/praat). NV versions of these words were generated using Tiger CIS version 1.05.02 developed by Qian-Jie Fu (http://tigerspeech.com). A naturally-spoken word was first passed through a series of bandpass filters from 200-7000 Hz. Default filter slopes of 24 dB/octave and center frequencies based on the Greenwood function were used. The number of channels was fixed at 1, 2, 4, 8, 16, 32, 64 or 128. The envelope was extracted for each channel with the lowpass cut-off frequency fixed at 160 Hz. The carrier type of the vocoder was set to white noise; in this mode, the extracted envelope from a given channel is used to modulate a wideband noise. This modulated wideband noise was then bandpass filtered with a center frequency equal to that of the analysis channel. The number of contiguous bandpass noises used for resynthesis equaled the number of channels used for analysis. The bandpass-filtered, modulated noises were then summed to yield the vocoded version of the word. NV words were stored as wav files that were later converted to

16-bit integer files at a sampling rate of 50 kHz using Adobe Audition in order to be played through Tucker-Davis System II modules. The natural and vocoded words had durations of approximately 500 ms. Example waveforms and spectrograms for the word “sit” are illustrated in Figure 1. Example envelopes extracted from the waveforms by half-wave rectification and low-pass filtering with a cutoff frequency of 100 Hz are illustrated in Figure 2.

Testing Procedures

Chinchillas were placed into a testing cage located in a single-walled sound attenuating chamber. A pellet dispenser was located at one end of the cage with a reward chute attached to a response lever. A loudspeaker was located next to the pellet dispenser approximately 30^{o} to the right of center at a distance of 6” in front of the chinchilla. The sound pressure level was fixed at 73 dB SPL (A-weighted) for all sounds. A standard sound was presented continually in 500-ms bursts at a rate of one per second, regardless of whether or not a trial was initiated. The 2-channel version was used as the standard instead of the 1-channel version in order to avoid introducing overall spectral shape as a confounding variable (see Shofner, 2014). Chinchillas initiated a trial by pressing down on the response lever; the holdtime varied randomly for each trial ranging from 1.15-8.15 seconds for 4 chinchillas and from 1.15-6.15 seconds for a 5th chinchilla. After the lever was depressed for the required holdtime, two 500-ms bursts of a selected sound were presented for that trial. The response window was coincident with the duration of the two 500-ms bursts, but began 150 ms after the onset of the first burst. Thus, the duration of the response window was 1850 ms.

The sounds presented during the response window could be signals, test sounds, or standards. A signal trial consisted of two bursts of the word the animal was trained to discriminate (e.g. naturally-spoken word). If the animal released the lever during the response window of a signal trial, then this positive response was treated as a hit and was rewarded with a food pellet. A standard trial consisted of two additional bursts of the 2-channel standard. If the animal continued to depress the lever throughout the response window of a standard trial, then this negative response was treated as a correct rejection. Food pellet rewards for correct rejections were generally not necessary to reinforce continued lever depression for 4/5 chinchillas; one animal did receive a food pellet for a correct rejection. A test trial consisted of two bursts of a test sound which was generally a NV version of the naturally-spoken word based on 4- to 128-channels. Chinchillas did not receive food pellet rewards for positive responses to test stimuli.

Chinchillas were tested in blocks of either 40 or 10 trials. Trials were presented randomly with a given block of trials. Within a block of trials, the signal (e.g. naturally-spoken “cut”) was presented on 60% of the trials, the standard (e.g. 2-channel NV “cut”) was presented on 20% of the trials, test stimulus #1 (e.g. 32-channel NV “cut”) was presented on 10% of the trails, and test stimulus #2 (e.g. 4-channel NV “cut”) was presented on 10% of the trials. Responses were collected for a minimum of 2000 total trials, which results in a minimum of 200 trials for each test sound. The two test stimuli were changed after the minimum number of trials was completed. Animals completed all 6 test versions for one word before moving on to the next word. Responses obtained from the stimulus generalization task are in terms of percent. Because percent is not normally distributed, percent responses were converted into percent rationalized arcsine units (RAUs) using the formula described by Sherbecoe and Studebaker (2004) in order to carry out inferential statistical analyses on the responses.

Human listeners sat in front of a computer keyboard and monitor in a double-walled sound attenuating chamber and were tested in a single-interval, forced-choice procedure. At the beginning of a block of 40 test trials, listeners first heard the naturally-spoken target word and were instructed to reply “yes” when they recognized that specific target word. Listeners then heard a Gaussian noise and were instructed to reply “no” if they did not recognize that target word. NV versions of 1, 2, 4, 8, 16, 32, 64 and 128 channels of the target word were presented randomly 5 times each at 50 dB SL through Koss ESP/950 electrostatic earphones monaurally to the left ear. Listeners did not receive feedback for their responses. Recognition functions were collected for 200-520 total trials for each target word, which results in 25-65 trials for each test word.

Experiment 1: Responses to NV-test words

Introduction

The purpose of this experiment was to determine whether chinchillas show a perceptual equivalence between naturally-spoken words and NV words. That is, are the behavioral responses to vocoded test words similar to their responses to naturally-spoken words?

Methods

Chinchillas were trained to discriminate naturally-spoken words from the 2-channel NV versions. In this experiment, a naturally-spoken word was the signal, the 2-channel NV version was the standard, and test sounds were 4-, 8-, 16-, 32-, 64-, and 128-channel NV versions of the word. The following six monosyllabic words were tested: “ball”, “cut”, “hot”, “meat”, “sit” and “wet”.

Results

The mean responses for individual chinchillas to the naturally-spoken words are high, ranging from 93.7-98.9%, whereas the mean responses to the 2-channel versions are low, ranging from 0.7-7.3% (Fig.3). However, the responses to the NV test words are highly variable across animals. For example, the responses obtained from C12 and C47 to the vocoded test words are generally below 20% even when the number of vocoder channels is as high as 32 or above. In these cases, there are no systematic increases in behavioral responses as the number of channels increases. The generalization gradients are relatively “flat” for these 2 chinchillas for the 6 words tested. In contrast, for C24, C36 and C15, the generalization gradients obtained for some of the words show a systematic increase as the number of channels increases (e.g. “ball”), but the gradients obtained for other words are relatively “flat” (e.g. “meat”). That is, the responses are more variable within these animals. Note that for these animals, the responses to vocoded test words based on 4-16 channels are generally below 40% for C24 and C36, and below 60% for C15. The largest responses obtained for C15, C24 and C36 are around 50-80% for some, but not all, vocoded test words based on 32, 64 and 128 channels (Fig. 3). Thus, the generalization gradients obtained from these 5 chinchillas indicate that some chinchillas can show relatively large behavioral responses to some NV monosyllabic words, but only for a high number of vocoder channels (i.e. 32-128 channels). However, even the largest responses obtained to vocoded test words are generally well below the responses to the naturally-spoken words.

The responses to NV monosyllabic test words averaged across animals and across words appears to show a shallow gradient with a logarithmic increase in response as the number of channels increases from 2 to 128 (Fig. 4A). The responses averaged across animals and words ranges from 8% for 4-channels to 31.7% for 128-channels and fall well below the averaged response of 96.9% obtained to the naturally-spoken words. A repeated-measures analysis of variance (ANOVA) based on the chinchilla RAUs (Fig. 4B) shows a significant effect of stimulus, F(7, 21) = 26.9, $p < 0.001$, $\eta^2 = 0.72$. A regression analysis was applied to the averaged data covering the range of 2 to 128 channels. The regression line had the form of $Y = m*\log_{10}(X) + b$, where Y is the percent RAU, m is the slope, X is the number of channels and b is the Y-intercept. The slope of the mean generalization gradient is 26.08, and a two-tailed t-test showed this is significantly different from a slope of 0 (t = 5.815; p = 0.0021).

There is a substantial difference between the average vocoded word recognition function obtained from human listeners and the generalization gradient obtained from chinchillas (Fig. 4A). The percent responses averaged across individuals and across words for these specific words show that recognition is around 90% with as few as 8 channels for the human listeners which is well above the percent responses averaged across chinchillas and across words. The goals of the remaining experiments are to determine specifically the reasons for the differences in performance between chinchillas and humans.

Experiment 2: Testing the role of listening experience

Introduction

The purpose of this experiment is to test the role of listening experience in chinchillas, but how can chinchillas acquire additional listening experience with words? Clearly, it is not feasible to provide chinchillas with experience that is comparable to the lifetime of experience that human listeners have with words. One would have to breed and raise chinchillas and then attempt to provide the young animals with linguistic experience in a manner similar to that provided to Panzee (Heimbauer et al., 2011). Moreover, if chinchillas use the waveform period of the vowel (i.e. vowel 1/F0) as the acoustic cue controlling their behavioral responses when the signal in the generalization task is the naturally-spoken version (see Shofner, 2014), it is unlikely that additional experience with the naturally-spoken words will change the use of periodicity cues to the use of phonetic structure cues. What is necessary is to provide chinchillas with listening experience in terms of the phonetic structure of words by reducing the pitch cues.

Methods

Chinchillas were re-trained to discriminate 64-channel NV words from the 2-channel versions. There was no fixed number of training trials. Once an animal reached a performance of approximately 85%, training with a specific 64-channel word lasted over 560-4200 trials in order to be assured that the animal had a sufficient amount of listening experience with the degraded 64-channel word prior to testing in the generalization task. Following re-training, 64-channel NV words were used as the signals in this generalization task. The same six monosyllabic words used in Experiment 1 were used in Experiment 2. Figure 1 illustrates example spectrograms for the naturally-spoken and 64-channel NV versions of "sit." It can be observed that the harmonic structure is reduced in the 64-channel NV version compared to that of the naturally-spoken version. Although the first few harmonics are still represented in the 64-channel version, chinchillas are unable to use this spectral information (Shofner and Chaney, 2013). Consequently, this means that the periodicity cues for pitch are reduced in the 64-channel NV words. However, the wide, dark bands indicating the formant frequencies appear to be clearly present in the 64-channel NV words. Because 64-channel NV words lack the strong harmonic structures and waveform periodicities of the naturally-spoken words, this re-training should provide animals with additional experience listening to the speech cues of the words when pitch cues are reduced.

Results

Responses from three chinchillas averaged across words and animals are shown in Figure 5A. When 64-channel NV words are used as the signal, responses from chinchillas to other vocoded versions as well as responses to the naturally-spoken versions are high and less variable (red open circles and red lines in Fig. 5A) than when the naturally-spoken versions were used as the signal (Fig. 4A). Note that not only are the responses of chinchillas to 8 or more vocoder channels higher than before (≥ 87%), but they are now similar to those of human listeners (Fig. 5A). This similarity in the responses of chinchillas and humans argues that the differences observed previously between chinchillas and humans (Fig. 4A) are not simply a reflection of different behavioral tasks. That is, chinchillas do give large behavioral responses to NV words in the stimulus generalization task under the appropriate stimulus conditions. The differences in behavioral responses by individual chinchillas when naturally-spoken words and 64-channel NV words are used as signals are shown in Figure 5B-D in terms of RAUs. A 2-factor ANOVA was carried out for RAUs obtained for 4-, 8, 16-, 32, 64- and 128- NV-words and showed significant effects of the signal used (i.e. naturally-spoken vs. 64-channel), significant effects of the stimulus, and significant interactions (see Table 2). The significant effect of stimulus appears to largely be due to the lower RAUs generally observed for 4- and 8-channel NV words. The significant effect observed for the signal suggests differences in the acoustic cues controlling the behavioral responses when the signal is the naturally-spoken word or the 64-channel NV word.

That is, the acoustic cues controlling chinchilla behavioral responses are different in the context of the naturally-spoken words and in the context of the 64-channel NV words. Thus, listening experience with reduced pitch cues does appear to lead to improvement in degraded word recognition in chinchillas.

Experiment 3: Testing the role of phonetic structure

Introduction

The purpose of this experiment was to determine specifically if the acoustic cue controlling the behavioral response when the 64-channel NV words were used as the signals was phonetic structure. The data described above in Experiment 2 suggest listening experience with reduced F0 cues improves degraded word recognition by chinchillas to an extent that it is similar to that of humans. However, it has been argued that similarity in performance does not equate to similarity of mechanisms (e.g. Trout, 2001). For example, although the responses between humans and chinchillas are similar after chinchillas were re-trained using 64-channel NV words as signals (see Fig. 5A), it is possible chinchillas could be processing different cues than humans use and thus could be using different mechanisms (Trout, 2001). Consequently, the results of Experiment 2 do not demonstrate that behavioral responses in the chinchilla are controlled by phonetic structure. Saberi and Perrott (1999) showed that speech is unintelligible when segments of speech 200 ms or longer are time reversed. Time reversing a word changes its phonetic structure and consequently changes the perception of the word.

Methods

In the generalization task used in this experiment, 64-channel NV words were used as signals and the 2-channel versions were used as the standards. Monosyllabic words used were "ball", "hot" and "cut". Test words included 16-, 32- and 128-channel NV versions of the words and time-reversed versions of 16-, 32-, 64- and 128-channel NV versions. Examples of waveforms as well as the temporal envelopes for time-normal and time-reversed 64-channel "cut" are illustrated in Figure 6, and example spectrograms are illustrated in Figure 7.

Results

We verified in a group of 8 human listeners from the original 16 using a single-interval task that when the vocoded words are time reversed (approximately 500 ms), they are no longer recognized as the time-normal vocoded words. The responses to 16-, 32-, 64- and 128-channel time-normal NV words are high (see red filled triangles and red dashed lines in left-hand column in Fig. 8), because listeners recognize the vocoded versions as being perceptually equivalent to the naturally-spoken target word. However, responses to time-reversed versions of the NV words are virtually 0 (see blue filled circles and blue dashed solid line in left-hand column in Fig. 8). Listeners do not recognize time-reversed vocoded versions of words as being perceptually equivalent to the naturally-spoken words, because time reversing alters the detailed phonetic structure of the words. If chinchillas show similar behavioral responses in the stimulus generalization paradigm, it would argue that the acoustic cues controlling their behavioral responses to the vocoded words are the phonetic structures of the words.

Three chinchillas were tested in the stimulus generalization task using the time-normal 64-channel version as the signal with both time-normal and time-reversed NV words as test stimuli (see right-hand column of Fig. 8). If chinchillas are responding to the phonetic structure of the words, in the context of the 64-channel signal, then responses to the 16-, 32-, and 128-

channel versions should be high, whereas responses to the time-reversed 16-, 32-, 64- and 128-channel versions should be low. That is, the time-reversed versions should not be perceptually equivalent to the time-normal versions. Clearly, behavioral responses of chinchillas to the time-reversed versions are high and appear to be well above the responses to the 2-channel versions (right-hand column in Fig. 8), but they do not appear to be equal to those of the time-normal words. The mean responses to the time-reversed NV “ball”, “cut” and “hot” averaged across channels and animals are 88%, 70.2%, and 72.4%, respectively, whereas the mean responses to the time-normal NV “ball”, “cut” and “hot” are 96.2%, 95.6%, and 97.1%, respectively. A two-factor ANOVA showed that the differences between RAUs for each of time-normal and time-reversed words were significant (see Table 3). Because there was not a significant effect of the number of channels (Table 3), the responses of time-reversed 16-, 32-, 64- and 128-channel were combined, and a one-tailed t-test showed that the mean RAUs for the time-reversed NV test words was significantly larger than the mean RAUs for the 2-channel standards (see Table 4). The effect sizes in terms of Cohen’s d (Table 4) were converted into η^2 using formulae described by Cohen (1988) in order to compare to the effect sizes given in Table 3. It can be observed that the effect sizes for the differences between RAUs for time-reversed words and the 2-channel standards (Table 4) are larger than the effect sizes for the differences in RAUs between time-reversed and time-normal words (Table 3). Thus, unlike the human listeners, the large responses to time-reversed words by chinchillas indicates that there is some degree of perceptual equivalence between the time-reversed and time-normal NV words in the chinchillas suggesting that the detailed phonetic structure of the words is not the acoustic cue controlling their behavioral responses.

Experiment 4: Testing with naturally-spoken words

Introduction

Experiment 1 and our previous work (Shofner, 2014) suggest that fundamental frequency (F0) of naturally-spoken vowels is a dominant acoustic cue. Experiments 2 and 3 above suggest that following additional listening experience with the 64-channel NV versions, chinchillas may be responding to some speech cues, but not phonetic structure per se. The purpose of this experiment was to test the saliency of pitch cues over speech cues in chinchillas.

Methods

Chinchillas were re-trained to discriminate naturally-spoken “cap” from the 2-channel NV version. Chinchillas had not been previously exposed to the naturally-spoken “cap.” Chinchillas were then tested in the generalization task using the naturally-spoken “cap” as the signal and the 2-channel NV “cap” as the standard; test stimuli were naturally-spoken “cut”, “hot”, “ball” and “wet” as well as the musical note G^b3 as played on a piano and cello. Musical notes were obtained from the University of Iowa Electronic Music Studios (http://theremin.music.uiowa.edu/). Although there was some variation in stimuli in terms of phonetic structures and formant structures, all test sounds had approximately the same F0 and fundamental waveform period (Table 1). Although the specific F0 difference limen (F0DL) is unknown for 193 Hz F0, which is the F0 of the naturally-spoken “cap”, the F0DL in chinchillas for a closely comparable 250-Hz F0 harmonic tone complex is 30 Hz (Shofner, 2000). Note, however, that the differences between the F0 for “cap” and the F0s of the other stimuli are less than 30 Hz. Thus, because the F0 differences between the signal and test stimuli are below the apparent F0DL, it is unlikely that F0 can be used by chinchillas as a cue to discriminate among these sounds.

Results

Chinchillas gave high responses to all of the test sounds, including the two musical notes, although there is some individual variability (Fig. 9A). For example, the responses of C12 are consistently > 90% for all test stimuli. In contrast, although the responses of C24 are generally high to the test words, they are lower for the two musical notes (Fig. 9A). A repeated-measures ANOVA on the RAUs averaged across the 3 chinchillas (Fig. 9B) showed a significant effect of stimulus, $F(7, 14) = 33.39$, $p < 0.001$, $\eta^2 = 0.872$. The 95% confidence intervals indicate that responses to the natural-spoken "cap" signal and all test sounds are well above those of the 2-channel "cap" standard. Although the confidence intervals for the cello G^b3 fall short of overlapping with those of the "cap" signal, the confidence intervals for all other test sounds, including piano G^b3, do overlap with those of the signal (Fig. 9B). Thus, there is a perceptual equivalence among the signal and test sounds, in general.

Discussion

Motivation in the context of each stimulus generalization task

In the generalization task, animals are trained to discriminate a specific signal from a standard. There will be some acoustic cue(s) which will control the behavioral responses. Animals are presented with the signal for most trials and receive a food reward for correct responses. Thus, animals will learn to expect the signal and receive reinforcement. Test stimuli are presented infrequently. Given the expectation of a signal and reinforcement, if an animal perceives that the test stimulus sufficiently contains the acoustic cue, it will be motivated to respond, expecting a food reward. If the animal does not perceive the acoustic cue to be sufficient in the test stimulus, there is no motivation for it to respond, because it would have learned that these sounds are not reinforced with food. Thus, in the context of a specific signal and standard, an animal will presumably respond to the acoustic cue that will maximize its food reinforcement while minimizing its effort. Consequently, the stimulus generalization paradigm allows us to deduce what acoustic cue is controlling the behavioral response. For each of the generalization experiments, we want to deduce the cue being used by the chinchillas in the context of discriminating the signal from the standard that will presumably maximize their food reward and minimize their effort.

In the context of discriminating a naturally-spoken word from the 2-channel version (Experiment 1), the responses obtained to NV test versions were generally low and were not close to those obtained for the naturally-spoken version (Fig 4B). That is, the chinchillas showed virtually no perceptual equivalence of the vocoded words to the naturally-spoken words, suggesting that the phonetic structure is not the acoustic cue controlling the behavioral responses. It is also unlikely that the temporal envelope is the acoustic cue, because the shapes of the temporal envelopes for a naturally-spoken word and the 2-channel version are extremely similar (e.g. Fig. 2). The formant structure for vocoded words having a high number of channels (e.g. Fig. 1) would suggest that the responses should be higher to these test words if the formant structure was the acoustic cue. However, our previous work with speech tokens (Shofner, 2014) and NV harmonic tone complexes (Shofner and Chaney, 2013) argues that in this context, the periodicity of the F0 is the most likely acoustic cue being used. The variability observed for some chinchillas (e.g. C 24 and C36 in Fig. 3) which show ‘flat’ responses for some words, but gradients in responses to other words, suggests that the acoustic cue(s) controlling behavioral responses in these animals may be F0 for some words but a combination of F0 and formant structure for other words. Thus, in the context of the naturally-spoken word, chinchillas recognize the strong harmonic structures or more specifically the fundamental periods of the vowels as well as voiced consonants (e.g. /m/, /l/, and /j/). Since this stimulus is presented most of the time, chinchillas learn to expect this sound and receive food reinforcement. When the fundamental period is weak or absent, as in the case of the vocoded test sounds, there is little or no motivation for the chinchillas to respond, because they have learned that these sounds are not reinforced.

When the chinchillas are re-trained to discriminate 64-channel noise-vocoded words from the 2-channel version, then the responses to the other NV test words and the naturally-spoken versions are high (Experiment 2). That is, there is now a perceptual equivalence among vocoded words and the naturally-spoken words. In this case, the acoustic cue controlling the behavioral response must be strongly similar among the vocoded test words and 64-channel signal in order for the chinchillas to be motivated to respond. Given the similarity between the envelopes of the 2-channel standards and the 64-channel signals (e.g. Fig. 6), it is unlikely that the temporal envelope is the acoustic cue. Since the harmonic structures of the vowels in the 64-channel versions are weak, we argue that the acoustic cue is not likely to be F0, but rather is either the phonetic structures of the words or the formant structures of the vowels. This was tested by time-reversing the vocoded test words (Experiment 3).

Time-reversing changes the phonetic structure and the time-domain envelope (Fig. 6), but it can also change the formant structure. For naturally-spoken words, the vocal tract resonance can change over time resulting in dynamic changes in the trajectories of formants which are then turned around when the word is time-reversed. Thus, if phonetic structure is the acoustic cue, then chinchillas should recognize the difference between the time-reversed test words and the time-normal 64-channel signal. Consequently, they should not be motivated to release the lever, because they will learn that they will receive no food reward for this stimulus. Likewise, if chinchillas are responding to dynamic changes in formant structure, they should recognize the difference between time-normal and time-reversed words and not be motivated to respond to the time-reversed words. However, if the dynamic changes in formant frequencies for the specific words tested are too small to be detected by the chinchillas, such that the chinchillas only recognize the *average* formant structure, then chinchillas will not recognize the difference between the time-reversed test words and the time-normal 64-channel signal. Consequently, they will be motivated to release the lever, because they will be expecting a food reward. Hence, we conclude that average formant structure is the dominant acoustic cue controlling behavioral responses for most trials in the context of the 64-channel signals, but acknowledge that chinchillas may recognize dynamic changes in formant structure for some trials.

Finally, when chinchillas are re-trained using a naturally-spoken “cap” as the signal and tested with other naturally-spoken words as well as a musical note played by two different musical instruments (Experiment 4), the responses to all test sounds were high and not substantially different from the signal as indicated by the 95% confidence intervals. In this context, the phonetic structure, formant structures (Table 1) and temporal envelopes of the test words are all clearly different from those of the signal. If any of these features are the acoustic cue(s) being used to control behavioral responses, then the chinchillas should recognize the difference between the test word and the signal. Consequently, they will not be motivated to release the lever, because they will have learned that no food reward will be given for “non-cap” sounds. However, all of these words have similar fundamental period of voice pitch. If fundamental period of the vowels (i.e. 1/F0) is the acoustic cue, then the chinchillas should recognize the test words to be equivalent to the signal word and thus, will be motivated to release the lever expecting a food reward. We conclude based on the high responses to test words that the fundamental period is a highly salient cue that is controlling the behavioral responses in the context of naturally-spoken words. This conclusion is reinforced by the high responses to the piano and cello notes. However, we acknowledge that other phonemes may also play a secondary role in some chinchillas.

Comparison with Panzee and humans: Implications for the evolution of speech perception

The current results are in contrast to results obtained from a chimpanzee, Panzee, in which recognition of 7-channel NV words paralleled that of a group of human listeners (Heimbauer et al., 2011). The recognition for the chinchillas averaged across animals and all six 8-channel NV words tested was around 15%, whereas the recognition averaged across all 7-channel NV words tested in Panzee was around 55% (Heimbauer et al., 2011). A major difference between these two studies is that the words presented to the chinchillas clearly lacked any linguistic meaning, whereas Panzee was a linguistically trained chimpanzee (see Heimbauer et al., 2011 for details) and the words presented to Panzee did have linguistic meaning. Note, however, that increasing the *listening* experience of chinchillas by re-training on a 64-channel NV word does not equate to *linguistic* experience. Linguistic experience implies that the word has meaning to the listener and is part of a listener's lexicon. For example, "ball" is simply a complex sound and possesses no meaning for the chinchillas.

Chinchillas can discriminate vowels (Burdick & Miller, 1975), and appear to respond to the average formant structure of NV words when they are re-trained using 64-channel NV words in the generalization paradigm. However, in the context of naturally-spoken words, the formant cues appear to be less salient than the F0 cues in chinchillas. Even when F0 cues are reduced, as when chinchillas were re-trained using 64-channel NV words, subsequent testing with time-reversed NV words suggests that chinchillas were responding to the average formant structure, not to the detailed phonetic structure. We conclude that the acoustic cues available in these words that are being used by chinchillas are different from those used by humans and Panzee. Whether or not chinchillas can learn to use phonetic structure as a cue is certainly an interesting question and would be the next logical step in this line of research.

Chinchillas are phylogenetically further removed from humans (Huchon et al., 2002) than are chimpanzees (Uddin et al., 2004). Because the common mammalian ancestor of chinchillas and humans is more distant than that of humans and chimpanzees, the perceptual and cognitive mechanisms underlying speech perception by chinchillas will be based on mechanisms far more ancestrally shared than those common to chimpanzees and humans. Thus, we contend that chinchilla behavioral responses illustrate degraded speech recognition based solely on the general auditory and perceptual mechanisms that are common among mammals. It has been argued that the ability of humans to recognize highly degraded speech cannot be taken as evidence for the existence of speech-specific mechanisms without corresponding data establishing that this ability is absent in animals (Fitch, 2011). The lack of equivalent generalization among vocoded words to their naturally-spoken versions reported in the present study for chinchillas suggests an inability as compared to humans. More importantly, the results of the present study show that chinchillas appear to use different cues than human listeners. Chinchillas appear to learn to use formant structure as a cue when F0 cues are reduced (Experiments 2 and 3), but in the present study chinchillas never learned to use phonetic structure as a cue. The use of different cues by chinchillas argues that general auditory and perceptual mechanisms are insufficient to fully account for the ability to recognize degraded speech that is observed for human listeners (and Panzee).

The present findings provide evidence that linguistic experience is critical for degraded word recognition, consistent with the conclusions of Heimbauer et al. (2011). Whereas Panzee had linguistic experience and could recognize degraded words, the chinchillas lacked linguistic experience and could not recognize degraded words. Shannon (2005) concludes that "…speech recognition is primarily a top-down process of pattern recognition that is highly overlearned from a lifetime of experience." One might argue that this top-down processing reflects mechanisms that exists beyond the general auditory mechanisms and is an adaptation for learning to recognize the phonetic structure of words. The results of the present study suggest that these top-down processing mechanisms important for word learning are inadequate or non-existent in chinchillas. Although this conclusion would seem to be consistent with the 'speech-is-special' hypothesis, it does not imply that the mechanisms important for word recognition are unique to humans. Indeed, this adaptation also appears to have evolved in chimpanzees (Heimbauer et al., 2011) and domestic dogs (e.g. Pilley & Reid, 2011). Given that chimpanzees are the closest relative to humans genetically (Uddin et al., 2004), one could argue that this adaptation presumably appeared in the common ancestor of humans and chimpanzees (Heimbauer et al., 2011); that is, these top-down mechanisms are homologous in humans and chimpanzees. The evolution of domestic dogs parallels the evolution of humans (Wang et al., 2013), but it is likely that these top-down adaptations are the result of convergent evolution. That is, these mechanisms presumably evolved independently in dogs and humans. Thus, this adaptation for top-down processing may not be unique to humans, but its existence appears to be only in mammals that have demonstrated abilities to associate words with objects.

Limitations of the present study

The design of the present study is one that is typically used in psychoacoustics studies, namely a repeated measures design in which a limited number of subjects are each tested with a fixed set of stimulus conditions. This general approach has been criticized as lacking independent sampling and has been referred to as pseudoreplication (Hurlbert, 1984). Although this approach does have limitations (Kroodsma, 1990), inferential statistical analysis can still provide useful information (Oksanen, 2001). Indeed, Oksanen (2001) questions whether pseudoreplication is a spurious issue. In order to avoid pseudoreplication for both stimuli and subjects in the current study, we would need a total number of subjects of 6 words x N word-samples (i.e. groups) per word x n subjects per group. If, for example, we use 4 different samples of each specific word (e.g. recorded from two female speakers and two male speakers), and have 3 subjects in each group, then this design would avoid pseudoreplication by providing independent sampling in terms of both stimuli and subjects, and thus, would be a stronger test theoretically (see Kroodsma, 1990; see also Macgregor, 2000). However, this example would require a total of 72 trained chinchillas and would be untenable for the current study given the 3-4 months typically required to train a naïve chinchilla, the per diem animal costs and housing requirements for 72 chinchillas, the cost of buying 72 chinchillas and the fact that not every naïve chinchilla is capable of learning the behavioral task. Thus, a compromise must often be made in animal psychoacoustics experiments. The choice is either to recognize and accept some level of pseudoreplication in an effort to obtain some insight into the question (as in the present study) or not to attempt to answer the question at all (given that the alternative approach is untenable). As stated by Kroodsma (1990) “There is, of course, nothing technically 'wrong' with doing experiments that have no replication of treatments, but one must be aware that the hypotheses actually being tested are about specific exemplars…” In fact, the hypotheses tested in the present study are limited to the specific words used.

Finally, one other limitation is in regards to the behavioral procedures used. Caution should be exercised in comparing directly the responses of humans and chinchillas due to differences in the behavioral tasks. Humans responded in a single-interval, forced choice task whereas chinchillas responded in the stimulus generalization task. We have previously described that when humans are tested in the generalization task, their responses are similar to those obtained in the single interval task (Shofner, 2014). The low responses obtained from chinchillas when the naturally-spoken words were used as signals does not mean that the generalization task will inherently produce low responses to test stimuli, because high responses were obtained to the same test stimuli when 64-channel NV words were used as the signals. The main difference between tasks, however, is that humans were verbally instructed to respond when they recognize the vocoded test word as the naturally-spoken target word. As such, we know that the responses obtained from humans reflect recognition of these specific vocoded words based on phonetic structure. Of course, we cannot instruct the chinchillas to respond to phonetic structure, but rather must deduce what acoustic cue is controlling the behavioral responses of the chinchillas in the generalization task in the context of the signal and standard.

Thus, in the present study, a limited number of animals and a limited set of monosyllabic words were tested. The specific word-stimuli used were not meant to represent ideal exemplars of those words, but rather were used as samples in which specific acoustic features could be analyzed. What we are deducing in the present study are the acoustic cues that chinchillas are using under the specific conditions in each generalization task. The hypotheses tested and the conclusions reached are limited to understanding the acoustic cues controlling behavioral responses of chinchillas to these specific word stimuli and comparing whether these cues are the same as the cues known to be used by human listeners for the identical stimuli.

References

Burdick, C.K., & Miller, J.D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*, 58: 415-427. https://doi.org/10.1121/1.380686

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition. Lawrence Erlbaum Associates.

Fitch, W.T. (2011) Speech perception: A language-trained chimpanzee weighs in. *Current Biology*, R544-R546. http://dx.doi.org/10.1016/j.cub.2011.06.035

Guttman, N. (1963). Laws of behavior and facts of perception. In S. Koch (Ed.), *Psychology: A study of a science: vol. 5. The process areas, the person, and some applied fields: Their place in psychology and in science* (pp. 114-178). New York, NY: McGraw-Hill.

Heffner, R.S., & Heffner, H.E. (1991). Behavioral hearing range of the chinchilla. *Hearing Research*, 52,13-16. http://dx.doi.org/10.1016/0378-5955(91)90183-A

Heimbauer, L.A., Beran, M.J. & Owren, M.J., (2011) A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues. *Current Biology*, 21, 1210-1214. http://dx.doi.org/10.1016/j.cub.2011.06.007

Huchon, D., Madsen, O., Sibbald, M.J.J.B., Ament, K., Stanhope, M.J., Catzeflis, F., de Jong, W.W., & Douzery, E.J.P. (2002). Rodent phylogeny and a timescale for the evolution of glires: Evidence from an extensive taxon sampling using three nuclear genes. *Molecular Biology and Evolution,* 19, 1053-1065. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004164

Hulse, S.H. (1995). The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals. In G.M. Klump, R.J. Dooling, R.R. Fay, & W.C. Stebbins (Eds.), *Methods in Comparative Psychoacoustics* (pp. 319-330). Basel, Switzerland: Birkhauser-Verlag. http://dx.doi.org/10.1007/978-3-0348-7463-2_27

Hurlbert, S.H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs*, 54, 187-211. http://www.jstor.org/stable/1942661

Institute of Medicine (2011). *Chimpanzees in biomedical and behavioral research: Assessing the necessity.* Washington, DC: The National Academies Press.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing,16,* 470-481. http://dx.doi.org/10.1097/00003446-199510000-00004

Kroodsma, D.E. (1990). Using appropriate experimental designs for intended hypotheses in ‘song’ playbacks, with examples for testing effects of song repertoire sizes. *Animal Behavior,* 40, 1138-1150. https://doi.org/10.1016/S0003-3472(05)80180-0

Kuhl, P.K., & Miller, J.D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science,* 190, 69-72. http://dx.doi.org/10.1126/science.1166301

Kuhl, P.K., & Miller, J.D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63, 905-917. http://dx.doi.org/10.1121/1.381770

Macgregor, P.K. (2000). Playback experiments: Design and analysis. *Acta Ethologica*. 3*,* 3-8. DOI: 10.1007/s102110000023.

Malott, R. W., & Malott, M. K. (1970). Perception and stimulus generalization. In W. C. Stebbins (Ed.), *Animal psychophysics: The design and conduct of sensory experiments* (pp. 363–400). New York, NY: Appleton-Century-Crofts.

Nelson, D.A. & Kiester, T.E. (1978). Frequency discrimination in the chinchilla. *Journal of the Acoustical Society of America*, 64, 114-126. https://doi.org/10.1121/1.381977

Niemiec, A.J., Yost, W.A., & Shofner, W.P. (1992). Behavioral measures of frequency selectivity in the chinchilla. *Journal of the Acoustical Society of America*, 92: 2636-2649. https://doi.org/10.1121/1.404380

Ohlemiller, K.K., Jones, L.B., Heidbreder, A.F., Clark, W.W., & Miller, J.D. (1999). Voicing judgements by chinchillas trained with a reward paradigm. *Behavioural Brain Research*, 100, 185-195. https://doi.org/10.1016/S0166-4328(98)00130-2

Oksanen, L. (2001). Logic of experiments in ecology: is pseudoreplication a pseudoissue? *Oikos*, 94, 27–38. https://doi.org/10.1034/j.1600-0706.2001.11311.x

Pilley, J.W., & Reid, A.K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, 86: 184-195. https://doi.org/10.1016/j.beproc.2010.11.007

Ranasinghe, K.G., Vrana, W.A., Matney, C.J., & Kilgard, M.P. (2012). Neural mechanisms supporting robust discrimination of spectrally and temporally degraded speech. *Journal of the Association for Research in Otolaryngology*, 13: 527-542. https://doi.org/10.1007/s10162-012-0328-1

Remez, R.E., Rubin, P.E., Pisoni, D.B., & Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science*, 212, 847-950. http://dx.doi.org/10.1126/science.7233191

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S. & Lang, J.M. (1994). On the perceptual organization of speech. *Psychological Review*, 101, 129-156. http://dx.doi.org/10.1037/0033-295X.101.1.129

Rock, I., Lasker, A., & Simon, J. (1969). Stimulus ‘generalization’ as a process of recognition. *American Journal of Psychology*, 82, 1-22. http://www.jstor.org/stable/1420604

Saberi, K., & Perrott, D.R. (1999). Cognitive restoration of reversed speech. *Nature*, 398: 760. doi:10.1038/19652

Shannon, R.V. (2005) Speech and music have different requirements for spectral resolution. *International Review of Neurobiology*, 70, 121-134. http://dx.doi.org/10.1016/S0074-7742(05)70004-0

Sherbecoe, R.L. & Studebaker , G.A. (2004). Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units. *International Journal of Audiology*, 43, 442-448. DOI: 10.1080/14992020400050056

Shofner, W.P. (2000). Comparison of frequency discrimination thresholds for complex and single tones in chinchillas. *Hearing Research*, 149, 106-114. https://doi.org/10.1016/S0378-5955(00)00171-4

Shofner, W.P. (2011). Perception of the missing fundamental by chinchillas in the presence of low-pass masking noise. *Journal of the Association for Research in Otolaryngology*, 12, 101-112. DOI: 10.1007/s10162-010-0237-0

Shofner, W.P. (2014). Perception of degraded speech sounds differs in chinchilla and human listeners. *Journal of the Acoustical Society of America*, *1*35, 2065-2077. http://dx.doi.org/10.1121/1.4867362

Shofner, W.P. and Chaney, M. (2013). Processing Pitch in a Nonhuman Mammal (Chinchilla laniger). *Journal of Comparative Psychology*, 127, 142–153. http://dx.doi.org/10.1037/a0029734

Shofner, W.P. & Sheft, S. (1994). Detection of bandlimited noise masked by wideband noise in the chinchilla. *Hearing Research*, 77, 231-235, https://doi.org/10.1016/0378-5955(94)90271-2

Shofner, W.P. & Yost, W.A. (1997). Discrimination of rippled-spectrum noise from flat-spectrum noise by chinchillas: evidence for a spectral dominance region. *Hearing Research*, 110, 15-24. https://doi.org/10.1016/S0378-5955(97)00063-4

Shofner, W.P., Yost, W.A. & Sheft, S. (1993). Increment detection of bandlimited noises in the chinchilla. *Hearing Research*, 66, 67-80. https://doi.org/10.1016/0378-5955(93)90261-X

Teoh, S.W., Pisoni, D.B. & Miyamoto, R.T. (2004a). Cochlear implantation in adults with prelingual deafness. Part I. Clinical results. *Laryngoscope*, 114, 1536–1540. https://doi.org/10.1097/00005537-200409000-00006

Teoh, S.W., Pisoni, D.B. & Miyamoto, R.T. (2004b). Cochlear implantation in adults with prelingual deafness. Part II. Underlying constraints that affect audiological outcomes. *Laryngoscope*, 114, 1714-1719. https://doi.org/10.1097/00005537-200410000-00007

Trout, J.D. (2001). The biological basis of speech: What to infer from talking to animals. *Psychological Review*, 3, 523-549. http://dx.doi.org/10.1037/0033-295X.108.3.523

Uddin, M., Wildman, D.E., Lui, G., Xu, W., Johnson, R.M., Hof, P.R., Kapatos, G., Grossman, L.I., & Goodman, M. (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brane gene expression profiles. *Proceedings of the National Academy of Sciences, 101,* 2957-2962. http://dx.doi.org/10.1073/pnas.0308725100

Wang, G-d., Zhai, W., Yang, H-c., Fan, R-x., Cao, A., Zhong, L., Wang, L., Liu, F., Wu, H., Cheng, L-g., Poyarkov, A.D., Poyarlov,. N.A., Tang, S-s., Zhao, W-m., Gao., Y., Lv, X-m., Irwin, D.M., Savolainen, P., Wu, C-I., & Zhang, Y-p. (2013). The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nature Communications,* 4, 1-9. DOI: 10.1038/ncomms2814

Figure Legends

Figure 1. Time-domain waveforms (top) and narrowband spectrograms (bottom) are illustrated for the word "sit". Panels show the naturally-spoken word and noise-vocoded versions based on 2, 8 and 64 vocoder channels. The bands in the spectrograms corresponding to the first, second and third formant frequencies of the vowel are indicated by F1, F2 and F3, respectively.

Figure 2. Examples of temporal envelopes are illustrated for "cut" (A) and "sit" (B). Black lines illustrate the 2-channel NV words; red lines illustrate the naturally-spoken words. Envelopes were extract using Praat by half-wave rectification and low-pass filtering with an upper cut-off frequency of 100 Hz.

Figure 3. Behavioral responses obtained for individual chinchillas in the stimulus generalization task are shown for each of the 6 test words. Different symbols represent different test words. Labels along the abscissa indicate the naturally-spoken word (nat) and the number of channels for vocoded words. Percent responses are the number of positive responses divided by the number of trials multiplied by 100.

Figure 4. A: Mean responses obtained from chinchillas (red solid line & red filled circles) and human listeners (blue dashed line & blue filled squares) are compared. Naturally-spoken words were used as the signal in the stimulus generalization paradigm. Responses are averaged across individual listeners and across the 6 words. Error bars indicate ± 1 standard deviation of the mean. B: Mean RAUs obtained from chinchillas (red solid line & red filled squares) for C12,

C24, C36 and C47. Data for C15 were incomplete and not included in the ANOVA. Open blue triangles show mean RAUs when the data for C15 were also included. Error bars indicate 95% confidence intervals.

Figure 5. Responses obtained when 64-channel NV words were used as signals in the generalization task. A. Responses of 3 chinchillas averaged across animals and words are illustrated by the solid red line and red open circles. Error bars show ± 1 standard deviation. For comparison, the average responses from 16 human listeners are shown by the blue dashed line and blue filled squares. RAUs averaged across words are shown for C12 (B), C24 (C) and C47 (D). Red filled squares and red solid lines show responses obtained when the naturally-spoken words were used as signals in the generalization task. Blue filled circles and blue dashed lines show responses obtained when the 64-channel NV words were used as signals.

Figure 6. Example waveforms are illustrated for time-normal "cut" for 64-channel (A) and 2-channel (B) noise-vocoded versions and for 64-channel time-reversed "cut" (C). Corresponding envelopes (D) are shown for the 2-channel time-normal word (2-ch) by black solid line, for the 64-channel time-normal version (64-ch) by the red solid line and for the 64-channel time-reversed version (64r-ch) by the blue solid line.

Figure 7. Spectrograms are illustrated for time-normal (A) and time-reversed (B) 64-channel NV "cut". F1, F2, and F3 illustrate formant frequencies of the vowel.

Figure 8. Responses of 8 humans (left-hand column) and 3 chinchillas (right-hand column) to

time-normal and time-reversed noise-vocoded words. X-axis shows the number of vocoder channels. Red filled triangles and red dashed lines show responses to time-normal words; blue filled circles and blue solid line show response to time-reversed words. The gray dotted line in the right-hand column shows the responses obtained for the 2-channel standard in the generalization task.

Figure 9. (A) Bar graph illustrating responses of 3 chinchillas in the generalization task when the naturally-spoken "cap" was the signal. Stimuli presented are shown on the X-axis. The standard was the 2-channel NV "cap" (2-chan cap) and the signal was the naturally-spoken "cap" (cap) as illustrated by the diagonally striped bars. The responses to other naturally-spoken test words and two musical instruments are indicated in the graph by the filled bars: C12 in blue; C47 in green; C24 in red. (B) Bar graph illustrating mean RAUs from 3 chinchillas in the generalization task when the naturally-spoken "cap" was the signal. The standard was the 2-channel NV "cap" (2-chan cap) and the signal was the naturally-spoken "cap" (cap) as illustrated by the blue diagonally striped bars. The responses to other naturally-spoken test words and two musical instruments are indicated in the graph by the blue filled bars. Error bars show 95% confidence intervals.

Table 1.

Formant and fundamental frequencies of the unmodified, non-vocoded sounds

	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F0 (Hz)	1/F0 (ms)
‘ball’	846	1156	3025	4193	180.8	5.53
‘cap’	944	1862	2034	4889	193.2	5.18
‘cut’	800	2144	3223	4441	186.5	5.36
‘hot’	963	1655	2758	4310	184.5	5.42
‘meat’	395	3029	3099	4856	192.5	5.19
‘sit’	610	1203	2325	3286	200.5	4.99
‘wet’	1067	2407	3470	4410	197.8	5.06
Piano G^{b}3	753	1327	1871	2485	186.2	5.37
Cello G^{b}3	951	1960	2574	3889	187.7	5.33

Table 2.

2-factor analysis of variance for the effect of signal for C12, C24, C47

	Factor a Effect of signal (natural vs. 64-channel)	Factor b Effect of stimulus	Interaction a x b
C12	F(1, 60) = 1158.0 $p < 0.001$ $\eta^2 = 0.54$	F(1, 60) = 69.6 $p < 0.001$ $\eta^2 = 0.23$	F(1, 60) = 61.4 p = 1.52588E-05 $\eta^2 = 0.199$
C24	F(1, 60) = 532.4 $p < 0.001$ $\eta^2 = 0.35$	F(1, 60) = 99.9 $p < 0.001$ $\eta^2 = 0.47$	F(1, 60) = 27.1 p = 1.52588E-05 $\eta^2 = 0.126$
C47	F(1, 36) = 1256.7 $p < 0.001$ $\eta^2 = 0.55$	F(1, 36) = 65.9 $p < 0.001$ $\eta^2 = 0.202$	F(1, 36) = 73.2 p = 1.52588E-05 $\eta^2 = 0.225$

Table 3.

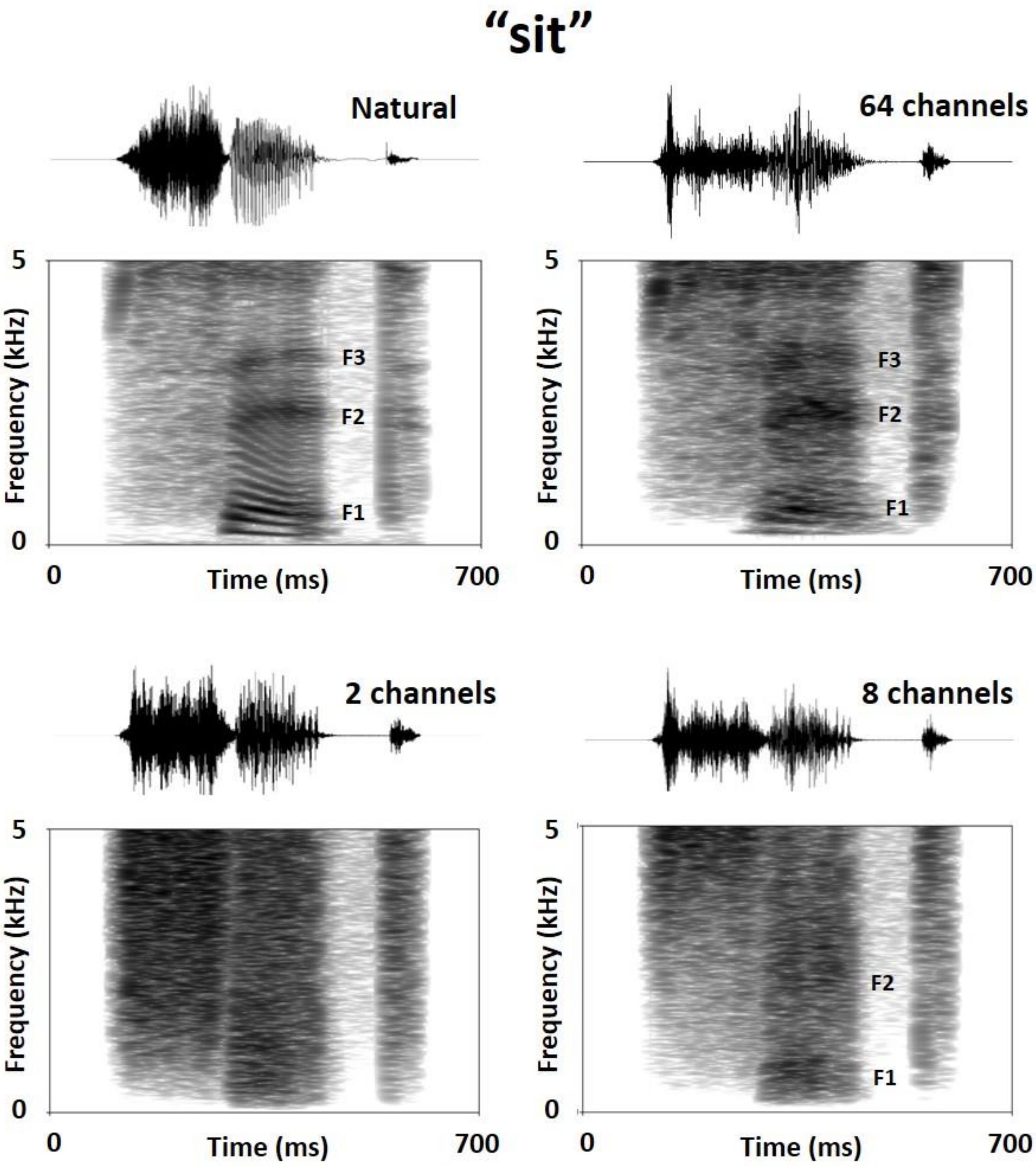
2-factor analysis of variance for time-reversed noise-vocoded words

	Factor a Effect of time reversing	Factor b Effect of number of channels	Interaction a x b
‘ball’	F(1,16) = 4.73 $p = 0.04$ $\eta^2 = 0.178$	F(1,16) = 1.14 $p = 0.30$ $\eta^2 = 0.128$	F(1,16) = 0.81 $p = 0.38$ $\eta^2 = 0.091$
‘cut’	F(1,16) = 47.9 $p < 0.001$ $\eta^2 = 0.584$	F(1,16) = 4.15 $p = 0.06$ $\eta^2 = 0.152$	(F(1,16) = 1.89 $p = 0.19$ $\eta^2 = 0.069$
‘hot’	F(1,16) = 38.2 $p < 0.001$ $\eta^2 = 0.638$	F(1,16) = 0.54 $p =0.48$ $\eta^2 = 0.027$	F(1,16) = 1.36 $p = 0.26$ $\eta^2 = 0.068$

Table 4.

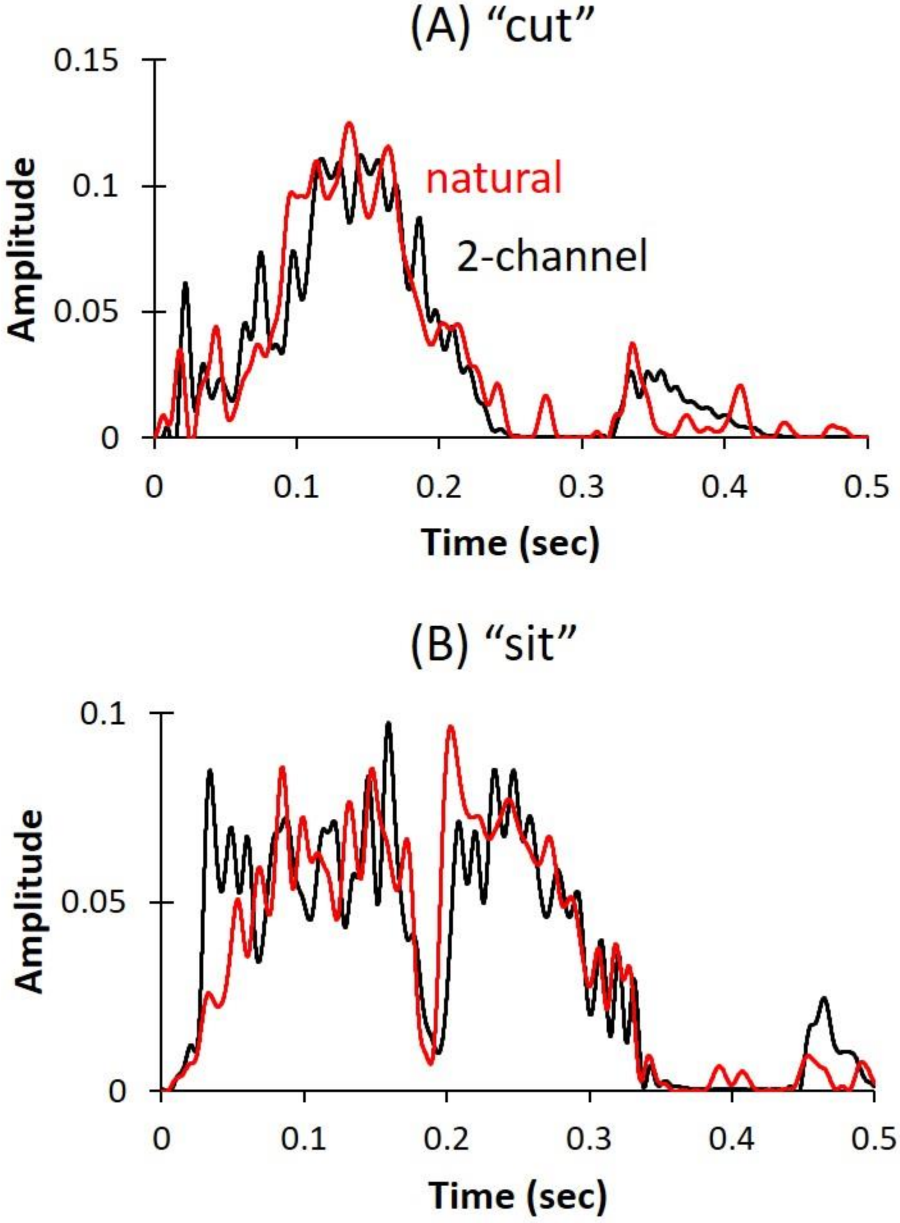
One-tailed t-test for time-reversed noise-vocoded words

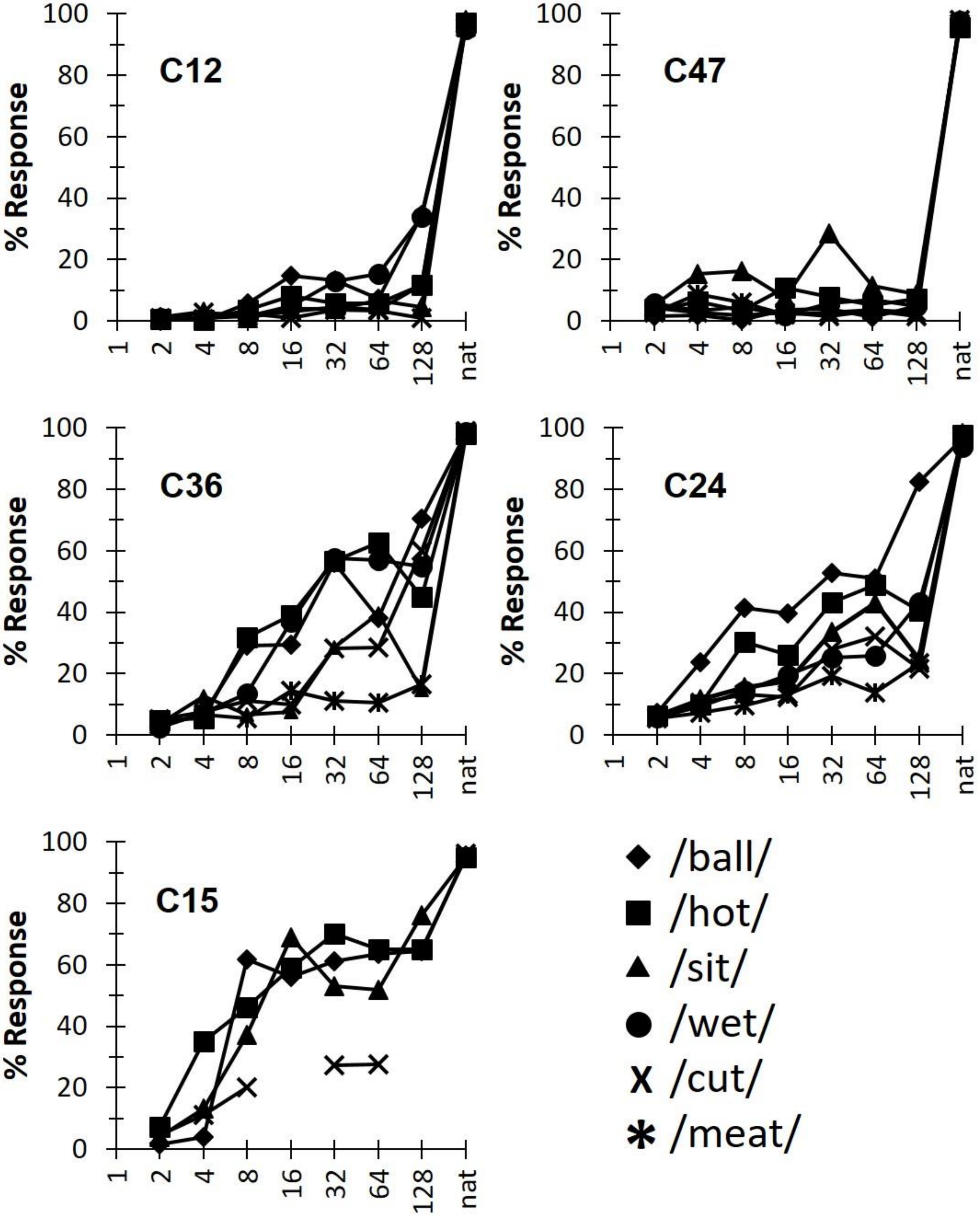
‘ball’	$t = 9.20$, $p < 0.001$	Cohen’s d = 6.00 ($\eta^2 = 0.9$)
‘cut’	$t = 6.90$, $p < 0.001$	Cohen’s d = 4.47 ($\eta^2 = 0.8332$)
‘hot’	$t = 7.10$, $p < 0.001$	Cohen’s d = 4.58 ($\eta^2 = 0.8398$)

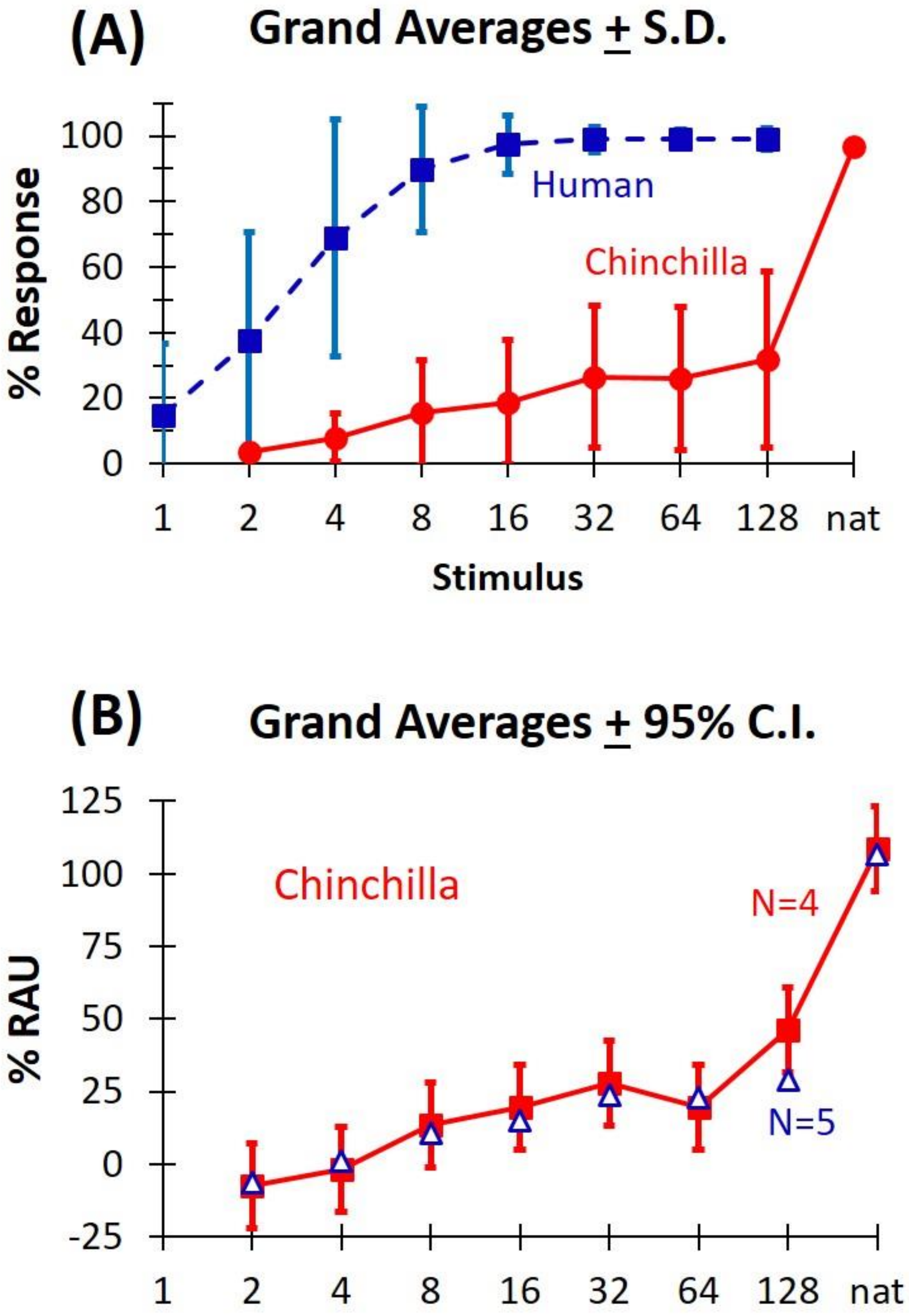

"sit"
Natural
64 channels
2 channels
8 channels
5
Frequency (kHz)
0
0
Time (ms)
700
F3
F2
F1

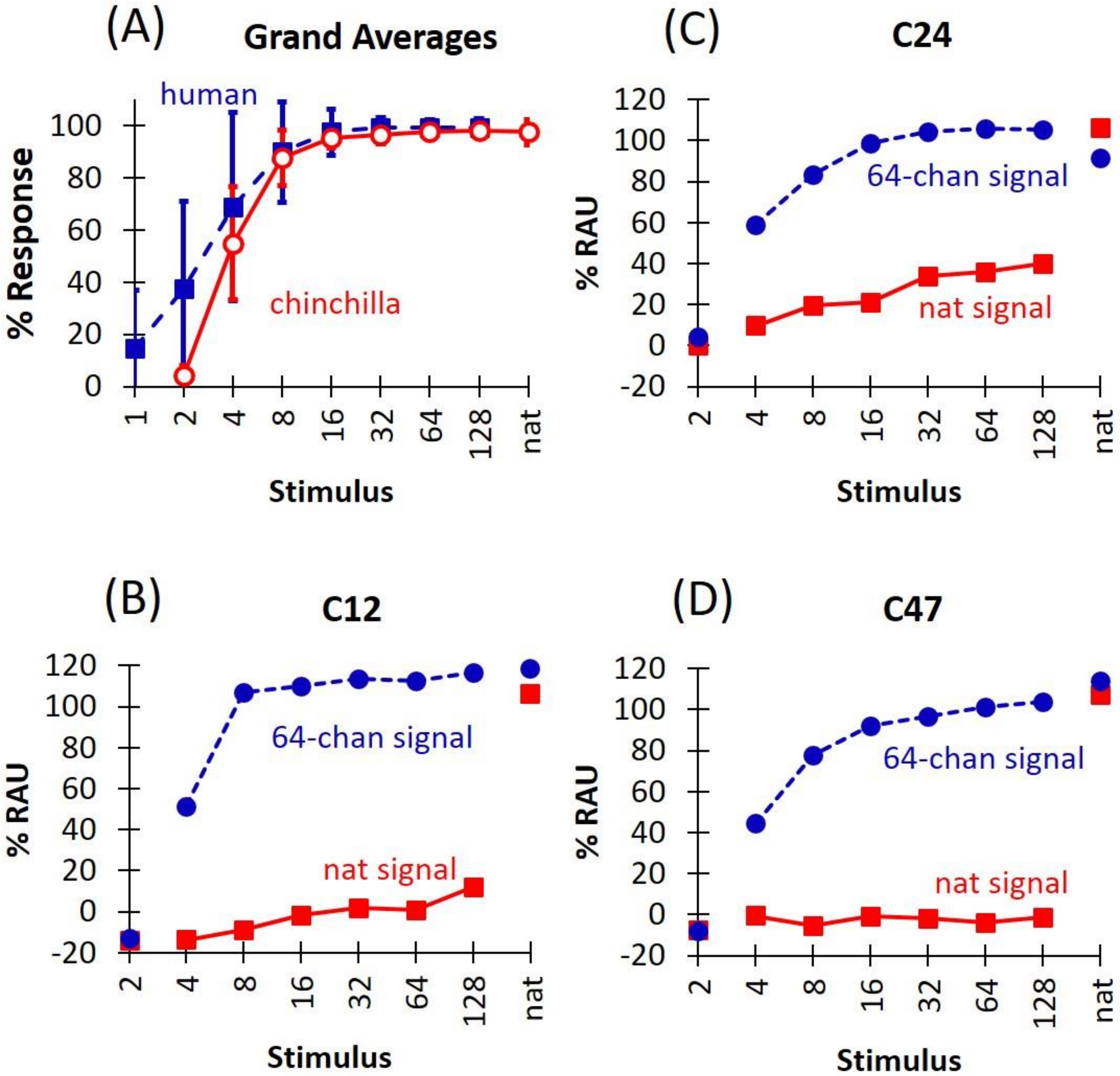
866

867


(A) “cut”
0.15
0.1
0.05
0
Amplitude
natural
2-channel
0
0.1
0.2
0.3
0.4
0.5
Time (sec)
(B) “sit”
0.1
0.05
0
Amplitude
0
0.1
0.2
0.3
0.4
0.5
Time (sec)

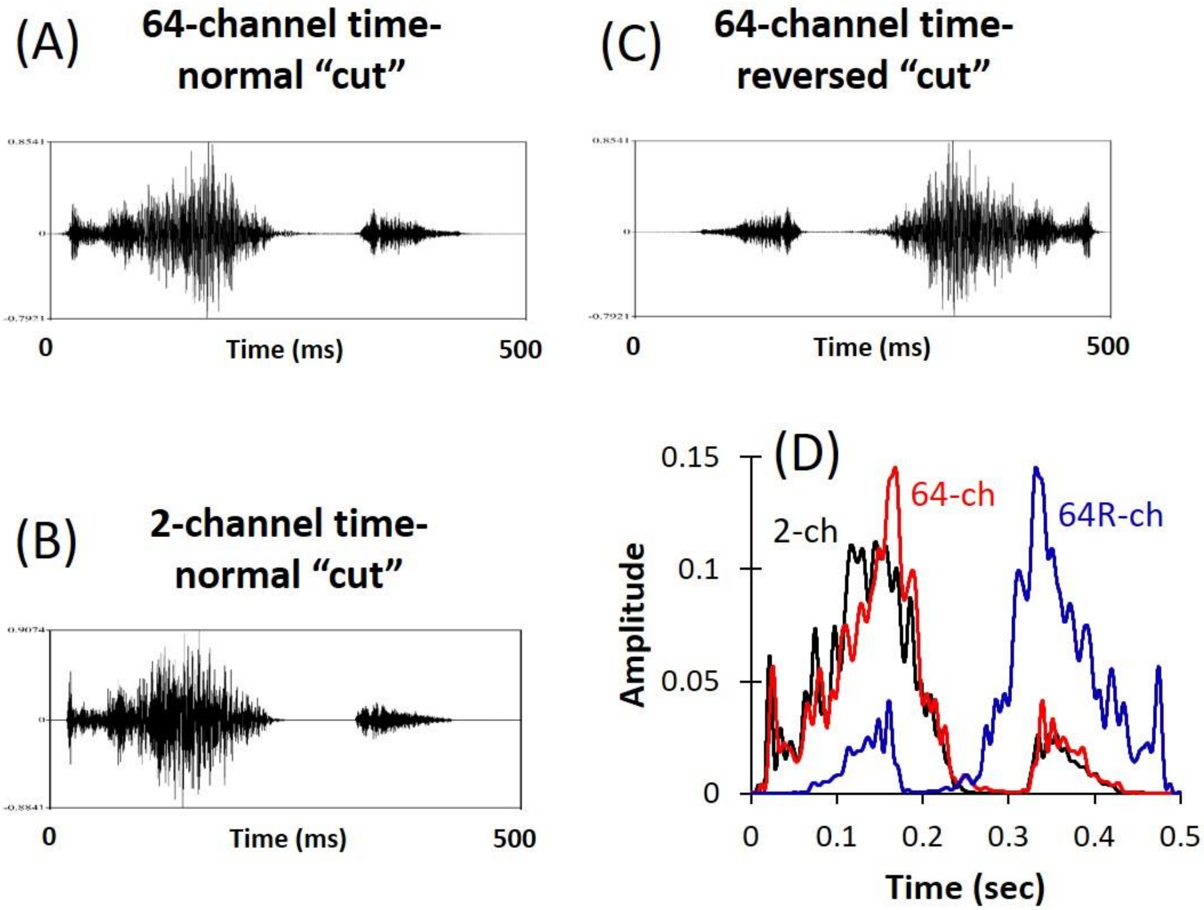

C12
C47
C36
C24
C15
% Response
0
20
40
60
80
100
1
2
4
8
16
32
64
128
nat
◆ /ball/
■ /hot/
▲ /sit/
● /wet/
X /cut/
✱ /meat/

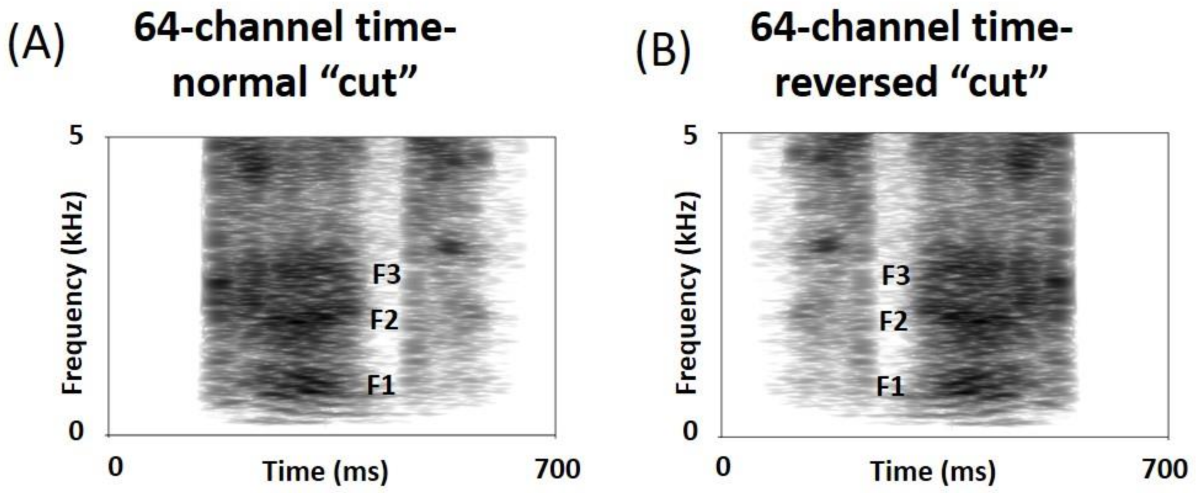

(A)
Grand Averages + S.D.
% Response
100
80
60
40
20
0
Human
Chinchilla
1
2
4
8
16
32
64
128
nat
Stimulus
(B)
Grand Averages + 95% C.I.
% RAU
125
100
75
50
25
0
-25
Chinchilla
N=4
N=5
1
2
4
8
16
32
64
128
nat

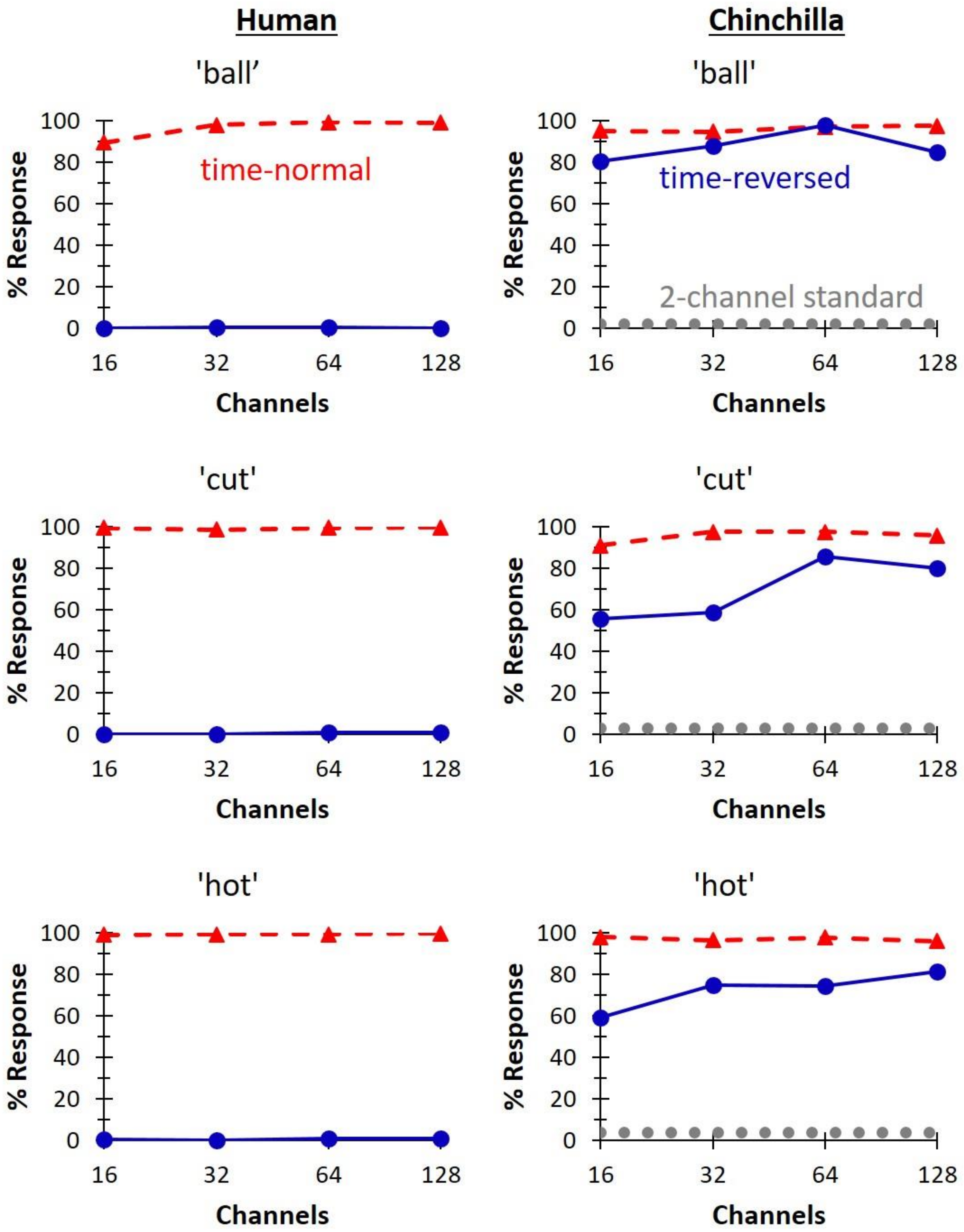

(A)
Grand Averages
human
chinchilla
% Response
100
80
60
40
20
0
1
2
4
8
16
32
64
128
nat
Stimulus
(C)
C24
120
100
80
60
40
20
0
-20
% RAU
64-chan signal
nat signal
2
4
8
16
32
64
128
nat
Stimulus
(B)
C12
120
100
80
60
40
20
0
-20
% RAU
64-chan signal
nat signal
2
4
8
16
32
64
128
nat
Stimulus
(D)
C47
120
100
80
60
40
20
0
-20
% RAU
64-chan signal
nat signal
2
4
8
16
32
64
128
nat
Stimulus

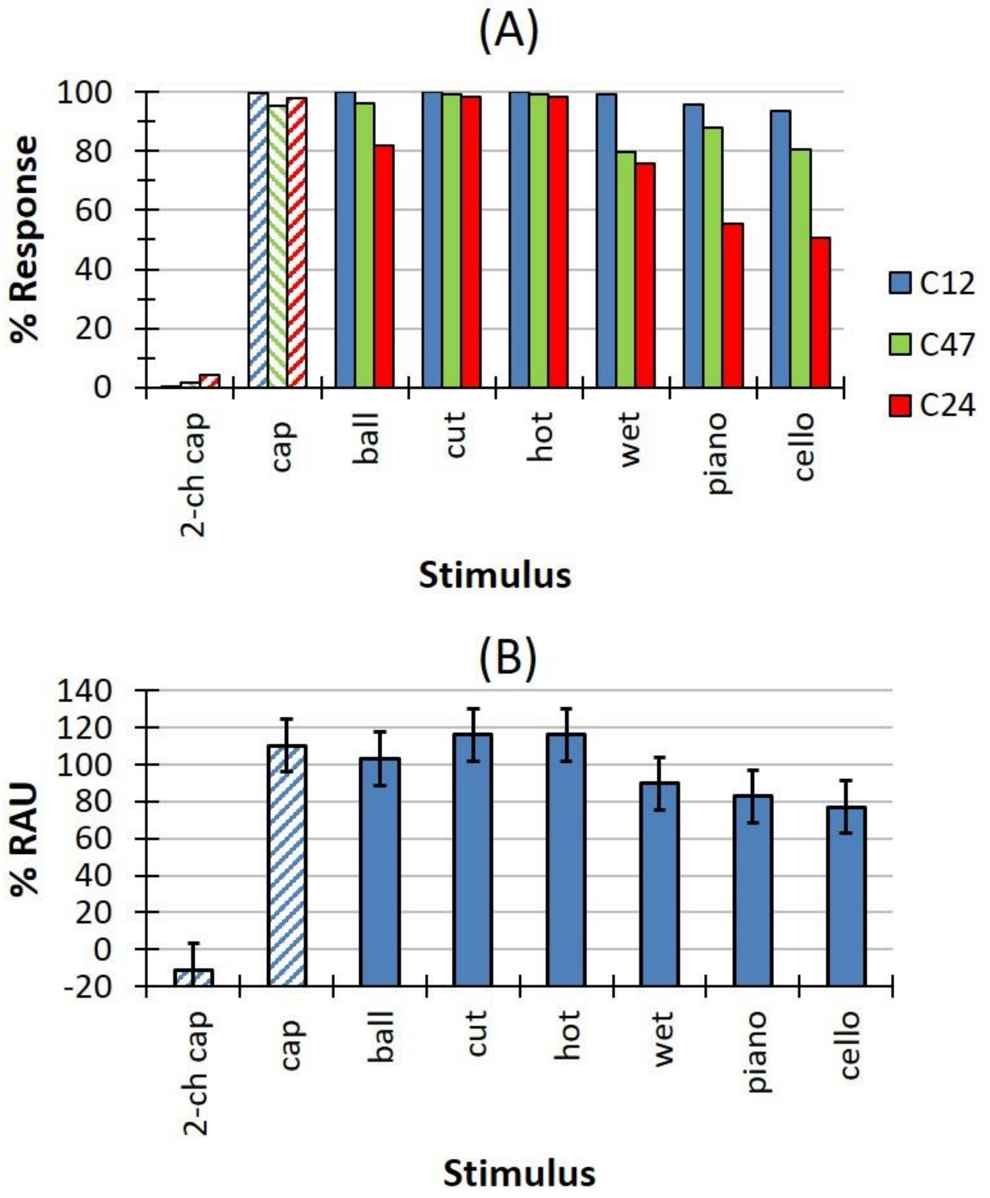
874

875


(A)
64-channel time-normal "cut"
0.8541
0
-0.7921
0
Time (ms)
500
(B)
2-channel time-normal "cut"
0.9074
0
-0.8841
0
Time (ms)
500
(C)
64-channel time-reversed "cut"
0.8541
0
-0.7921
0
Time (ms)
500
(D)
2-ch
64-ch
64R-ch
Amplitude
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
Time (sec)


(A)
64-channel time-normal “cut”
5
Frequency (kHz)
F3
F2
F1
0
0
Time (ms)
700
(B)
64-channel time-reversed “cut”
5
Frequency (kHz)
F3
F2
F1
0
0
Time (ms)
700


Human
Chinchilla
'ball’
'ball'
'cut'
'cut'
'hot'
'hot'
% Response
100
80
60
40
20
0
16
32
64
128
Channels
time-normal
time-reversed
2-channel standard


(A)
% Response
100
80
60
40
20
0
C12
C47
C24
2-ch cap
cap
ball
cut
hot
wet
piano
cello
Stimulus
(B)
% RAU
140
120
100
80
60
40
20
0
-20
2-ch cap
cap
ball
cut
hot
wet
piano
cello
Stimulus