Mechanisms of masking by Schroeder-phase harmonic tone complexes in the budgerigar (*Melopsittacus undulatus*)

Kenneth S. Henry [a,b,c,*], Yingxuan Wang [b], Kristina S. Abrams [c], Laurel H Carney [a,c]

Departments of [a] Otolaryngology, [b] Biomedical Engineering, and [c] Neuroscience, University of Rochester, Rochester, NY 14642, USA

Keywords: Budgerigar; Inferior colliculus; Operant conditioning; Schroeder masking

* Corresponding author. *E-mail address:* kenneth_henry@urmc.rochester.edu (K.S. Henry).

Abstract

Schroeder-phase harmonic tone complexes have a flat temporal envelope and either rising or falling instantaneous frequency sweeps within F0 periods, depending on the polarity of the phase-scaling parameter *C*. Human thresholds for detection of a pure-tone target added to a Schroeder HTC can be 10-15 dB lower for the positive Schroeder (downward-sweeping) compared to negative (upward-sweeping), potentially due to the impulse response of the basilar membrane though this hypothesis remains controversial. Specifically, the upward-gliding impulse response observed empirically in several studies is expected to produce a “peakier” envelope response for positive Schroeder HTCs, thereby reducing masking compared to negative Schroeder HTCs. Birds provide an interesting animal model for studying Schroeder masking because prior studies suggest less behavioral threshold difference between Schroeder maskers of opposite polarity. However, neurophysiological mechanisms have not been explored and previous behavioral studies mostly used Schroeder HTCs with relatively low F0s. To gain further insight into underlying Schroeder masking mechanisms, we performed operant-conditioning experiments in budgerigars (*Melopsittacus undulatus*), a parakeet species, using a wide range of F0 and C values. Awake neurophysiological recordings in the avian inferior colliculus, a major envelope processing station in the midbrain, characterized neural encoding of behavioral test stimuli. Behavioral Schroeder masking thresholds increased with increasing F0 and showed minimal difference between opposite Schroeder polarities for all conditions, similar to prior studies in this and other bird species. Neural recordings showed prominent temporal and average-rate-based encoding of Schroeder F0, and in many neurons, marked response

asymmetries between opposite Schroeder polarities. While response bias varied considerably across neurons, average bias of the population shifted from negative to positive with increasing characteristic frequency, with a breakpoint at ~2-3 kHz. Response bias was weakly associated with lower Schroeder-masked tone detection of neurons for the 400-Hz F0 only. These results challenge the common assumption that a peakier masker response results in a lower threshold for target tone detection with this paradigm.

1. Introduction

Schroeder's algorithm can produce harmonic tone complexes (HTCs) with a flat temporal envelope and either upward- or downward-sloping instantaneous frequency sweeps within periods of the fundamental frequency (F0), depending on the polarity of the phase-scaling parameter *C* (Schroeder, 1970) (Fig. 1). Early on, it was discovered that HTCs generated with positive (downward sweeping frequency) and negative (upward sweeping frequency) *C* values differ in the extent to which they mask behavioral thresholds for pure-tone detection in humans, with up to 20 dB lower thresholds for the positive *C* value compared to negative (Kohlrausch & Sander, 1995; Smith et al., 1986). Less masking by Schroeder HTCs with positive *C* values is not explainable by traditional power-spectrum models of masking (Fletcher, 1940) because Schroeder HTCs generated with opposite *C* values have identical magnitude spectra and share the same waveform when one signal is time reversed.

While the underlying reason for masking differences between Schroeder *C* polarities remains uncertain, one possibility is that the difference originates from asymmetry of mechanical filtering by the mammalian cochlea (Smith et al., 1986). Specifically, recordings of basilar membrane motion in response to clicks show that the impulse response of the cochlea exhibits an upward-sweeping instantaneous frequency glide in response to clicks – a response property associated with asymmetry in the shape of pure-tone frequency tuning curves (Recio et al., 1998; Strube, 1985). Filters with an upward-sweeping instantaneous frequency glide of the impulse response are be expected to produce more prominent response envelope fluctuations (i.e., a "peakier" temporal responses profile) for downward-sweeping Schroeder HTCs with positive *C* values compared upward-sweeping Schroeder HTCs with negative *C* values. The deeper and longer low-amplitude periods in the cochlear response to positive-*C* Schroeder maskers could potentially provide ampler opportunity for detection of an added low-level tone, thereby leading to the observed lower masked threshold under the positive-*C* condition.

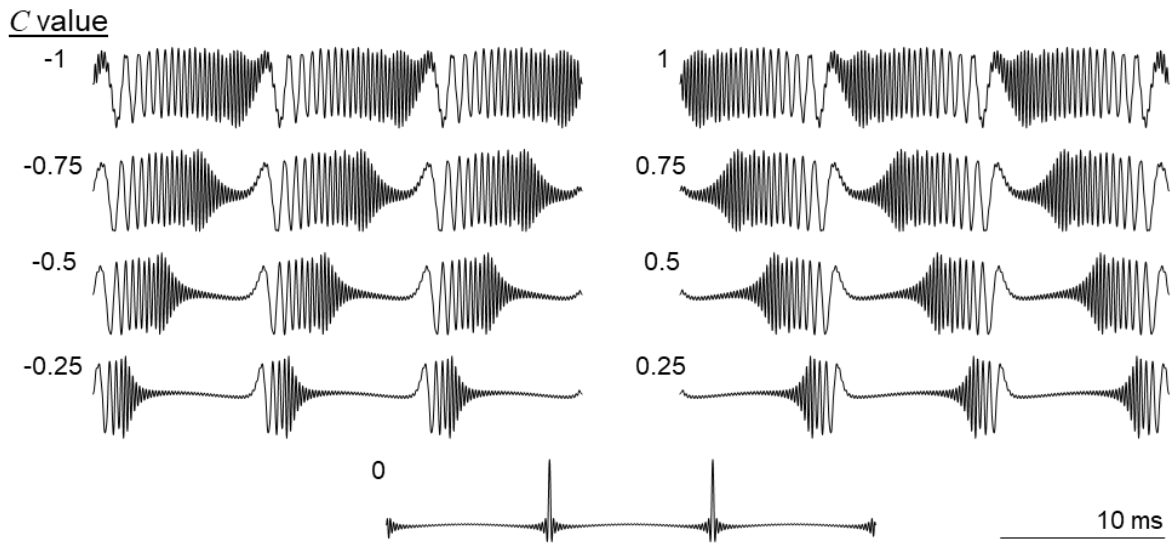

Figure 1. Waveforms of Schroeder-phase harmonic tone complexes (HTCs) produces with the same fundamental frequency (F0; 100 Hz) and values of the phase-scaling parameter, *C*, ranging from -1 to 1. Positive *C* values produce downward-sloping instantaneous frequency sweeps within F0 periods whereas negative *C* values produce upward-sloping sweeps. Lower absolute *C* value decrease the duty cycle of the waveform with *C* of zero producing an impulse train.

Animal models have been used to test the hypothesized contribution of cochlear filtering asymmetries to Schroeder masking differences between opposite *C* polarities, but so far have yielded an incomplete understanding of underlying mechanisms. Studies in chinchilla and guinea pig recorded basilar membrane motion in response to Schroeder HTCs to test for the predicted difference in response envelope fluctuations between opposite C polarities (Recio & Rhode, 2000; Summers et al., 2003). Results were in some respects consistent with an explanation based on cochlear impulse responses, showing the expected, larger fluctuations for responses to Schroeder HTCs with positive C values compared to negative. However, measurements were restricted to basal cochlear regions with characteristic frequencies (CFs) above the frequency range of target signals typical of human behavioral studies. Note that for lower CFs below 1 kHz, auditory-nerve fiber recordings in chinchilla and cat show a transition to downward-sweeping frequency glides of the cochlear impulse response (Carney et al., 1999; Recio-Spinoso et al., 2005). Filters with downward-gliding impulse responses should theoretically produce a peakier response to Schroeder HTCs with negative C values rather than positive, and consequently less masking for the negative-*C* condition. However, less masking by Schroeder HTCs with negative *C* values has not been observed in any human behavioral study (Kohlrausch & Sander, 1995; Smith et al., 1986). While it remains possible that chinchillas show lower thresholds for the negative-*C* Schroeder masking condition at test frequencies below 1 kHz, animal behavioral studies of Schroeder masking have not been conducted in chinchillas or

any other nonhuman mammal to our knowledge. In summary, it remains unknown whether asymmetry of cochlear filtering can explain observed behavioral masking differences between Schroeder *C* polarities.

Birds provide an interesting animal model for studies of Schroeder masking because many species use vocalizations containing rapid frequency changes as part of their social communication system and can be readily trained to perform complex auditory detection and discrimination tasks (Dooling et al., 2000). In contrast with human results, the few behavioral studies of Schroeder masking conducted so far in birds have found minimal threshold differences for tones masked by Schroeder HTCs with opposite *C* values. In budgerigars, thresholds were determined for detection of 1, 2.8, and 4-kHz tones masked by Schroeder HTCs with positive or negative *C* values, for comparison with human thresholds determined using the same stimuli (Leek et al., 2000). Whereas human subjects showed the expected lower threshold of up to 15-20 dB for positive-*C* Schroeder maskers compared to negative *C* values, with greatest masking differences at high sound levels, budgerigar thresholds were either similar between positive and negative *C* values or slightly lower for Schroeder HTCs generated with negative *C* values. In two follow-up studies, budgerigars, zebra finches, and canaries were tested across a wider range of stimulus conditions including higher F0s up to 400 Hz (Dooling et al., 2001) and multiple, intermediate *C* values between -1 and 1 (Lauer et al., 2007). Note that lower absolute *C* values produce stimuli with reduced duty cycle and, consequently, faster instantaneous frequency sweeps within F0 periods, with *C* of zero producing an impulse train associated with all components in cosine phase (Fig. 1). Even over this wider range of conditions, including faster rates of instantaneous frequency change potentially better suited to the mechanics of the avian inner ear, birds showed little appreciable difference in masking between Schroeder HTCs generated with opposite *C* values (Dooling et al., 2001; Lauer et al., 2007).

The reason for less masking difference between Schroeder HTCs of opposite *C* polarity in birds compared to humans is unclear due to limited physiological investigations. One possibility is that the avian cochlea differs from that of mammals in its impulse response, perhaps lacking a prominent instantaneous frequency glide [see barn owl auditory-nerve recordings potentially consistent with this interpretation (Fontaine et al., 2015)], or showing a glide that is faster than the rate of frequency change found in Schroeder maskers from existing behavioral studies. Despite basic physiological similarities between birds and mammals in auditory-nerve frequency tuning, phase-locking limits, and other response properties of auditory-nerve fibers (Manley et al., 1985; Sachs et al., 1974), the sensory epithelium is considerably

shorter and wider in birds than in mammals with dozens of hair cells spanning its width in apical low-CF regions (Takasaka & Smith, 1971). If these anatomic difference result in a cochlear impulse response without a prominent frequency glide (see Fontaine et al., 2015, for auditory-nerve impulse responses in barn owl), this could explain less masking difference between Schroeder HTCs of opposite polarity. While Dooling and colleagues founded stronger compound auditory-nerve action potentials for negative-C Schroeder HTCs in several bird species (Dooling et al., 2002), these gross potentials are strongly influenced by across-channel neural synchrony and are therefore not directly comparable to the single-channel hypothesis based on impulse-response frequency glides proposed to explain human masking asymmetry.

To gain further insight into mechanisms of Schroeder masking, and into possible reasons for minimal Schroeder masking asymmetry in birds, we performed parallel behavioral and neurophysiological experiments in budgerigars using tones masked by Schroeder HTCs. In the first experiment, behavioral studies were conducted in animals trained using operant-conditioning procedures over a wide range of F0 and *C* values, to improve the possibility of detecting masking differences between Schroeder *C* polarities if they exist. In contrast to prior studies, stimuli were presented using a roving-level paradigm for which the overall masker level was randomly selected on each trial from a 20-dB range centered on 80 dB SPL (a typical level used in prior studies), thereby diminishing the extent to which overall loudness and/or single-channel energy cues could be used to perform the task. In the second experiment, neurophysiological recordings were made in awake passively listening animals of extracellular activity in the central nucleus of the avian inferior colliculus (IC), which is a large and nearly obligatory midbrain processing center in the ascending auditory pathway. The IC shows pronounced temporal and rate-based encoding of acoustic envelope fluctuations at amplitude-modulation frequencies up to several hundred Hz (Henry et al., 2017; Joris et al., 2004), encompassing the range of envelope frequencies found in Schroeder HTCs. We obtained neural responses to Schroeder HTCs across widely ranging F0 and C values and estimated neural thresholds for Schoeder-masked detection of tones presented at the CF of individual recording sites.

2. Methods

2.1. Behavioral experiments

Behavioral experiments were conducted in four adult budgerigars of either sex using operant conditioning procedures. All procedures were approved by the University of Rochester Committee on Animal Resources.

2.1.1. Stimuli

Stimuli were Schroeder-phase HTCs presented with or without a simultaneously gated 2.8-kHz pure-tone signal. Schroeder HTCs were generated with component phases scaled according to Schroeder's (1970) algorithm,

$$\theta_n = C\pi n(n+1)/N$$

where θ_n is the phase of the nth harmonic, N is the total number of harmonics, and C is a scaling parameter. Harmonic components ranged in frequency from 2F0 to 5000 Hz and were presented with equal level. The overall level of the masker was 80 dB SPL prior to application of random level variation for some conditions (i.e., roving-level paradigm; see below). Consequently, for F0s of 100, 200, and 400 Hz, the number of components included in the Schroeder masker was 49, 24, and 11, respectively, and the level of individual masker components prior to roving was 63.10, 66.20, and 69.59 dB SPL, respectively.

Schroeder maskers and pure-tone targets were simultaneously gated on target trials for a duration of 260 ms with 20-ms raised cosine onset and offset ramps. The phase of the 2.8-kHz pure tone target was equal to that of the masker frequency component at the same frequency. The masker alone was presented on standard trials. Threshold signal-to-masker ratios for tone detection are is expressed in dB relative to the level of single frequency components of the masker, as in previous studies (Kohlrausch & Sander, 1995; Lauer et al., 2007; Smith et al., 1986).

The first behavioral experiment assessed Schroeder masking in three animals over a wide range of masking conditions using F0s of 100, 200 and 400 Hz and C values of -1, -0.5, 0, 0.5, and 1. Testing was conducted using a roving-level paradigm for which the overall level of the masker was randomly selected on each trial from a uniform distribution of values ranging from 70 to 90 dB SPL with 1-dB resolution. A second behavioral experiment compared Schroeder-masked behavioral thresholds between conditions with fixed and roving masker level in four animals, using F0 of 100 Hz and C values of -1, 0 and 1. Testing alternated between fixed- and roving-level conditions within tracking sessions as described in detail below.

2.1.3. Apparatus

Behavioral tests were conducted in four identically constructed small-animal acoustic isolation chambers (0.3 m^3) lined with sound-absorbing foam. Chambers contained a house light, overhead loudspeaker (Polk Audio MC60; Baltimore, MD), seed dispenser (ENV-203 mini; Med Associates, St. Albans, VT) and feeding trough, and small wire-mesh cage with a perch and three illuminated response switches. The loudspeaker was positioned 20 cm above the location of the animal's head during testing. Behavioral tests were controlled using custom MATLAB software (The MathWorks, Natick, MA) running on a PC. Stimulus waveforms were generated digitally with 50-kHz sampling frequency and converted to analog (National Instruments PCI-6151; Austin, TX) before power amplification (Crown D-75A). Custom built electronic equipment was used to process input from the response switches and control the house light and seed dispenser. Stimuli were calibrated using a pre-emphasis filter generated based on the output of a precision microphone (Brüel and Kjær type 4134, Marlborough, MA) in response to 249 log-spaced tone frequencies from 50-15,100 Hz. The pre-emphasis filter corrected the stimulus spectrum for the frequency responses of the loudspeaker system prior to analog conversion.

2.1.4. Procedure

Budgerigars were trained using operant-conditioning procedures to perform a single-interval, two-alternative auditory discrimination task. Birds initiated a trial by pecking the center observing response switch, resulting in presentation of a single stimulus. The stimulus was either a standard Schroeder HTC masker or target with a pure-tone signal added to the masker. Standard and target stimuli were presented with equal probability. The correct response to the standard stimulus was the right switch and the correct response to the target stimulus was the left switch. Animals had 3 seconds to report a response following stimulus onset. Correct responses were reinforced with delivery of hulled millet seeds whereas incorrect responses resulted in a 5-s timeout during which the house light was turned off. The trial was aborted in rare instances for which animals did not report a response (<0.2% of trials). The number of millet seeds was adaptively varied (1 or 2 seeds) using an automated procedure to control response bias. Bias was calculated based on the most recent 50 trials as 0.5 times the sum of the Z scores of the hit rate and false alarm rate (Macmillan & Creelman, 1991). Additional seeds were delivered for correct responses made with the switch the animal was biased against. Testing was conducted 6-7 days per week during morning and afternoon blocks separated by four hours, and consisting of several hundred trials each.

Animals were initially trained to behaviorally discriminate the target stimulus from the standard at a relatively high signal-to-masker ratio of 10 dB. Once 90% correct discrimination performance was attained, masked signal-detection thresholds were tested repeatedly using two-down one-up adaptive tracking procedures (Levitt, 1971) . Behavioral tracking sessions began at a signal-to-masker ratio of 10 dB. The signal level was reduced following every two consecutive correct responses to target trials (hits) with the same signal-to-masker ratio, and increased following every incorrect response to a target trial (miss). Responses on standard trials (false alarms and correct rejections) were reinforced but did not affect the signal level of the track. The step size of the tracking procedure was varied based on the number of reversals (upward vs. downward) of the track plotting signal-to-masker ratio across target trials. The initial step size of 3 dB was decreased to 2 dB after 2 reversals and to 1 dB after 4 reversals. Tracks continued for at least 15 reversals until the signal-to-masker ratio of the final eight track reversal points met two stability criteria. Both (1) the standard deviation of the final eight reversal values and (2) the mean difference between the first and second four reversal values of the final eight were required to be less than 3 dB.

For the first experiment, the mean of the final eight reversal values was taken as the track threshold. For the second experiment, extended-duration tracks (Wang et al., 2021) were used to evaluate the effect of roving stimulus level on Schroeder-masked signal detection within sessions, and fewer tracks were completed daily. For extended tracks, the masker level remained fixed at 80 dB SPL until the stability criteria described above were satisfied. This initial portion of the track provided reversal points for calculation of the fixed-level masked threshold. Thereafter, the track was allowed to continue for an additional 10 reversals under the roving-level condition (±10 dB variation in masker level across trials) until the track stability criteria were once again satisfied. This second portion of the track was used to calculate the roving-level threshold (based on the final eight reversal points as above). Note that neither the signal level nor the step size of the track were altered following the onset of the roving-level portion of the track.

For both experiments, tracking sessions were conducted repeatedly on the same stimulus condition until at least 13 unbiased thresholds (or threshold pairs for experiment 2) were obtained (absolute response bias less than 0.3) and two across-session stability criteria were satisfied. Both (1) the standard deviation of the final six unbiased track thresholds and (2) the mean difference between the first and second half of the final six unbiased thresholds were required to be less than 3 dB. Sessions for which the overall absolute bias exceeded 0.3 were excluded from further analysis. For both experiments, stimulus conditions were tested in

different random sequences across animals. Animals were tested 2-4 times on each condition, with each testing period separated by testing on the other conditions.

2.1.4. Analysis

The last 25 unbiased track thresholds were averaged to calculate masked thresholds of individual animals on each condition. Thresholds were analyzed in R (v. 4.1.3) using linear mixed-effect models (Bates et al., 2015). Models incorporated random intercepts for each animal subject to account for repeated measures, and fixed effects of F0, C, and roving level (for experiment two). Two-way interactions were included and dropped when not significant ($p > 0.05$) in order of decreasing p value. Degrees of freedom for post-hoc t-tests were calculated based on the Satterthwaite approximation.

2.2 Neurophysiological recordings

Neurophysiological recordings were made from the IC in four adult budgerigars of either sex. All procedures were approved by the University of Rochester Committee on Animal Resources.

2.2.1 Neural implantation procedure

Procedures for implanting microelectrodes into the avian IC have been described previously (Henry et al., 2016, 2017; Wang et al., 2021). Briefly, a probe assembly consisting of 1-2 microelectrodes mounted to a miniature microdrive was implanted into the right IC under anesthesia using a stereotaxic approach. Microelectrodes were either single iridium wires or four-channel silicon probes with impedance ranging from 1.5-3 MΩ. Anesthesia was induced with subcutaneous injection of ketamine (3-5 mg’kg) and dexmedetomidine (0.08-0.1 mg/kg), and maintained throughout the 2-3 h implantation surgery with slow subcutaneous infusion of the same agents (ketamine: 6-10 mg/kg/h; dexmedetomidine: 0.16-0.27 mg/kg/h) in lactated ringers solution. Whenever possible, electrodes were initially implanted superficially in the dorsal region of the IC tuned to low frequencies. This method allowed sampling of responses tuned to higher frequencies by advancing the electrodes deeper into the IC steps with the microdrive control screw. Electrodes were typically advanced 30 um between recording sessions. Recordings were made from 142 multi-unit sites and 8 sites with well-isolated single-unit activity

2.2.2. Apparatus

Neurophysiological recordings were made with animals perched in a small wire-mesh located inside a walk-in acoustic isolation chamber. The chamber was lined with sound-absorbing foam and a loudspeaker (MC60, Polk Audio) was positioned 0.45 m from the normal location of the animal's head in the same horizontal plane. A closed-circuit video system was monitored to ensure that animals remained perched and facing the loudspeaker throughout recording sessions.

Stimuli were generated in MATLAB with a sampling frequency of 50 kHz and converted to analog at full scale using a data acquisition card (610V; PCIe-6251, National Instruments). Signals were then attenuated to the specified level (PA5; Tucker-Davis Technologies) followed by power amplification (D-75A, Crown Audio). Stimulus calibration was accomplished with a digital "pre-emphasis" filter that compensated for the frequency response of the system. The filter was designed based on the output of a 0.25" precision microphone (type 4938, Brüel & Kjær) placed at the location of the animal's head in response to 249 log-space tone pips ranging in frequency from 0.05-15.1 kHz.

Neural recordings were made using a multi-channel recording system (RHD2132 amplifier chip and C3100 USB interface board, Intan Technologies), referenced to one of the anchor screws for the microdrive assembly. Recordings were high-pass filtered at 150 Hz and sampled on the amplifier chip at 30 kHz for storage on the computer hard drive. In post processing, recordings were resampled at 50 kHz and band-pass filtered from 0.75-10 kHz (1000-point finite impulse response) to reduce local field potentials. Thereafter, a third-order Teager-energy operation (Choi et al., 2006; Wang et al., 2021) was applied for spike detection based on a visually determined threshold criterion. Sessions with consistent spike shape and amplitude through the recording period and less than 1% of inter-spike intervals less than 1 ms were classified as single-unit sessions.

2.3.3. Stimuli

Stimuli were pure tones for determination of CF, Schroeder-phase HTCs, and Schroeder-masked tones with frequency equal to the estimated CF. Responses to Schroeder HTCs were obtained for F0s of 50, 100, 200, and 400 Hz and C values from -1 to 1 with a step size of 0.25. Stimuli were 0.4 s in duration with 25-ms raised cosine onset and offset ramps, presented for 20

repetitions in random sequence with 0.2 s of silence between stimuli. Overall stimulus level was 60 dB SPL. Responses to Schroeder-masked CF tones were obtained for F0s of 100, 200, and 400 Hz and C values form -1 to 1 with a step size of 0.5. Tone and masker were simultaneous gated with a duration of 0.26 s, 20-ms onset and offset ramps, and 0.25 s of silence between stimuli. Signal-to-masker component ratios ranged from -33 to 0 dB and included a masker with no tone added. The overall level of the masker was 60 dB SPL. Each stimulus was presented for 20 repetitions in random sequence.

2.3.4. Analyses

Neural responses to Schroeder HTC were quantified with an average-rate analysis based on the total number of spikes during the stimulus duration and with a temporal metric calculated as the peak-to-peak amplitude of single-trial period histograms. This temporal metric quantifies the amplitude of fluctuations in the neural response at F0, and was calculated after first integrating each single-trial spike-train response over a duration equal to $1/25^{th}$ of the F0 period for F0s up to 200 Hz and over 0.2 ms for F0 of 400 Hz. Response bias was calculated for both average rate and F0 rate fluctuation as the mean difference between response to positive and negative Schroeder polarities divided by the pooled standard deviation. Hence, positive (negative) values indicate a greater response amplitude for the positive (negative) Schroeder stimulus and the magnitude of the effect indicate separation in terms of standard deviation, as in a d-prime metric.

3. Results and Discussion

3.1. Behavioral Results

3.1.1. Effects of F0 and C on Schroeder-masked thresholds

Budgerigars were trained with operant-condition procedures to detect a 2.8-kHz tone signal added to a Schroeder HTC masker. Masked detection thresholds were tested in three animals using two-down one-up tracking procedures (Fig. 2A) and a ±10-dB roving-level paradigm that reduces the usefulness of single-channel energy cues for performing the task. Maskers had F0

of 100, 200, or 400 Hz and C values of -1, -0.5, 0, 0.5, or 1 (15 conditions). In total, animals conducted 850-950 test sessions over 8-9 months to complete the experiment.

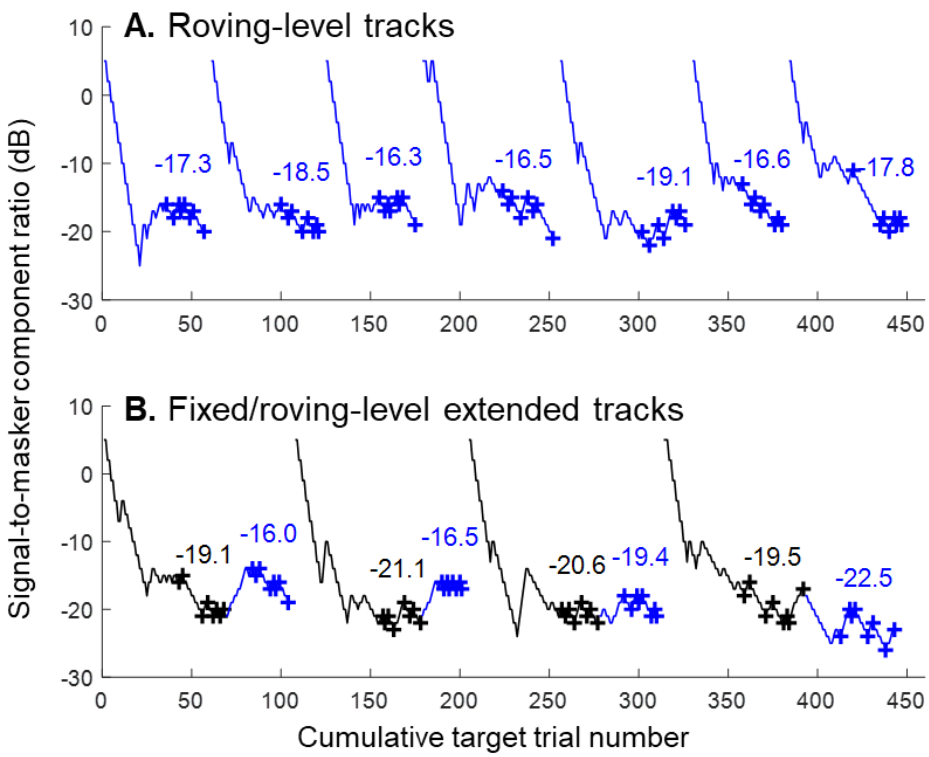

Figure 2. Representative results of two-down one-up tracking sessions in trained budgerigars. Roving-level tracks (A) were obtained with ±10-dB masker-level variation that reduces the usefulness of single-channel energy cues for performing the task. Extended tracks (B) transition from a fixed- (black) to the roving-level (blue) test paradigm within test sessions. Crosses indicate final reversal points used to calculate track thresholds given above in dB. F0 is 100 Hz and C is +1.

Masked thresholds of individual animals were calculated for each condition (Fig. 3A-C) based on the final 25 unbiased tracks. Masked thresholds increased for higher F0 values and for greater absolute C values, with C values of ±0.5 producing intermediate thresholds between C of zero, for which thresholds were lowest, and C of ±1, for which thresholds were highest. A mixed-effects model analysis of masked thresholds showed significant effects of F0 ($F_{2,28}$=34.14, p<0.0001), C ($F_{4,28}$=17.59, p<0.0001), and the F0-by-C interaction ($F_{8,28}$=2.76, p=0.0220). The interaction was due to greater elevation of thresholds with increasing F0 for C of zero and ±0.5 compared to ±1. No significant differences in threshold were observed between negative and positive polarities of the same C value at F0s of 100 Hz (*C* ±1: -1.77 ±1.85 dB; t_{28}=-0.96, p=0.35; *C* ±0.5: -2.55 ±1.85 dB; t_{28}=-1.38, p=0.18), 200 Hz (*C* ±1: 2.02 ±1.85 dB; t_{28}=1.09, p=0.28; *C* ±0.5: 1.39 ±1.85 dB; t_{28}=0.75, p=0.46), or 400 Hz (*C* ±1: 1.12 ±1.85 dB; t_{28}=0.60, p=0.55; *C* ±0.5: -0.46 ±1.85 dB; t_{28}=-0.25, p=0.80; pairwise comparison of least-

squared mean differences ±1 SE), suggesting no asymmetry of Schroeder masking in this species.

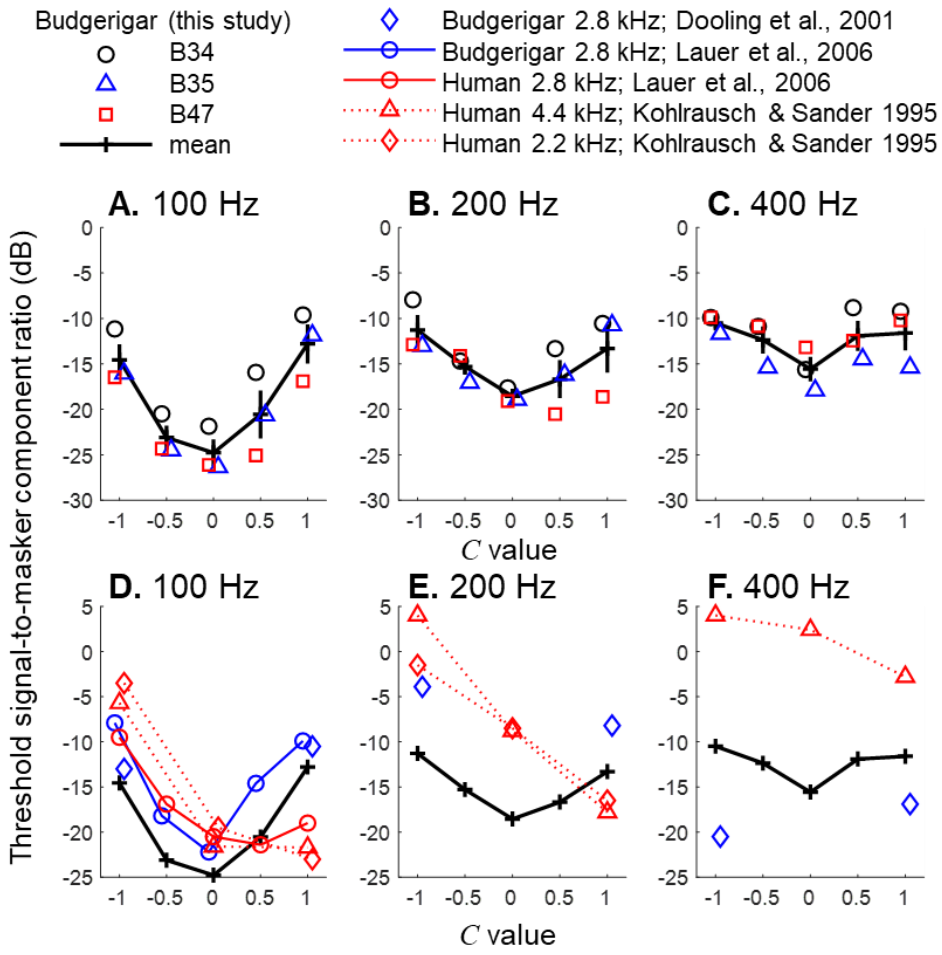

Figure 3. Schroeder-masked behavioral tone-detection thresholds of animals from the current experiment (A-C) and compared to previously published results (D-F) from humans (red) and other budgerigar studies (blue; see legend). Trend lines and error bars indicate means and across-subject standard deviations, respectively. Budgerigars show less masking difference between Schroeder HTCS of opposite polarity, in contrast to human results

Several previous studies quantified Schroeder masking in budgerigars and humans using similar stimulus parameters (Fig 3D-F). Studies by Dooling et al. (2001; budgerigars) and Lauer et al. (2006; budgerigars and humans) used a signal frequency of 2.8 kHz as in the present study, whereas Kohlrausch and Sander (1995; humans) used signal frequencies of 2.2 and 4.4 kHz. In general, the budgerigar thresholds reported previously for F0 of 100 Hz (Dooling et al., 2001; Lauer et al., 2007) appear similar to those from the present study, showing the same V-shaped pattern with lowest thresholds for C of approximately zero (Fig. 3D; blue symbols). In contrast, prior budgerigar thresholds at 200 Hz (Fig. 3E) and 400 Hz (Fig. 3F) are higher and lower, respectively, than those of the present study for unknown reasons. Compared

to previously published human results (Kohlrausch & Sander, 1995; Lauer et al., 2007) (Fig. 3D-F, red symbols), our findings suggest less threshold difference in budgerigars between Schroeder maskers of opposite C polarity. Furthermore, budgerigars appear to show less increase in masked thresholds for higher F0 values than has been reported in humans.

3.1.2. Roving-level paradigm

One difference between our study and previous Schroeder masking studies in budgerigars and humans is that we used a roving-level paradigm for which masker level ranged from 70-90 dB SPL from trial to trial, while previous studies fixed the masker level, generally at 80 dB SPL. A second experiment was conducted in four animals to evaluate the extent to which roving stimulus level impacted thresholds. C values of -1, zero, and 1 were tested using F0 of 100 Hz. The rove effect was assessed by transitioning from a fixed-level (80 dB SPL masker level) to roving-level test procedure (70-90 dB SPL masker level) partway through individual tracking sessions (Fig. 2B). In total, animals conducted 200-300 sessions over 2-3 months to complete the experiment.

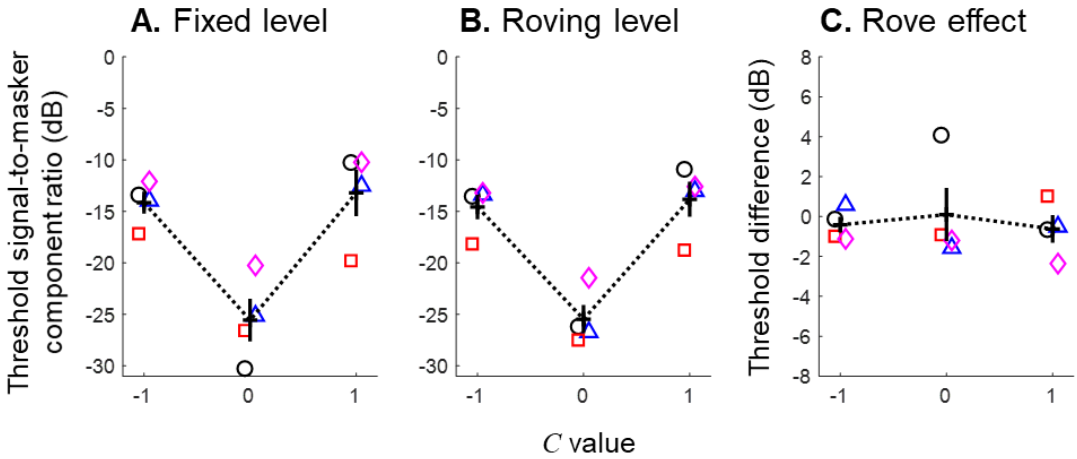

Figure 4. Behavioral threshold from the second experiment comparing performance under fixed- and roving-level test conditions. Trend lines and error bars indicate means and standard deviations, respectively. F0 is 100 Hz. Budgerigars showed little of no effect of roving stimulus level on Schroeder masked tone detection.

Behavioral tracks showed decreasing signal-to-masker level over the first 10-20 trials before stabilizing at the animal's fixed-level thresholds, similar to roving-level tracks from experiment 1. Furthermore, most tracks showed no apparent change in level following the onset of the roving-level paradigm (Fig. 2B). Fixed-level masked thresholds calculated based on the final 25 unbiased tracking sessions (portions of tracks prior to the rove onset) were lower for C of zero than for C of ±1 and similar between C values of -1 and 1, consistent with results of the

first behavioral experiment (Fig. 4A). Masked thresholds calculated from the roving-level part of tracks showed essentially the same pattern (Fig. 4B), suggesting minimal impact of this stimulus perturbation that makes single-channel energy cues less informative. Finally, roving-level threshold shifts calculated from individual test sessions were generally near zero (Fig. 4C), though one animal (B34) showed a mean roving-level shift of 4 dB for C of zero only.

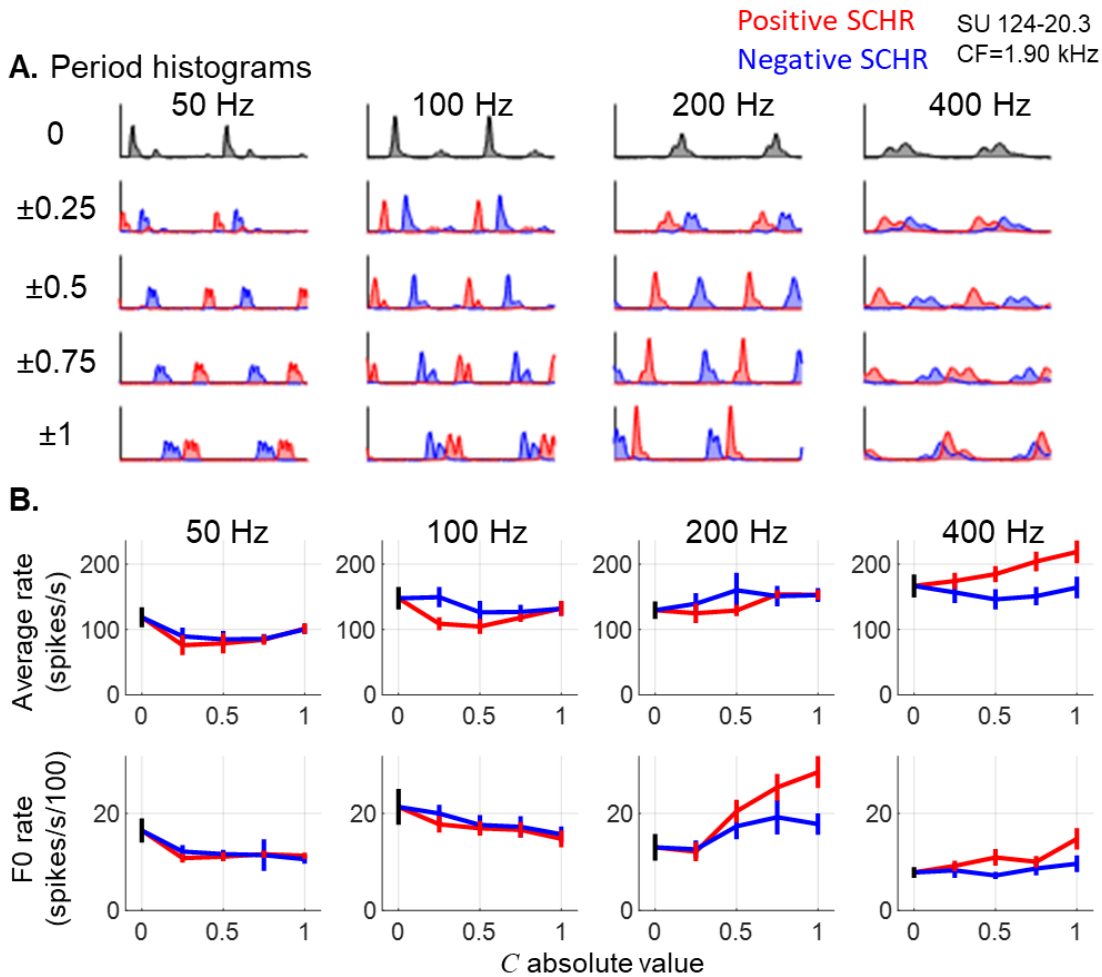

Figure 5. Representative neural responses to Schroeder-phase HTC. Period histograms (A) illustrate variation in average discharge rate over two cycles of F0. F0 is indicated at the top of each column at C value is given to the left. (B) Single-trial average discharge rate (top) and F0 rate plotted as a function of C value for different F0s. Trend lines and error bars show means and across trial standard deviation. Note bias of average discharge rate and F0 rate toward the positive Schroeder for high F0 and absolute C values. CF was 1.90 kHz.

A mixed-effects model analysis of fixed- and roving-level thresholds showed a significant effect of C value ($F_{2,15}$=69.57, p<0.0001) but no significant impact of roving level ($F_{1,15}$=0.12, p=0.74) or the C value by roving-level interaction ($F_{2,15}$=0.054, p=0.95). Thresholds were lower for C of zero than for C of -1 (-11.15 ±1.14 dB; t_{15}=-9.82, p<0.0001) or 1 (-12.00 ±1.14 dB; t_{15}=-10.57, p<0.0001; pairwise comparison of least-squared mean differences ±1 SE), and similar

between C of -1 and 1 (-0.85 ±0.75 dB; t_{15}=-0.75, p=0.47). The estimated effect of roving stimulus level was -0.32 ±0.93 dB (t_{15}=-0.34, p=0.74). An ANOVA of within-track threshold shifts during the roving-level portion of tracks showed no significant intercept (-0.42 ±0.90 dB; $t_{8.9}$=-0.46, p=0.65) or effect of C ($F_{2,6}$=0.18, p=0.84).

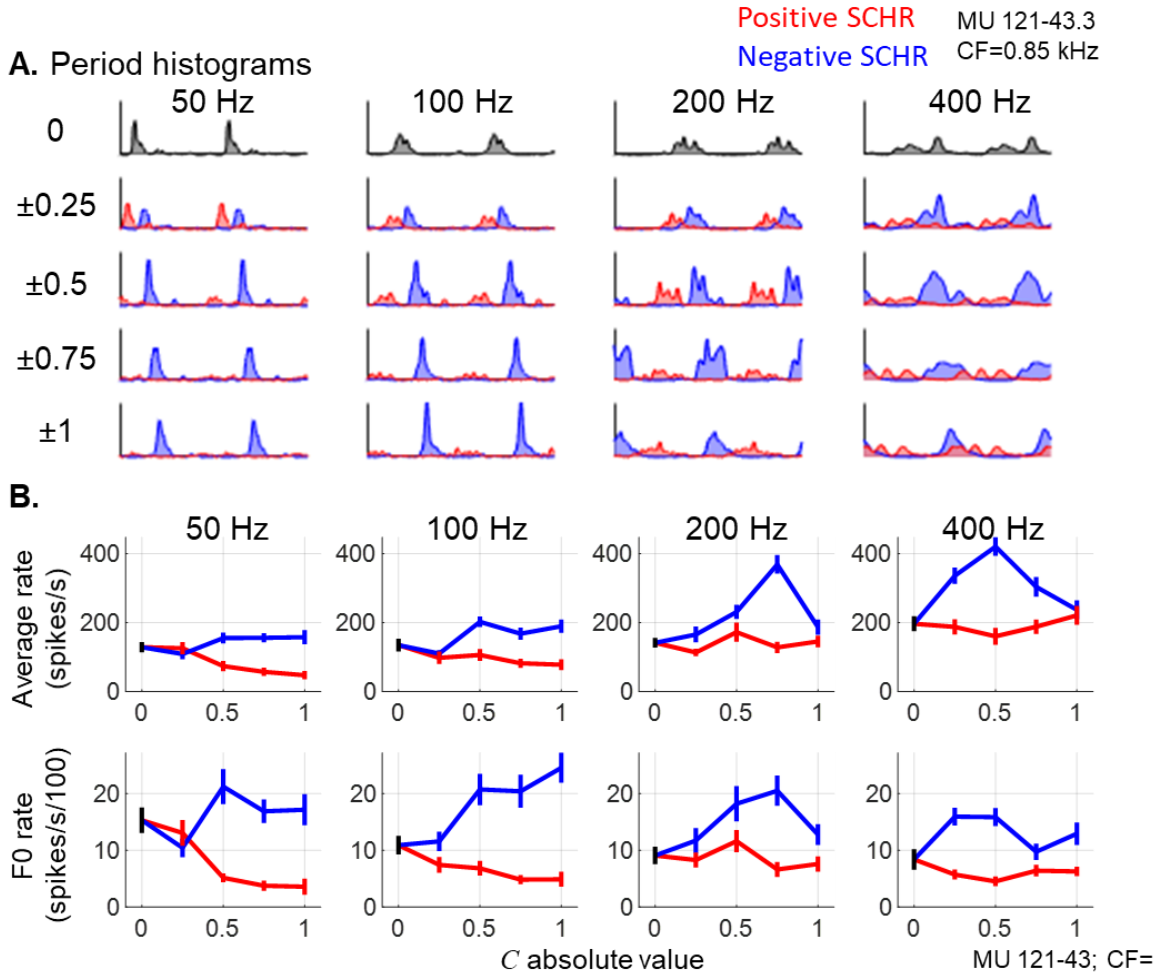

Figure 6. Representative neural responses as in Figure 4, but for a different unit.. Note widespread bias of average discharge rate and F0 rate toward the negative Schroeder, most prominently for high absolute C values. CF was 0.85 kHz.

In summary, results from the two behavioral experiments support prior findings that Schroeder HTCs generated with positive and negative C values produce similar masking in budgerigars, contrasting with prior human studies showing lower thresholds for Schroeder HTCs generated with positive C values. Budgerigars also showed less threshold elevation with increasing masker F0 in experiment 1 than has been reported in humans. Finally, experiment 2 showed that budgerigar detection performance was largely unaffected by roving stimulus level, suggesting that animals use cues other than single-channel energy level to perform the masked detection task.

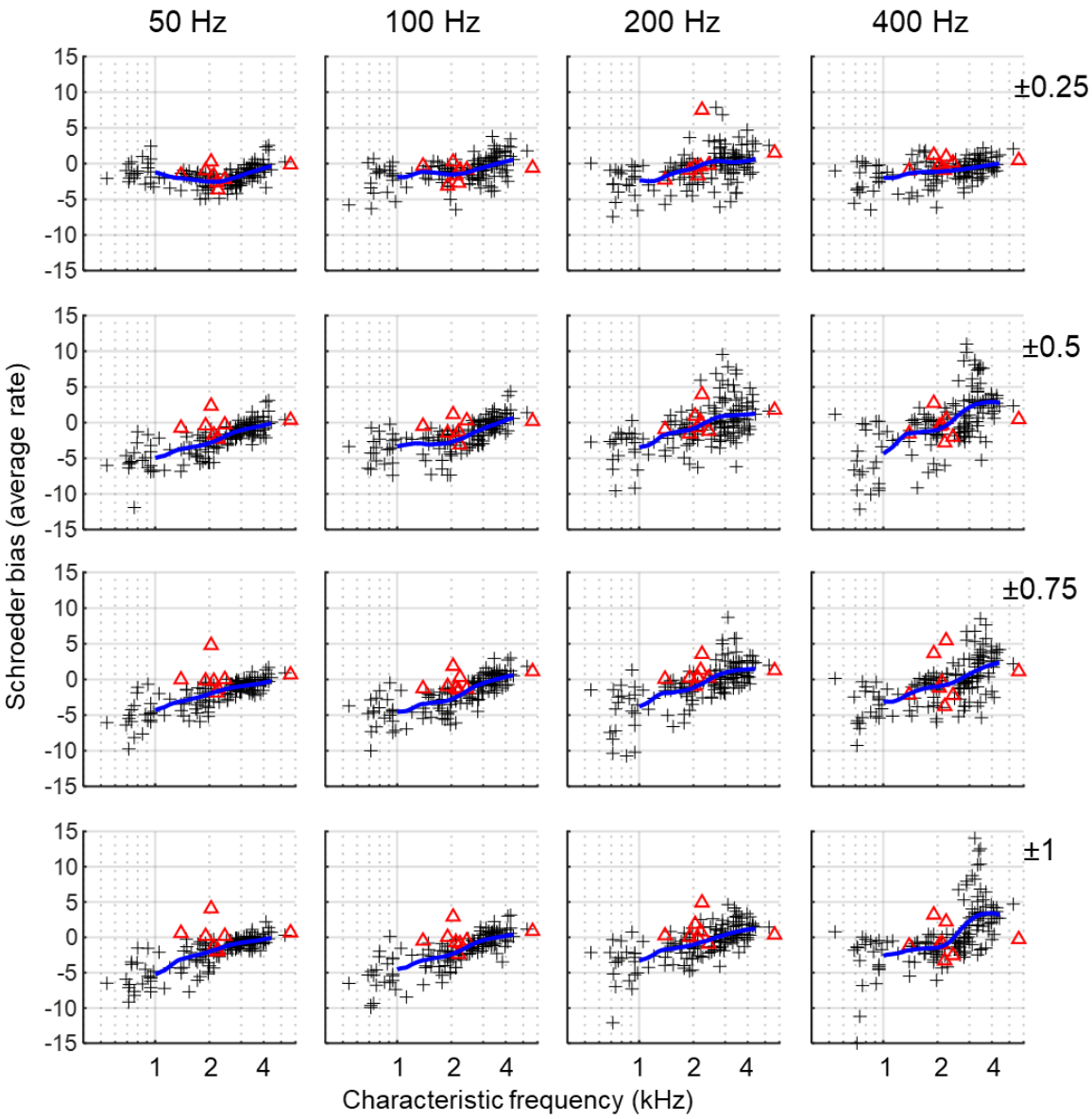

Figure 7. Schroeder bias of average discharge rate, shown as a function of CF for all stimulus conditions. F0 is indicated at the top of each column and absolute C values are give to the right. Red triangles denote data from single-unit recordings.

3.2. Neurophysiological Results

3.2.1. Effects of F0 and C on neural responses to Schroeder-phase HTCs

Neurophysiological recordings were made from the budgerigar IC in awake animals to obtain neural responses to Schroeder HTCs of varying F0 and C value. Units showed V-shaped frequency tuning curves in response to tones with thresholds at CF typically less than 20 dB

SPL. CFs ranging from 0.54-5.6 kHz across the neural population (median: 2.59 kHz; interquartile range: 1.73-3.33 kHz).

Neurophysiological recordings made in response to Schroeder HTCs showed robust fluctuation in instantaneous discharge rate with F0 across the full range of F0 and *C* values studied. Responses of two representative units are shown in Figures 5 and 6. The neural response shown if Figure 5 shows similar average rate and F0 rate between opposite Schroeder polarities for most conditions, but with positive Schroeder bias of F0 rate for higher absolute C values at 200 Hz and positive Schroeder bias of both average rate and F0 rate for higher absolute C values at 400 Hz F0. CF was 1.9 kHz. In contrast, responses illustrated in Figure 6 show prominent negative Schroeder bias of both average rate and F0 under most conditions. CF was 0.85 kHz.

To examine Schroeder bias across the neural population, we calculated a Schroeder bias metric is the mean difference in response between positive and negative Schroeder polarities divided by the pooled standard deviation. Schroeder bias calculated based both average rate (Fig. 7) and F0 rate (Fig. 8) showed similar patterns across the neural population. Whereas Schroeder bias was variable across individual neural recordings of similar CF, the mean Schroeder bias of the neural population increased with increasing CF, a trend that increased in prominence for higher F0 and absolute C values. In general, the transition from negative bias at low CFs to positive bias at higher CF occurred within the CF range from 2-3 kHz. The relatively striking degree of CF-dependent Schroeder bias found here is surprising in light of minimal behavioral differences in masking between C values of opposite polarity, challenging the common assumption that a peakier masker response results in a lower behavioral threshold for target tone detection with this paradigm.

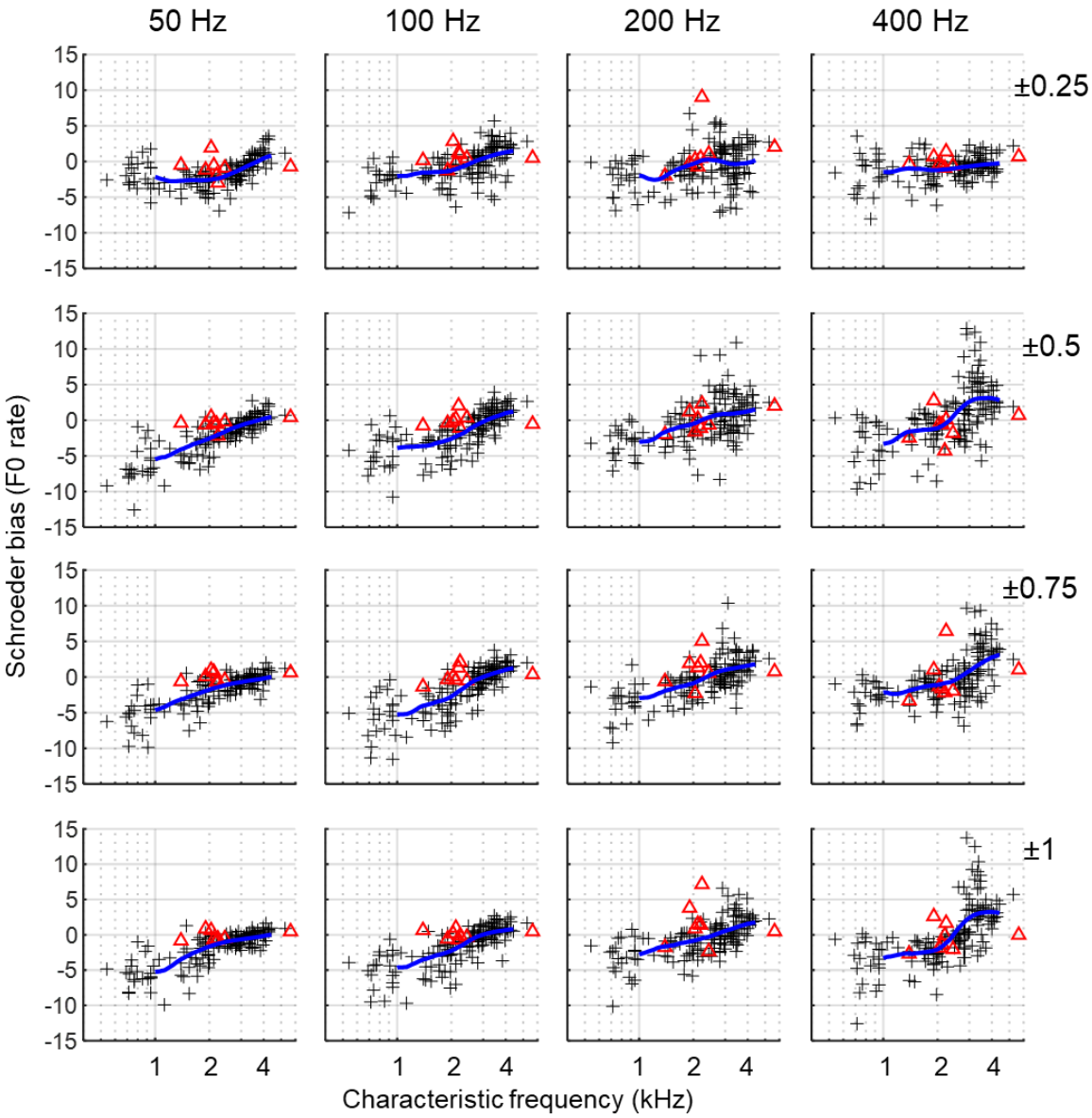

Figure 8. Schroeder bias of F0 rate, shown in the same manner as Figure 7.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carney, L. H., McDuffy, M. J., & Shekhter, I. (1999). Frequency glides in the impulse responses of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *105*(4), 2384–2391. https://doi.org/10.1121/1.426843

Choi, J. H., Jung, H. K., & Kim, T. (2006). A new action potential detector using the MTEO and its effects on spike sorting systems at low signal-to-noise ratios. *IEEE Transactions on Biomedical Engineering*, *53*(4), 738–746. https://doi.org/10.1109/TBME.2006.870239

Dooling, R. J., Dent, M. L., Leek, M. R., & Gleich, O. (2001). Masking by harmonic complexes in birds: behavioral thresholds and cochlear responses. *Hearing Research*, *152*(1–2), 159–172. https://doi.org/10.1016/S0378-5955(00)00249-5

Dooling, R. J., Leek, M. R., Gleich, O., & Dent, M. L. (2002). Auditory temporal resolution in birds: discrimination of harmonic complexes. *The Journal of the Acoustical Society of America*, *112*(2), 748–759. http://www.ncbi.nlm.nih.gov/pubmed/12186054

Dooling, R. J., Lohr, B., & Dent, M. L. (2000). Hearing in Birds and Reptiles. In R. J. Dooling, R. R. Fay, & A. N. Popper (Eds.), *Comparative Hearing: Birds and Reptiles* (pp. 308–359). Springer.

Fletcher, H. (1940). Rev. Mod. Phys. 12, 47 (1940): Auditory Patterns. *Reviews of Modern Physics*.

Fontaine, B., Köppl, C., & Peña, J. L. (2015). Reverse Correlation Analysis of Auditory-Nerve Fiber Responses to Broadband Noise in a Bird, the Barn Owl. *Journal of the Association for Research in Otolaryngology*, *16*(1), 101–119. https://doi.org/10.1007/s10162-014-0494-4

Henry, K. S., Abrams, K. S., Forst, J., Mender, M. J., Neilans, E. G., Idrobo, F., & Carney, L. H. (2017). Midbrain Synchrony to Envelope Structure Supports Behavioral Sensitivity to Single-Formant Vowel-Like Sounds in Noise. *JARO - Journal of the Association for Research in Otolaryngology*, *18*(1), 165–181. https://doi.org/10.1007/s10162-016-0594-4

Henry, K. S., Neilans, E. G., Abrams, K. S., Idrobo, F., & Carney, L. H. (2016). Neural correlates of behavioral amplitude modulation sensitivity in the budgerigar midbrain. *Journal of Neurophysiology*, *115*(4), 1905–1916. https://doi.org/10.1152/jn.01003.2015

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, *84*(2), 541–577. https://doi.org/10.1152/physrev.00029.2003

Kohlrausch, A., & Sander, A. (1995). Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *The Journal of the Acoustical Society of America*, *97*(3), 1817–1829. https://doi.org/10.1121/1.413097

Lauer, A. M., Dooling, R. J., Leek, M. R., & Poling, K. (2007). Detection and discrimination of simple and complex sounds by hearing-impaired Belgian Waterslager canaries. *The Journal of the Acoustical Society of America*, *122*(6), 3615–3627. https://doi.org/10.1121/1.2799482

Leek, M. R., Dent, M. L., & Dooling, R. J. (2000). Masking by harmonic complexes in budgerigars (Melopsittacus undulatus) . *The Journal of the Acoustical Society of America*, *107*(3), 1737–1744. https://doi.org/10.1121/1.428455

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. CUP Archive.

Manley, G. A., Gleich, O., Leppelsack, H. J., & Oeckinghaus, H. (1985). Activity patterns of cochlear ganglion neurones in the starling. *Journal of Comparative Physiology. A, Sensory, Neural, and Behavioral Physiology*, *157*(2), 161–181. https://doi.org/10.1007/BF01350025

Recio, A., & Rhode, W. S. (2000). Basilar membrane responses to broadband stimuli. *The Journal of the Acoustical Society of America*, *108*(5), 2281–2298. https://doi.org/10.1121/1.1318898

Recio, A., Rich, N. C., Narayan, S. S., & Ruggero, M. A. (1998). Basilar-membrane responses to clicks at the base of the chinchilla cochlea. *The Journal of the Acoustical Society of America*, *103*(4), 1972. https://doi.org/10.1121/1.421377

Recio-Spinoso, A., Temchin, A. N., van Dijk, P., Fan, Y.-H., & Ruggero, M. A. (2005). Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *Journal of Neurophysiology*, *93*(6), 3615–3634. https://doi.org/10.1152/jn.00882.2004

Sachs, M. B., Young, E. D., & Lewis, R. H. (1974). Discharge patterns of single fibers in the pigeon auditory nerve. *Brain Research*, *70*, 431–447.

Schroeder, M. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (Corresp.). *IEEE Transactions on Information Theory*, *16*(1), 85–89. https://doi.org/10.1109/TIT.1970.1054411

Smith, B. K., Sieben, U. K., Kohlrausch, A., & Schroeder, M. R. (1986). Phase effects in masking related to dispersion in the inner ear. *The Journal of the Acoustical Society of America*, *80*(6), 1631–1637. https://doi.org/10.1121/1.394327

Strube, H. W. (1985). A computationally efficient basilar-membrane model. *Acoustica*, *58*(4), 207–214. https://www.ingentaconnect.com/content/dav/aaua/1985/00000058/00000004/art00006

Summers, V., de Boer, E., & Nuttall, A. L. (2003). Basilar-membrane responses to multicomponent (Schroeder-phase) signals: Understanding intensity effects. *The Journal of the Acoustical Society of America*, *114*(1), 294–306. https://doi.org/10.1121/1.1580813

Takasaka, T., & Smith, C. (1971). The Structure and Innervation of the Pigeon ' s Basilar Papilla 1. *Journal of Ultrastructure Research*, *35*, 20–65.

Wang, Y., Abrams, K. S., Carney, L. H., & Henry, K. S. (2021). Midbrain-level neural correlates of behavioral tone-in-noise detection: dependence on energy and envelope cues. *The Journal of Neuroscience*, *41*(34), JN-RM-3103-20. https://doi.org/10.1523/jneurosci.3103-20.2021