

Detection of Tones in Reproducible Noise Maskers by Rabbits and Comparison to Detection by Humans

YAN GAI^{1,2}, LAUREL H. CARNEY^{1,2,3}, KRISTINA S. ABRAMS², FABIO IDROBO⁴,
J. MICHAEL HARRISON⁴, AND ROBERT H. GILKEY^{5,6},

¹*Department of Biomedical and Chemical Engineering, Syracuse University, Syracuse, NY 13244, USA*

²*Institute for Sensory Research, Syracuse University, Syracuse, NY 13244, USA*

³*Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244, USA*

⁴*Department of Psychology, Boston University, Boston, MA 02215, USA*

⁵*Department of Psychology, Wright State University, Dayton, OH 45435, USA*

⁶*Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Dayton, OH 45433, USA*

Received: 26 April 2007; Accepted: 10 August 2007; Online publication: 25 September 2007

ABSTRACT

Processing mechanisms used for detection of tones in noise can be revealed by using reproducible noise maskers and analyzing the pattern of results across masker waveforms. This study reports detection of a 500-Hz tone in broadband reproducible noise by rabbits using a set of masker waveforms for which human results are available. An appetitive-reinforcement, operant-conditioning procedure with bias control was used. Both fixed-level and roving-level noises were used to explore the utility of energy-related cues for detection. An energy-based detection model was able to partially explain the fixed-level results across reproducible noise waveforms for both rabbit and human. A multiple-channel energy model was able to explain fixed-level results, as well as the robust performance observed with roving-level noises. Further analysis using the energy model indicated a difference between species: human detection was influenced most by the noise spectrum surrounding the tone frequency, whereas rabbit detection was influenced most by the noise spectrum at frequencies above that of the tone. In addition, a temporal envelope-based model predicted detection by humans as well as the single-channel energy model did, but the envelope-based model failed to predict detection

by rabbits. This result indicates that the contributions of energy and temporal cues to auditory processing differ across species. Overall, these findings suggest that caution must be used when evaluating neural encoding mechanisms in one species on the basis of behavioral results in another.

Keywords: energy, envelope, operant, bias, masking

INTRODUCTION

Detection of tones in noise is a fundamental problem in auditory psychophysics and physiology. One approach to this problem is based on the finding that the detectability of a tone varies in a consistent way in response to different reproducible (frozen) noise maskers for human listeners (Pfaflin and Mathews 1966, Gilkey et al. 1985; Isabelle and Colburn 1991; Evisizer et al. 2002; Davidson et al. 2006). Although the variation in detectability across reproducible noise maskers can be partially explained by several models for psychophysical detection (Gilkey and Robinson 1986; Carney et al. 2002; Davidson et al. 2006), a complete understanding of the neural mechanisms used by the mammalian auditory system requires physiological studies, which are typically conducted

Correspondence to: Laurel H. Carney · Department of Neurobiology and Anatomy · University of Rochester · 601 Elmwood Ave., Box 603, Rochester, NY 14642, USA. Telephone: +1-585-2752591; fax: +1-585-7565334; email: laurel_carney@urmc.rochester.edu

in nonhuman animals. However, tone detection by humans and experimental animals may differ for various reasons, even when the same set of reproducible noise maskers is used. For example, the peripheral frequency tuning of most animals is likely to be broader than that of humans (for review, see Shera et al. 2002; but see Ruggero and Temchin 2005), which affects both energy- and timing-based features of neural responses to complex sounds. Central processing of tones in noise may also differ among humans and other mammalian species.

In the present study, the same reproducible noise maskers that were used in previous human psychophysical studies (Evilsizer et al. 2002; Davidson et al. 2006) were used in rabbit behavioral testing, allowing a direct comparison between species. Predictions of energy- and envelope-based detection models were also compared to the psychophysical detection results of both species.

This work builds from a previous study of detection in reproducible noise by rabbit, which used a Pavlovian-conditioning paradigm (Early et al. 2001; Zheng et al. 2002). However, problems with the previous study were identified by Mason et al. (2003), who showed that detection thresholds obtained with Pavlovian conditioning were elevated by undesired association of the masker with the unconditioned stimulus on tone-plus-noise trials and by extinction of conditioning on false-alarm trials. These effects also reduce the consistency of performance over time, making it difficult to obtain a stable pattern of detectability across reproducible maskers. Operant conditioning provides an alternative paradigm for studying detection in animals. Several previous studies have used “go/no-go” procedures for detection tasks (e.g., Peterson et al. 1977; Costalupes 1983; May et al. 1995). However, in our preliminary studies, these paradigms did not successfully maintain stimulus control in rabbit. Therefore, an operant-conditioning paradigm with a two-alternative-choice discrimination procedure was developed to provide reliable estimates of detection performance.

We describe a detailed analysis of rabbit behavioral results with fixed-level and roving-level reproducible-noise maskers. Detection threshold and detection patterns across a set of reproducible-noise maskers are compared to human results. Finally, we examine possible underlying detection mechanisms, based on the predictions of energy- and envelope-based models for masked detection.

EXPERIMENTAL METHODS

Animal subjects

Four female Dutch-Belted rabbits (R4, R10, R11, and R13) were used in this study. R4 (8 years old, 1.7 kg)

had experience in the previous Pavlovian-conditioning detection study (Zheng et al. 2002). The other three rabbits (1.5–2.5 years old, 1.8–2.0 kg) did not participate in the previous study. However, all four rabbits participated in several pilot studies of paradigms that led to the present operant procedure.

Food deprivation started several weeks before training to achieve 80% of *ad lib* feeding weight. Unrestricted water and a small amount of hay (Timothy) were provided daily, in addition to the food provided as reinforcement during testing (Purina Mills Test Diet: MRabbit Tablet 97 mg). Subjects were housed in an environment that had an average temperature of 19°C, 40–70% humidity, and a 12-h-on/12-h-off light cycle. Each subject ran a 1–2 h session every day between 8 AM and 4 PM. Typically, the subjects received their daily allotment of food (about 32–36 g) as reinforcement during the experiment; the session was ended when the subject consumed the daily allotment of food or when the session reached 2 h, whichever came first. When necessary, supplemental food (Scotts Distributing, Inc.: Prolab, Hi-Fiber Rabbit) was provided after the behavioral sessions to maintain approximately 80% *ad lib* feeding weight. These procedures were approved by the Syracuse University Institutional Animal Care and Use Community (IACUC).

Sound stimuli

On each trial, a Gaussian noise was presented either alone or with a 500-Hz tone that was gated on and off simultaneously with the noise. Both the noise and the tone were 300 ms in duration and had 10-ms cosine-squared onset/offset ramps. The noise on each trial was drawn from a set of 25 reproducible maskers that were used in previous studies of human detection (Evilsizer et al. 2002; Davidson et al. 2006). These noises had a frequency range of 100 to 3,000 Hz, and were presented at an average spectrum level of 40 dB SPL (corresponding to an overall noise level of 74.6 dB SPL). Note that, with assumed auditory-filter bandwidths for rabbits (see below), the choice of the 40-dB SPL spectrum level means the noise at the output of any auditory filter within the noise frequency range should have been audible (based on the rabbit audiogram; Heffner and Masterton 1980).

A Tucker-Davis Technologies (TDT, Gainesville, FL, USA) System III, under the control of Matlab (Mathworks, Natick, MA, USA), was used to generate sound stimuli that were presented via a Morel tweeter (MDT 33). An acoustic calibration was performed using a 1/2-inch condenser microphone (Brüel & Kjaer, type 4134) positioned in the empty cage approximately midway between the locations of the rabbit's two ears during a trial. Based on this calibra-

tion, the stimuli were digitally compensated such that the effective frequency response of the system was flat within the 100- to 3,000-Hz frequency range at the position of the animal's head. In the following text, "stimulus" refers to the compensated sound that was presented to the animal (neglecting changes in the acoustic waveform that may have been introduced by small head movements, etc.).

Experimental setup

The animal was placed in a wire hardware-cloth cage in a double-walled soundproof booth (Fig. 1, left). Three nose-poke response holes were arranged horizontally along the front panel of the cage. The speaker was located behind the front panel. A pair of infrared sensors (Emitter-Detector pair, Radio Shack) was positioned across the opening of each hole. A response was registered when the animal poked its nose in the hole. Food was delivered from a pellet dispenser (MED Associates), which was mounted behind the speaker, and transported through a plastic tube to a dish on the floor of the cage, beneath the center nose-poke hole. A lamp outside the cage provided the house light that was turned off during timeouts (see below).

Figure 1 (right) illustrates the front panel as seen by the animal during the experiment. An observing response (OR) consisted of a nose poke in the middle hole, which initiated a trial. The OR guaranteed an approximately consistent head position when the stimulus was presented. A reporting response (RR) was generated by poking the nose in the left or right hole. The LEDs and auditory clickers at the corners of the panel generated visual and audio signals based on the type of stimulus (noise or tone-plus-noise) that was just presented and the animal's response (described below).

Training

A single-interval two-alternative (nonforced) choice paradigm was used. Training took place in five stages. (1) Magazine training: the middle hole was blocked and a food pellet was dispensed whenever the animal made a RR by poking in the left or right hole. This training stage was completed within 5–60 min depending on the behavioral experience of each rabbit. (2) The middle hole was still blocked. A low-level random noise (e.g., 20-dB SPL spectrum level) or an 80-dB SPL tone was initiated and played continuously until the animal made a RR. A correct response, i.e., a response in the left hole for noise or a response in the right hole for a tone, was reinforced with a food pellet. An intertrial interval of 5 s was used for this stage of training. Animals were trained on this stage for 2 days. (3) The middle hole was unplugged. The stimulus was initiated by a nose poke in the middle hole; i.e., an OR. The stimulus remained on until a RR was produced. Response reinforcement was the same as in training stage 2. To prevent the animal from initiating a stimulus while still chewing food from the previous trial, a 1.5-s interval was added after food was dispensed, during which ORs were ignored. No additional intertrial interval was imposed. Animals were trained on this stage for 6–11 days. (4) The duration of the stimuli was decreased to 300 ms. A correct RR within a reaction-time window of 3 s after the offset of a stimulus was required for food reinforcement. An incorrect RR within the reaction-time window or a RR outside of the reaction-time window (called an R-in-quiet) caused a 3-s timeout, during which the house light was extinguished. After 80% correct performance was reached, which required approximately 2 days of training, a low-level noise was added to the tone stimulus, and the noise spectrum level was gradually increased to 40 dB SPL (i.e., to match the level of the noise-alone stimulus that would be used during testing).

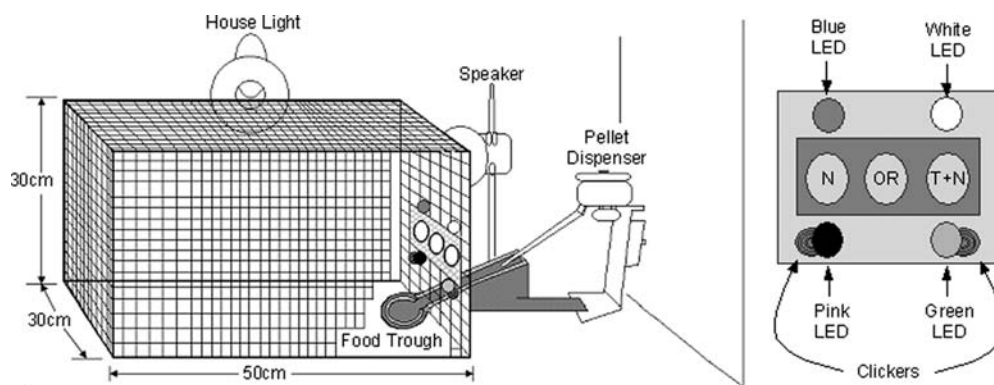


FIG. 1. Behavioral apparatus. *Left*, schematic diagram of the setup. Each response hole in the front panel has an inner diameter of 4 cm. *Right*, view of the front panel. OR, observing response. RR, reporting response. N, noise alone. T + N, tone plus noise. The clickers were used to generate auditory feedback.

The animal was trained to the 80% correct criterion after each noise-level increment (6–10 dB). This training stage required 1–2 months. (5) Finally, the tone level was varied using a two-down-one-up (2D1U) tracking procedure (Levitt 1971). Correct and incorrect RRs were followed by different visual and/or audio signals (described below). A set of 25 randomly generated reproducible noise maskers, which differed from the set of noise waveforms used later in the testing sessions, were used in stages 4 and 5 of the training. This training stage required approximately 2 months.

Definitions

A “hit” was a RR within the reaction-time window to the right hole for a tone-plus-noise trial; a “miss” was a RR made to the left hole within the reaction time window for a tone-plus-noise trial. A “correct rejection” was a RR within the reaction-time window to the left hole on a noise-alone trial; a “false alarm” was a RR to the right hole within the reaction-time window on a noise-alone trial. In the nonforced choice paradigm, if the subject failed to respond during the reaction-time window, a non-response was recorded, and a new trial could be initiated by a subsequent OR. Nonresponse trials represented only 0.6–1.0% of all trials across the four rabbits; these trials were uniformly distributed across masker waveforms.

The hit rate, P_H , was defined as the number of hits divided by the total number of hits and misses (the total number of hits and misses was not necessarily equal to the number of tone-plus-noise trials because the animal was not forced to respond). The false-alarm rate, P_{FA} , was defined as the number of false alarms divided by the total number of false alarms and correct rejections.

Testing

The testing procedure was generally the same as the final stage of training. A major difference was that the 25 reproducible noise waveforms used in previous human studies were used during testing. A 2D1U tracking procedure maintained approximately 71% correct performance. (The percent correct was defined as the number of correct responses, or hits plus correct rejections, adivided by the number of trials with responses.) A procedure for bias control is described in the next section.

Figure 2A shows a flow chart of the paradigm. An OR produced either a noise-alone or a tone-plus-noise trial with equal probability. The noise was randomly selected, without replacement, from the set of 25 reproducible-noise maskers; after each set of 25 trials, the sequence of noises was rerandomized. At the beginning of each session, the tone level was set slightly above the estimated mean level of reversals for each animal (Levitt 1971). After a miss, the tone level was increased by 0.5 dB. After two consecutive hits, the tone level was decreased by 0.5 dB. The tone level was not affected by responses on noise-alone trials or by nonresponse trials. When the animal did not respond on a trial, the stimulus was not repeated on the next trial; instead, another noise masker was selected with or without the tone. This procedure led to a slight variation in the total number of trials for each noise. Correct responses (hits and correct rejections) were reinforced with food pellets. A 5-s timeout followed each incorrect response (a miss or a false alarm), as well as each R-in-quiet. The length of the timeout was longer than that used in training to match the 5-s audio/visual signals delivered after each trial, described below.

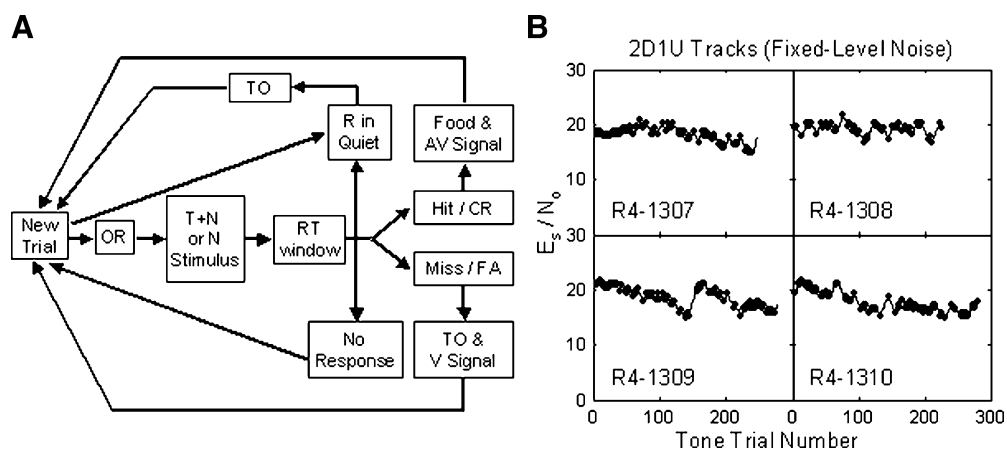


FIG. 2. **A** Schematic diagram of the final testing paradigm. *OR*, observing response. *RT*, reaction time. *CR*, correct rejection. *FA*, false alarm. *TO*, timeout. *AV signal*, audio/visual signal. *V signal*, visual signal. **B** Examples of four 2D1U tracks with fixed-level noise.

The dataset ID number is shown at the *bottom left of each panel*. The *lines* represent the E_s/N_0 values in the tone-plus-noise trials. Each symbol represents a reversal of the tone level. Noise-alone trials and Rs-in-quiet are not plotted.

Distinct audio and visual signals (5 s in duration) were given after different types of correct or incorrect RRs to provide information about the stimulus-response combination for the trial that was just completed. After a hit, a green LED on the lower-right corner of the front panel (Fig. 1, right) flashed for 5 s while a 5-s click train played. After a miss, the white LED on the upper-right corner flashed for 5 s while the house light was off during the timeout. After a correct rejection, the pink LED on the lower-left corner flashed for 5 s while a 5-s click train played. After a false alarm, the blue LED on the upper-left corner flashed for 5 s during the timeout. In short, the illuminated LED was always on the side where the correct RR should have been. The two click trains were not only in different locations but also had different temporal patterns for hits and for correct rejections, and the flashing LEDs also had different temporal patterns for different trial types. These signals were provided after each trial in an effort to increase the differences between the response consequences for different trial types, particularly those for which similar behavioral responses were involved. Increasing this difference has been suggested as a way to improve discriminability (Davison and Nevin 1999). These visual and audio signals were not provided after an R-in-quiet. ORs and RRs were ignored during the 5-s post-trial signal, which also served to allow time for the animal to eat when the previous trial was reinforced (the 1.5-s post-reinforcement interval used during later training stages was removed).

Bias control

The final training and testing sessions included a bias-control procedure. Bias is a possible uncontrolled outcome in any single-interval two-alternative choice test with 100% reinforcement because responses made on only one side would yield reinforcement for half the trials. Bias was defined as

$$B = -0.5[\mathbf{z}(P_H) + \mathbf{z}(P_{FA})] \quad (1)$$

(MacMillan and Creelman 2005). $\mathbf{z}(\cdot)$ is the z -score transformation.

We anticipated a bias to the left side (more correct rejections than hits, and more misses than false alarms) because noise was always present in the stimulus. To discourage bias to a particular side, the amount of food reinforcement was systematically varied depending upon a running estimate of bias (Davison and Tustin 1978). When the animal showed no bias, one pellet of food was dispensed for a correct RR on either side. When the animal was biased to one side, a percentage of correct RRs to the opposite side were reinforced with two pellets of

food. The two-pellet reinforcement ratio, R_f , was the percentage of tone-plus-noise trials with two pellets when $R_f > 0$, or the percentage of noise-alone trials with two pellets when $R_f < 0$. After each block of 50 trials, the bias was adjusted as follows: If the bias of the previous two blocks (100 trials) exceeded B_0 (i.e., a bias to the noise side), R_f was incremented by 5% (i.e., reinforcement for tone-plus-noise trials was increased). If the bias was below $-B_0$ (i.e., a bias to the tone-plus-noise side), R_f was decreased by 5% (i.e. reinforcement for noise-alone trials was increased). The limiting bias, B_0 , was chosen individually for each subject. B_0 was 0.4 for R4 and R13 and 0.3 for R10 and R11. (The larger B_0 value was used for the rabbits with naturally low bias because they were observed to be very sensitive to this manipulation.) If the absolute value of the bias was larger than 0.9, R_f was increased or decreased by 10%. In this manner, the total amount of reinforcement for the less preferred side was adjusted throughout the course of each session. As shown below, this strategy was effective in minimizing bias during testing.

Data analysis

Overall detection performance was quantified by the mean signal-to-noise ratio (E_s/N_0) for the reversals in each adaptive track, allowing direct comparison to previous human studies (Evliszler et al. 2002; Davidson et al. 2006). The detection threshold, E_s/N_0 , was computed as the average of the tone levels at the reversals (L_m) normalized by the noise spectrum level, N_0 , and the stimulus duration (300 ms) was referenced to a 1-s standard duration (Green and Swets 1966):

$$E_s/N_0(\text{dB}) = L_m(\text{dB SPL}) - N_0(\text{dB SPL}) \\ + 10 \log_{10}(0.3 \text{ s}/1 \text{ s}).$$

The P_H and P_{FA} for each of the 25 reproducible noise maskers were calculated to study the patterns of detectability across the set of tone-plus-noise and noise-alone waveforms; these patterns are referred to as detection patterns (Davidson et al. 2006). It is important to understand the meaning of the sample-dependent fluctuations in P_H and P_{FA} . Although it is possible to use P_H and P_{FA} to compute statistics such as d' and B on a per sample basis, these numbers have very different meanings than the same statistics computed across the ensemble of noise-alone and signal-plus-noise samples. For example, B will vary from sample to sample, but this *does not* indicate that the subject is altering their criterion on a sample-by-sample basis. Instead, we assume that the subject establishes a “fixed” criterion relative to the ensemble of samples (with some long-term variation as described

above, which is unrelated to the individual samples). The positions of the individual samples along the decision axis relative to criterion is determined, not by movement of the criterion (i.e., because of a sample-by-sample change in response bias), but because the physical statistics of stimuli vary from sample to sample, making them “sound” more like or less like they contain a signal. Our goal here was to use the positions of the individual noise-alone and signal-plus-noise samples relative to the criterion to determine which physical statistics the subject was using to perform the detection task.

A χ^2 test was used to determine if the observed differences in P_H and P_{FA} across waveforms were greater than would be expected by random variation alone (Siegel and Colburn 1989). All trials of all test sessions were included in the analyses presented below because detection patterns were found to be robust across tone level for the relatively small range of tone levels presented in the 2D1U tracks (see the example tracks in Figure 2B, described below).

Roving-level paradigm

In the conditions described above, the long-term average spectrum level of the masker was set at 40 dB SPL. However, if the level of the masker is varied across trials, a more stringent test of models is achieved. In particular, the roving-level paradigm provides an empirical test of the energy model because this manipulation makes the energy of a single filter an unreliable cue (Kidd et al. 1989). Two rabbits (R4 and R13) were tested with the roving-level paradigm after they completed the fixed-noise-level condition. In each tone-plus-noise or noise-alone trial, the noise spectrum level was randomly selected from a predetermined range (40±6 dB SPL, 1-dB step) for each trial. Consequently, the tone level normalized by the roved noise spectrum level was tracked by the 2D1U track. The E_s/N_0 , the mean-reversal tone level normalized by the roved noise level and by the stimulus duration, was compared to that of the fixed-level condition. The similarity between the detection patterns for fixed-level and roving-level conditions was quantified by the correlation of the hit or false-alarm rates between the two conditions.

Because it was possible that different cues were learned during the roving-level condition (Richards 1992), the fixed-level condition was repeated for several sessions after completion of the roving-level condition. If a subject had learned a different cue during the roving-level conditions, the detection threshold and/or detection patterns in the second test with fixed-level noise might have differed from those observed for the initial testing.

Human testing

Detection threshold and detection patterns based on the combined results of eight human subjects from previous studies (four from Evilsizer et al. 2002 and four from Davidson et al. 2006), referred to as the “average human subject”, were compared to the results in rabbit. However, there were several differences in the experimental designs between the human and rabbit studies. In the previous human studies, the tone level was fixed at a level that maintained a d' of approximately unity and no feedback was provided, whereas in the rabbit behavioral study, a 2D1U track with audio and/or visual signals after each response was used. In the previous human study, sound was delivered diotically using headphones with flat frequency responses over the noise bandwidth, whereas in this study, free-field sound was used with acoustic compensation to achieve a flat frequency response. To test whether human responses in the rabbit behavioral setup were consistent with previous results for human listeners, the first author of this study (H1) performed the detection task using the rabbit setup for four sessions, each with 600 trials. The human subject was positioned at the right side of the rabbit enclosure and responded manually. The audio/visual signals that followed responses were shortened, and the food reinforcement and time-outs were removed to reduce the duration of the sessions. The results were analyzed in the same manner as for the rabbits.

EXPERIMENTAL RESULTS

Detection threshold and detection patterns

Examples of four fixed-noise-level tracks are shown in Figure 2B. Each filled symbol indicates a reversal of the tone level that was determined by the 2D1U track. The detection threshold (E_s/N_0), bias (B), and two-pellet reinforcement ratio (R_t) for the fixed-level condition are shown in Table 1 for each rabbit. An ANOVA showed that there were significant differences among the threshold values [$F(3,94)=119.7$, $p<0.001$]. (Although analyses indicated that the underlying variances were heterogeneous, subsequent Monte Carlo simulations with the same sample sizes and degree of heterogeneity suggested only a slight impact on alpha). Scheffe tests indicated that all pairwise differences among rabbits were significant ($p<0.005$), except between R10 and R11, $p=0.36$. R4 showed the lowest threshold (18.8 dB), and R11 showed the highest threshold (25.0 dB). Compared to the range of thresholds observed for the eight previous human subjects (9.8–13.8 dB), the thresholds of the rabbits were elevated by 5 dB or more in all cases.

TABLE 1

Detection performance in rabbit with fixed-level noise				
	R4 (N=28)	R10 (N=15)	R11 (N=15)	R13 (N=40)
E_s/N_0 (dB)	18.8±0.7	24.2±1.2	25.0±1.2	22.8±1.4
B	-0.03±0.06	-0.02±0.10	-0.01±0.08	0.06±0.11
R_f (%)	0.6±4.1	28.9±7.6	44.8±2.7	1.7±2.9

E_s/N_0 , mean-reversal level normalized by the noise spectrum level and stimulus duration. R_f , two-pellet reinforcement ratio used for bias control. N , the number of sessions included. B , bias. The numbers are mean±standard deviations across sessions.

All four rabbits were unbiased when the bias-control procedure was used (their bias values were not significantly different from zero; t ranged from 0.1 to 0.5, df ranged from 14 to 39, $p > 0.05$). R10 and R11 consistently required two-pellet reinforcement for the tone-plus-noise side (the mean values of the R_f were 28.9 and 44.8% for R10 and R11, respectively), whereas the R_f values were not significantly different from zero for the other two rabbits ($t(27)=0.1$ for R4, $t(39)=0.6$ for R13, $p > 0.05$). However, even though the controlled bias was not significant over the course of the entire session, during portions of some sessions (e.g., over a stretch of 100 consecutive trials) the bias could vary from zero. For R10 and R11, there was a significant negative correlation between the detection threshold and the amount of food reinforced on the tone-plus-noise side normalized by the total amount of food received for whole sessions [$r^2=0.61$ and 0.55 for R10 and R11, respectively ($p < 0.05$)]. All correlations are presented as r^2 , the coefficient of determination, to be consistent with the quantification of model performance. The significance of r^2 was tested using the unsquared value of r .] When more reinforcement was provided on the tone-plus-noise side, the detection threshold tended to be lower. However, as shown in Table 1, the standard deviations of the threshold across sessions were comparable between the two rabbits that were biased (R10 and R11) and the two rabbits that were unbiased (R4 and R13), suggesting that changes in the rate of reinforcement on the tone-plus-noise side did not introduce significant variation in the threshold as compared to that caused by other unknown factors, such as internal variability of detection decisions. The correlation between threshold and R_f was not significant for either R4 or R13 ($r^2=0.15$ and 0.08 , $p > 0.05$).

Figure 3 shows the detection patterns in terms of hit and false-alarm rates for each rabbit. The growth of the χ^2 statistic with N_r , the averaged number of trials for each reproducible noise, was also plotted to show the consistency of the detection patterns across trials. An increase of χ^2 with trials indicated that the detection patterns were consistent from day to day; otherwise, the χ^2 would have remained the same or decreased over time. Notice that even though the χ^2

values grew with N_r for all rabbits, χ^2 increased at different rates for different rabbits (R4>R13>R11>R10). For example, R4's χ^2 values for both hits and false alarms reached the significance point ($df=24$, $p < 0.05$, marked by arrows, Figure 3, right panel) when N_r reached 10 trials, whereas R10's χ^2 values for hits were not significant until N_r reached 110 trials.

Another metric to quantify the consistency of the reproducible noise patterns is the split-half coefficient of determination, which is the square of correlation coefficient between detection patterns based on the first half and the second half of all the trials of all sessions. All split-half coefficients of determination were significant (Table 2), indicating that the patterns were consistent over time.

Table 3 shows the similarity of detection patterns across subjects (both rabbits and human), which was quantified by the inter-subject coefficient of determination. The r^2 values of hit rates across rabbits were all significant ($p < 0.05$) except between R10 and R11, which were weakly but not significantly correlated. For rabbits, the intersubject coefficients of determination for false-alarm rates were generally lower. Compared to the high intersubject coefficients of determination observed for human listeners (Evilsizer et al. 2002; Davidson et al. 2006), the rabbits showed greater diversity in their detection patterns. Part of the diversity may be because of the different detection mechanisms, and part may be a reflection of the greater variability in the rabbit behavioral results.

Comparisons with previous human results

Four sessions ($N_r=48$) obtained from H1 in the modified rabbit setup were compared to the average human subject ($N_r=64-96$ for each of the eight subjects) to examine potential differences in results between studies caused by the experimental setup. (H1 was not a participant in the previous human studies.) The E_s/N_0 for H1 (9.4 ± 0.7 dB) was near the low end of the range of thresholds measured in the previous human studies. As shown in Table 3, the coefficient of determination between H1's detection pattern and that of the average human subject in the previous

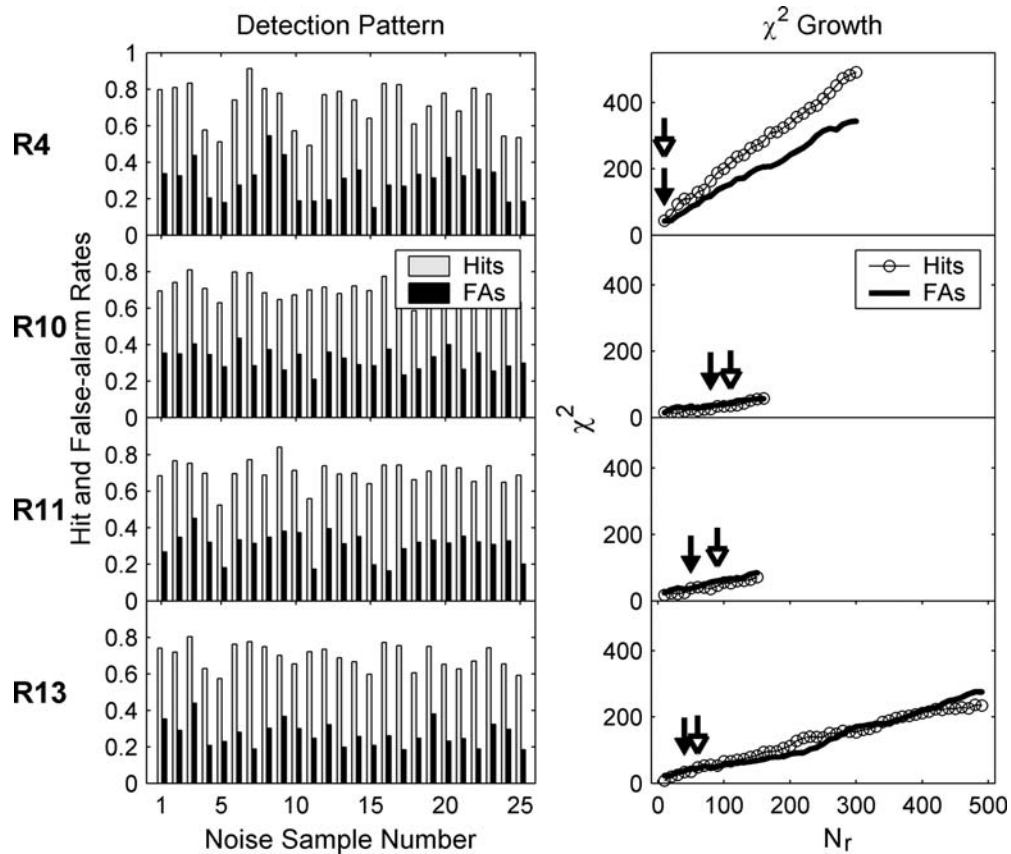


FIG. 3. Detection patterns and χ^2 growth for each individual rabbit. For each rabbit, the left column shows hit and false-alarm rates for the 25 noise waveforms; the right column shows the χ^2 values for hit and false-alarm rates. N_r , the averaged number of trials

for each reproducible noise waveform. The x axis resolution is 10 trials per noise. The arrows indicate when the χ^2 became significant ($df=24, p<0.05$) for hits (open arrows) and false alarms (filled arrows). FAs, false alarms.

studies was significant ($r^2=0.55$ for hits and 0.64 for false alarms, $df=23, p<0.001$). Based on these results, we conclude that the compensated acoustic waveforms used in the rabbit setup were well matched to those used in the previous headphone study in humans, and that the two paradigms were generally similar enough to yield the same detection patterns, at least for a human listener.

Table 3 also shows the coefficient of determination between each rabbit and the average human subject for hit and false-alarm rates. None of the r^2 values was significant ($p>0.05$). The r^2 values for comparisons between the rabbits and H1 were also not significant ($p>0.05$). To examine possible rea-

sons for this discrepancy, such as differences in the peripheral tuning of the two species, detection models were used to quantitatively study the detection patterns for both species (see below).

Roving-level results

There was little difference in detection results between the fixed- and roving-level conditions. Table 4 compares the detection thresholds between the fixed-level and roving-level conditions for R4 and R13. In one of the two comparisons, there was a small but significant increase in threshold (0.9 dB for R4; $t(38)=2.4, p<0.05$) for the roving-level condition. The mean

	R4 ($N_r=296$)	R10 ($N_r=154$)	R11 ($N_r=144$)	R13 ($N_r=484$)
r^2 (hits)	0.76 ^a	0.53 ^a	0.38 ^a	0.56 ^a
r^2 (false alarms)	0.60 ^a	0.40 ^a	0.38 ^a	0.81 ^a

N_r , the averaged number of trials for each reproducible noise waveform.

^aStatistically significant ($df=23, p<0.05$).

TABLE 3

Intersubject coefficient of determination, r^2 , of detection patterns for hits (FAs) with fixed-level noise					
	R10	R11	R13	H_{AVE}	H1
R4	0.39 ^a (0.08)	0.47 ^a (0.26 ^a)	0.51 ^a (0.13)	0.07 (0.00)	0.01 (0.05)
R10		0.14 (0.12)	0.44 ^a (0.08)	0.09 (0.06)	0.03 (0.06)
R11			0.23 ^a (0.27 ^a)	0.00 (0.00)	0.00 (0.00)
R13				0.05 (0.00)	0.01 (0.00)
H_{AVE}					0.55 ^a (0.64 ^a)

H_{AVE} , average human.

^aStatistically significant ($df=23$, $p<0.05$).

E_s/N_0 for R13 did not change significantly between the fixed-level and roving-level conditions ($t(50)=0.3$, $p>0.05$). In the second run of the fixed-level condition, thresholds were not significantly different from those of the original fixed-level condition for either R4 or R13 ($t(30)=1.4$ for R4, $t(43)=2.0$ for R13, $p>0.05$). The bias and R_f values for both rabbits were not significantly different from zero in any of the conditions (t test, $p>0.05$).

The coefficients of determination for comparisons of the patterns for roving-level and fixed-level results were 0.83 (hits) and 0.81 (false alarms) for R4, and 0.67 (hits) and 0.64 (false alarms) for R13. The coefficients of determination of the patterns between the first and second fixed-level conditions were 0.77 (hits) and 0.69 (false alarms) for R4, and 0.50 (hits) and 0.50 (false alarms) for R13. All of these r^2 values were significant ($df=23$, $p<0.001$). Because only four and five sessions were obtained in the second fixed-level condition for R4 and R13, respectively, the last four and five sessions of the first test were used in the comparison between the two fixed-level conditions. These results suggested that the rabbits used the same detection mechanism or two mechanisms with very similar outcomes in the fixed and roving-level conditions, and that if they used a single mechanism, it was a rove-resistant mechanism.

DETECTION MODELS AND RESULTS

Energy model

A single-channel energy-based model (Fletcher 1940; Gilkey and Robinson 1986; Davidson et al. 2006) was used to predict the detection patterns and thresholds obtained from each human and rabbit. As shown in Figure 4A, the output of the energy model was calculated as the root-mean-square (RMS) value of the noise-alone or noise-plus-tone waveforms after passing through a fourth-order gammatone filter. The variation in detection patterns across rabbits (Table 3) suggested that the parameters of detection models must vary substantially to account for individual performance; therefore, a wide range of model filter bandwidths and center frequencies were tested. The model was tested with the same 25 reproducible-noise waveforms, but with fixed-level tones. The RMS of the filter output was recorded as E_m or E_n , for a tone-plus-noise or noise-alone trial, respectively. An individual subject's threshold determined the tone level used by the model to obtain predicted detection patterns for that subject. In this study, the j^{th} component of a detection pattern for hits or false alarms was the E_m or E_n for the j^{th} reproducible noise waveform, with or without the tone ($j \in 1, 2, \dots, 25$), respectively. That is, the model detection patterns were composed of

TABLE 4

Detection performance in rabbit with fixed-level and roving-level noise					
Condition	No rove (1st)		Rove	No rove (2nd)	
	(N=28)	(N=40)	(N=12)	(N=4)	(N=5)
R4	E_s/N_0 (dB)	18.8±0.7	19.7±1.2	19.2±0.5	
	B	-0.03±0.06	0.01±0.11	0.01±0.06	
	R_f (%)	0.6±4.1	0.9±2.4	0.4±2.7	
R13	E_s/N_0 (dB)		22.8±1.4	22.7±0.9	21.8±1.0
	B		0.06±0.11	0.07±0.14	-0.01±0.1
	R_f (%)		1.7±2.9	3.1±2.5	4.5±4.4

No rove, fixed-level noise. Rove, roving-level noise. E_s/N_0 , mean-reversal level normalized by the noise spectrum level and referenced to a standard 1-s stimulus duration. R_f , two-pellet reinforcement ratio. B , bias. N , the number of sessions included. The numbers are mean values ± standard deviations across sessions.

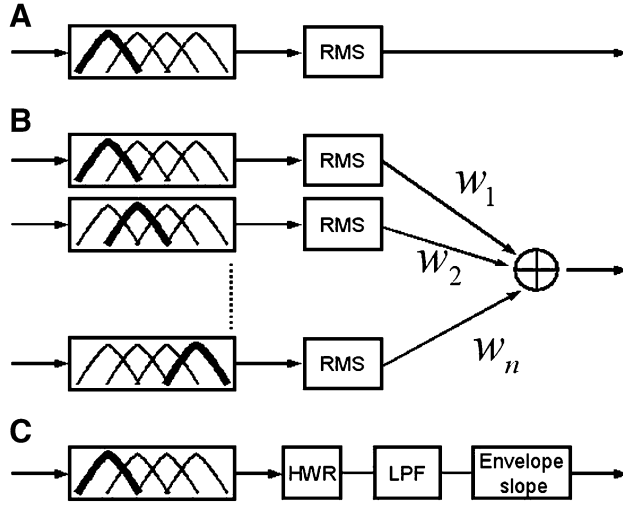


FIG. 4. Schematic diagram of the single-channel energy (A), multiple detector (B), and envelope-slope (C) models (replotted from Davidson et al. 2006, their Fig. 2, with slight modifications). In A and C, a bank of fourth-order gammatone filters with different CFs is plotted (*thin and thick lines*), but only one filter (*thick lines*) was used by the model. In B, the model output was a weighted sum of the energy from filters with different CFs. RMS, root mean square. HWR, half-wave rectification. LPF, low-pass filter with 250-Hz cutoff frequency.

decision variables, not hit or false-alarm rates, to avoid having to make assumptions about the level of internal variability (i.e., internal noise) and choose a detection criterion for the model.

Because the detection patterns obtained from the behavioral data were represented by hit and false-alarm rates, a z -score transformation was applied to each component of the empirical detection patterns to enable comparison with the model detection patterns. The similarity between the model results and data was quantified by the coefficient of determination (r^2) between the model detection patterns and the transformed-data detection patterns (e.g., Davidson et al. 2006), which represents the proportion of variance in the empirical detection patterns that was predicted by the model.

The model responses were computed using gammatone filters with center frequencies (CFs) that varied from 400 to 800 Hz (in 5-Hz steps) and for filter bandwidths that varied from 50 to 800 Hz (in 0.1-octave steps), to avoid making assumptions about detection mechanisms or the sharpness of peripheral tuning in either species. The coefficients of determination between model and data detection patterns for different model parameters provide information about the range of frequencies in the stimulus that influenced the behavioral responses. The CF and bandwidth of the filter that yielded the best prediction of the empirical detection pattern was compared to the tone frequency (500 Hz) and the presumed bandwidth of peripheral tuning for rabbits and human.

The bandwidth of a gammatone filter is determined by the parameter τ , which is related to the equivalent rectangular bandwidth (ERB) of the auditory filter (Patterson et al. 1988) as follows:

$$\tau = 1/(2\pi \times 1.019 \times \text{ERB}) \quad (2)$$

The presumed human ERB at a center frequency of 500 Hz is 75 Hz, derived from psychophysical studies of masked detection (Glasberg and Moore 1990). For rabbits, a behavioral estimate of auditory-filter bandwidth is not available; however, the peripheral tuning in rabbit can be inferred from the tuning curve of auditory nerve fibers. The sharpness of the tuning curve is measured by Q_{10} , which is the CF divided by the bandwidth (BW_{10}) of the tuning curve 10 dB above the minimum threshold (i.e., the threshold at CF). Q_{10} for rabbits at 500 Hz is 1.75, derived from a linear fit (Eq. 3) to tuning curves for auditory nerve fibers in rabbit (Borg et al. 1988).

$$Q_{10} = 10^{0.35 \log_{10}(\text{CF}/1000) + 0.35} \quad (3)$$

The relationship between Q_{10} and τ can be derived by fitting the gammatone filter to the tuning curve of auditory nerve fibers (Zhang et al. 2001):

$$\tau = 1/(2\pi \times 0.5 \times BW_{10}) \quad (4)$$

Based on the fit described in Eq. 3, and the relationships between Q_{10} and τ , the gammatone filter that describes rabbit physiological tuning at 500 Hz has $\tau=1.1$ ms. Using the relationship between τ and ERB for human (Eq. 2), we can estimate a presumed rabbit ERB at 500 Hz of 140 Hz. This estimate of a presumed rabbit ERB from measures of physiological tuning was based on several assumptions about the relationship between ERBs and peripheral tuning (which are not known for either rabbit or human), but these assumptions did not affect any of the model outcomes or interpretations presented in this study. The use of ERBs for both species allowed a convenient presentation of model results in terms of filter bandwidths (in Hertz).

The model detection threshold was also estimated based on responses of a single filter with CF=500 Hz and the presumed human or rabbit ERB, unless otherwise specified. The discriminability, or d' , was calculated as

$$d' = \frac{M_{\text{tn}} - M_{\text{n}}}{\sqrt{(s_{\text{tn}}^2 + s_{\text{n}}^2)/2}}, \quad (5)$$

where M_{tn} and M_{n} are the mean values of E_{tn} and E_{n} , respectively, and s_{tn}^2 and s_{n}^2 are the sample variances of E_{tn} and E_{n} , respectively. The tone level that yielded

$d'=1$, normalized by the noise spectrum level and stimulus duration (E_s/N_0), was defined as the model detection threshold and was compared to the threshold estimated for each subject. It is important to note that this analysis is distinct from the sample-by-sample analysis based on the detection patterns. The thresholds for the model and the subject are determined across the ensemble of samples. It is quite possible for the model to make accurate predictions at the ensemble level and inaccurate sample-by-sample predictions, or vice versa. Indeed, because we have not included any internal variability in the models, they should (and do) consistently underestimate the subjects' thresholds. Moreover, because of this, when the model detection patterns were compared to the subject detection patterns, E_s/N_0 was set to the subject threshold, not the model threshold. In this way, the comparison was made using the same stimuli for the model and the subject.

Figure 5 shows the proportion of the variance in the human detection patterns that was predicted by the energy model (top, H1; bottom, the average human subject; left, hits; right, false alarms). The filter CF and bandwidth (ERB) were varied along the x axis and y axis, respectively. The lightness of the color indicates how well the energy filter with a certain CF and bandwidth explained the variation in the reproducible-noise detection pattern for each subject. The

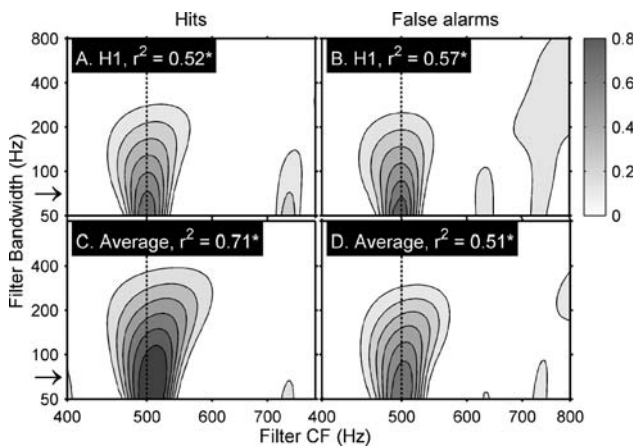


FIG. 5. Proportion of variance (r^2) of human detection patterns predicted by the energy model (hits, *left* and false alarms, *right*). The lightness of the shading represents the r^2 value, as indicated by the grayscale bar at the right. The x axis and y axis are the center frequency (CF) and bandwidth of the auditory filter, respectively. The frequency resolution on the x axis and y axis are 5 Hz and 0.1 octave, respectively. The *arrows* on the y axis indicate the presumed human ERB (75 Hz) for CF=500 Hz. The *vertical dashed line* marks the center frequency of 500 Hz. The r^2 value shown at the top of each panel is the maximum coefficient of determination between model and human detection patterns as a function of CF and bandwidth. *H1*, the human subject in this study. *Average*, the average human subject (Davidson et al. 2006). *Asterisk*, statistically significant ($df=23$, $p<0.001$).

maximum r^2 across filter CF and bandwidth (shown at the top of each panel) was significant for all four panels ($p<0.001$). The maximum r^2 values were obtained with filter CFs close to the tone frequency (500 Hz), but at filter bandwidths that tended to be smaller than the presumed human ERB (which is shown by the arrows, Fig. 5). These results suggested that the noise spectrum surrounding the tone frequency contributed to the human detection patterns more than the noise spectrum farther away from the tone frequency, as would be expected if the subjects were “attending” to the tone frequency. The implications of this sharp effective tuning are discussed in the next section. The model threshold (E_s/N_0) was 7.6 dB, predicted using a filter with 500-Hz CF and 75-Hz ERB, which was 2.2 to 6.2 dB lower than individual human thresholds. The model threshold would have been higher if internal variability were included in the model (Siegel and Colburn 1989; Davidson et al. 2006).

Figure 6 shows the proportion of variance in the rabbit detection patterns that was predicted by the energy model, with varying model CF and bandwidth. The maximum r^2 values were all significant, except with the false alarm data of R10 that were not explained by the energy in any filter. Note that because the rabbits had lower split-half coefficients of determination, the total amount of predictable variance (V_p , Ahumada and Lovell 1971) of the rabbit data (0.75–0.95) was generally smaller than that of the human data (0.91–0.95). In other words, the consistency of the rabbits' detection patterns set an upper limit on the model's ability to predict them.

It is interesting that, for rabbit, the highest r^2 values were not achieved based on the output of a filter with CF=500 Hz (vertical dashed line), which was the frequency of the tone. Instead, all four rabbits showed consistent shifts to the higher frequency side (except R10 on noise-alone trials; recall that her false alarms were not significantly correlated with model results). The CFs of the filters that yielded the best predictions for hits were 575, 555, 580, and 580 Hz for R4, R10, R11, and R13, respectively. The CFs of the filters that yielded the best predictions for false alarms were 640, 590, and 590 Hz for R4, R11 and R13, respectively. For R10, it was not clear why the energy model partially predicted the detection patterns of hits (Fig. 6C), but not false alarms, as seen from the absence of a systematically varied pattern of r^2 values with CF and bandwidth (Fig. 6D). Different subjects also showed different broadness of tuning (ranging from approximately 70 to 350 Hz), which could be larger (R4) or smaller (R10, R11, and R13) than the presumed rabbit ERB based on physiological tuning (140 Hz, as indicated by the arrow on the y axis, Fig. 6).

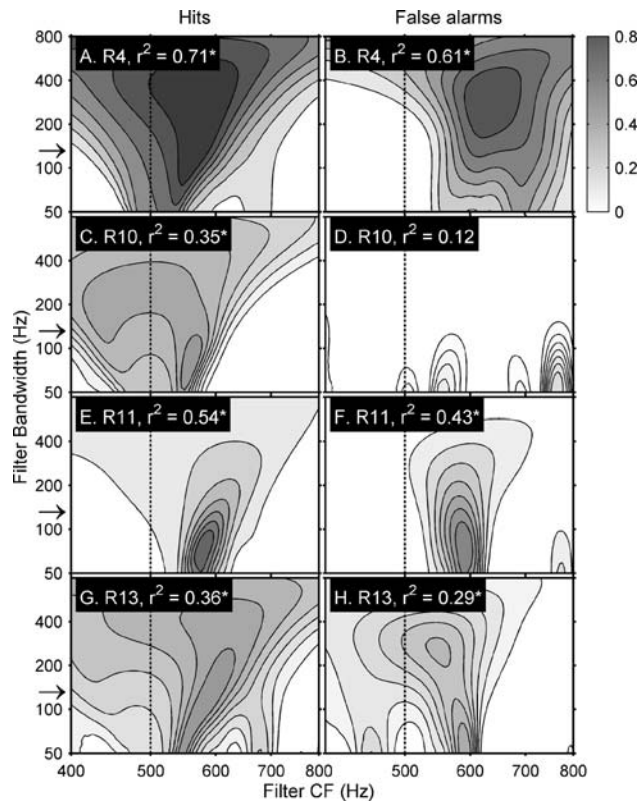


FIG. 6. Proportion of variance (r^2) in rabbit detection patterns predicted by the energy model (same format as Fig. 5). The arrows on the y axis indicate the presumed rabbit ERB (140 Hz) for CF=500 Hz. Asterisk, statistically significant ($df=23$, $p<0.05$).

The model threshold (E_s/N_0) was 8.2 dB using a 500-Hz-CF and 140-Hz-ERB filter, which was substantially lower than the rabbit thresholds (mean E_s/N_0 values were 18.8, 24.2, 25, and 22.8 dB, for R4, R10, R11, and R13, respectively; Table 1). The high thresholds obtained from the rabbits can be partly explained by the shift of the filter CF away from the tone frequency. Using the CFs and bandwidths that yielded the best predictions of detection patterns for hits for R4, R10, R11, and R13 (rather than the 500-Hz-CF 140-Hz-ERB filter), the energy model predicted thresholds of 10.7, 11.5, 16, and 12.4 dB. (Recall also, that the model did not include any internal variability.)

As stated previously, the roving-level paradigm provides a more stringent test of the energy model. For the two rabbits that participated in the roving-level condition, no more than 0.9-dB threshold (E_s/N_0) elevation was found when compared to the fixed-level condition. According to Green (1988), a single-channel energy model requires an increase of a quarter of the rove range in the level of the tone-plus-noise stimulus to maintain its detection performance when the noise level is roved. Increasing the level of the tone-plus-noise stimulus (after peripheral filtering) by 3 dB (a quarter of the 12-dB rove range)

corresponds to increasing the E_s/N_0 by 10.4 dB. (This value is similar to the value from Green's Figs. 1–4, and was confirmed by simulations of the single-channel model in the present study). Thus, the very small changes in threshold observed for the roving-level condition were not consistent with the single-channel energy model. In the next section, a modified multiple-detector model was explored as a possibility to account for the rove-resistance seen in the rabbit behavioral data.

Modified multiple-detector model

In addition to various single-channel energy-related models, Gilkey and Robinson (1986) employed a multiple-detector (MD) model. Figure 4B illustrates an MD model that forms a weighted combination of the energy values at the output of a bank of auditory filters with adjacent CFs. The best-fitting filter weights are typically positive at the signal frequency but negative for nearby filters (Gilkey and Robinson 1986; Davidson et al. 2006). This MD model predicts human detection patterns better than the single-channel energy model. Moreover, because the negatively weighted filters provide a rough estimate of the noise level (the quality of the estimate depends, in part, on the distance of the negatively weighted filters from the tone frequency), the MD model is less affected by the roving-level paradigm (Davidson et al. 2006). Note also that the best filter bandwidth for the single-channel energy model when fit to the human behavioral data was sharper than the presumed ERB (see Fig. 5). This result is compatible with the MD model, which predicts effective sharpening of the central positively weighted filter due to nearby negatively weighted filters.

However, in contrast to the results for the human data, the MD model did not yield appreciably more accurate predictions of the rabbit detection patterns compared to the single-channel model, and sizable negative weights were not obtained when fitting the MD model to the data (results not shown here). Recall, however, that the single-channel model predicted dramatically higher thresholds in the roving-level condition, whereas there was little or no impact of the rove on the rabbit thresholds. In an attempt to predict both the detection patterns and the minimal impact of the roving paradigm on detection threshold, a modified multiple-detector model (MMD) was developed using a bank of weak negatively weighted filters that were centered relatively far from the positively weighted filter. With this configuration, the overall model results were dominated by the positively weighted filter (i.e., it made predictions similar to those of the single-channel energy model), but showed the desired rove-resistance.

Figure 7 shows the linear weighting function of the MMD model with a single positively weighted filter and 15 negatively weighted filters. Each symbol indicates a single-channel energy filter. The ERB of the negatively weighted filters were set equal to the ERB of the positively weighted filter. Note that the weighting function was not derived with the fitting algorithm used in the work of Gilkey and Robinson (1986). Rather, the weights were selected arbitrarily such that they summed to zero (the sum of the weights must be approximately zero for the model to be robust in the roving-level paradigm; Davidson et al. 2006), and the negative weights were all equal. The CFs of the two negatively weighted filters surrounding the positively weighted filter were two ERBs away from the CF of the positively weighted filter, so that the responses of the adjacent filters were not highly correlated to the positively weighted filter. The difference in CF between adjacent negatively weighted filters was 1 ERB, and their CFs were limited to the band of the noise spectrum (100–3,000 Hz). The larger the number of negatively weighted filters, the closer the overall model results were to that of the positively weighted filter. A relatively small number (15) of negatively weighted filters that yielded reasonable model results (Fig. 8) was chosen.

Figure 8 shows the proportion of variance of the rabbit detection patterns that was predicted by the MMD model. The CF of the MMD model, CF_{MMD} , was defined as the CF of the positively weighted filter. In general, the MMD model predicted a comparable

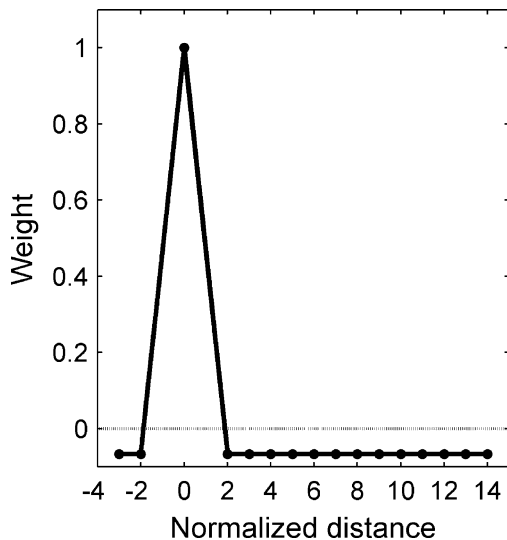


FIG. 7. Weighting function for the MMD model with one positively weighted and 15 negatively weighted channels. Each symbol represents a filter. The ERB of all filters was the same, matched to that of the positively weighted filter. The x axis is the difference in CF between each filter and the positively weighted filter, divided by ERB.

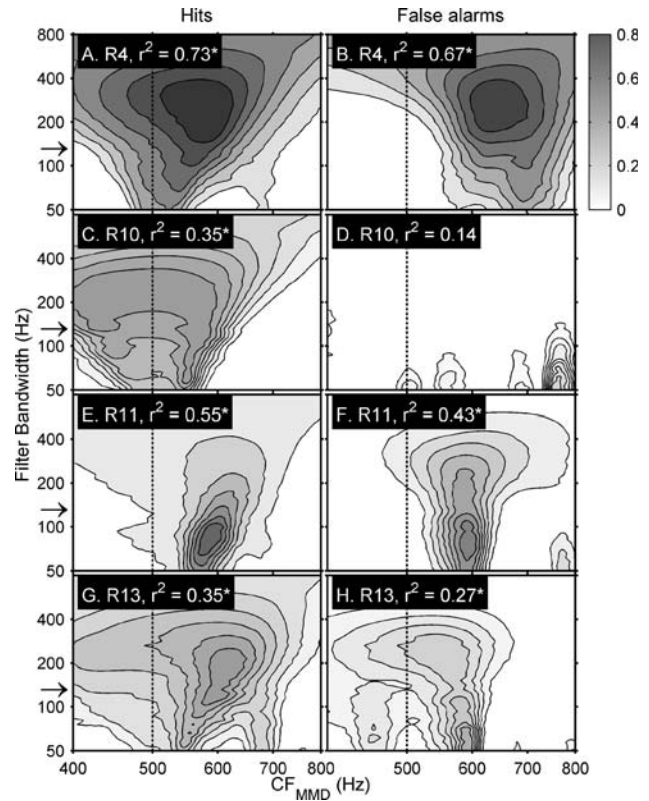


FIG. 8. Proportion of variance (r^2) of rabbit detection patterns predicted by the MMD model (same format as Fig. 6). CF_{MMD} is the CF of the positively weighted filter in the MMD model. Asterisk, statistically significant ($df = 23$, $p < 0.05$).

amount of the variance in the detection patterns as the single-channel energy model (Fig. 6). The detection threshold (E_s/N_0) of the MMD model with $CF_{MMD} = 500$ Hz was 8.9 dB.

Simulations of the MMD model with a 12-dB rove in the noise level showed an approximately 0.9-dB threshold elevation, which was comparable to that observed for the two rabbits and much lower than the threshold elevation for the roving-level condition predicted by the single-channel energy model. These simulation results suggested that predictions of an energy model that includes an estimation of the noise spectrum level were generally consistent with the rabbit results for both fixed-level and roving-level results.

Envelope-slope model

Because the addition of a tone to a noise makes the stimulus envelope flatter, the amount of fluctuation of the envelope can potentially be used as a cue for detection. An envelope-slope model developed by Richards (1992) was found to be efficient in extracting the fluctuation of the envelope. This envelope-slope model was tested with the rabbit and human behavioral results in the same way as the energy model (Fig. 4C). After passing through the simulated periph-

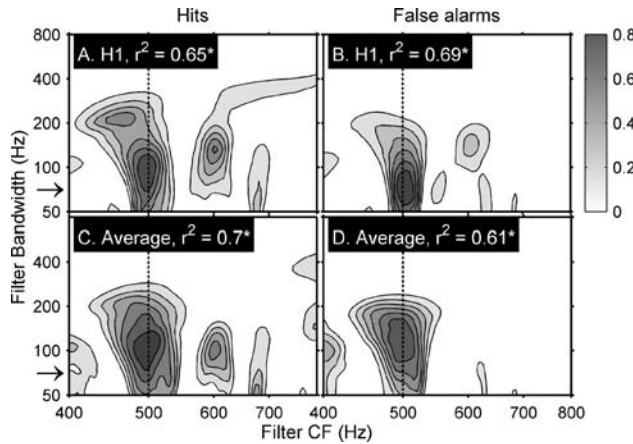


FIG. 9. Proportion of variance (r^2) of human detection patterns predicted by the envelope-slope model (same format as Fig. 5). Asterisk, statistically significant ($df = 23$, $p < 0.001$).

eral filter, the envelope of the stimulus was extracted by half-wave rectification and low-pass filtering (cut-off frequency = 250 Hz; Davidson et al. 2006). The envelope-slope statistic was the sum of the absolute difference between each pair of adjacent samples of the envelope, normalized by the sum of all the samples of the envelope. A relatively small envelope-slope statistic indicated the presence of a tone. The simulation of the envelope-slope model was generally the same as that for the single-channel energy model, except that the envelope-slope statistic, rather than the stimulus RMS, was used to make decisions.

Figure 9 shows the proportion of variance in the human detection patterns that was predicted by the envelope-slope model with varying CF and bandwidth. For human detection patterns, the envelope-slope model yielded predictions that were generally comparable to the energy model predictions, with both the threshold ($E_s/N_0 = 8.3$ dB) and the maximum r^2 values similar to those of the energy model. Some differences can be observed in the details of the correlation patterns and the best filter bandwidths. For example, the best prediction of the envelope-slope model was obtained with a filter bandwidth larger than the presumed human ERB of 75 Hz (Fig. 9), whereas for the energy model the best filter bandwidth was smaller (Fig. 5). Note that in the left two panels of Fig. 9, the peaks in the r^2 values at approximately CF = 600 Hz correspond to negative r values caused by the phase properties of the filter.

The envelope-slope model's predictions of rabbit detection patterns (Fig. 10) did not change systematically with filter CF or bandwidth, in contrast to the results for the single-filter energy model (Fig. 6) and the MMD model (Fig. 8). Although all maximum r^2 values were statistically significant ($df = 23$, $p < 0.05$), they were generally not as robust as for the energy-model or MMD models.

It should be noted that after passing through the peripheral filter, the envelope flattening caused by the tone occurred most reliably when the CF of the peripheral filter was equal to the tone frequency. When the response of an off-frequency filter was considered, the envelope cue was sometimes unreliable. Although the envelope-slope model predicted a comparable threshold ($E_s/N_0 = 8.8$ dB) to that of the single-channel energy model using the 500-Hz-CF filter, it failed to predict a reliable threshold (i.e., thresholds were either high or nonexistent, depending on filter CF), for filters with the CFs suggested by the energy model for each rabbit (555–580 Hz for hits). Therefore, if the rabbits indeed used frequencies above the tone frequency to detect the 500-Hz tone in noise, the envelope-slope cue would not have been effective.

DISCUSSION

This study is the first to present detailed evidence of cross-species differences in a behavioral task based on both empirical results and tests of specific models for auditory detection. The use of reproducible noise maskers to study tone-in-noise detection allows detailed testing of detection models (Gilkey and

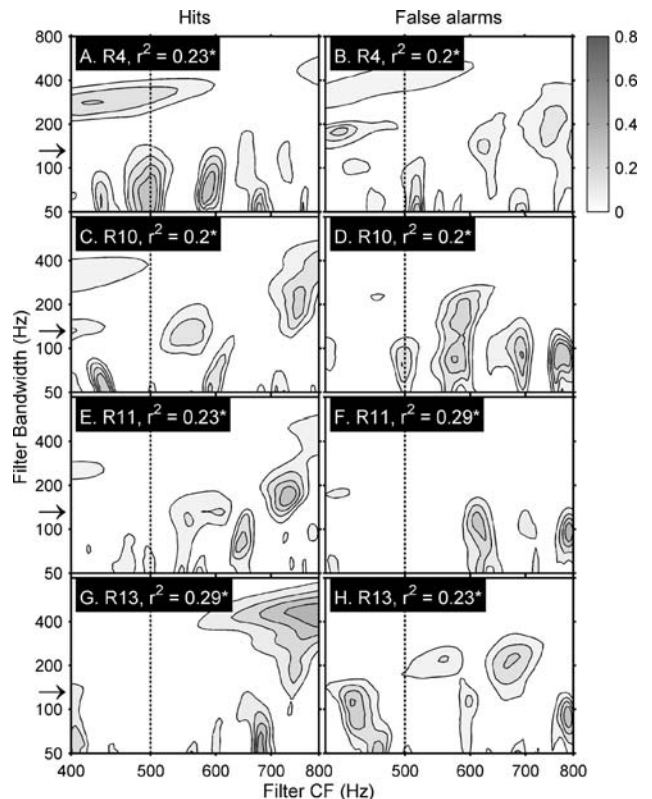


FIG. 10. Proportion of variance (r^2) of rabbit detection patterns predicted by the envelope-slope model (same format as Fig. 6). Asterisk, statistically significant ($df = 23$, $p < 0.05$).

Robinson 1986; Carney et al. 2002; Davidson et al. 2006). We found that although the detection patterns for both species were reliable (Table 2 and Davidson et al. 2006), the detection patterns for rabbits were not correlated with those of the average human subject for the same set of reproducible-noise maskers (Table 3). Comparisons between detection patterns from these two species and those of several models suggested that rabbits and humans might use different detection mechanisms.

First, the detection performance of rabbits was worse than that of humans in terms of threshold and consistency. Although peripheral tuning in rabbit is presumably broader than in human, simulations of the energy model showed that broadening of tuning alone can account for only a small portion of the threshold difference (the 500-Hz CF single-channel energy model had thresholds of 7.6 and 8.2 dB using the presumed human and rabbit ERBs, respectively; see Green and Swets (1966) for relationships between threshold and bandwidth). Therefore, the detection mechanism used by the rabbits was less efficient than that of the human subjects, at least in the context of the present experimental design. Possible factors causing the higher thresholds include off-frequency listening. Using the best CF and bandwidth for each rabbit, the energy model predicted thresholds of 10.7–16 dB. In addition, it seems likely that the rabbits had more internal variability (e.g., frequency uncertainty, criterion fluctuation) than human subjects, which would raise these estimates closer to the observed thresholds.

Second, although several detection models had similar partial success when predicting the human detection patterns, only the energy-based models were able to partially predict the rabbit detection patterns. For example, the envelope-slope model and the energy model were equally successful in predicting the human results, but the envelope-slope model did not predict the rabbit results. We also tested two types of fine-structure-based detection models, the phase-opponency model (Carney et al. 2002) and the zero-crossing model (Richards 1992) (not shown). Neither of these models explained the rabbit results (r^2 values not significant, $p > 0.05$), although they partially explained human results (r^2 values smaller than those presented here but significant, $p < 0.05$).

Third, human detection patterns can be best predicted by the energy in a small frequency range centered at the tone frequency. However, all four rabbits appeared to focus on, or be influenced by, spectral components at frequencies higher than the tone frequency. The test with one human subject (H1) using the rabbit apparatus showed that this was unlikely to be an artifact that resulted from sound calibration or distortion. To affirm this finding, we

reanalyzed the results from a previous Pavlovian-conditioning study that used the first 10 reproducible-noise maskers of the 25 maskers used here, with stimuli delivered by earphones (Zheng et al. 2002). Those detection patterns were less consistent for each rabbit (especially for false alarms), and correlations between the energy model and data were weaker than the correlations obtained in this study. Again, the detection patterns for hits were most strongly affected by frequencies higher than the tone frequency (565–585 Hz), but the detection patterns for false alarms were most strongly affected by lower frequencies (430–460 Hz). The difference in false-alarm results across studies is puzzling; however, the nature of false alarms in the Pavlovian-conditioning paradigm differs greatly from that in the operant paradigm. Overall, the analysis of the Pavlovian-conditioning results further suggested that the off-frequency-listening observed in this study, at least for hits, was not caused by the experimental design or setup used in the operant paradigm.

It was not clear whether there was any benefit in this “off-frequency listening”; several possible factors are considered here. Ahumada et al. (1975) suggested that if a subject monitors several auditory filters and makes decisions based on the filter with the maximum response, it is possible that the subject focuses on off-frequency filters when the tone is absent. However, because different noise waveforms have different spectral characteristics, the averaged effect would be a broader effective filter for false alarms rather than a shifted filter. Another possible cause for listening at a frequency above 500 Hz is the decreasing trend in threshold with increasing frequency for detection of tones in quiet in this frequency range, as shown by audiograms for rabbit (Heffner and Masterton 1980). The estimated tone-in-quiet threshold at 586 Hz (the mean value of the best CF across subjects for both hits and false alarms) from the published audiograms is approximately 4 dB lower than the threshold at 500 Hz. The lowest threshold, achieved at 2 kHz, was 24 dB lower than the threshold at 500 Hz. This increased sensitivity at high frequencies might introduce a natural tendency to focus at high frequencies, even if the consequences are detrimental for a particular task. However, note that differences in sensitivity across frequency alone cannot explain the apparent off-frequency listening. For example, higher sensitivity at 600 Hz would apply to both the masker and the tone, and would not result in any signal-detection benefit. The best strategy for detection of the 500-Hz tone is to use the output of the 500-Hz CF filter, even if sensitivity varies across CF.

It is possible that off-frequency listening was a reflection of physiological properties that differ

between rabbits and human. For example, at 500 Hz, the peripheral filters of rabbits might be more asymmetrical than those of humans. The simple gammatone filter used in this study was symmetrical. Therefore, we tested the energy model with an asymmetrical filter (for cat, Zilany and Bruce 2006; results not shown), which allows more high-frequency components to pass through a 500-Hz filter than the gammatone filter does. This asymmetrical filter predicted a smaller shift in the best CF for each rabbit compared to the gammatone filter but was unable to account for the total amount of shift observed. Cochlear amplification, which is associated with the compression and suppression nonlinearities in the periphery, is also a factor that may differ across species. These peripheral nonlinearities interact with filter asymmetry to create shifts in best frequency with sound level (Carney 1999; Tan and Carney 2003). Future studies of detection at other frequencies may provide more insight into the off-frequency listening observed in this study. For example, the direction and amount of asymmetry of peripheral filtering changes with CF (in cat, Carney et al. 1999); thus, if asymmetrical tuning is a factor, different frequency shifts would be predicted for low-, high-, and mid-frequency tones.

The modeling results also indicate that the rabbits are more diverse in the bandwidth of effective tuning compared to humans. Differences in both CF and bandwidth of the effective tuning across rabbits, as well as substantial internal variability, presumably led to relatively weak intersubject correlations, as compared to the strong correlations found among human subjects (Davidson et al. 2006).

Finally, although the detection pattern predictions of the MMD model were not appreciably better than those of the single-channel energy model, the MMD model with combined positive and negative weights that summed to zero was rove-resistant. The negatively weighted filters for human subjects are typically close to the center filter and contribute to the detection patterns (Ahumada and Lovell 1971, Gilkey and Robinson 1986; Davidson et al. 2006). However, it was not possible to predict the rabbit detection patterns with strong negatively weighted filters close to the positively weighted filter. In fact, the widely spaced frequency channels with negative weights that predicted the rabbit results were similar to weighting functions suggested by profile-analysis studies (Green et al. 1983; Bernstein and Green 1987). It should also be noted that these negatively weighted filters do not necessarily require excitatory or inhibitory mechanisms at relatively low stages in the auditory pathway but could reflect excitatory (Davidson et al. 2006), inhibitory, or network processing at higher auditory centers.

In conclusion, this study showed that operant conditioning with a discrimination task is a useful procedure to study detection in rabbit. This procedure yielded consistent reproducible-noise detection patterns that were useful for evaluating detection mechanisms. Because tone-plus-noise and noise-alone trials were presented with equal probability, this discrimination procedure was effective in studying false-alarm rates as well as hit rates, whereas in “go/no-go” procedures, the false-alarm rates are typically kept low to ensure good stimulus control of behavior. Subjects may develop a response bias associated with one type of sound stimulus in a choice discrimination task such as that used in this study. This bias might be caused by the nature of the stimulus presentation (e.g., in this study, bias tended to be toward the noise-alone response because the noise was present in every trial), or an innate tendency of the subject, which differs across individuals. An effective bias-control mechanism here was to adaptively adjust the amount of positive reinforcement.

The obtained reproducible-noise detection patterns for rabbit and human subjects were quite different. The two species may use different detection cues, e.g., temporal versus energy cues, or different weighting mechanisms applied to the same cue. These findings suggest that studies involving comparisons between physiology in one species and psychophysics in another must take into account possible differences in neural processing or detection mechanisms between the species.

ACKNOWLEDGEMENTS

We received important assistance from Sean Davidson with data analysis and psychophysical detection models. Paul Nelson, Christine Mason, J. Anthony Nevin, and Steven Colburn offered valuable suggestions. Susan Early provided editorial assistance. We also thank Bill Dossert for help with the experimental setup and the staff of Laboratory Animal Resources of Syracuse University. This study was supported by NIH NIDCD-01641.

REFERENCES

- AHUMADA A JR, LOVELL J. Stimulus features in signal detection. *J. Acoust. Soc. Am.* 49:1751–1756, 1971.
- AHUMADA A JR, MARKEN R, SANDUSKY A. Time and frequency analyses of auditory signal detection. *J. Acoust. Soc. Am.* 57:385–390, 1975.
- BERNSTEIN LR, GREEN DM. The profile analysis bandwidth. *J. Acoust. Soc. Am.* 81:1888–1895, 1987.
- BORG E, EGNSTROM B, LINDE G, MARKLUND K. Eighth nerve fiber firing features in normal-hearing rabbits. *Hear. Res.* 36:191–201, 1988.
- CARNEY LH. Temporal response properties of neurons in the auditory pathway. *Curr. Opin. Neurobiol.* 9:442–446, 1999.

- CARNEY LH, HEINZ MG, EVILSIZER ME, GILKEY RH, COLBURN HS. Auditory phase opponency: a temporal model for masked detection at low frequencies. *Acta Acustica united with Acustica* 88:334–347, 2002.
- CARNEY LH, McDUFFY MJ, SHEKHTER I. Frequency glides in the impulse responses of auditory-nerve fibers. *J. Acoust. Soc. Am.* 105:2384–2391, 1999.
- COSTALUPES JA. Broadband masking noise and behavioral pure tone thresholds in cats. I. Comparison with detection thresholds. *J. Acoust. Soc. Am.* 74:758–764, 1983.
- DAVIDSON SA, GILKEY RH, COLBURN HS, CARNEY LH. Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results. *J. Acoust. Soc. Am.* 119:2258–2275, 2006.
- DAVISON M, NEVIN JA. Stimuli, reinforcers, and behavior: an integration. *J. Exp. Anal. Behav.* 71:439–482, 1999.
- DAVISON MC, TUSTIN RD. The relation between the generalized matching law and signal-detection theory. *J. Exp. Anal. Behav.* 29:331–336, 1978.
- EARLY SJ, MASON CR, ZHENG L, EVILSIZER M, IDROBO F, HARRISON JM, CARNEY LH. Studies of binaural detection in the rabbit (*Oryctolagus cuniculus*) with Pavlovian conditioning. *Behav. Neurosci.* 115:650–660, 2001.
- EVILSIZER ME, GILKEY RH, MASON CR, COLBURN HS, CARNEY LH. Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human. *J. Acoust. Soc. Am.* 111:336–345, 2002.
- FLETCHER H. Auditory patterns. *Rev. Mod. Phys.* 12:47–65, 1940.
- GILKEY RH, ROBINSON DE. Models of auditory masking: a molecular psychophysical approach. *J. Acoust. Soc. Am.* 79:1499–1510, 1986.
- GILKEY RH, ROBINSON DE, HANNA TE. Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise. *J. Acoust. Soc. Am.* 78:1207–1219, 1985.
- GLASBERG BR, MOORE BCJ. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47:103–138, 1990.
- GREEN DM. *Profile Analysis: Auditory Intensity Discrimination.* Oxford Psychology Series No. 13. Oxford, New York, 1988.
- GREEN DM, KIDD G JR, PICARDI MC. Successive versus simultaneous comparison in auditory intensity discrimination. *J. Acoust. Soc. Am.* 73:639–643, 1983.
- GREEN DM, SWETS JA. *Signal Detection Theory and Psychophysics.* John Wiley & Sons, New York, 1966.
- HEFFNER H, MASTERTON B. Hearing in Glires: domestic rabbit, cotton rat, feral house mouse, and kangaroo rat. *J. Acoust. Soc. Am.* 68:1584–1599, 1980.
- ISABELLE SK, COLBURN HS. Detection of tones in reproducible narrow-band noise. *J. Acoust. Soc. Am.* 89:352–359, 1991.
- KIDD G JR, MASON CR, BRANTLEY MA, OWEN GA. Roving-level tone-in-noise detection. *J. Acoust. Soc. Am.* 86:1310–1317, 1989.
- LEVITT H. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49:467–477, 1971.
- MACMILLAN NA, CREELMAN CD. *Detection Theory: A User's Guide.* Lawrence Erlbaum Associates, Mahwah, NJ, 2005.
- MASON CR, IDROBO F, EARLY SJ, ABIBI A, ZHENG L, HARRISON JM, CARNEY LH. CS-dependent response probability in an auditory masked-detection task: considerations based on models of Pavlovian conditioning. *Q. J. Exp. Psychol. B* 56:193–205, 2003.
- MAY BJ, HUANG AY, ALESZCZYK CM, HIENZ RD. Design and conduct of sensory experiments for domestic cats. In: Dooling RJ, Fay RR, Klump GM, and Stebbins WC (eds) *Methods in Comparative Psychoacoustics.* Birkhäuser Verlag, Basel, pp. 95–108, 1995.
- PATTERSON RD, NIMMO-SMITH I, HOLDSWORTH J, RICE P. Implementing a gammatone filter bank. *SVOS Final Report: The Auditory Filter Bank*, 1988.
- PETERSON MR, PROSEN CA, MOODY DB, STEBBINS WC. Technical note: operant conditioning in the guinea pig. *J. Exp. Anal. Behav.* 27:529–532, 1977.
- PFÄFFLIN SM, MATHEWS MV. Detection of auditory signals in reproducible noise. *J. Acoust. Soc. Am.* 39:340–345, 1966.
- RICHARDS VM. The detectability of a tone added to narrow bands of equal-energy noise. *J. Acoust. Soc. Am.* 91:3424–3435, 1992.
- RUGGERO MA, TEMCHIN AN. Unexceptional sharpness of frequency tuning in the human cochlea. *Proc. Natl. Acad. Sci. U. S. A.* 102:18614–18619, 2005.
- SHERA CA, GUINAN JJ JR, OXENHAM AJ. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci. U. S. A.* 99:3318–3323, 2002.
- SIEGEL RA, COLBURN HS. Binaural processing of noisy stimuli: internal/external noise ratios for diotic and dichotic stimuli. *J. Acoust. Soc. Am.* 86:2122–2128, 1989.
- TAN Q, CARNEY LH. A phenomenological model for the responses of auditory-nerve fibers: II. Nonlinear tuning with a frequency glide. *J. Acoust. Soc. Am.* 114: 2007–2020, 2003.
- ZHANG X, HEINZ MG, BRUCE IC, CARNEY LH. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.* 109:648–670, 2001.
- ZHENG L, EARLY SJ, MASON CR, IDROBO F, HARRISON JM, CARNEY LH. Binaural detection with narrowband and wideband reproducible noise maskers: II. Results for rabbit. *J. Acoust. Soc. Am.* 111:346–356, 2002.
- ZILANY MS, BRUCE IC. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J. Acoust. Soc. Am.* 120:1446–1466, 2006.