Gradual decay and sudden death of short-term memory for pitch

Samuel R. Mathias, Leonard Varghese, Christophe Micheyl, and Barbara G. Shinn-Cunningham

---

ARTICLES YOU MAY BE INTERESTED IN

Gradual decay and sudden death of short-term memory for pitch

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:

The ability to discriminate frequency differences between pure tones declines as the duration of the interstimulus interval (ISI) increases. The conventional explanation for this finding is that pitch representations gradually decay from auditory short-term memory. Gradual decay means that internal noise increases with increasing ISI duration. Another possibility is that pitch representations experience "sudden death," disappearing without a trace from memory. Sudden death means that listeners guess (respond at random) more often when the ISIs are longer. Since internal noise and guessing probabilities influence the shape of psychometric functions in different ways, they can be estimated simultaneously. Eleven amateur musicians performed a two-interval, two-alternative forced-choice frequency-discrimination task. The frequencies of the first tones were roved, and frequency differences and ISI durations were manipulated across trials. Data were analyzed using Bayesian models that simultaneously estimated internal noise and guessing probabilities. On average across listeners, internal noise increased monotonically as a function of increasing ISI duration, suggesting that gradual decay occurred. The guessing rate decreased with an increasing ISI duration between 0.5 and 2 s but then increased with further increases in ISI duration, suggesting that sudden death occurred but perhaps only at longer ISIs. Results are problematic for decay-only models of discrimination and contrast with those from a study on visual short-term memory, which found that over similar durations, visual representations experienced little gradual decay yet substantial sudden death.
© *2021 Acoustical Society of America.* https://doi.org/10.1121/10.0002992

I. INTRODUCTION

To discriminate two sounds separated in time, a listener must remember the relevant features of the first sound until they have encoded the corresponding features of the second sound. Typically, discrimination performance declines as the duration of the interstimulus interval (ISI) increases (e.g., Berliner and Durlach, 1973; Clément *et al.*, 1999; Harris, 1952; Kinchla and Smyzer, 1967; Mathias *et al.*, 2020; Wickelgren, 1969).[1] Importantly, since this decline in performance occurs when the ISI is silent, it is not caused by interference from irrelevant sounds (cf. Deutsch, 2013; Mathias and von Kriegstein, 2014; Mercer and McKeown, 2010a,b; Ries *et al.*, 2010; Semal and Demany, 1991, 1993; Starr and Pitt, 1997). These findings suggest that representations of sound features are stored within—and eventually, forgotten from—auditory short-term memory (ASTM).

The conventional view of forgetting from ASTM is that representations experience *gradual decay*, becoming weaker or less precise over time. For example, according to the influential model proposed by Kinchla and Smyzer (1967), representations fluctuate over time according to Wiener

diffusion (Wiener, 1923).[2] Diffusion can be considered to increase *internal noise* (Swets *et al.*, 1959). The model by Kinchla and Smyzer (1967) assumes that the variance of internal noise is proportional to the ISI—or more precisely, the stimulus-onset asynchrony (SOA)—whereas other models assume that this relationship is nonlinear (e.g., Wickelgren, 1969). Historically, it has proven difficult to determine the exact gradual-decay function and adjudicate between different gradual-decay models (Laming and Scheiwiller, 1985). In other words, the results of many previous studies are broadly consistent with gradual decay.

An alternative view of forgetting is that representations are maintained in ASTM without decaying until at some point they disappear without a trace. In vision, this has been called *sudden death* (Zhang and Luck, 2009). During a trial in a discrimination experiment, if the representation of the first sound suddenly dies before the second representation is formed, the listener's response on that trial must be a complete guess, provided that care has been taken to rule out alternative decision strategies relying on long-term memory (Dai and Micheyl, 2012). Sudden death is a random event that may or may not occur on any given trial but is more likely to occur when the ISI is longer. Thus, sudden death predicts a gradual decline in performance (when the outcomes of many trials are averaged) as a function of

---

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

increasing ISI duration. Results from studies in which discrimination performance declined as a function of ISI duration (e.g., Berliner and Durlach, 1973; Clément et al., 1999; Harris, 1952; Kinchla and Smyzer, 1967; Mathias et al., 2020; Wickelgren, 1969) may be explained either by gradual decay or sudden death or perhaps a combination of both.

One trick to disentangling gradual decay and sudden death is to estimate internal noise and guessing probabilities simultaneously, using either *delayed-estimation tasks* (Prinzmetal et al., 1998; Wilken and Ma, 2004) or *psychometric functions* (Klein, 2001). A number of visual short-term memory (VSTM) studies have employed the former approach. Briefly, an observer sees a transient visual stimulus (e.g., a colored square) and estimates its relevant feature using a response array (e.g., a color wheel). On each trial, the error (e.g., angular separation between the true and estimated colors) is recorded. The distribution of errors over trials typically resembles a mixture of a normal-like distribution, reflecting errors on non-guess trials, and a uniform-like distribution, reflecting errors on guess trials. Internal noise is related to the spread of the normal-like distribution, whereas guessing probability is related to the mixture weighting. Perhaps surprisingly, results from some visual delayed-estimation tasks suggest that color and shape representations experience almost no gradual decay and substantial sudden death (Zhang and Luck, 2009). Whereas delayed-estimation tasks have been developed for the auditory domain (Kumar et al., 2013; Teki and Griffiths, 2014), as discussed later, these tasks may have some limitations and may not be strongly analogous to tasks used in vision studies (see Sec. IV).

A complementary approach to estimating internal noise and guessing probabilities simultaneously is to use psychometric functions. In a typical two-interval, two-alternative forced-choice (2I-2AFC) discrimination experiment, internal noise is more likely to cause a listener to make the wrong decision on a trial in which the stimuli are difficult to discriminate than on a trial in which the stimuli are highly discriminable. When proportions of responses belonging to one response category are plotted as a function of the differences in physical properties of the stimuli (i.e., a psychometric function; see Fig. 1), greater internal noise leads to a shallower slope. However, the lower and upper asymptotes of the psychometric function do not change. By contrast, guesses should occur randomly on any trial in a 2I-2AFC experiment regardless of discriminability. Therefore, increases in guessing probability draw the asymptotes of a psychometric function toward the middle. Previous studies have used similar approaches to estimate the extent to which psychometric functions are contaminated by guesses due to lapses in attention (Dai and Micheyl, 2011; Klein, 2001; Mathias et al., 2020; Prins, 2012; Wichmann and Hill, 2001). Here, we adopt a similar approach to disentangle gradual decay and sudden death in 2I-2AFC frequency-discrimination experiments.

In the present study, we investigated whether pitch representations experience gradual decay, sudden death, or a
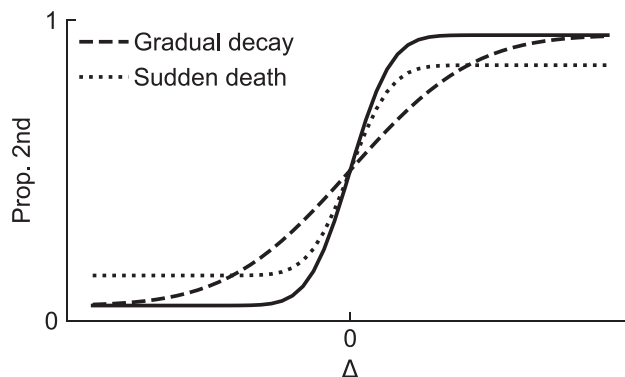

FIG. 1. Three possible psychometric functions from a 2I-2AFC discrimination experiment. On each trial, the listener chooses either the first or second stimulus (e.g., "which was higher?"). Curves show the proportion of times that the second stimulus is chosen as a function of Δ, denoting the relevant physical property of the second stimulus minus that of the first stimulus. The solid curve shows a reference psychometric function. The dashed line shows a psychometric function with greater internal noise relative to the reference as predicted by the gradual-decay view of forgetting. This psychometric function has a shallower slope but similar values at the asymptotes (extreme values of Δ). The dotted line shows a psychometric function with a greater guessing probability relative to the reference as predicted by the sudden-death view. The asymptotes of this psychometric function are closer to the middle of the y axis.

combination of both. In two 2I-2AFC pure-tone frequency-discrimination experiments, the frequency of the first tone was randomized per trial over a wide range, making it likely that listeners based their decisions on pitch representations stored in ASTM rather than in long-term memory (cf. Harris, 1952; Mathias et al., 2020). The frequency differences between the tones, denoted by Δ, and the durations of the silent ISIs were manipulated across blocks of trials.

Data were analyzed in two ways. First, we fitted a model that simultaneously estimated the listeners' internal noises and guessing probabilities (along with other parameters of the psychometric function) at each ISI duration. If listeners' pitch representations experienced gradual decay, there should be more internal noise, on average, on trials with longer ISIs. If listeners' pitch representations experienced sudden death, guessing should be more likely, on average, on trials with longer ISIs. We also analyzed the data by fitting additional models that made more explicit predictions concerning the potential effects of gradual decay and sudden death, and we explored which of these best explained the data.

II. METHOD

A. Experiment 1

Eight listeners (L0–7) participated in experiment 1 (four female, 19–29 years old). Each had a ≤ 15 dB hearing level for frequencies at octave steps between 250 and 8000 Hz and at least some degree of musical experience. Musicians rather than nonmusicians were used to avoid the need for extensive training in pure-tone frequency discrimination prior to the experiment (Micheyl et al., 2006). None had experience in psychoacoustical experiments, all were

naive to the aims to 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 of the listeners reported having absolute pitch.

On each trial, listeners heard two 100-ms pure tones, presented at 70 dB sound pressure level, gated on and off with 20-ms raised-cosine amplitude ramps to reduce spectral splatter, and separated by silent ISIs of 0.5, 2, 5, or 10 s. 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). On each trial, the frequency of the first tone was roved uniformly on the $\log_2$ scale over a three-octave range (400–3200 Hz). The frequency difference between the tones, Δ, was selected from 13 values, which were roughly logarithmically spaced on the $\log_2$ scale (-0.75, -0.38, -0.19, -0.1, -0.05, -0.03, 0, 0.03, 0.05, 0.1, 0.19, 0.38, and 0.75 semitone). Trials were organized into blocks containing $\sim$50 trials each. The ISI was fixed for all trials within a block, and blocks were randomly ordered. All listeners but L0 completed 40 trials of each combination of ISI and Δ. Due to a programming error, L0 completed 38–50 trials per condition.

Listeners were tested alone in a sound-attenuating chamber (IAC, North Aurora, IL). Task instructions were always the same: indicate which tone had the higher pitch by pressing "1" or "2" on the computer keyboard. Response times were unlimited. After each trial, listeners were given visual feedback about the response accuracy in the form of green or red text on the computer monitor (none of the listeners were colorblind). On trials in which $\Delta = 0$, neither response was considered correct, and listeners always received negative feedback.[3] The next trial began 2 s after a response. Listeners completed the experiment over several sessions, completed on different days, and each session lasted $\sim$2 h.

B. Experiment 2

Values of Δ exceeding a listener's difference limen capture the extremities of their psychometric function and are highly informative for estimating the guessing probability. However, for some listeners in experiment 1, the most extreme Δ values ($\pm$0.75 semitone) were not much larger than their difference limens. Therefore, their guessing probabilities may not have been estimated precisely, which, in turn, could have influenced estimation of the standard deviation of their internal noises and influenced overall results. Experiment 2 was performed to address this potential limitation of experiment 1.

Four listeners from experiment 1 (L0–3) and three new listeners (L8–10) participated in experiment 2 (seven total; four female; age range 19–29 years old). Like the other listeners, the three new listeners had normal hearing, some degree of musical experience, and no experience in

psychoacoustical experiments. They were also naive to the aims of the study, paid for their participation, and provided informed consent via approved documents. None reported having absolute pitch.

Experiment 2 was identical in design to experiment 1 except that Δ values were linearly spaced and spanned twice that of the previous range. The possible values were -1.5, -1.25, -1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1, 1.25, and 1.5 semitones. All listeners except L3 completed 40 trials of each combination of ISI and Δ; due to a programming error, L3 completed 40 trials of each 0.5-s ISI condition and 45 trials of all other conditions. As in experiment 1, trials were organized into blocks of $\sim$50 trials, the ISI was fixed for all trials within a block, and listeners completed the experiment over several sessions on different days with each session lasting $\sim$2 h. Experiment 2 was conducted after experiment 1.

C. Data analysis

1. Agnostic model

In a previous study, we developed a Bayesian model (Gelman, 2014; Kruschke, 2015; Lee and Wagenmakers, 2013) to explain listeners' responses in 2I-2AFC experiments that estimated internal noise and guessing probabilities simultaneously (Mathias et al., 2020). Here, we used a modified version of that model to test whether internal noise and guessing probabilities differed between ISI durations.

The model assumed that on each trial, there was non-zero probability that a listener responded completely at random regardless of the stimuli. We call this behavior *guessing*. We let g denote the guessing probability and a denote the probability of choosing the second tone while guessing. The model further assumed that on non-guessing trials, listeners formed two noisy representations corresponding to the tones' pitches. Pitch representations were corrupted by internal noise that was normally distributed on the musical scale and had equal variance across all frequencies. We let n denote the standard deviation of the ensemble internal noise (i.e., the combined noise affecting both pitch representations per trial) in semitones. Listeners chose the first tone if the first pitch representation—plus an amount to account for response bias on non-guessing trials, denoted by b—was greater than the second pitch representation, and chose the second tone otherwise. These assumptions were combined to form a psychometric function with four parameters,

$$p = ag + (1 - g)\Phi\left(\frac{\Delta - b}{n}\right), \tag{1}$$

where p is the probability that a listener chose the second tone, Φ is the cumulative distribution function of the standard normal distribution, and Δ is the frequency of the first tone minus the frequency of the second tone. Given the similarities between this model and the one described by Mathias et al. (2020), we have omitted its derivation.

J. Acoust. Soc. Am. **149** (1), January 2021

Mathias *et al.*   261

Each parameter of the psychometric function (a, b, n, and g) was the sum of a stochastic random variable representing the baseline value of that parameter per listener and a stochastic random variable representing the fixed or group-level effect of ISI duration on that parameter. Group-level effects were coded using reduced-rank dummy coding with ISI = 0.5 s as the reference condition; this meant that, per parameter, there were variables representing the group-level differences between ISI = 0.5 s and ISI = 2 s, ISI = 0.5 s and ISI = 5 s, and ISI = 0.5 s and ISI = 10 s. Bayes factors (BFs; Jeffreys, 1998; Kass and Raftery, 1995), approximated via the Savage–Dickey method (Dickey and Lientz, 1970; Wagenmakers et al., 2010), were used to test whether these variables differed from zero. This involved fitting skew normal distributions (Azzalini, 1985) to the variable's prior and posterior samples, computing their probability densities at zero, and then dividing the prior probability density at zero by the posterior probability at zero. BFs were interpreted using the scheme proposed by Jeffreys (1998) and later modified by Lee and Wagenmakers (2013).

Mildly informative priors were assigned to random variables. See Appendix A for the complete specification.

We describe the above model as *agnostic* because it did not assume that either gradual decay or sudden death actually occurred—it was possible for the group-level differences in n and g across ISI durations to be zero or even negative. Moreover, the model allowed a and b to differ between ISI durations, although neither view of forgetting made any predictions about these parameters.

2. Prescriptive models

We modified the agnostic model to create three *prescriptive* models, called the *decay-and-death* model, the *decay-only* model, and the *death-only* model. Under these models, the ensemble variance of internal noise, n^2, was decomposed into independent variance components: *sensory noise*, the reflecting error in the initial encoding of representations, and *memory noise*, which accumulated over time according to Wiener diffusion (Kinchla and Smyzer, 1967; Wiener, 1923). Wiener diffusion may be considered to be a simple model of gradual decay, and because previous studies comparing different gradual-decay models have produced somewhat equivocal results (Laming and Scheiwiller, 1985), we consider it to be a reasonable choice. Following our previous work (Mathias et al., 2020), we assume that sensory noise was equal for both tones per trial and memory noise only affected the first tone. Therefore,

$$n^2 = 2s^2 + m^2, \tag{2}$$

where s is the standard deviation of sensory noise and m is the standard deviation of memory noise. Wiener diffusion means that the memory-noise variance is proportional to the SOA or the ISI plus the duration of the first tone (0.1 s). Therefore,

$$m^2 = d\mathrm{SOA}, \tag{3}$$

where d is the decay rate in semitones squared per second.

The prescriptive models further assumed that the guessing probability, g, was the union of two nonexclusive probabilities: the probability of a lapse in attention, denoted by l, and the probability of sudden death, denoted by q. We define lapses in attention as guesses that occur only at the response stage after the computation of the decision variable, which are not influenced by ISI duration. Guessing probability is, therefore,

$$g = l + q - lq. \tag{4}$$

We are not aware of any previous study proposing that sudden death follows a specific function. We reasoned that a sensible assumption, analogous to the assumption of Wiener diffusion for gradual decay, is that the probability of sudden death occurring within a 1-s interval is constant. Letting u denote this probability, it follows that the probability of sudden death not occurring in a 1-s interval is $1 - u$. It also follows that the probability of sudden death not occurring in a variable interval is the intersection of all infinitesimal non-overlapping subintervals or $(1 - u)^t$, where t is the interval in seconds. Therefore, the probability of sudden death occurring within t is $1 - (1 - u)^t$, and q as a function of the SOA is given by

$$q = 1 - (1 - u)^{\mathrm{SOA}}. \tag{5}$$

Under all three prescriptive models, psychometric-function parameters a, b, l, and s were estimated separately per listener but assumed to be the same on every trial. The models differed in terms of d and u: under the decay-and-death model, they were both estimated per listener, whereas under the decay-only and death-only models, u and d, respectively, were always zero. To improve the plausibility of the prescriptive models, stochastic random variables were assigned informative, hierarchical, multivariate priors (see Appendix B).

The prescriptive models were fitted to the data separately and their goodness-of-fits were compared using leave-one-out cross validation (LOO) with Pareto-smoothed importance sampling (Vehtari et al., 2017). The LOO information criterion (IC) quantifies the out-of-sample predictive accuracy of a Bayesian model while penalizing for its complexity; when LOO ICs are reported on a log scale (in contrast to a deviance scale), a model with a greater score may be considered to be a better model than a model with a smaller score. Like other ICs, absolute LOO ICs are not meaningful. They are commonly interpreted in terms of the difference in IC between models divided by the standard error of the difference; in the social and psychological sciences, differences on the order of two standard errors or greater are usually considered to be substantial.

3. Sampling and checking

Joint posterior distributions were estimated using no-U-turn sampling (NUTS; Hoffman and Gelman, 2014) implemented in PyMC3 (Salvatier et al., 2016). Per model,

three independent chains of 15 000 samples were collected. The first 5000 samples per chain were used for tuning and then discarded. Chains were inspected for good sampling and convergence using various diagnostic metrics before being concatenated. Prior and posterior distributions of all variables were compared visually to ensure that appropriate priors were chosen.

4. Additional analyses, data, and code availability

We examined the consequences of numerous modifications to the models described above. We fitted the models to the experiments separately and to both experiments combined[4] with the two poorest-performing listeners excluded (see Sec. III A). We created versions of the agnostic model where a, b, g, and/or n were fixed across ISI durations. We fitted versions of the prescriptive models with different kinds of priors (nonhierarchical and hierarchical, univariate and multivariate). With the exception of one ancillary result, none of these modifications yielded any substantive changes in the findings. Therefore, we have omitted them for the sake of brevity. However, all models and results can be found online.[5]

III. RESULTS

A. Raw data

As shown in Fig. 2, all listeners exhibited clear psychometric functions, although L6 and L7 (who performed experiment 1 only) had noticeably different psychometric functions from the other listeners. These two listeners found the experiment boring and were perhaps less motivated than the others. On these grounds, one could make a case for excluding their data; however, for the sake of transparency, we opted to include them. We re-performed all analyses with these listeners excluded, and none of the results changed qualitatively (see Sec. II C 4). As expected, the wider range of Δ values in experiment 2 resulted in capturing more of the asymptotes of listeners' psychometric functions than were captured in experiment 1.

B. Results from the agnostic model

The agnostic model produced good, convergent posterior samples and fitted the data well. The Bayesian fractions of missing information (BFMIs; Betancourt, 2016) for the three chains were 0.997–1.038, all rank-normalized $\hat{R}$ (Vehtari *et al.*, 2020) were ≤ 1.000, and all effective sample sizes (ESS, calculated using the bulk method; Vehtari *et al.*, 2020) for stochastic random variables were $\geq 1.05 \times 10^4$. Median Bayesian R^2 (Gelman *et al.*, 2019) was 0.977 (standard deviation = 0.019), and visual posterior predictive checks (Fig. 2) revealed a high correspondence between the model predictions and data.

If representations experienced gradual decay, there should have been more internal noise (i.e., larger n) on trials with longer ISIs than on trials with shorter ISIs on average. Posterior distributions of the relevant stochastic random variables are summarized in Fig. 3. For all three variables, posterior means were positive, and the 95% highest density intervals did not include zero. Furthermore, BFs for all three corresponding hypothesis tests revealed "extreme" evidence against the null hypothesis ((BF for 0.5-s versus 2-s ISI = 2.99×10^4; BF for 0.5-s versus 5-s ISI = 2.72×10^{16}; BF for 0.5-s versus 10-s ISI = 4.55×10^{20}). Together, these results suggest that internal noise monotonically increased with increasing ISI duration.

If representations experienced sudden death, there should have been a greater guessing probability (i.e., larger g) on trials with longer ISIs than on trials with shorter ISIs on average. Unexpectedly, the mean and entire 95% highest density interval of the variable representing the difference in guessing probability between ISI = 0.5 s and ISI = 2 s was negative (Fig. 3), indicating a smaller guessing probability on trials with the longer ISI. The corresponding BF was 2254.988, suggesting extreme evidence against the null hypothesis of no difference. For the variable representing the difference between ISI = 0.5 s and ISI = 5 s, the 95% highest density interval included zero, and the corresponding BF was 0.114, suggesting "moderate" evidence *for* the null hypothesis of no difference. Finally, for the variable representing the difference between ISI = 0.5 s and ISI = 10 s, the mean and 95% highest density interval were greater than zero and the BF was 3.04×10^4, suggesting extreme evidence against the null hypothesis. To summarize, somewhat contrary to the predictions of sudden death, the guessing probability was a nonmonotonic function of ISI duration: the guessing probability appeared to decline when the ISI increased from 0.5 to 2 s but increased with further increases in ISI duration.

Although we did not make *a priori* predictions concerning them, the agnostic model allowed us to examine the differences between guessing preference, a, and bias, b, as a function of ISI duration. There appeared to be a shift toward greater preference for choosing the second tone while guessing at ISI = 2 s and 10 s. We are cautious about placing too much importance on these results because guessing trials reflect the minority of trials for all listeners.[6] There also appeared to be a negative shift in bias (greater bias toward choosing the second tone) at ISI = 5 s and 10 s. This result could reflect the consistent drift of pitch representations over time.

C. Results from the prescriptive models

All three prescriptive models produced good, convergent posterior samples. The BFMI was 0.800–0.855 for all models and chains, all rank-normalized $\hat{R}$ were ≤ 1.001, and all ESSs were ≥ 7727.928.

The LOO scores (Table I) showed that the decay-only model outperformed the death-only model. However, the decay-and-death model outperformed both of them, suggesting that both kinds of forgetting were necessary to best explain the data. The difference in the LOO score between

J. Acoust. Soc. Am. **149** (1), January 2021
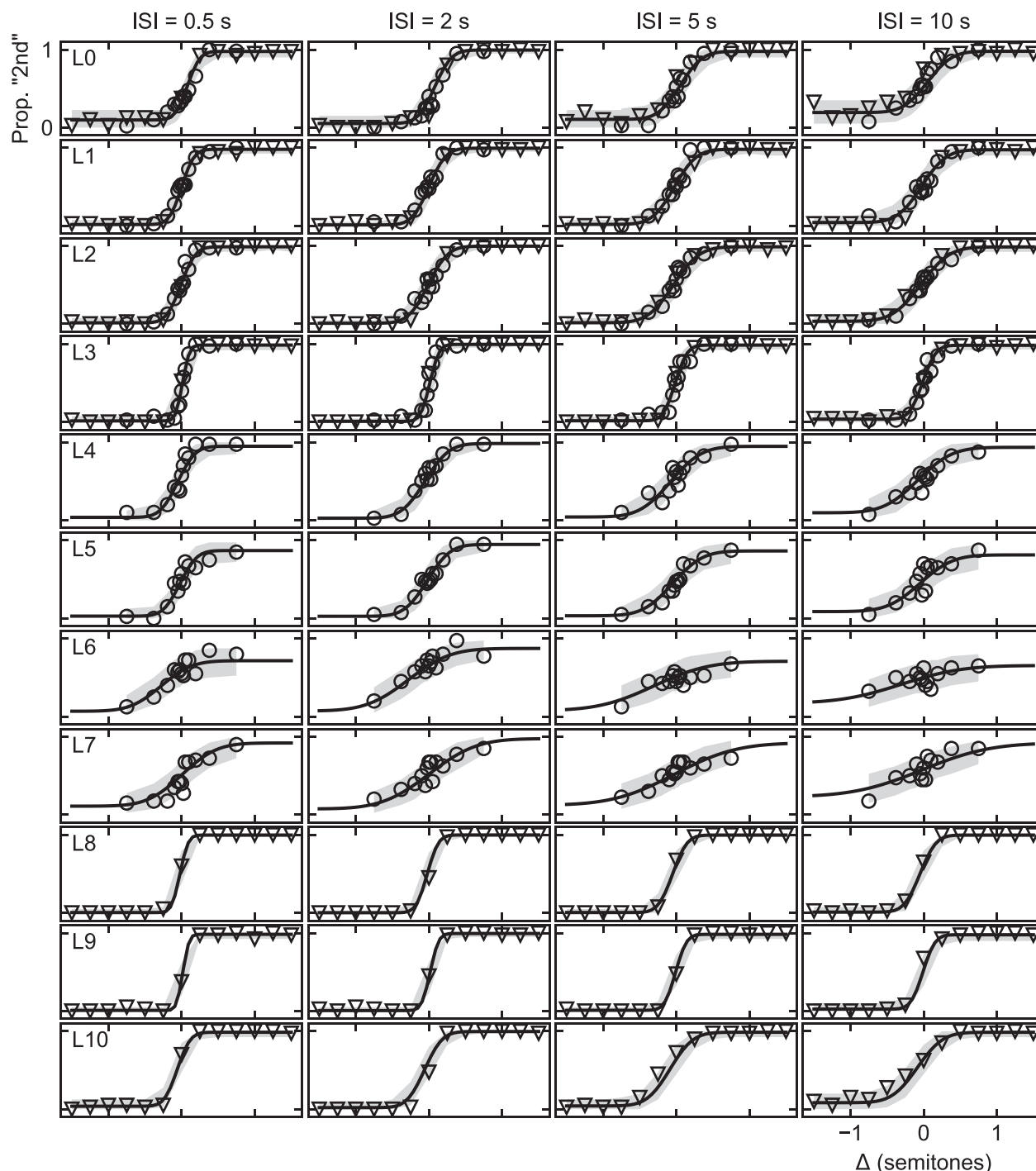
Mathias *et al.* 263

FIG. 2. Symbols represent the proportion of second responses for a given experiment (circles for experiment 1 and triangles for experiment 2), listener (row panels), ISI (column panels), and Δ (40 trials per symbol, on average). Curves are the posterior mean values of p, and shaded regions are the ranges between the 1st and 99th centiles of the posterior predictive distribution under the agnostic model.

the decay-and-death model and the decay-only model was substantial at 2.9 times larger than its standard error.

IV. DISCUSSION

A. Summary of the findings

The present study examined how pitch representations are forgotten from ASTM. One possibility is that they experience gradual decay, becoming weaker over time. Another is that they experience sudden death, disappearing without a trace. Over two experiments, 11 listeners discriminated frequency differences between pairs of pure tones separated by silent ISIs of 0.5, 2, 5, or 10 s. We analyzed the data in two ways. First, we used a model that estimated the standard deviation of internal noise and guessing probabilities as a function of ISI duration. Under this model, internal
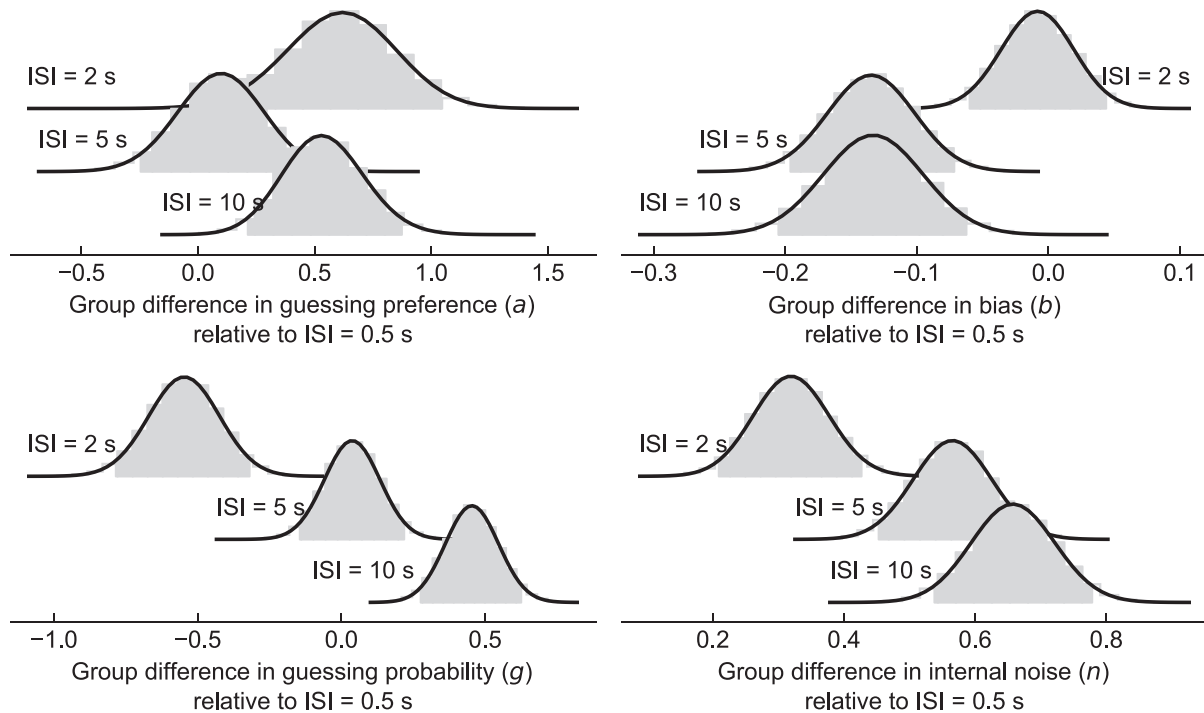
FIG. 3. Marginal posterior distributions of the variables from the agnostic model, which represent the group-level differences between the ISI conditions. Although the units are arbitrary, a value of zero represents no difference between conditions, a negative value represents a reduction from ISI = 0.5 s to the longer ISI duration, and a positive value represents an increase from ISI = 0.5 s to the longer ISI duration. The shaded regions are 95% of the highest density intervals.

noise appeared to increase with increasing ISI duration, across all ISI durations, consistent with gradual decay. Guessing probabilities decreased with increasing ISI duration when ISI <2 s and increased with increasing ISI duration for longer ISIs. These results suggest that sudden death also occurred, at least for longer ISIs. In the second round of analysis, we compared models that assumed specific gradual decay and/or sudden death functions. While the decay-only model outperformed the death-only model, the decay-and-death model outperformed both of them. Considering all results together, we conclude that the data provide evidence that both gradual decay and sudden death occurred. However, we also note that the evidence for sudden death was somewhat weaker than that for gradual decay, and sudden death may be pronounced only at relatively long ISI durations.

B. Implications for ASTM

Many previous studies have found that the ability to discriminate differences between sequential sounds declines

with increasing ISI duration, even in the absence of interfering sounds, as long as the ISI is longer than a few hundred ms (e.g., Berliner and Durlach, 1973; Clément et al., 1999; Harris, 1952; Kinchla and Smyzer, 1967; Mathias et al., 2020; Wickelgren, 1969). Most of these studies assumed that worsening performance was caused by gradual decay. However, the present study suggests that performance in these studies may have been influenced by sudden death as well as gradual decay at long ISIs, which may have implications for their conclusions. For example, Clément et al. (1999) measured the listeners' difference limens for frequency and intensity when the ISI was 0.5 s, and then measured the sensitivity (d'; Green and Swets, 1988) for the frequency and intensity differences of these magnitudes when the ISIs were in the range 0.5–10 s. The authors found that d' decreased more rapidly as a function of the increasing ISI for the intensity than for the frequency differences, suggesting that loudness representations decayed more rapidly than the pitch representations. Alternatively, the results by Clément et al. (1999) could be explained by loudness representations being more susceptible to sudden death than are pitch representations, although it remains to be seen whether discrimination of other sound features besides pitch are affected by sudden death. Given that sudden death was difficult to observe when ISI <5 s, we speculate that it plays a relatively small role in most studies of frequency discrimination (besides those cited above) because it is rare for such studies to employ very long ISIs. Furthermore, outside of the laboratory, it is extremely rare to encounter such long intervals of silence; therefore, it seems likely that the role of

TABLE I. Results of the LOO model comparison. Higher information criterion (IC) indicates a better model. N_p is the effective number of parameters, se( ) is the standard error of the estimate, and Δ is the difference from the best model.

Model	IC	N_p	ΔIC	se(IC)	se(ΔIC)
Decay and death	−1553.261	71.070		36.016	
Decay only	−1586.578	63.581	33.317	38.294	12.914
Death only	−1625.497	78.757	72.236	38.611	14.009

sudden death in real-world frequency discrimination is also small.

Previous studies have proposed mathematical descriptions of the gradual-decay function. For example, Kinchla and Smyzer (1967) proposed that gradual decay follows Wiener diffusion, which implies that the variance of internal noise increases linearly with increasing ISI duration. Other researchers have proposed alternative functions (for a discussion, see Laming and Scheiwiller, 1985). The present findings suggest that the goal of selecting the correct gradual-decay function is even more difficult than previously thought because the possibility of sudden death further increases the complexity of the problem. To our knowledge, no previous study has proposed any kind of function for sudden death. In our prescriptive models, we assumed a simple function in which the probability of sudden death per second was constant over the ISI. We chose this function mostly for convenience so that the death-only model was roughly equal in terms of overall complexity to the decay-only model. However, the observation that the guessing probability was not monotonically related to ISI duration may call this assumption into question. Future experimentation could attempt to adjudicate between more complex gradual-decay and sudden-death functions simultaneously.

C. Limitations and open questions

An important limitation of the present work is that it can be difficult to accurately and simultaneously measure the slope and asymptotes of a psychometric function because they exert somewhat similar influences on the middle of the function. If the range of stimulus values is too narrow, such that not enough of the extremes of the psychometric function are captured, there is a risk that the parameters controlling the slope and asymptotes may "trade-off." We recognized this issue during data collection for experiment 1, which is why we performed experiment 2 with a wider range of frequency differences. However, as pointed out by Prins (2012), trade-offs and misestimation may occur even with very wide stimulus ranges. This was the motivation behind fitting versions of the agnostic model where some variables were fixed across different ISI durations (see Sec. II C 4). Overall, these models did not produce results that were markedly different from those reported in this paper. For example, in the model in which the guessing probability was fixed across ISIs, one might have expected, due to trade-off, that the relationship between internal noise and ISI duration would be nonmonotonic, resembling the relationship between the guessing probability and ISI duration in other models; instead, this relationship was monotonic and almost identical to the relationship in the main agnostic model.

It is unclear why the guessing probability decreased between trials with 0.5-s ISIs and trials with 2-s ISIs. It seems likely that some other phenomenon besides sudden death influenced the listeners' guessing behavior, which was strongest at relatively short ISI durations. One possibility is that this phenomenon was related to trials being blocked by the ISI condition. It could be that the listeners' strategies were in some way modified by the ISI being predictable. This is probably not a simple practice or fatigue effect because ISI durations were varied randomly across blocks of trials. More experimentation is needed to discover why this occurred. For example, a future study could have listeners perform the experiment with ISIs randomly varied across trials within the same block.

In all of our models, we assumed that for a given listener, the standard deviation of sensory noise was always the same. This assumption meant that under the agnostic model, the standard deviation of ensemble internal noise [n in Eq. (1)] was fixed per listener and ISI, and under the prescriptive models, the standard deviation of sensory noise [s in Eq. (2)] was fixed per listener. However, at least three factors may have caused trial-by-trial variation in n or s. First, the listeners' difference limens for frequency are not perfectly constant in semitones across the roving range we employed (for a review and meta-analysis, see Micheyl et al., 2012). Second, there may have been a resolution-edge effect in which internal noise is lower when tone frequencies are close to the limits of the roving range (e.g., Berliner et al., 1977). Third, there may have been sequential interference in which the second tone from one trial interferes with the representation of the first tone on the next trial (e.g., Cowan et al., 1997). We believe that, overall, trial-by-trial variability in n or s was quite small and roughly constant across the different ISI durations, meaning that it was unlikely to have made a considerable impact on our results and conclusions.

Future work could examine the role of trial-by-trial variability in roved frequency discrimination, although doing this within the current analytic framework may prove to be rather challenging. One approach is to relax the assumption that sensory noise is normally distributed on the semitone scale. A sensible alternative assumption is that it follows a student t distribution, which is equivalent to assuming that it is Gaussian, but from one trial to the next, its variance follows the inverse-gamma probability distribution. Unfortunately, this model involves convolving and marginalizing over highly complex nonstandard probability distributions. An alternative approach is to explicitly model the sources of trial-by-trial variability in internal noise. For example, the standard deviation of sensory noise on a given trial could be made to be a function of the frequency of the first tone (e.g., Micheyl et al., 2012) or of frequencies visited on previous trials in the experiment (e.g., Arzounian et al., 2017). We attempted to implement these kinds of models here, but they did not yield acceptable diagnostic metrics. We speculate that many more trials per listener may be necessary to detect such effects.

D. Comparison to VSTM

Our results concerning ASTM contrast with those of previous studies on VSTM, most notably with the study by

Zhang and Luck (2009). In the first of two experiments, each trial presented three colored squares simultaneously for 100 ms, followed by an empty retention interval of 1, 4, or 10 s, followed by a probe to indicate the target square and a color wheel. Observers estimated the color of the target square using the color wheel. In the second experiment, the colored squares were replaced by "s"-like objects whose shapes were determined by a function that took as inputs values from the circle, and the color wheel was replaced by 180 possible shapes arranged in a circle. In both experiments, the authors fitted a mixture of two circular probability distributions to observers' errors (true color or shape in degrees minus reported color or shape in degrees). The mixture comprised a circular uniform distribution and a von Mises distribution (an approximation to the wrapped normal distribution). From this model, the authors estimated the spread of the von Mises distribution (roughly analogous to n in our models) and the mixture weighting (roughly analogous to g) for each listener and retention interval. The authors concluded that in both experiments, the spread of the von Mises distribution did not differ across the different retention intervals. By contrast, the weighting of the circular uniform distribution increased significantly as a function of the increasing retention interval. These results were interpreted as suggesting that representations of visual color and shape experience only sudden death and no gradual decay from VSTM.[7]

What could account for this apparent discrepancy between findings concerning ASTM and VSTM? One obvious difference is that the experiments of Zhang and Luck (2009) involved remembering three visual objects during the retention interval, whereas in our experiments, listeners needed to remember just one tone's pitch during the ISI. We are aware of three other studies in which observers performed delayed-estimation tasks with one visual object per trial and a variable retention interval, but none of them considered the possibility of sudden death (Nilsson and Nelson, 1981; Pertzov et al., 2017; Shin et al., 2017).

Another possibility is that delayed-estimation tasks and 2I-2AFC tasks produce different estimates of internal noise and guessing probabilities, leading to different results. One way to resolve this potential issue would be to use both kinds of tasks in the same sensory domain and compare the results. We are not aware of any vision studies that employed 2I-2AFC tasks to investigate gradual decay and sudden death. Two previous studies have employed delayed-estimation tasks in audition. In the first (Kumar et al., 2013), listeners heard sequences of one, two, or four tones. At the end of each sequence, one of the tones was probed via a number on the screen. A random-frequency probe tone was then played, whose pitch listeners adjusted using a dial to match that of the remembered target. On trials in which listeners heard one tone, the average standard deviation of the listeners' errors (equivalent to n in our models) was ~0.9 semitone. This value is much larger than the n estimates obtained at 0.5-s ISI in our experiments, suggesting that the task by Kumar et al. (2013) did not estimate the

same kind of internal noise as our 2I-2AFC discrimination tasks did. Before conducting the present experiments, we informally tried a similar delayed-estimation procedure but quickly abandoned it. Anecdotally, this task did not feel like a faithful analog of VSTM delayed estimation because the tone to be adjusted tended to strongly interfere with the memory of the target tone (cf. Deutsch, 2013) in a way that may not happen in VSTM delayed estimation. In another study (Teki and Griffiths, 2014), listeners estimated the durations of intervals formed between clicks. However, one could argue that because the intervals were actually silent and merely demarcated by clicks, ASTM may not have been used to remember the interval durations.

A final possibility is that analogous stores within auditory and visual memory operate on different time scales. Previous vision studies have often failed to find effects of ISI duration during 2I-2AFC discrimination tasks of various visual features when ISIs range from 1 to 60 s (e.g., Blake et al., 1997; Greenlee et al., 1995; Greenlee et al., 1993; Magnussen and Greenlee, 1992; Magnussen et al., 1990, 1991; Magnussen et al., 1996; Magnussen et al., 1998; Magnussen et al., 1985). These results contrast sharply with auditory studies in which effects of ISI duration are typically robust. In vision, there is strong evidence for a distinction between VSTM and *iconic memory* (Sperling, 1960). It could be that sudden death is more frequent in VSTM than in iconic memory or VSTM experiences less gradual decay than does iconic memory. In one study, Demany et al. (2008) suggested that there may be an auditory equivalent of iconic memory, a high-capacity auditory memory storing raw spectral information. However, its duration is at least 2 s, much longer than the duration of iconic memory (cf. Phillips, 1974). It is possible that what we call ASTM here is actually the analog of iconic memory and not VSTM.

ACKNOWLEDGMENTS

APPENDIX A: SPECIFICATION OF THE AGNOSTIC MODEL

The vector $\boldsymbol{w} = [w_0, w_1, \ldots w_{779}]^T$ denoted the counts of trials on which the second tone was chosen as binned by listener, ISI, and Δ. Elements of $\boldsymbol{w}$ followed the probability distribution

$$\boldsymbol{w} \sim \text{Binomial}(\boldsymbol{r}, \boldsymbol{p}), \tag{A1}$$

where $\boldsymbol{r}$ is the vector containing the numbers of trials per bin (40 in most cases) and $\boldsymbol{p}$ is the vector of probabilities [see Eq. (1)]. Vectors of parameters from the psychometric function were transformations of vectors of latent variables as follows:

$$\boldsymbol{a} = \text{logistic}(\sqrt{2}\boldsymbol{\alpha}), \tag{A2}$$

$$\boldsymbol{b} = \boldsymbol{\beta}/5, \tag{A3}$$

$$g = \text{logistic}\left(\sqrt{2}(\gamma - 1)\right), \tag{A4}$$

$$n = \exp(v - 2). \tag{A5}$$

These transformations enabled the sampling of stochastic random variables to be performed on the real number line while ensuring informative priors on psychometric-function parameters. Vectors of pretransformed variables were given by

$$[\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, v] = \boldsymbol{X\Theta}, \tag{A6}$$

where $\boldsymbol{X}$ was a 14×780 design matrix [created using the R/patsy formula "$0 + C(\text{listener}) + C(\text{isi})$"], and $\boldsymbol{\Theta}$ was a 4×14 matrix of independent stochastic random variables with univariate standard normal priors,

$$\boldsymbol{\Theta} \sim \text{Normal}(0, 1). \tag{A7}$$

APPENDIX B: SPECIFICATION OF THE PRESCRIPTIVE MODELS

Equations (A1)–(A3) also applied to the prescriptive models. Additional psychometric-function parameters were given by

$$l = \text{logistic}\left(\sqrt{2}(\lambda - 1)\right), \tag{B1}$$

$$s = \exp(\varsigma - 2). \tag{B2}$$

Under the decay-and-death and decay-only models,

$$d = \exp(\boldsymbol{\delta} - 5), \tag{B3}$$

whereas under the death-only model, $d = 0$. Under the decay-and-death and death-only models,

$$u = \text{logistic}\left(\sqrt{2}(v - 2)\right), \tag{B4}$$

whereas under the decay-only model, $u = 0$. Vectors of the pretransformed variables were given by

$$[\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \lambda, v] = \boldsymbol{Z\Omega}, \tag{B5}$$

where $\boldsymbol{Z}$ was an 11×780 design matrix [created using the R/patsy formula "$0 + C(\text{listener})$"] and

$$\boldsymbol{\Omega} = \boldsymbol{\mu} + (\boldsymbol{LA})^T, \tag{B6}$$

where $\boldsymbol{\mu}$ was a five-item vector of stochastic random variables, representing the hyperprior means,

$$\boldsymbol{\mu} \sim \text{Normal}(0, 1). \tag{B7}$$

$\boldsymbol{A}$ was an 11×5 matrix of independent stochastic random variables, representing the offsets of listener-level parameters,

$$\boldsymbol{A} \sim \text{Normal}(0, 1), \tag{B8}$$

and $\boldsymbol{L}$ was the Cholesky decomposition of the variance-covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{LL}^T$, and

$$\boldsymbol{\Sigma} = \boldsymbol{RDR}, \tag{B9}$$

where $\boldsymbol{R}$ was a stochastic correlation matrix assigned the prior distribution proposed by Lewandowski *et al.* (2009) with flat priors on all nondiagonal entries,

$$\boldsymbol{R} \sim \text{LKJ}(1), \tag{B10}$$

and $\boldsymbol{D}$ was a diagonal matrix of stochastic standard deviations,

$$\boldsymbol{D} = \text{diag}(\boldsymbol{\sigma}), \tag{B11}$$

$$\boldsymbol{\sigma} \sim \text{Exponential}(1). \tag{B12}$$

[1]This is not always the case. When the stimuli are very brief and ISIs are shorter than a few hundred ms, longer ISIs actually improve discrimination performance (e.g., Carbotte, 1973; Demany *et al.*, 2005; Demany and Semal, 2005; Massaro and Idson, 1977; Small and Campbell, 1962; Taylor and Smith, 1975).

[2]Kinchla and Smyzer (1967) actually considered gradual decay to be a Gaussian random walk with arbitrarily small steps. However, it is equivalent and more rigorous mathematically (because it bypasses the need to define a step size) to take the random process to the limit and consider gradual decay to be the Wiener process, which is a Gaussian random walk with infinitely small steps (Donsker, 1951).

[3]This feature was inherited from our previous study and had no discernible bearing on its results; see footnote 1 of Mathias *et al.*, 2020, for a discussion.

[4]For listeners who participated in both experiments, all of their data were used to estimate a single set of listener-specific variables (i.e., they were not treated as separate listeners between experiments).

[5]See https://github.com/sammosummo/ForgettingPitchPublic (Last viewed Dec. 4, 2020).

[6]Moreover, these group-level differences were no longer present when L6 and L7 were removed.

[7]These findings were disputed in a conference abstract by Fougnie *et al.* (2013); however, to the best of our knowledge, a full article describing this work was never published.

Arzounian, D., de Kerangal, M., and de Cheveigné, A. (**2017**). "Sequential dependencies in pitch judgments," J. Acoust. Soc. Am. **142**(5), 3047–3057.

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

Berliner, J. E., and Durlach, N. I. (**1973**). "Intensity perception. IV. Resolution in roving-level discrimination," J. Acoust. Soc. Am. **53**(5), 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**(6), 1577–1585.

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

Blake, R., Cepeda, N. J., and Hiris, E. (**1997**). "Memory for visual motion," J. Exp. Psychol. Hum. Percept. Perform. **23**(2), 353–369.

Carbotte, R. M. (**1973**). "Retention of time information in forced-choice duration discrimination," Percept. Psychophys. **14**(3), 440–444.

Clément, S., Demany, L., and Semal, C. (**1999**). "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**(5), 2805–2811.

Cowan, N., Saults, J. S., and Nugent, L. D. (**1997**). "The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons," Psychon. Bull. Rev. **4**(3), 393–397.

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.

Demany, L., Montandon, G., and Semal, C. (**2005**). "Internal noise and memory for pitch," in *Auditory Signal Processing*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 136–144.

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**(8), 1376–1383.

Demany, L., Trost, W., Serman, M., and Semal, C. (**2008**). "Auditory change detection: Simple sounds are not memorized better than complex sounds," Psychol. Sci. **19**(1), 85–91.

Deutsch, D. (**2013**). "The processing of pitch combinations," in *The Psychology of Music* (Elsevier, New York), pp. 249–325.

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

Donsker, M. D. (**1951**). "An invariance principle for certain probability limit theorems," Mem. Am. Math. Soc. **6**, 12.

Fougnie, D., Suchow, J. W., and Alvarez, G. A. (**2013**). "Gradual decay and death by natural causes in visual working memory," J. Vision **13**(9), 19.

Gelman, A. (**2014**). *Bayesian Data Analysis*, 3rd ed., Chapman and Hall/CRC Texts in Statistical Science (CRC Press, Boca Raton, FL).

Gelman, A., Goodrich, B., Gabry, J., and Vehtari, A. (**2019**). "*R*-squared for Bayesian regression models," Am. Stat. **73**(3), 307–309.

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

Greenlee, M., Lang, H., Mergner, T., and Seeger, W. (**1995**). "Visual short-term memory of stimulus velocity in patients with unilateral posterior brain damage," J. Neurosci. **15**(3), 2287–2300.

Greenlee, M., Rischewski, J., Mergner, T., and Seeger, W. (**1993**). "Delayed pattern discrimination in patients with unilateral temporal lobe damage," J. Neurosci. **13**(6), 2565–2574.

Harris, J. D. (**1952**). "The decline of pitch discrimination with time," J. Exp. Psychol. **43**(2), 96–99.

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**, 1593–1623.

Jeffreys, H. (**1998**). "Oxford classic texts in the physical sciences," in *Theory of Probability*, 3rd ed. (Oxford University Press, New York).

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

Kinchla, R. A., and Smyzer, F. (**1967**). "A diffusion model of perceptual memory," Percept. Psychophys. **2**(6), 219–229.

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

Kruschke, J. K. (**2015**). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed. (Academic, Boston, MA).

Kumar, S., Joseph, S., Pearson, B., Teki, S., Fox, Z. V., Griffiths, T. D., and Husain, M. (**2013**). "Resource allocation and prioritization in auditory working memory," Cogn. Neurosci. **4**(1), 12–20.

Laming, D., and Scheiwiller, P. (**1985**). "Retention in perceptual memory: A review of models and data," Percept. Psychophys. **37**(3), 189–197.

Lee, M. D., and Wagenmakers, E.-J. (**2013**). *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge, UK).

Lewandowski, D., Kurowicka, D., and Joe, H. (**2009**). "Generating random correlation matrices based on vines and extended onion method," J. Multivar. Anal. **100**(9), 1989–2001.

Magnussen, S., and Greenlee, M. W. (**1992**). "Retention and disruption of motion information in visual short-term memory," J. Exp. Psychol. Learn. Mem. Cogn. **18**(1), 151–156.

Magnussen, S., Greenlee, M. W., Asplund, R., and Dyrnes, S. (**1990**). "Perfect visual short-term memory for periodic patterns," Eur. J. Cogn. Psychol. **2**(4), 345–362.

Magnussen, S., Greenlee, M. W., Asplund, R., and Dyrnes, S. (**1991**). "Stimulus-specific mechanisms of visual short-term memory," Vis. Res. **31**(7-8), 1213–1219.

Magnussen, S., Greenlee, M. W., and Thomas, J. P. (**1996**). "Parallel processing in visual short-term memory," J. Exp. Psychol. Hum. Percept. Perform. **22**(1), 202–212.

Magnussen, S., Idås, E., and Myhre, S. H. (**1998**). "Representation of orientation and spatial frequency in perception and memory: A choice reaction-time analysis," J. Exp. Psychol. Hum. Percept. Perform. **24**(3), 707–718.

Magnussen, S., Landrø, N. I., and Johnsen, T. (**1985**). "Visual half-field symmetry in orientation perception," Percept **14**(3), 265–273.

Massaro, D. W., and Idson, W. L. (**1977**). "Backward recognition masking in relative pitch judgments," Percept. Mot. Skills **45**(1), 87–97.

Mathias, S. R., Varghese, L., Micheyl, C., and Shinn-Cunningham, B. G. (**2020**). "On the utility of perceptual anchors during pure-tone frequency discrimination," J. Acoust. Soc. Am. **147**(1), 371–380.

Mathias, S. R., and von Kriegstein, K. (**2014**). "Percepts, not acoustic properties, are the units of auditory short-term memory," J. Exp. Psychol. Hum. Percept. Perform. **40**(2), 445–450.

Mercer, T., and McKeown, D. (**2010a**). "Interference in short-term auditory memory," Q. J. Exp. Psychol. **63**(7), 1256–1265.

Mercer, T., and McKeown, D. (**2010b**). "Updating and feature overwriting in short-term memory for timbre," Atten. Percept. Psychophys. **72**(8), 2289–2303.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res. **219**(1-2), 36–47.

Micheyl, C., Xiao, L., and Oxenham, A. J. (**2012**). "Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level," Hear. Res. **292**(1-2), 1–13.

Nilsson, T. H., and Nelson, T. M. (**1981**). "Delayed monochromatic hue matches indicate characteristics of visual memory," J. Exp. Psychol. Hum. Percept. Perform. **7**(1), 141–150.

Pertzov, Y., Manohar, S., and Husain, M. (**2017**). "Rapid forgetting results from competition over time between items in visual working memory," J. Exp. Psychol. Hum. Percept. Perform. **43**(4), 528–536.

Phillips, W. A. (**1974**). "On the distinction between sensory storage and short-term visual memory," Percept. Psychophys. **16**(2), 283–290.

Prins, N. (**2012**). "The psychometric function: The lapse rate revisited," J. Vis. **12**(6), 25.

Prinzmetal, W., Amiri, H., Allen, K., and Edwards, T. (**1998**). "Phenomenology of attention: I. Color, location, orientation, and spatial frequency," J. Exp. Psychol. Hum. Percept. Perform. **24**(1), 261–282.

Ries, D. T., Hamilton, T. R., and Grossmann, A. J. (**2010**). "The effects of intervening interference on working memory for sound location as a function of inter-comparison interval," Hear. Res. **268**(1-2), 227–233.

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

Semal, C., and Demany, L. (**1991**). "Dissociation of pitch from timbre in auditory short-term memory," J. Acoust. Soc. Am. **89**(5), 2404–2410.

Semal, C., and Demany, L. (**1993**). "Further evidence for an autonomous processing of pitch in auditory short-term memory," J. Acoust. Soc. Am. **94**(3), 1315–1322.

Shin, H., Zou, Q., and Ma, W. J. (**2017**). "The effects of delay duration on visual working memory for orientation," J. Vis. **17**(14), 1–24.

Small, A. M., and Campbell, R. A. (**1962**). "Temporal differential sensitivity for auditory stimuli," Am. J. Psychol. **75**(3), 401–410.

Sperling, G. (**1960**). "The information available in brief visual presentations," Psychol. Mon. Gen. Appl. **74**(11), 1–29.

Starr, G. E., and Pitt, M. A. (**1997**). "Interference effects in short-term memory for timbre," J. Acoust. Soc. Am. **102**(1), 486–494.

Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1959**). "Multiple observations of signals in noise," J. Acoust. Soc. Am. **31**(4), 514–521.

Taylor, M. M., and Smith, S. M. (**1975**). "Monaural detection with contralateral cue. V. Interstimulus interval in MDCC and amplitude discrimination," J. Acoust. Soc. Am. **57**(6), 1500–1511.

Teki, S., and Griffiths, T. D. (**2014**). "Working memory for time intervals in auditory rhythmic sequences," Front. Psychol. **5**, 1329.

Vehtari, A., Gelman, A., and Gabry, J. (**2017**). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," Stat. Comput. **27**(5), 1413–1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (**2020**). "Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC," Bayesian Anal. (published online).

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (**2010**). "Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method," Cognit. Psychol. **60**(3), 158–189.

Wichmann, F. A., and Hill, N. J. (**2001**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**(8), 1293–1313.

Wickelgren, W. A. (**1969**). "Associative strength theory of recognition memory for pitch," J. Math. Psychol. **6**(1), 13–61.

Wiener, N. (**1923**). "Differential-space," J. Math. Phys. **2**(1-4), 131–174.

Wilken, P., and Ma, W. J. (**2004**). "A detection theory account of change detection," J. Vis. **4**(12), 11.

Zhang, W., and Luck, S. J. (**2009**). "Sudden death and gradual decay in visual working memory," Psychol. Sci. **20**(4), 423–428.