JASA ARTICLE

On the utility of perceptual anchors during pure-tone frequency discrimination

Samuel R. Mathias,[1,a)] Leonard Varghese,[2] Christophe Micheyl,[3] and Barbara G. Shinn-Cunningham[4]
[1]*Department of Psychiatry, Boston Children's Hospital, Boston, Massachusetts 02115, USA*
[2]*Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, Dartmouth, Massachusetts 02747, USA*
[3]*Starkey France, 94000 Créteil, France*
[4]*Neuroscience Institute, Carnegie Mellon University, Pittsburg, Pennsylvania 15213, USA*

ABSTRACT:
Perceptual anchors are representations of stimulus features stored in long-term memory rather than short-term memory. The present study investigated whether listeners use perceptual anchors to improve pure-tone frequency discrimination. Ten amateur musicians performed a two-interval, two-alternative forced-choice frequency-discrimination experiment. In one half of the experiment, the frequency of the first tone was fixed across trials, and in the other half, the frequency of the first tone was roved widely across trials. The durations of the interstimulus intervals (ISIs) and the frequency differences between the tones on each trial were also manipulated. The data were analyzed with a Bayesian model that assumed that performance was limited by sensory noise (related to the initial encoding of the stimuli), memory noise (which increased proportionally to the ISI), fluctuations in attention, and response bias. It was hypothesized that memory-noise variance increased more rapidly during roved-frequency discrimination than fixed-frequency discrimination because listeners used perceptual anchors in the latter condition. The results supported this hypothesis. The results also suggested that listeners experienced more lapses in attention during roved-frequency discrimination. https://doi.org/10.1121/10.0000584

(Received 30 July 2019; revised 26 November 2019; accepted 18 December 2019; published online 29 January 2020)
[Editor: Jennifer J. Lentz] Pages: 371–380

I. INTRODUCTION

Models of discrimination commonly assume that sensory representations are stored in more than one form of memory. For example, Durlach and Braida (1969) proposed two modes of auditory intensity discrimination. The first, called *trace mode*, is used to discriminate the intensities of sounds separated by up to a few seconds and involves storing accurate but transient representations in short-term memory. The second, called *context-coding mode*, involves longer lasting but vaguer representations coded in terms of their distances from stable referents stored in long-term memory. The term *perceptual anchor* was introduced to describe long-term referents in a later study (Braida *et al.*, 1984). Models of visual discrimination sometimes employ similar constructs. For example, Donkin *et al.* (2015) proposed a model of color discrimination in which individuals store certain hues in short-term memory and assign verbal labels to others (e.g., "greenish blue"), storing those hues along with their labels in long-term memory. Thus, regardless of differences in terminology, auditory and visual scientists agree that short-term and long-term representations are combined to improve discrimination (for more examples, see Hafter *et al.*, 1998; Jackson and Raymond, 2008; Lin and Luck, 2009; Sorkin, 1987; Spencer and Hund, 2002). The precise durations of these representations are debatable and may differ between sensory modalities and stimulus features. Here, we consider short-term representations to be those that persist for up to several seconds and are not likely to be remembered after the end of a trial in a typical psychophysical experiment. By contrast, we consider long-term representations, or perceptual anchors, to persist for the entire duration of a typical experiment.

The present study investigated whether listeners use perceptual anchors during auditory frequency discrimination. In an early study on this topic, Harris (1952a) estimated listeners' difference limens for frequency (DLFs) for pure tones in a two-interval, two-alternative forced-choice (2I-2AFC) experiment. During one set of conditions, the first tone on each trial was always 1000 Hz (fixed). During another set of conditions, the first tone varied across trials between 950 and 1050 Hz (roved). The interstimulus interval (ISI) separating the tones was also manipulated. On average, DLFs were larger at longer ISIs. Moreover, the rate of increase in DLFs as a function of ISI duration was greater during roved conditions than fixed conditions. Harris reasoned that, during fixed conditions, listeners formed a perceptual anchor at the repeated frequency, which allowed them to compare representations of the second tones to the perceptual anchor rather than to representations of the first tones. Since perceptual anchors are stable over time, ISI duration had only a weak influence on DLFs measured

a)Also at: Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA. Electronic mail: samuel.mathias@childrens.harvard.edu

during fixed conditions. By contrast, during roved conditions, listeners were unable to use perceptual anchors, resulting in a stronger influence of ISI duration on roved-frequency DLFs. A possible limitation of Harris's study is that the 100-Hz roving range may have been too narrow to minimize the utility of perceptual anchors during roved conditions (Dai and Micheyl, 2012).

In the present study, listeners also performed a 2I-2AFC pure-tone frequency-discrimination experiment similar to the one described by Harris (1952a). In one half of the experiment, the frequency of the first tone was fixed across trials, and in the other half, the frequency of the first tone was roved. ISI durations and the frequency differences between the tones on each trial were also manipulated. In contrast to Harris, we employed a wide roving range, making it unlikely that listeners used perceptual anchors on roved trials. However, the major novelty of the present study was its analytic approach. We devised a model that explicitly quantified several variables that may have influenced listeners' decisions in the experiment. We hypothesized that one of these variables—the *decay rate*, or rate of increase of memory-noise variance over time—would be greater on roved trials than fixed trials because listeners used perceptual anchors in the latter condition. Although we did not expect other variables in the model to be influenced by the use of perceptual anchors, we hypothesized that they could have differed between conditions for other reasons. We therefore estimated all variables separately for fixed and roved trials.

II. METHOD

A. Listeners

Ten listeners (six female, 19–29 years old) participated in the experiment. All of them had ≤ 15 dB hearing level for frequencies at octave steps between 250 and 8000 Hz and at least some degree of musical experience. Musicians were used to avoid the need for extensive training in pure-tone frequency discrimination prior to the experiment (see Micheyl *et al.*, 2006). None had prior experience in psychoacoustical experiments, all were naive to the aims of the study, and all were paid for their participation. All listeners provided informed consent via documents approved by the Boston University Charles River Campus Institutional Review Board. None reported having absolute pitch.

B. Stimuli

On each trial, listeners heard two pure tones. All tones were 100 ms in duration, presented at 70 dB sound pressure level, and gated on and off with 20-ms raised-cosine amplitude ramps. Tones were generated digitally and delivered diotically via headphones (Sennheiser HD 580, Hannover, Germany) using a 24-bit digital-to-analog converter at a sampling rate of 44 100 Hz (MOTU Microbook, Cambridge, MA). The frequency of the first tone on each trial was either selected randomly (roved) from a rectangular probability distribution defined on the equal-temperament musical scale with a three-octave range (400–3200 Hz) or fixed at the center of the roving range ($\sim$1130 Hz). The frequency of the second tone on each trial differed from the first by an amount denoted by Δ. The possible values of Δ were -1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, and 1 semitone. The duration of the silent ISI was either 2, 5, or 10 s. The corresponding stimulus onset asynchronies (SOAs) were 2.1, 5.1, and 10.1 s.

C. Procedure

Listeners were tested individually in a sound-attenuating chamber (IAC, North Aurora, IL) in two or more sessions lasting $\sim$2 h each, completed on different days. The task instructions were always the same: indicate which tone had the higher pitch by pressing "1" or "2" on a keyboard. Response times were unlimited, and listeners were given visual feedback about response accuracy in the form of green or red text on the monitor (none of the listeners were colorblind). On trials in which $\Delta = 0$, neither response was considered correct, and listeners always received negative feedback.[1] The next trial began 2-s after a response.

The experiment comprised 24 blocks of 90 trials. Blocks contained ten repetitions of each Δ value presented in random order. The duration of the ISI was fixed within blocks but was shuffled between blocks. For half of the listeners, the frequency of the first tone was fixed on trials in the first 12 blocks, and roved on trials in the second 12 blocks. For the other half of the listeners, the opposite was true. After completing a block, listeners were given an opportunity to take a break, and written instructions indicated the details of the next block (e.g., "roved; ISI = 10 s.").

D. Data analysis

1. Modeling the decision process

We assumed that responses were governed by a simple decision process similar to the one assumed by signal detection theory (SDT; Green and Swets, 1988; Macmillan and Creelman, 2005) in the analysis of data from 2I-2AFC experiments but with several important differences. First, we distinguished between two independent sources of internal noise, *sensory noise* and *memory noise*. Sensory noise was assumed to be normally distributed on the equal-temperament musical scale. Sensory-noise variance was assumed to be constant across the roving range, across ISI durations, and across both tones on a trial. The standard deviation of sensory noise in semitones associated with the representation of a single tone frequency was denoted by s.

Memory noise was also assumed to be normally distributed on the musical scale. Memory-noise variance was assumed to be constant across the roving range but differed across ISI durations. Following Kinchla and Smyzer (1967; see also Kinchla and Allan, 1969), we assumed that memory noise followed the Wiener process (Wiener, 1923). This assumption meant that memory-noise variance associated with the representation of a single tone frequency, denoted by m^2, was proportional to the SOA such that $m^2 = d\text{SOA}$,

https://doi.org/10.1121/10.0000584

where d was the rate of increase in memory-noise variance in semitones squared per s. For brevity, we call d the *decay rate*. We assumed that the second tone per trial was not influenced by memory noise.

Another difference between our model and classic SDT was that we allowed for *lapses in attention*. A lapse was defined as a response made independently of the stimuli and task demands. Psychophysical models that do not allow for lapses consider errors made on trials with highly discriminable stimuli to be extremely unlikely, which can severely distort estimates of other variables (Dai and Micheyl, 2011; Klein, 2001; Prins, 2012; Wichmann and Hill, 2001). The two variables related to lapses were l, the lapse probability, and a, the probability of responding "second" on a lapse trial.

Finally, we assumed that decisions on non-lapse trials may have been influenced by *response bias*. SDT models of yes/no experiments account for bias in choosing either "yes" or "no" by incorporating a criterion (Green and Swets, 1988; Macmillan and Creelman, 2005) or equivalently, by shifting observations by a constant value (e.g., DeCarlo, 2010). Similarly, SDT models of 2I-2AFC experiments can account for bias in choosing either "first" or "second" (sometimes called interval bias). Although this kind of bias is commonly assumed to be small and usually ignored (Green and Swets, 1988; Macmillan and Creelman, 2005), this assumption may not hold for many experiments and, similar to lapses, may distort estimates of other variables if ignored (García-Pérez and Alcalá-Quintana, 2011; Yeshurun *et al.*, 2008). Response bias on non-lapse trials, quantified in semitones, was denoted by b.

We combined the above assumptions to derive an analytic expression of a psychometric function (Appendix A). This function yielded the probability that a given listener responded second on a given trial in the experiment. The inputs to the function were two *stimulus variables*, Δ and SOA (defined in Sec. II B), and five *psychological variables*, denoted by a, b, d, l, and s (defined above and summarized in Table I).

2. Parameterization

The five psychological variables were allowed to differ between listeners and between fixed and roved trials. This was achieved by applying what may be described as nonlinear (or generalized) mixed-effects models (Lindstrom and Bates, 1990) to each psychological variable (Appendix B).

TABLE I. Descriptions of the five psychological variables from the psychometric function.

Symbol	Interpretation	Units
a	Second on lapse trials	Probability
b	Bias on non-lapse trials	Semitones
d	Decay rate	Semitones^2/s
l	Lapses	Probability
s	Sensory noise	Semitones

Briefly, each psychological variable was defined as a monotonic transformation of a corresponding *latent variable* (e.g., a was the logistic transform of α). Each latent variable was defined as the sum of a fixed effect and a random effect. The fixed effect was trial type (fixed or roved), and the random effect was listener. This approach exploited the repeated-measures design of the experiment (Baayen *et al.*, 2008). A separate *stochastic variable* was created for each latent variable and effect level (2 levels of fixed effect, 10 levels of random effect), resulting in 60 stochastic variables (5 latent variables $\times$ 12 effect levels).

Five *deterministic variables*, one for each psychological/latent variable, were defined as the difference between the two corresponding fixed-effect stochastic variables (roved minus fixed). These variables, denoted by Λ_α, Λ_β, Λ_δ, Λ_λ, and Λ_ς, were created for hypothesis testing (see Sec. II D 4). For example, if Λ_α were larger than 0, it would imply that a was larger during roved trials, on average, across all listeners.

3. Model fitting and evaluation

Model fitting was done within a Bayesian framework (Gelman *et al.*, 2013; Kruschke, 2014). Observations were assigned a binomial prior distribution, and stochastic variables were assigned informative normal priors that induced appropriate implicit priors on all other variables (see Appendix B). The joint posterior distribution was estimated using the no-U-turn sampling Markov chain Monte Carlo algorithm (Hoffman and Gelman, 2014). Sampling was done in Python (Python Software Foundation[2]) using PyMC3 (Salvatier *et al.*, 2016). Two independent chains of 11 000 samples were collected. The first 1000 samples per chain were used for tuning and then discarded. Chains were inspected for convergence and autocorrelation using Gelman–Rubin $\hat{R}$ (Brooks and Gelman, 1996; Gelman and Rubin, 1992), effective sample size (N_{eff}; Gelman *et al.*, 2013), and Bayesian fraction of missing information (BFMI; Betancourt, 2016). Model goodness of fit was evaluated using Bayesian R^2 (Gelman *et al.*, 2018) and posterior predictive checking (PPC; Gelman *et al.*, 2013).

4. Hypothesis tests

Bayes factors (Jeffreys, 1998; Kass and Raftery, 1995) were used to test five hypotheses, namely that the five deterministic variables differed from zero. The Savage–Dickey method (Dickey and Lientz, 1970; Wagenmakers *et al.*, 2010) was used to approximate Bayes factors. This involved fitting a skew normal distribution (Azzalini, 1985) to its marginal posterior samples, computing the probability density at zero, then dividing the prior probability density at zero by this value. Since the five deterministic variables all had standard normal priors (see Appendix B), prior density at zero was always 0.399.

5. Data and code availability

All data and code are publicly available.[3]

III. RESULTS

A. Raw data

Figures 1 and 2 show the proportions of second responses on fixed trials and roved trials, respectively. For all listeners, the data showed clear sigmoidal psychometric functions. The slopes of these functions were shallower on roved trials than fixed trials, on average. Since the slope of a listeners' psychometric function is inversely proportional to their DLF, this observation is consistent with those made in previous studies, namely that listeners' DLFs tend to be larger during roved-frequency discrimination than fixed-frequency discrimination (e.g., Amitay *et al.*, 2005; Demany and Semal, 2005; Harris, 1952b; Jesteadt and Bilger, 1974; Mathias *et al.*, 2011; Mathias *et al.*, 2010). Slopes were also shallower at longer ISI durations than shorter ones, generally, and this trend was more pronounced on roved trials. This observation is equivalent to the one made originally by Harris (1952a), which led him to conclude that listeners use perceptual anchors during fixed-frequency discrimination.

B. Results from the model

1. Diagnostics and goodness of fit

The two chains of posterior samples did not diverge for any variable (all $\hat{R} > 0.999$) and exhibited low autocorrelation for most variables. For variables with moderate autocorrelation, N_{eff} was satisfactory (all $N_{\mathrm{eff}} \geq 2593$). BFMI was 1.02, which was good. Median Bayesian R^2 was 0.978, and PPC (Figs. 1 and 2) revealed extremely small discrepancies

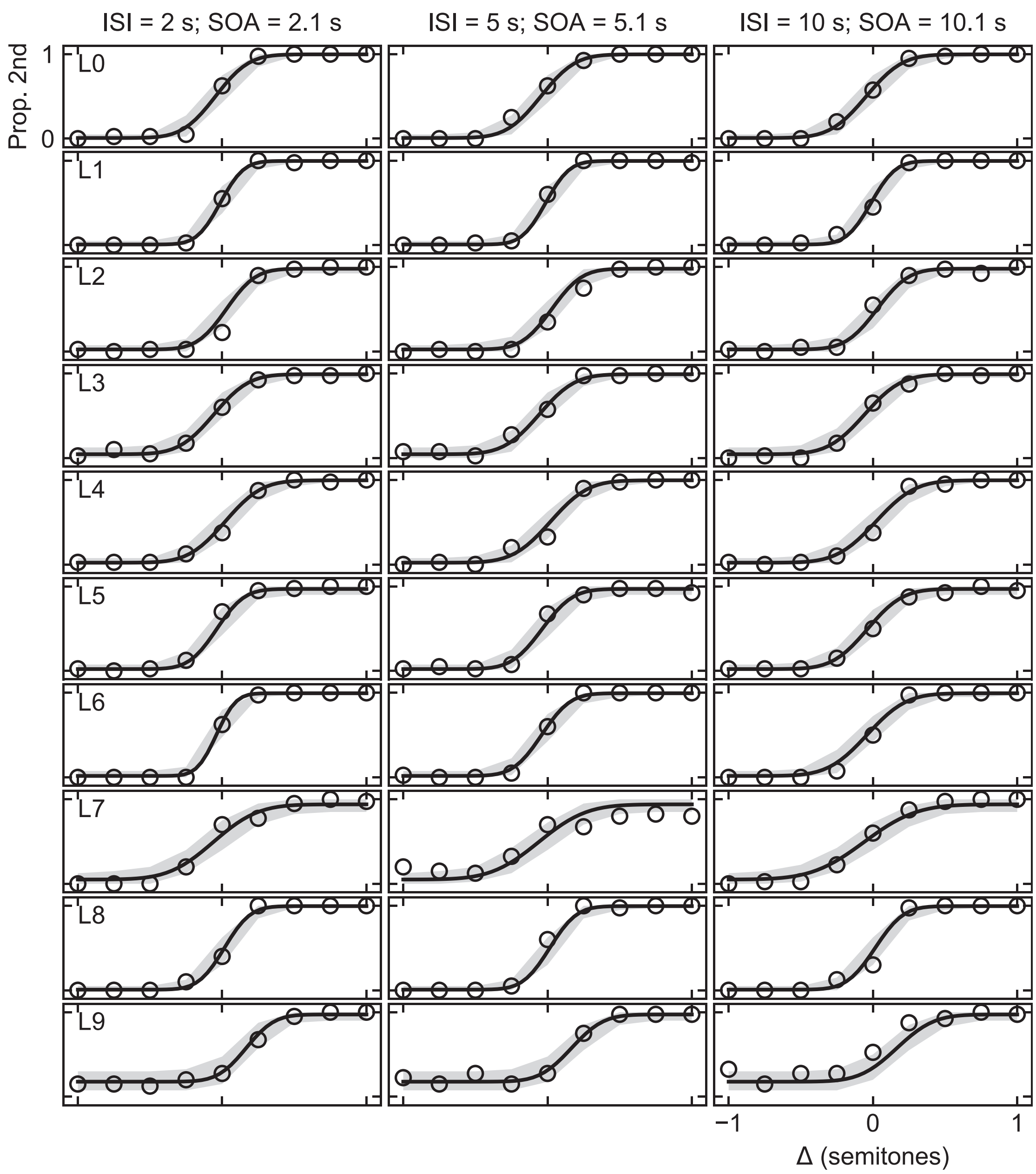

FIG. 1. Raw data from fixed trials and model predictions. Symbols represent the proportion of second responses for a given listener, ISI, and Δ (40 trials per symbol). Curves are posterior mean values of p. Shaded regions are in the range between the 2.5 and 97.5 centiles of simulated data via PPC.

https://doi.org/10.1121/10.0000584

between the predictions of the model and the observed data. Taken together, these findings suggest that the model was an excellent fit to the data.

2. Sensory noise

Table II and Fig. 3 summarize the marginal posterior means of s and s^2, respectively. The group mean posterior value of s was 10.9 cents (1 cent = 0.01 semitone) on fixed trials and 5.6 cents on roved trials. The relevant deterministic variable for testing whether s differed between fixed and roved trials was Λ_ς. As shown in Fig. 4, the 95% credible interval of the marginal posterior distribution of Λ_ς was shifted below zero, and the corresponding Bayes factor (Table III) provided "extreme" evidence for the hypothesis that s was smaller on roved trials.

This result can be interpreted as suggesting that, on average, the standard deviation of listeners' sensory noise was lower on roved trials than fixed trials. At first glance, this result may appear to be counterintuitive because DLFs should be larger, not smaller, during roved-frequency discrimination (e.g., Amitay *et al.*, 2005; Demany and Semal, 2005; Harris, 1952b; Jesteadt and Bilger, 1974; Mathias *et al.*, 2011; Mathias *et al.*, 2010). We return to this point in the discussion (Sec. IV).

3. Memory noise

If listeners used perceptual anchors during fixed-frequency discrimination in the present experiment, they should have showed larger d on roved trials than fixed trials. Typically, across listeners, d was 0.139 cent^2/s on fixed trials and 2.81 cents^2/s on roved trials, which is more than a

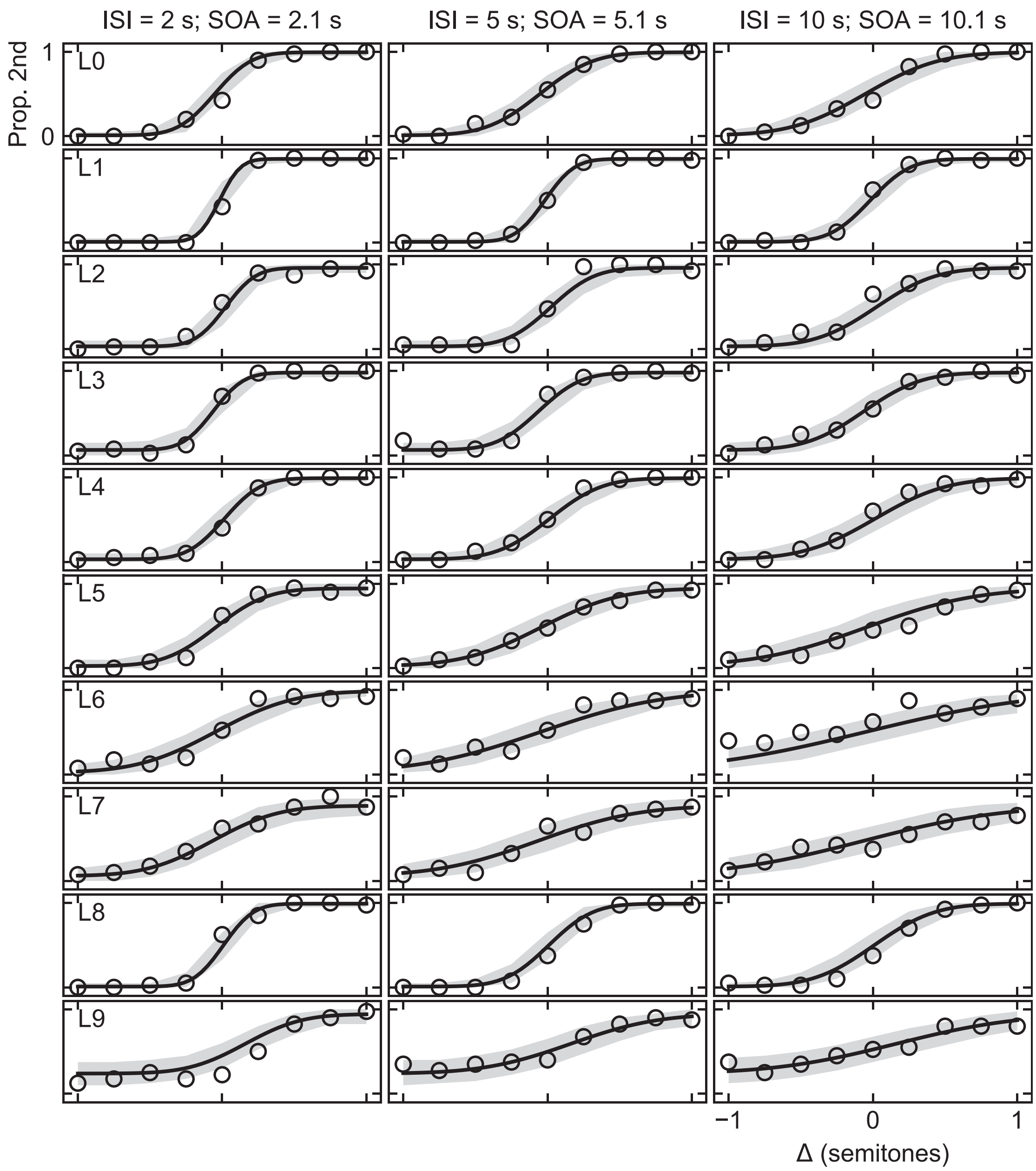

FIG. 2. Same as Fig. 1 but for roved trials.

https://doi.org/10.1121/10.0000584

TABLE II. Marginal posterior means of psychological variables.

	a		*b*		*d*		*l*		*s*	
Listener	Fixed	roved	Fixed	Roved	Fixed	Roved	Fixed	Roved	Fixed	Roved
L0	0.744	0.649	−0.0454	−0.0513	7.48×10^{-4}	0.0152	0.0101	0.0164	0.125	0.0644
L1	0.623	0.513	−0.0125	−0.0184	2.13×10^{-4}	0.00432	0.011	0.0178	0.0888	0.0458
L2	0.557	0.445	0.0295	0.0236	4.93×10^{-4}	0.00991	0.0444	0.0707	0.11	0.0567
L3	0.839	0.769	−0.0563	−0.0623	4.34×10^{-4}	0.0088	0.0529	0.0838	0.133	0.0685
L4	0.823	0.748	0.021	0.0151	6.93×10^{-4}	0.014	0.0248	0.04	0.138	0.0714
L5	0.415	0.311	−0.0374	−0.0433	0.0017	0.0345	0.0518	0.0821	0.0976	0.0503
L6	0.783	0.697	−0.0393	−0.0452	0.00425	0.086	0.0213	0.0345	0.0346	0.0179
L7	0.457	0.349	−0.0603	−0.0662	0.00272	0.0552	0.11	0.169	0.163	0.0842
L8	0.614	0.503	0.0157	0.00971	5.13×10^{-4}	0.0103	0.0113	0.0184	0.0972	0.0501
L9	0.868	0.807	0.166	0.16	0.00212	0.0429	0.205	0.296	0.0986	0.0509
Group mean	0.672	0.579	−0.00189	−0.00783	0.00139	0.0281	0.0543	0.0829	0.109	0.056

20-fold difference. Figure 3 illustrates how this difference influenced the memory-noise variance, m^2, at different ISIs.

The relevant deterministic variable for testing the hypothesis that d differed between fixed and roved trials was Λ_δ. As shown in Fig. 4, the 95% credible interval of the marginal posterior distribution of this variable was shifted above zero. The corresponding Bayes factor (Table III) provided extreme evidence for the hypothesis that d was larger on roved trials.

4. Lapses

On average, listeners lapsed on 5.43% of fixed trials and 8.28% of roved trials (Table II). Figure 4 shows that the 95% credible interval of the marginal posterior distribution of Λ_λ was shifted above zero. The corresponding Bayes factor (Table III) provided "very strong" evidence for the hypothesis that l was larger on roved trials.

Commonly, listeners responded second on lapse trials 67.2% of the time on fixed trials and 57.9% of the time on roved trials (Table II). However, the results are equivocal concerning whether these values are meaningfully different since the 95% credible interval of the marginal posterior distribution of Λ_α included zero (Fig. 4), and the corresponding Bayes factor (Table III) provided "anecdotal" evidence of a difference in a between fixed and roved trials.

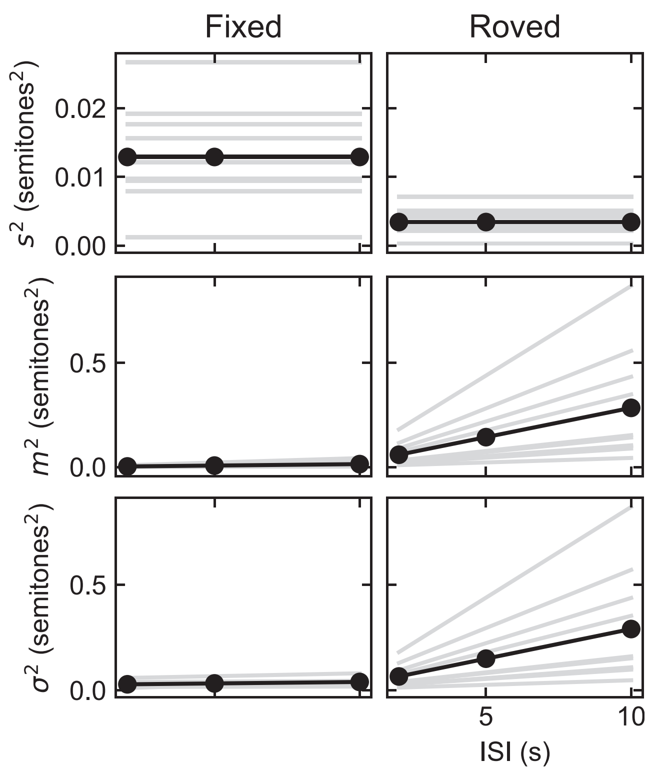

FIG. 3. Posterior mean variances of sensory noise (s^2), memory noise ($m^2 = d\text{SOA}$), and internal noise ($\sigma^2 = 2s^2 + m^2$). Gray lines are posterior means for individual listeners.

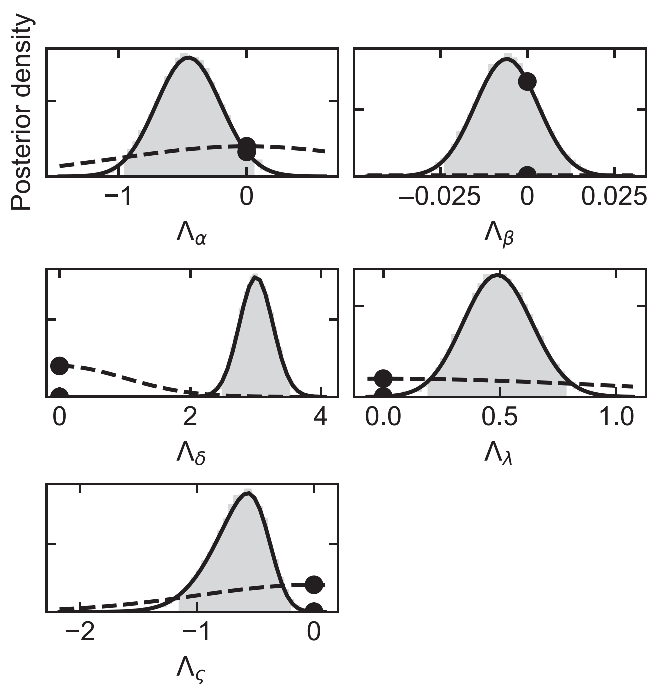

FIG. 4. Histograms of posterior samples from the five deterministic variables. Shaded regions represent the 95% highest posterior density interval or credible interval. Solid curves are skew normal approximations fitted to posterior samples. Dashed curves are prior densities. Points are point probabilities used to approximate Bayes factors.

TABLE III. Summaries of marginal posterior distributions for the five deterministic variables, their proposed interpretations, Bayes factors, and strength of evidence for these interpretations according to the scheme for interpreting Bayes factors originally proposed by Jeffreys (1998) and modified by Lee and Wagenmakers (2014). Entries in the "interpretation" column (except the first row) complete the sentence, "On average, listeners experienced on roved trials…."

Variable	Posterior mean (95% Credible interval)	Interpretation	Bayes factor	Evidence
Λ_α	−0.451 (−0.947, 0.0504)	—	1.23	Anecdotal
Λ_β	−0.00594 (−0.0236, 0.0122)	Similar bias	0.0112	Very strong
Λ_δ	3.013 (2.53, 3.513)	Greater decay rate	1.53×10^{30}	Extreme
Λ_λ	0.493 (0.196, 0.781)	More lapses	43.1	Very strong
Λ_ς	−0.662 (−1.15, −0.209)	Less sensory noise[a]	168	Extreme

[a]More precisely, listeners experienced less sensory noise, on average, across all frequencies within the roving range on roved trials than at ~1130 Hz on fixed trials.

5. Response bias

The results suggest that, on average, listeners were not more biased toward responding first or second on non-lapse roved trials than non-lapse fixed trials. This statement can be made because the corresponding Bayes factor was less than one (Table III), providing very strong evidence for the null hypothesis.

IV. DISCUSSION

The present study investigated whether listeners use perceptual anchors during pure-tone frequency discrimination, as they appear to do when discriminating other features of auditory and visual stimuli (e.g., Donkin *et al.*, 2015; Hafter *et al.*, 1998; Jackson and Raymond, 2008; Lin and Luck, 2009; Spencer and Hund, 2002). Ten amateur musicians completed a 2I-2AFC frequency-discrimination experiment, in which the frequency of the first tone per trial was either fixed or roved over trials within blocks, and durations of the silent ISIs and values of Δ were manipulated. Our analytic approach was intended to disentangle the effects of sensory noise, memory noise, lapses in attention, and bias on performance. We initially hypothesized that listeners would show a greater decay rate (rate of increase in memory-noise variance) on roved trials than fixed trials, and found clear evidence supporting this hypothesis. We also found that listeners experienced more lapses in attention, and possibly less sensory noise, on roved trials. Listeners were not more or less biased during roved-frequency discrimination than fixed-frequency discrimination.

In a much earlier study, Harris (1952a) conducted a very similar experiment and also concluded that listeners use perceptual anchors during fixed-frequency discrimination. However, a possible limitation of this study was that the roving range was too narrow to rule out the possibility that perceptual anchors were useful during roved-frequency discrimination as well (Dai and Micheyl, 2012). This limitation applies to more recent studies on this topic also (e.g., Ahissar *et al.*, 2006). Here, we found evidence for the utility of perceptual anchors during fixed-frequency discrimination while employing a considerably wider roving range on roved trials.

On the basis of our results, we speculate that other features related to perceptual anchors may be found via experiments of pure-tone frequency discrimination. One such feature is the so-called *resolution-edge effect*, where during roved discrimination internal noise is reduced when the stimulus feature to be discriminated falls close to one of the limits of the roving range. Previously, the resolution-edge effect has been observed during pure-tone intensity discrimination (Berliner and Durlach, 1973; Berliner *et al.*, 1977). Here, we attempted to find evidence of the resolution-edge effect via extensions of our model in which the variables of the psychometric function were influenced by the frequency of the first tone on each trial. Unfortunately, these models were complex, and we were unable to develop one which yielded acceptable diagnostic metrics. We speculate that the experiment did not contain enough roved trials, and we intend to explore this topic further in a future study involving more data.

One of the ancillary findings of the present study was that listeners were more likely to experience lapses in attention on roved trials. It is possible that listeners were more attentive during fixed trials than roved trials for some reason. However, this finding more likely reflects a difference in listening strategy than listener attention. Suppose that on a given trial, a listener lapsed during presentation of the first tone and, consequently, failed to form a representation of the frequency of the tone in short-term memory. Further, suppose that during the same trial, the listener regains attention soon enough to form a representation of the frequency of the second tone. Under these circumstances, the listener may be more likely to make a correct response on fixed trials than roved trials because in the former case the frequency of the second tone could be compared to the perceptual anchor instead of the non-existent short-term representation of the frequency of the first tone. On roved trials, when presumably there is no useful perceptual anchor, the listener has nothing with which to compare the frequency of the second tone. In other words, listeners may have been able to compensate for lapses in attention more successfully due to perceptual anchors during fixed trials. The design of the present experiment and model did not permit lapse-with-recovery trials to be distinguished from non-lapse trials. Regardless of whether listeners' states of attention truly differed between fixed trials and roved trials, this finding of a difference in observed lapse rates is important because, as pointed out earlier, psychophysical models that allow for lapses provide better estimates of other variables.

Another ancillary finding was that the standard deviation of listeners' sensory noise was smaller during roved trials than fixed trials, on average. At first glance, this result may seem counterintuitive because listeners usually show larger DLFs during roved-frequency discrimination than fixed-frequency discrimination (e.g., Amitay *et al.*, 2005; Demany and Semal, 2005; Harris, 1952b; Jesteadt and Bilger, 1974; Mathias *et al.*, 2011; Mathias *et al.*, 2010). However, in the present experiment, the variable s represented the average standard deviation of sensory noise *across all frequencies* visited on roved trials. It is possible that this average value was lower than the standard deviation of sensory noise at ∼1130 Hz, the frequency visited on fixed trials. Our model did not permit a direct comparison of sensory noise at ∼1130 Hz on roved trials and fixed trials, and attempts to fit a version of the model where s was dependent on the frequency of the first tone per trial failed (see the earlier point concerning the resolution-edge effect).

In conclusion, the results of the present study are broadly consistent with those of numerous other studies showing that individuals can boost their performance during sensory discrimination by learning from the context of the experiment (e.g., Donkin *et al.*, 2015; Hafter *et al.*, 1998; Jackson and Raymond, 2008; Lin and Luck, 2009; Sorkin, 1987; Spencer and Hund, 2002). It is interesting that, despite their different terminology, psychophysical models proposed by auditory and visual scientists to explain such effects in their respective domains have turned out to be quite similar (cf. Donkin *et al.*, 2015; Durlach and Braida, 1969). This observation provides support for the idea that the mechanisms of perception, sensory discrimination, and memory are broadly similar across the perceptual modalities.

ACKNOWLEDGMENTS

This work was supported by a National Institutes of Health grant awarded to B.G.S.C. (Award No. R01DC013825).

APPENDIX A: DERIVATION OF THE PSYCHOMETRIC FUNCTION

Let X denote whether a listener lapsed on a trial: if $X=0$, they lapsed, and if $X=1$, they did not. Let Y denote a listener's response on a lapse trial: if $Y=0$, they responded first, and if $Y=1$, they responded second. The probability that a listener lapsed and responded second on a given trial was

$$\Pr\{X = 0 \cap Y = 1\} = la. \tag{A1}$$

On a non-lapse trial, we assumed that a listener generated noisy internal representations of the frequencies of the tones. Let x_0 and x_1 denote the true frequencies of the first and second tones in semitones, respectively, and let ψ_0 and ψ_1 denote the corresponding internal representations. Random variable ψ_0 had the probability distribution

$$\psi_0 \sim \mathrm{Normal}\left(x_0, s^2 + m^2\right). \tag{A2}$$

Due to the assumption of Wiener diffusion,

$$m^2 = d\mathrm{SOA}. \tag{A3}$$

For simplicity, we assumed that ψ_0 and ψ_1 had the same sensory-noise variance, and ψ_1 was not affected by memory noise. Consequently, ψ_1 had the probability distribution

$$\psi_1 \sim \mathrm{Normal}\left(x_1, s^2\right). \tag{A4}$$

Let Z denote a listener's response on a non-lapse trial: if $Z=0$, they responded first, and if $Z=1$, they responded second. Formally,

$$Z = 0 \text{ if } \psi_0 + b > \psi_1, \tag{A5}$$

$$Z = 1 \text{ if } \psi_0 + b \leq \psi_1. \tag{A6}$$

Unlike the classic SDT 2I-2AFC model, this decision rule does not involve "differencing" (Macmillan *et al.*, 1977). However, as discussed by DeCarlo (2012), differencing is actually not necessary to derive d' under the SDT 2I-2AFC model. Likewise, the differencing assumption was not necessary here. From the above, the conditional probability that a listener responded second on a non-lapse trial was

$$\Pr\{Z = 1 \mid \psi_0 = y\} = 1 - F\left(y \mid x_1 - b, s^2\right), \tag{A7}$$

where y is a realization of ψ_0, and F is the normal cumulative distribution function. The corresponding unconditional probability was

$$\begin{aligned}\Pr\{Z = 1\} = \int &\left[1 - F\left(y \mid x_1 - b, s^2\right)\right] \\ &\times f\left(y \mid x_0, s^2 + d\mathrm{SOA}\right)\mathrm{d}y,\end{aligned} \tag{A8}$$

where f is the normal probability density function. Conveniently, the above simplifies to (see DeCarlo, 2012)

$$\begin{aligned}\Pr\{Z = 1\} &= 1 - F\left(0 \mid x_1 - b - x_0, 2s^2 + d\mathrm{SOA}\right) \\ &= \Phi\left(\frac{\Delta - b}{\sqrt{2s^2 + d\mathrm{SOA}}}\right),\end{aligned} \tag{A9}$$

where Φ is the standard normal cumulative distribution function. Finally, the probability that a listener responded second on any trial, denoted by p, was given by

$$\begin{aligned}p &= \Pr\{(X = 0 \cap Y = 1) \cup (X = 1 \cap Z = 1)\} \\ &= la + (1 - l)\Phi\left(\frac{\Delta - b}{\sqrt{2s^2 + d\mathrm{SOA}}}\right).\end{aligned} \tag{A10}$$

https://doi.org/10.1121/10.0000584

APPENDIX B: BAYESIAN MODEL

The vector $\boldsymbol{w} = [w_0, w_1, \ldots, w_k]^T$ contained counts of second responses binned by listener, trial type (fixed or roved), ISI, and Δ. Elements of $\boldsymbol{w}$ followed the probability distribution

$$w_i \sim \text{Binomial}(n, p_i), \tag{B1}$$

where $n = 40$ (see Sec. II C), and p_i was the ith element of $\boldsymbol{p} = [p_0, p_1, \ldots, p_k]^T$. Probabilities in $\boldsymbol{p}$ were related to vectors of physical variables and vectors of psychological variables according to Eq. (A10). Psychological variables were transformations of latent variables,

$$\boldsymbol{a} = \text{logistic}(\boldsymbol{\alpha}), \tag{B2}$$

$$\boldsymbol{b} = \text{identity}(\boldsymbol{\beta}), \tag{B3}$$

$$\boldsymbol{d} = \exp(\boldsymbol{\delta}), \tag{B4}$$

$$\boldsymbol{l} = \text{logistic}(\boldsymbol{\lambda}), \tag{B5}$$

$$\boldsymbol{s} = \exp(\boldsymbol{\varsigma}), \tag{B6}$$

and latent variables were related to stochastic variables by

$$\boldsymbol{\alpha} = \Gamma\boldsymbol{\zeta}_\alpha + \Theta\boldsymbol{\xi}_\alpha, \tag{B7}$$

$$\boldsymbol{\beta} = \Gamma\boldsymbol{\zeta}_\beta + \Theta\boldsymbol{\xi}_\beta, \tag{B8}$$

$$\boldsymbol{\delta} = \Gamma\boldsymbol{\zeta}_\delta + \Theta\boldsymbol{\xi}_\delta, \tag{B9}$$

$$\boldsymbol{\lambda} = \Gamma\boldsymbol{\zeta}_\lambda + \Theta\boldsymbol{\xi}_\lambda, \tag{B10}$$

$$\boldsymbol{\varsigma} = \Gamma\boldsymbol{\zeta}_\varsigma + \Theta\boldsymbol{\xi}_\varsigma, \tag{B11}$$

where Γ was a two-column design matrix containing one in the first column and zero in the second column for fixed trials, and vice versa for roved trials, Θ was a design matrix indicating the listener, and the rest were vectors of stochastic, independent, and identically distributed (i.i.d.) variables. Stochastic variables had the priors

$$\zeta_{\alpha_i}, \xi_{\alpha_i}, \zeta_{\beta_i}, \xi_{\beta_i}, \zeta_{\delta_i}, \xi_{\delta_i}, \zeta_{\lambda_i}, \xi_{\lambda_i}, \zeta_{\varsigma_i}, \xi_{\varsigma_i} \overset{i.i.d.}{\sim} \text{Normal}\left(0, \frac{1}{2}\right). \tag{B12}$$

These induced standard normal implicit priors on transformed variables, standard logit-normal priors on $\boldsymbol{l}$ and $\boldsymbol{a}$, standard normal priors on $\boldsymbol{b}$, and standard log-normal priors on $\boldsymbol{d}$ and $\boldsymbol{s}$. Five additional deterministic variables were

$$\Lambda_\alpha = \zeta_{\alpha_1} - \zeta_{\alpha_0}, \tag{B13}$$

$$\Lambda_\beta = \zeta_{\beta_1} - \zeta_{\beta_0}, \tag{B14}$$

$$\Lambda_\delta = \zeta_{\delta_1} - \zeta_{\delta_0}, \tag{B15}$$

$$\Lambda_\lambda = \zeta_{\lambda_1} - \zeta_{\lambda_0}, \tag{B16}$$

$$\Lambda_\varsigma = \zeta_{\varsigma_1} - \zeta_{\varsigma_0}, \tag{B17}$$

all of which had standard normal implicit priors.

[1]Technically, neither response was correct on such trials because the tones were identical. Anecdotally, the inclusion of these impossible trials seemed to increase listeners' attention to the task, especially in blocks where they would have made extremely few errors otherwise (e.g., 0.5-s ISI). However, we did not test this thoroughly and it may have been better to randomly provide positive or negative feedback on impossible trials. We included such trials in the experiment because we initially thought that they would be informative for estimating response bias, which may have differed between fixed trials and roved trials, although this turned out not to be the case (see Sec. III B 5). To determine whether inclusion of impossible trials influenced the results, we refitted the model while treating all such trials as missing. The new results were hardly discernible from those of the original analysis and none of the conclusions changed. We have chosen to omit these additional results from the paper for the sake of brevity, but interested readers can find them within the public repository (Sec. II D 5).

[2]See https://www.python.org (Last viewed 1/20/20).

[3]See https://github.com/sammosummo/PerceptualAnchorsPublic (Last viewed 1/20/20).

Ahissar, M., Lubin, Y., Putter-Katz, H., and Banai, K. (**2006**). "Dyslexia and the failure to form a perceptual anchor," Nat. Neurosci. **9**, 1558–1564.

Amitay, S., Hawkey, D. J. C., and Moore, D. R. (**2005**). "Auditory frequency discrimination learning is affected by stimulus variability," Percept. Psychophys. **67**(4), 691–698.

Azzalini, A. (**1985**). "A class of distributions which includes the normal ones," Scand. J. Stat. **12**(2), 171–178.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Berliner, J. E., and Durlach, N. I. (**1973**). "Intensity perception. IV. Resolution in roving-level discrimination," J. Acoust. Soc. Am. **53**, 1270–1287.

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1977**). "Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects," J. Acoust. Soc. Am. **61**, 1577–1585.

Betancourt, M. (**2016**). "Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo," arXiv:1604.00695.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Brooks, S. P., and Gelman, A. (**1996**). "General methods for monitoring convergence of iterative simulations," J. Comput. Graph. Stat. **7**(4), 434–455.

Dai, H., and Micheyl, C. (**2011**). "Psychometric functions for pure-tone frequency discrimination," J. Acoust. Soc. Am. **130**(1), 263–272.

Dai, H., and Micheyl, C. (**2012**). "Separating the contributions of primary and unwanted cues in psychophysical studies," Psychol. Rev. **119**(4), 770–788.

DeCarlo, L. T. (**2010**). "On the statistical and theoretical basis of signal detection theory and extensions: Unequal variance, random coefficient, and mixture models," J. Math. Psychol. **54**, 304–313.

DeCarlo, L. T. (**2012**). "On a signal detection approach to *m*-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation," J. Math. Psychol. **56**, 196–207.

Demany, L., and Semal, C. (**2005**). "The slow formation of a pitch percept beyond the ending time of a short tone burst," Percept. Psychophys. **67**, 1376–1383.

Dickey, J. M., and Lientz, B. P. (**1970**). "The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain," Ann. Math. Stat. **41**(1), 214–226.

Donkin, C., Nosofsky, R. M., Gold, J., and Shiffrin, R. M. (**2015**). "Verbal labeling, gradual decay, and sudden death in visual short-term memory," Psychonom. Bull. Rev. **22**, 170–178.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

García-Pérez, M. A., and Alcalá-Quintana, R. (**2011**). "Interval bias in 2afc detection tasks: Sorting out the artifacts," Atten. Percept. Psychophys. **73**(7), 2332–2352.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (**2013**). *Bayesian Data Analysis* (Chapman and Hall, New York).

Gelman, A., Goodrich, B., Gabry, J., and Vehtari, A. (**2018**). "*R*-squared for Bayesian regression models," Am. Stat. **73**(3), 307–309.

Gelman, A., and Rubin, D. B. (**1992**). "Inference from iterative simulation using multiple sequences," Stat. Sci. **7**(4), 457–511.

Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos Hills, CA).

Hafter, E. R., Bonnel, A.-M., Gallun, E., and Cohen, F. (**1998**). "A role for memory in divided attention between two independent stimuli," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London, UK), pp. 228–238.

Harris, J. D. (**1952a**). "The decline of pitch discrimination with time," J. Exp. Psychol. **43**, 96–99.

Harris, J. D. (**1952b**). "Pitch discrimination," J. Acoust. Soc. Am. **24**, 750–755.

Hoffman, M. D., and Gelman, A. (**2014**). "The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," J. Mach. Learn. Res. **15**, 1351–1381.

Jackson, M. C., and Raymond, J. E. (**2008**). "Familiarity enhances visual working memory for faces," J. Exp. Psychol. Hum. Percept. Perform. **34**(4), 556–568.

Jeffreys, H. (**1998**). *Oxford Classic Texts in the Physical Sciences Theory of Probability* (Oxford University Press, Oxford, UK).

Jesteadt, W., and Bilger, R. C. (**1974**). "Intensity and frequency discrimination in one- and two-interval paradigms," J. Acoust. Soc. Am. **55**, 1266–1276.

Kass, R. E., and Raftery, A. E. (**1995**). "Bayes factors," J. Am. Stat. Assoc. **90**(430), 773–795.

Kinchla, R. A., and Allan, L. G. (**1969**). "A theory of visual movement perception," Psychol. Rev. **76**(6), 537–558.

Kinchla, R. A., and Smyzer, F. (**1967**). "A diffusion model of perceptual memory," Percept. Psychophys. **2**, 219–229.

Klein, S. A. (**2001**). "Measuring, estimating, and understanding the psychometric function: A commentary," Percept. Psychophys. **63**(8), 1421–1455.

Kruschke, J. K. (**2014**). *Doing Bayesian Data Analysis*, *Second Edition: A Tutorial with R, JAGS, and Stan*, 2nd ed. (Academic, Cambridge, UK).

Lee, M. D., and Wagenmakers, E.-J. (**2014**). *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge, UK).

Lin, P.-H., and Luck, S. J. (**2009**). "The influence of similarity on visual working memory representations," Vis. Cogn. **17**(3), 356–372.

Lindstrom, M. J., and Bates, D. M. (**1990**). "Nonlinear mixed effects models for repeated measures data," Biometrics **46**(3), 673–687.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Dection Theory: A User's Guide* (Erlbaum, Mahwah, NJ).

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). "Psychophysics of categorical perception," Psychol. Rev. **84**(5), 452–471.

Mathias, S. R., Bailey, P. J., Semal, C., and Demany, L. (**2011**). "A note about insensitivity to pitch-change direction," J. Acoust. Soc. Am. **130**(4), EL129–EL134.

Mathias, S. R., Micheyl, C., and Bailey, P. J. (**2010**). "Stimulus uncertainty and insensitivity to pitch-change direction," J. Acoust. Soc. Am. **127**(5), 3026–3037.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res. **219**, 36–47.

Prins, N. (**2012**). "The psychometric function: The lapse rate revisited," J. Vis. **12**, 1–16.

Salvatier, J., Wiecki, T., and Fonnesbeck, C. (**2016**). "Probabilistic programming in python using PyMC3," PeerJ Comput. Sci. **2**, e55.

Sorkin, R. D. (**1987**). "Temporal factors in the discrimination of tonal sequences," J. Acoust. Soc. Am. **82**(4), 1218–1226.

Spencer, J. P., and Hund, A. M. (**2002**). "Prototypes and particulars: Geometric and experience-dependent spatial categories," J. Exp. Psychol. Gen. **131**(1), 16–37.

Wagenmakers, E., Lodewyckx, T., Kuriyal, K., and Grasman, R. (**2010**). "Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method," Cogn. Psychol. **60**, 158–189.

Wichmann, F. A., and Hill, N. J. (**2001**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Wiener, N. (**1923**). "Differential space," J. Math. Phys. **2**, 131–174.

Yeshurun, Y., Carrasco, M., and Maloney, L. T. (**2008**). "Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model," Vis. Res. **48**(17), 1837–1851.