Title

How rests and cyclic sequences influence performance in tone-scramble tasks

Journal

The Journal of the Acoustical Society of America, 147(6)

ISSN

0001-4966

Authors

Ho, Joselyn
Chubb, Charles

Publication Date

2020-06-01

DOI

10.1121/10.0001398

Peer reviewed

JASA ARTICLE

How rests and cyclic sequences influence performance in tone-scramble tasks

Joselyn Ho[a)] and Charles Chubb
Department of Cognitive Sciences, University of California at Irvine, Irvine, California 92697-5100, USA

ABSTRACT:
When classifying major versus minor tone-scrambles (random sequences of pure tones), most listeners (70%) perform at chance while the remaining listeners perform nearly perfectly. The current study investigated whether inserting rests and cyclic sequences into the stimuli could heighten sensitivity in such tasks. In separate blocks, listeners classified tone-scramble variants as major versus minor ("3" task) or fourth versus tritone ("4" task). In three "Fast" variants, tones were played at 65 ms/tone as a continuous, random stream ("FR"), or with a rest after every fourth tone ("FRwR"), or as a repeating sequence of four tones with a rest after every fourth tone ("FCwR"). In the "Slow" variant, tones were played at 325 ms/tone in random order. In both the 3 and 4 tasks, performance was ordered from best to worst as follows: FRwR > FR > FCwR > Slow. *Post hoc* analysis revealed that performance was suppressed in the Slow and FCwR task-variants due to a powerful bias inclining listeners to respond "major" or "fourth" ("minor" or "tritone") if the 4-note sequence defining the stimulus ended on a high (low) note. Overall, the results indicate that inserting regular rests into random tone sequences heightens sensitivity to musical mode.
 https://doi.org/10.1121/10.0001398

(Received 31 October 2019; revised 30 April 2020; accepted 26 May 2020; published online 12 June 2020)

[Editor: Tamara Smyth] Pages: 3859–3870

I. INTRODUCTION

The psychological reality of music's emotional expressiveness is well-documented. In western culture, music in the major diatonic scale tends to be associated with happy emotions, while music in the minor scale is associated with sad emotions (Crowder, 1984, 1985a,b; Gagnon and Peretz, 2003; Gerardi and Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner and Crowder, 1990; Temperley and Tan, 2013). Although this difference in perceived emotional quality might imply that judging the mode (major vs minor) of a melody comes naturally, several studies suggest that many listeners, including musicians, actually struggle with such tasks (Halpern, 1984; Halpern *et al.*, 1998; Leaver and Halpern, 2004).

Substantial evidence now shows that sensitivity to major vs minor musical modes is bimodally distributed across listeners. The first indication of this effect can be seen in the results of Crowder (1985b) who tested listeners in a task [replicating Blechner (1977)] in which they strove to classify triadic chords as major vs minor. Each stimulus was a 300-ms triad from the equal-tempered scale in either root position (note-order from low to high: tonic, third, fifth) or first inversion (note-order from low to high: third, fifth, tonic). Across trials, the tonic varied randomly (between 6 notes). The fifth of the triad was 7 semitones above the tonic, and the third of the triad was 1 to 9 logarithmic steps between the minor and major third, relative to the tonic. The task was to classify the triad according to whether the third was closer to the minor vs the major third. Although Crowder (1985b) only had 19 subjects, he observed that the psychometric functions (that related the level of the third in the triad to the probability that the subject responded "major") fell into two distinct groups. Either the psychometric function was very steep, suggesting that the listener was very sensitive to the major–minor difference, or utterly flat, suggesting that the listener had little or no sensitivity to the major–minor difference. Only three listeners fell in the middle between these two extremes.

A bimodal distribution in performance has also been observed in a task requiring listeners to classify major vs minor "tone-scrambles" (Chubb *et al.*, 2013; Dean and Chubb, 2017; Mednicoff *et al.*, 2018). In the basic major–-minor task, each tone-scramble contains thirty-two, 65-ms tones including 8 copies each of the notes G_5, D_6, G_6 (to establish G as the tonic of each stimulus), and a target note. The target note in major tone-scrambles is B_5 (the third degree of the G major scale), and the target note in minor tone-scrambles is $B\flat_5$ (the third degree of the G minor scale). On each trial, the listener hears a single tone-scramble and attempts, with feedback, to classify it as major or minor. Chubb *et al.* (2013) and subsequent studies (Dean and Chubb, 2017; Mednicoff *et al.*, 2018) have found that 70% of listeners perform near chance in this task while 30% perform near perfectly.

Dean and Chubb (2017) tested listeners in a range of tasks akin to the major–minor tone-scramble task but using different pairs of target notes. For example, in the "4" task, the target notes that differentiated the two stimulus types were C_6 (fourth scale degree of both the G major and G minor scales) and $D\flat_6$ (a tritone above G_5, included in

a)Electronic mail: joselyh@uci.edu

neither the G major nor G minor scales). The results were well-described by a bilinear model which proposes that sensitivity of listener k to the difference between the two types of tone-scramble stimuli used in task t ($d'_{k,t}$) is determined by the listener's amount of scale-sensitivity (R_k) and the strength with which scale-sensitivity facilitates performance in task t (F_t). Specifically,

$$d'_{k,t} = R_k F_t. \tag{1}$$

Dean and Chubb (2017) concluded that performance in all of the tasks, t, used in their study was determined predominantly by a single processing resource, R. As in Chubb *et al.* (2013), listeners' performance took the form of a bimodal distribution: 70% percent of listeners possessed levels of R near zero which yielded a near-chance performance in all tasks, while 30% of listeners possessed levels of R that yielded a higher performance. R also facilitated different tasks with different strengths. Since the target notes used in most of the tasks were unrelated to the difference between the major vs minor scales, Dean and Chubb (2017) concluded that R confers general sensitivity to variations in scale with a fixed tonic. They therefore called R "scale-sensitivity," proposing that listeners possess different levels of this resource which determines their ability to discriminate tone-scramble types.

It has also been shown that performance in the basic major–minor tone-scramble task does not improve if stimuli are presented more slowly (Mednicoff *et al.*, 2018). Listeners in this study were tested in major-vs-minor tone-scramble classification tasks in which the stimuli were played at different rates. In the slowest condition (which we refer to below as the Slow-3 task), each tone-scramble contained four, 520-ms tones: G_5, D_6, G_6, and a target note ($B\flat_5$ or B_5). Listeners who performed poorly in any condition performed poorly across all conditions. Surprisingly, performance was the worst in the slowest condition, and listeners' responses were strongly affected by the order of the four tones in each tone-scramble, regardless of whether the stimulus was major or minor. Specifically, even though note-order is irrelevant to the task, listeners were biased to respond major if a tone-scramble ended on a high note. In the slowest condition, the responses of more than half of all listeners were influenced more strongly by this shared bias than they were by whether the target note was $B\flat_5$ vs B_5.

Finally, musical training does not directly predict scale-sensitivity. While trained musicians tend to have a higher scale-sensitivity than non-musicians, Chubb *et al.* (2013), Dean and Chubb (2017), and Mednicoff *et al.* (2018) also observed many listeners with over 5 years of musical training with low scale-sensitivity, as well as many listeners with little or no musical training with high scale-sensitivity. The positive correlation between musicianship and scale-sensitivity may result from a self-selection bias, such that listeners with high scale-sensitivity are more likely to pursue musical training vs listeners with low scale-sensitivity. It should be noted, however, that only listeners with around 5 or more years of musical training reach the highest levels of scale-sensitivity. This suggests that although musical training does not suffice for high scale-sensitivity, training may be necessary to achieve the highest levels.

A. The current study

Does scale-sensitivity depend on temporal structure? If so, then perhaps scale-defined properties (e.g., majorness vs minorness) can be made more legible by introducing rhythmic and/or sequential structure into tone-scramble stimuli. It has long been recognized that chunking isolated pieces of information together can improve processing (Miller, 1956). Music typically comprises phrases, or subunits of a longer melody, that can be defined through temporal structure. For example, sequences of tones that occur within a rhythmic pattern are better remembered than sequences that span rhythmic patterns (Dowling, 1973). Further, sequences with regularly-occurring rests are better remembered than sequences that occur as a continuous stream, and the tones between each rest tend to be recalled or forgotten as a unit (Deutsch, 1982).

Therefore, introducing temporal structure into tone-scramble stimuli may heighten the sensitivity of listeners to the differences between the stimulus types. In the current study, we focused on the effects of rests and cyclic sequences (i.e., repeating sets of 4 tones).

We also sought to more deeply explore the relationship between scale-sensitivity and the note-order-specific response biases that tend to subvert performance in the Slow-3 task of Mednicoff *et al.* (2018). There was a strong, shared tendency to classify the Slow-3 stimuli as major if they ended on a high note (especially the high tonic). We speculated that this effect was provoked by the suggestion (made in the response prompt presented visually after each trial) to classify stimuli as major if they sounded "happy" and minor if they sounded "sad." The prevalence of the ending-on-a-high-note bias (across all listeners other than those with very high scale-sensitivity) suggests that, perhaps for some reason rooted in language-processing, the stimuli in the Slow-3 task of Mednicoff *et al.* (2018) naturally sound happier if they end on a high note vs a low note.

To test this possibility, we included four task conditions that might provoke sequence-specific biases. In two of these tasks, the stimuli differ in majorness vs minorness [as in the study of Mednicoff *et al.* (2018)]. In the other two tasks, the stimuli differ in a quality that might be described as harmoniousness vs dissonance. The target notes in the major–minor tasks are the third scale degrees of the major and minor scales. The target notes in the other task are the fourth scale degree (which is in both the major and minor scales) and the tritone (which is in neither).

II. METHODS

All methods were approved by the UCI Institutional Review Board.

https://doi.org/10.1121/10.0001398

A. Participants

Ninety-eight listeners participated in this study and were all undergraduate students at the University of California, Irvine, with self-reported normal hearing. Sixty-nine listeners reported having at least 1 year of formal musical training. The mean number of years of musical training across all 98 listeners was 4.5 (standard deviation: 4.9). All listeners received course credit for participating in the study.

B. Stimuli

The experiment used eight stimulus variants with two types each (determined by which target note it contained), for a total of 16 stimulus types (Table I). Stimuli were tone-scrambles, which are sequences of pure tones comprising equal numbers of a target note **T** plus three other notes from the standard equal-tempered chromatic scale: G_5 (783.99 Hz), D_6 (1174.66 Hz), and G_6 (1567.98 Hz). In the "3-task" variants, the target note **T** was $B\flat_5$ (932.33 Hz) for "low-target" (Minor) stimuli or B_5 (987.77 Hz) for "high-target" (Major) stimuli. In the "4-task" variants, **T** was C_6 (1046.50 Hz) for low-target (Fourth) or $D\flat_6$ (1108.73 Hz) for high-target (Tritone) stimuli. Thus, in all tasks, the high target note was a semitone higher in pitch than the low target note.

Stimuli in the six "Fast" task variants contained twenty, 65-ms tones. Tones in the FR-3 and FR-4 ("Fast Random") tasks were presented in a continuous stream. Tones in the FRwR-3 and FRwR-4 ("Fast Random with Rests") tasks were presented as five bursts of four tones. Each burst contained a random sequence of the notes G_5, D_6, G_6, and **T**, and bursts were separated by 130-ms rests. Tones in the FCwR-3 and FCwR-4 ("Fast Cyclic with Rests") tasks were presented in five repeating bursts of the same sequence of four tones (one each of G_5, D_6, G_6, and **T**), and bursts were separated by 130-ms rests.

The stimuli in the Slow-3 and Slow-4 tasks comprised one each of the notes G_5, D_6, G_6, and **T**, played in random order at 325 ms per tone.

In all tasks, each individual tone per stimulus was windowed by a raised cosine function with a 22.5-ms rise time.

C. Procedure

At the start of the experiment, listeners completed a brief survey to report (among other information) their number of years of musical training.

TABLE I. The temporal and sequential properties of the stimuli used in the 8 tasks. The number in the task name (i.e., 3 or 4) indicates that task's set of target notes ($B\flat_5/B_5$ and $C_6/D\flat_6$, respectively).

Tasks	Number of tones	Tone duration	Rests	Order
FR-3, FR-4	20	65 ms	no	random
FRwR-3, FRwR-4	20	65 ms	yes	random
FCwR-3, FCwR-4	20	65 ms	yes	cyclic
Slow-3, Slow-4	4	325 ms	no	random

Listeners were then tested in each of the FR-3, FRwR-3, FCwR-3, Slow-3, FR-4, FRwR-4, FCwR-4, and Slow-4 tasks. Task order was randomly generated for each listener.

At the start of each task, the listener heard eight example stimuli labeled as either "type 1" or "type 2." In 3-task variants, type 1 corresponded to high-target (major) stimuli; type 2 corresponded to low-target (minor) stimuli. In 4-task variants, type 1 corresponded to low-target (fourth) stimuli; type 2 corresponded to high-target (tritone) stimuli. These distinctions were not explicitly told to the listener. Then, on each trial, the listener heard a single stimulus and strove to judge which type was presented by entering "1" or "2" for their response. Correctness feedback was printed to the screen after each trial, and proportion correct was given at the end of each block.

Each task consisted of two blocks of 48 trials. Stimulus type (high- vs low-target) was determined randomly on each trial. To shorten the experiment duration, the number of trials in the first block of the FR-3, FR-4, FRwR-3, and FRwR-4 tasks was reduced to 24 for the last 70 listeners. For the basic analysis (Sec. III), we computed listeners' d' values from the second block of trials and treated the first block of trials as practice, as has been done in previous tone-scramble studies (Chubb *et al.*, 2013; Dean and Chubb, 2017; Mednicoff *et al.*, 2018). In analyzing the sequence-specific biases that occur in the FCwR-3, FCwR-4, Slow-3, and Slow-4 tasks (Sec. V), we used both blocks of 48 trials to increase statistical power.

The experiment took place in a quiet lab on a Windows Dell computer with a standard Realtek audio/sound card using MATLAB. Stimuli were presented at the rate of 50 000 samples/s, and listeners wore JBL Elite 300 noise-canceling headphones with the volume adjusted to their comfort level.

III. RESULTS

Listeners' d' values, our basic dependent measure, were computed using the last 48 trials of each task. The first block of 24 or 48 trials per task was treated as practice. If a listener was tested on n high-target (low-target) stimuli over the course of the last 48 trials and responded correctly on all of them, then the probability of a correct response was adjusted to $n - 0.5/n$ [as suggested by Macmillan and Kaplan (1985)]. This implies that d' values around 4.1 correspond to near-perfect performance on all 48 trials of a task.

Comparisons of the d' values achieved by listeners in different tasks suggest that tasks differed in difficulty. Table II lists the results of paired samples t-tests of the null hypothesis that the mean value of d' is equal for two tasks. Some of the main trends revealed by this table are: (1) performance in each of the FRwR-3, FR-3, FCwR-3, Slow-3 tasks is significantly better than performance in the corresponding 4-task, and (2) for each of $n = 3$, 4, performance in the Slow-n task is significantly worse than in the FRwR-n, FR-n, and FCwR-n tasks.

Performance showed a significant tendency to improve across tasks. Specifically, we computed the linear trend L_k

TABLE II. Results from paired samples t-tests of d' achieved by all listeners in each pair of tasks. The values shown are t-statistics for 2-tailed tests of the null hypothesis that the mean value of d' is equal for the two tasks. All t-statistics have 97 degrees of freedom. * $p < 0.05/28 = 0.0018$ (Bonferroni correction).

Task	FR-3	FRwR-3	FCwR-3	Slow-3	FR-4	FRwR-4	FCwR-4	Slow-4
FR-3	—	1.14	−0.34	5.52*	4.59*	1.40	4.20*	5.24*
FRwR-3		—	0.74	5.71*	5.76*	2.70	5.53*	6.27*
FCwR-3			—	6.41*	5.76*	2.08	6.04*	6.41*
Slow-3				—	0.23	3.35*	0.15	1.79
FR-4					—	4.56*	0.05	1.72
FRwR-4						—	4.14*	4.47*
FCwR-4							—	1.72
Slow-4								—

in the vector of eight d' values achieved by each listener k across the eight tasks in the order in which the listener was tested. The mean value of the L_k's was 0.18. A 1-tailed t-test of the null hypothesis that the true mean was 0 yielded $t_{97} = 2.30, p = 0.012$.

A. Bilinear model results

Using the bilinear model [Eq. (1)], we estimated F_t and R_k values. Following the analysis procedure of Mednicoff *et al.* (2018), we set the constraint that

$$\sum_{\text{tasks } t} F_t = 8 \quad \text{(where 8 is the number of tasks)}. \tag{2}$$

This constraint has convenient properties. First, if all tasks are equally facilitated by R, then F_t will be 1 for all tasks. Second, Eq. (2) makes R_k the average value of d' achieved by listener k across all 8 task-variants.

The estimated values of F_t for all tasks t are displayed in Fig. 1. As suggested by the d' results, and consistent with the results of Dean and Chubb (2017), F_t is higher for each

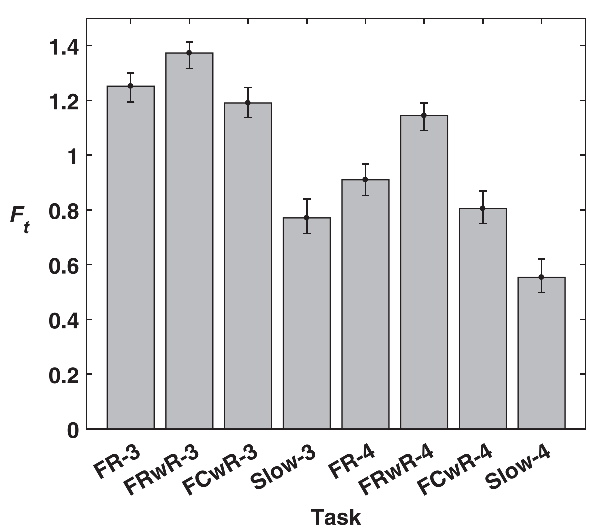

FIG. 1. Estimated values of F_t for the eight tasks. Error bars are 95% Bayesian credible intervals.

3-task variant t (t = FRwR-3, FR-3, FCwR-3, and Slow-3) than it is for the corresponding 4-task variant.

For $n = 3$, 4, $F_{FRwR-n} > F_{FR-n} \approx F_{FCwR-n} > F_{\text{Slow}-n}$. In particular, $F_{\text{Slow}-n}$ is much lower than F_t for each of the Fast conditions (t = FRwR-n, FR-n, FCwR-n). This is consistent with the finding of Mednicoff *et al.* (2018) that listeners perform worse when tone-scrambles are played more slowly.

The left panel of Fig. 2 displays the histogram of R_k estimated for the 98 listeners k. Similar to the histogram of R values observed by both Dean and Chubb (2017) and Mednicoff *et al.* (2018), this histogram is positively-skewed with most listeners possessing R values near 0. This histogram does not appear bimodal. However, as seen in the right panel of Fig. 2, the histogram of proportion correct that these listeners would be predicted to achieve in the FR-3 task (assuming they used optimal criteria) yields the bimodal distribution of performance that is typically observed for this task-variant (Chubb *et al.*, 2013; Dean and Chubb, 2017; Mednicoff *et al.*, 2018).

The results are well-described by the bilinear model. Figure 3 plots the estimates of $d'_{k,t}$ for each listener k in each task t against the values predicted by the bilinear model, and a strong relationship is observed. The bilinear model accounts for 73.8% of the variance in the values of $d'_{k,t}$ for the 98 listeners across the eight tasks.

B. Relationship with music training

Figure 4 plots each listener's R against his-or-her self-reported years of musical training, showing a significant correlation of 0.364 ($p < 0.01$). In the group of 37 listeners with at least 5 years of musical training, 21 listeners had R values below 1. Three of the listeners in this group of 21 had at least 15 years of musical training.

The highest R value attained by the 35 listeners with fewer than 2 years of musical training was 2.2. Among the six listeners who attained R values above 3, four listeners had at least 5 years of musical training. Therefore, listeners with high values of R tend to have more years of musical training, which follows the pattern observed by Dean and Chubb (2017) and Mednicoff *et al.* (2018).

IV. DISCUSSION

The current study explored the degree to which scale-sensitivity (Dean and Chubb, 2017) is modulated by basic variations in temporal structure, which were implemented through periodic rests and cyclic note sequences.

Similar to the results of Dean and Chubb (2017) and Mednicoff *et al.* (2018), performance was well-described by the bilinear model [Eq. (1)] across all listeners k in all tasks t, implying that performance on the tasks in this study is primarily determined by a single processing resource. The current study is linked to Mednicoff *et al.* (2018) and Dean and Chubb (2017) in the shared use of the FR-3 task. This commonality suggests that a single processing resource [called

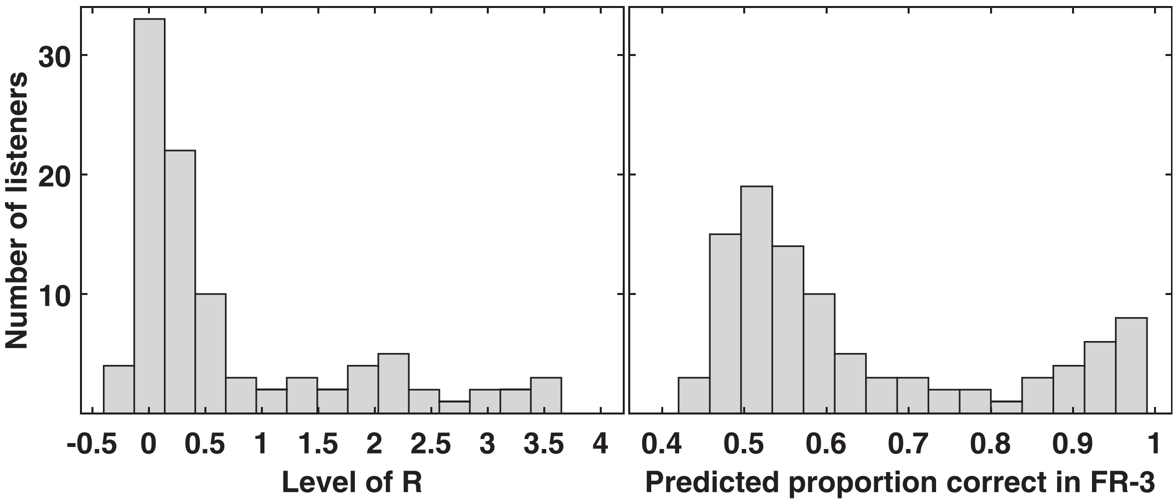

FIG. 2. Left panel: Histogram of estimated R levels for the 98 listeners. This histogram is positively-skewed with most listeners possessing R values near 0. Right panel: Histogram of predicted proportions correct in the FR-3 task corresponding to the R levels in the left panel. This histogram appears bimodal although the histogram of R levels does not.

scale-sensitivity by Dean and Chubb (2017)] underlies performance in all three studies.

We also note that for each of $n = 3$, 4, $F_{FRwR-n} > F_{FR-n}$. The current study does not clearly determine the basis of this effect because the stimuli in the FRwR-n task differ from those in the FR-n task in two ways. First, the FRwR-n task stimuli contained a rest between each burst of four notes. Second, each burst contained one each of the notes G_5, D_6, G_6, and **T**. Either or both of these features may have contributed to the heightened performance in the FRwR-n task, compared to in the FR-n task.

For each of $n = 3$, 4, $F_{FRwR-n} > F_{FCwR-n}$. In the FCwR-n task, the five bursts in each stimulus repeat the same randomly-ordered sequence of the four notes G_5, D_6, G_6, and **T**. By contrast, in the FRwR-n task, each of the five bursts in each stimulus contains a random sequence of the four notes. In Sec. V, we present evidence suggesting that the heightened difficulty of the FCwR-n task vs the FRwR-n task stems from systematic response biases associated with individual sequences of the four notes G_5, D_6, G_6, and **T**. These biases operate more strongly to subvert performance in the FCwR-n task than they do in the FRwR-n task.

V. NOTE-ORDER EFFECTS

In this section, we focus exclusively on the FCwR-3, FCwR-4, Slow-3, and Slow-4 tasks because a given stimulus in any of these tasks is completely determined by a single sequence of four notes. (This is not true in any of the other tasks.) In this section, we investigate whether specific permutations of these notes influence listeners' responses. Following Mednicoff *et al.* (2018), we first use a "Descriptive" model to capture the detailed structure in the data. We then show that

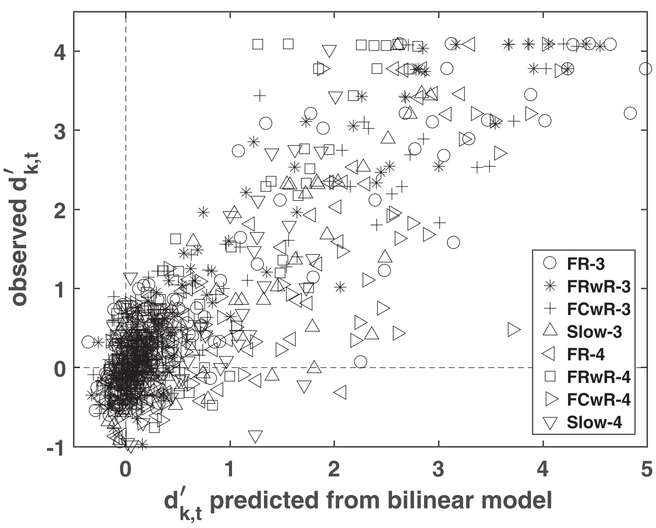

FIG. 3. Scatterplot of observed d' values against estimated d' values from the bilinear model. Each point corresponds to a given listener k in a given task t. Tasks are indicated by the symbols displayed in the legend.

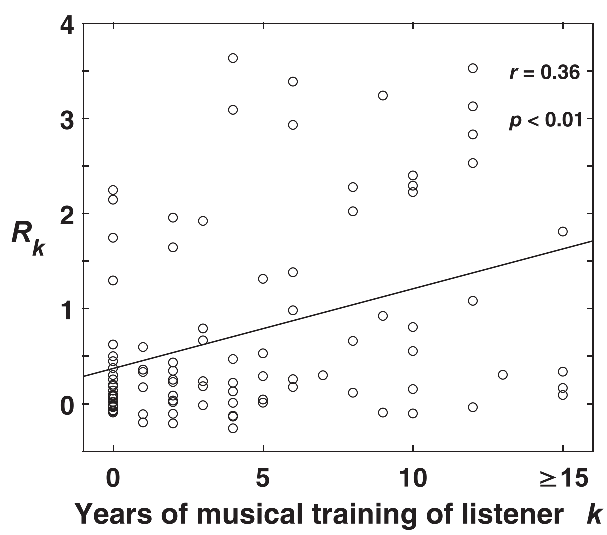

FIG. 4. The relationship between musical training and R_k.

the results from the Descriptive model can be captured by a much simpler "Note-function-biased" model.

A. Notation

We refer to the notes G_5, $B\flat_5$, B_5, C_6, $D\flat_6$, D_6, and G_6 by their respective pitch height values (1, 4, 5, 6, 7, 8, and 13), which represent the notes' locations in the chromatic scale starting at G_5.

In each of the FCwR-3, Slow-3, FCwR-4, and Slow-4 tasks, a given stimulus corresponds to a particular 4-note sequence. We refer to individual notes as "pips;" the symbol S refers to the four-pip sequence that determines a stimulus in a task. For $t = 1, 2, 3, 4$, $S(t)$ is the note assigned to pip t. Also in each task, all stimuli are constructed from a set $\text{Notes} = \{1, T^-, T^+, 8, 13\}$, where T^- denotes the lower of the two target notes and T^+ denotes the other (a semitone higher in pitch). In the FCwR-3 and Slow-3 Tasks, $T^- = 4$ and $T^+ = 5$, and in the FCwR-4 and Slow-4 Tasks, $T^- = 6$ and $T^+ = 7$. The symbol $\mathbf{T}$ refers to the target note in a stimulus. We call a stimulus with target note T^- (T^+) a low-target (high-target) stimulus.

A permutation of the four symbols "1," "8," "13," and "$\mathbf{T}$" produces a "note-order." Substituting "T^-" for $\mathbf{T}$ in a given note order Q yields a symbol string corresponding to a stimulus S_Q^-. Substituting T^+ for $\mathbf{T}$ yields a symbol string corresponding to a stimulus S_Q^+.

B. Modeling framework

In the general modeling framework, a listener k computes an internal statistic for each stimulus S and compares it to a criterion η_k which is fixed across all trials in a given task. Let $M_{k,S}$ be the expectation of this internal statistic. Then both of the models we consider assume

$$\text{Response of listener } k \text{ to stimulus } S = \begin{cases} \text{“high-target”} & \text{if } M_{k,S} + X > \eta_k \\ \text{“low-target”} & \text{if } M_{k,S} + X < \eta_k\,, \end{cases} \tag{3}$$

where X is a standard normal random variable.

C. Fitting procedures

In Secs. V D–V G, we fit the Descriptive and Note-function-biased models to the data in the FCwR-3, Slow-3, FCwR-4, and Slow-4 tasks. We derive maximum likelihood estimates of all parameters. In addition, we use a Bayesian fitting procedure to derive credible intervals around these estimates. Specifically, we assume a jointly uniform prior distribution with wide bounds on all model parameters. Then, using Markov chain Monte Carlo simulation, we extract a sample of vectors from the posterior joint density characterizing the parameters. In the figures in this section, line markers show maximum likelihood estimates of parameters, and error bars give the 0.025 and 0.975 quantiles of posterior, marginal parameter densities.

D. The descriptive model

For listeners k with high values of R_k, we expect $M_{k,s}$ to depend strongly on the value of

$$\tau_S = \begin{cases} 1 & \text{if } S \text{ is a high-target stimulus,} \\ -1 & \text{if } S \text{ is a low-target stimulus.} \end{cases} \tag{4}$$

As Mednicoff *et al.* (2018) discovered in the Slow-3 task, many listeners also exhibit shared, systematic, S-dependent response biases. In the Descriptive model, these biases are captured by free parameters β_S corresponding to all 48 possible stimuli S (24 note-orders $\times$ 2 target notes). The Note-function-biased model (described below) uses a simplified rule to predict the β_S values.

Both the Descriptive and Note-function-biased models assume that

$$M_{k,S} = f_\tau(R_k)\tau_S + f_\beta(R_k)\beta_S, \tag{5}$$

where the function $f_\tau(R)$ reflects the strength with which τ_S influences the response of a listener with scale-sensitivity R, and the function $f_\beta(R)$ reflects the strength with which β_S influences the response of a listener with scale-sensitivity R.

In order to uniquely specify the descriptive model, we impose several constraints on the parameters; these are described in the Appendix.

E. The note-function-biased model

The Note-function-biased model describes a simple theory of how the β_S's are computed. For a task with low and high target notes T^- and T^+, let $\text{Notes} = \{1, T^-, T^+, 8, 13\}$. Under the Note-function-biased model, there exist functions $f_{\text{note}} : \text{Notes} \rightarrow \mathbb{R}$ and $f_{\text{pip}} : \{1, 2, 3, 4\} \rightarrow \mathbb{R}$ such that

$$\beta_S = \sum_{t=1}^{4} f_{\text{note}}(S(t)) f_{\text{pip}}(t), \tag{6}$$

where $S(t)$ is the note occurring at pip t in the sequence defining S.

The "Pitch-height-biased" model used by Mednicoff *et al.* (2018) is a special case of the Note-function-biased model in which

$$f_{\text{note}} = f_{PH}(n) = n - M_{\text{Notes}} \quad \text{for all } n \in \text{Notes}, \tag{7}$$

where M_{Notes} is the mean of the notes $n \in \text{Notes}$.

In order to uniquely specify the Note-function-biased model, we impose several constraints on the parameters; these are described in the Appendix.

F. Modeling results

Instead of considering directly the parameters β_S for all sequences S, it is useful to focus instead on the equivalent, alternative parameters

$$\mu_Q = \frac{\beta_{S_Q^+} + \beta_{S_Q^-}}{2} \quad \text{and} \quad \delta_Q = \frac{\beta_{S_Q^+} - \beta_{S_Q^-}}{2}, \tag{8}$$

for all note-orders Q. For a given note-order Q, μ_Q reflects the bias injected by the note-order Q regardless of target note, and δ_Q reflects the difference in influence exerted by T^+ vs T^- in the context of Q. (Note that $\beta_{S_Q^+} = \mu_Q + \delta_Q$, and $\beta_{S_Q^-} = \mu_Q - \delta_Q$.)

Across the four tasks, many of the μ_Q values estimated from the descriptive model (plotted as black circles in Fig. 5) deviate significantly from 0, confirming that note-order exerts a strong influence on stimulus-specific biases regardless of the target note. By contrast, very few of the δ_Q values estimated from the descriptive model (black circles in Fig. 6) deviate significantly from 0 suggesting that note order does not strongly influence the relative influence exerted by T^+ vs T^-.

In Figs. 5 and 6, the gray triangles plot the results from the Note-function-biased model. As described in the Appendix, f_{note} uses only 3 degrees of freedom and f_{pip} uses only 2 as a result of the model constraints. Thus the Note-biased-function model uses only five degrees of freedom to account for all of the 24 μ_Q's and 24 δ_Q's. As reflected by the descriptive model fit, it captures the overall structure in the data remarkably well. In particular, for each of the four tasks, a likelihood ratio test (Hoel *et al.*, 1971; Wilks, 1944) fails to reject the nested Note-function-biased model in favor of the fuller Descriptive model. Under the null hypothesis, the test statistic X is chi-square distributed with 41 degrees of freedom. For the FCwR-3 task, $X = 44.7$, $p = 0.32$; for the Slow-3 task, $X = 26.7$, $p = 0.96$; for the FCwR-4 task, $X = 47.4$, $p = 0.23$; and for the Slow-4 task, $X = 54.8$, $p = 0.07$.

Figure 7 plots the functions f_τ (black) and f_β (gray). As expected, f_τ increases with R_k. For all four tasks, the Descriptive model estimates of f_β are, on average, significantly above 0 and roughly equal for listeners with scale-sensitivity levels in sextiles 4, 5, and 6. Thus, the biases reflected by the black circles in Fig. 5 operate with roughly equal strength across all listeners with scale-sensitivity greater than ≈ 0.3.

The relative influence of the biases β_S on the judgments of our listeners varies strongly across the four tasks. Consider a listener k with R_k in the sixth sextile. In the FCwR-3 task, $f_\tau(R_k)$ is around 6 times greater than $f_\beta(R_k)$; this implies that the identity of the target note exerts roughly 6 times more influence on the response of this listener than do the sequence-specific biases. At the other extreme, however, in the Slow-4 task, $f_\tau(R_k)$ is only around twice as great as $f_\beta(R_k)$. In the Slow-3 task, $f_\tau(R_k)$ is around 4 times greater than $f_\beta(R_k)$, and in the FCwR-4 task, $f_\tau(R_k)$ is around 3.5 times greater than $f_\beta(R_k)$.

The left panel of Fig. 8 plots the temporal weighting function f_{pip} for all four tasks. As described in the Appendix, f_{pip} is constrained to sum to 0 and to have $f_{\text{pip}}(4) > 0$; thus, all four functions rise up similarly.

The differences between the 3- and 4-task variants are concentrated in the note weights function f_{note} (right panel

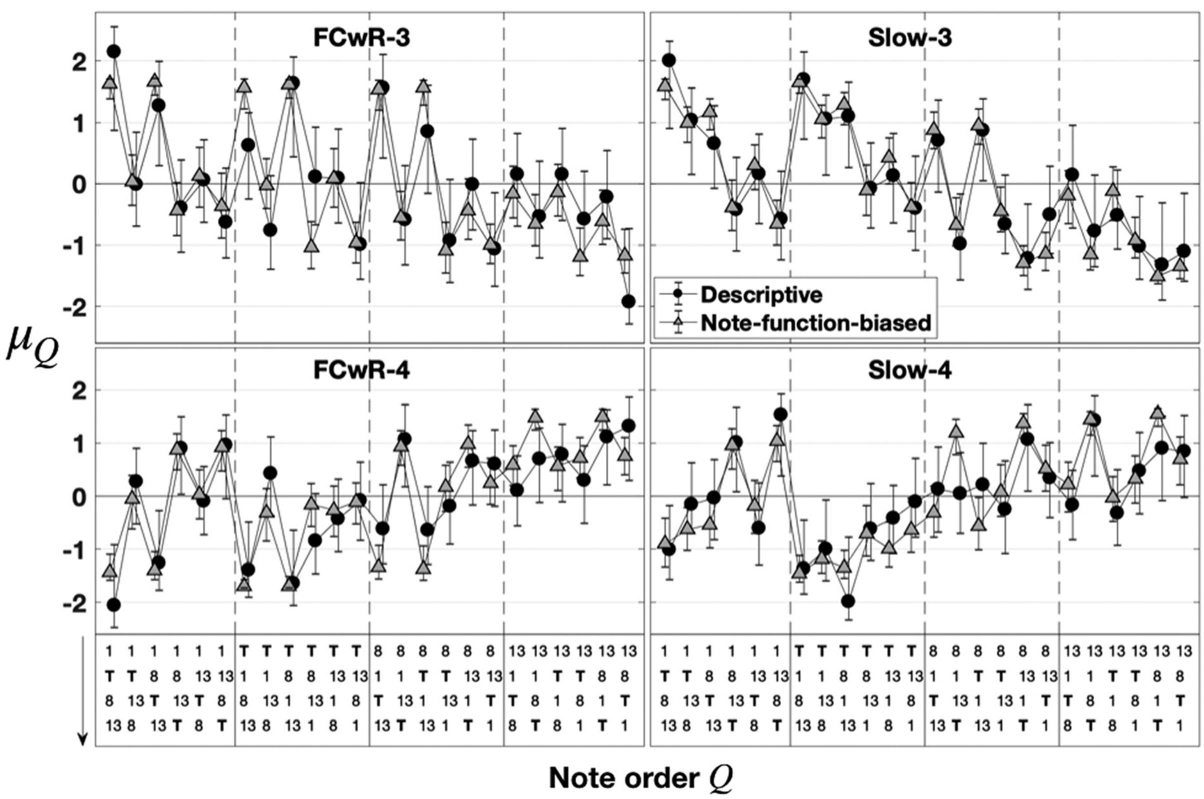

FIG. 5. The note-order-specific biases μ_Q for the 24 note-orders Q in each of the four tasks. The note-order Q of a given stimulus S is represented along the horizontal axis, running downward. Values estimated from the Descriptive model (Note-function-biased model) are plotted in black circles (gray triangles). Markers show maximum likelihood estimates; error bars show 95% Bayesian credible intervals.

https://doi.org/10.1121/10.0001398

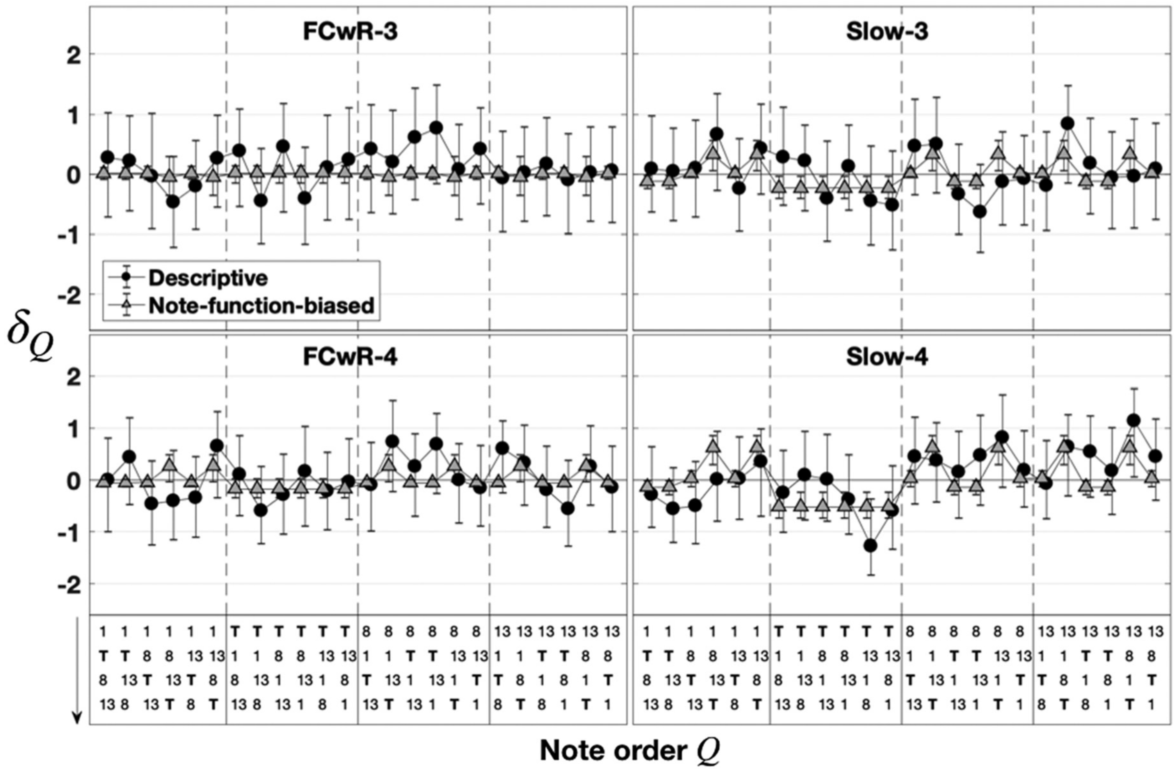

FIG. 6. The note-order-specific differences in influence exerted by T^+ vs T^- for the 24 note-orders Q in each of the four tasks. The note-order Q of a given stimulus S is represented along the horizontal axis, running downward. Values estimated from the Descriptive model (Note-function-biased model) are plotted in black circles (gray triangles). Markers show maximum likelihood estimates; error bars show 95% Bayesian credible intervals.

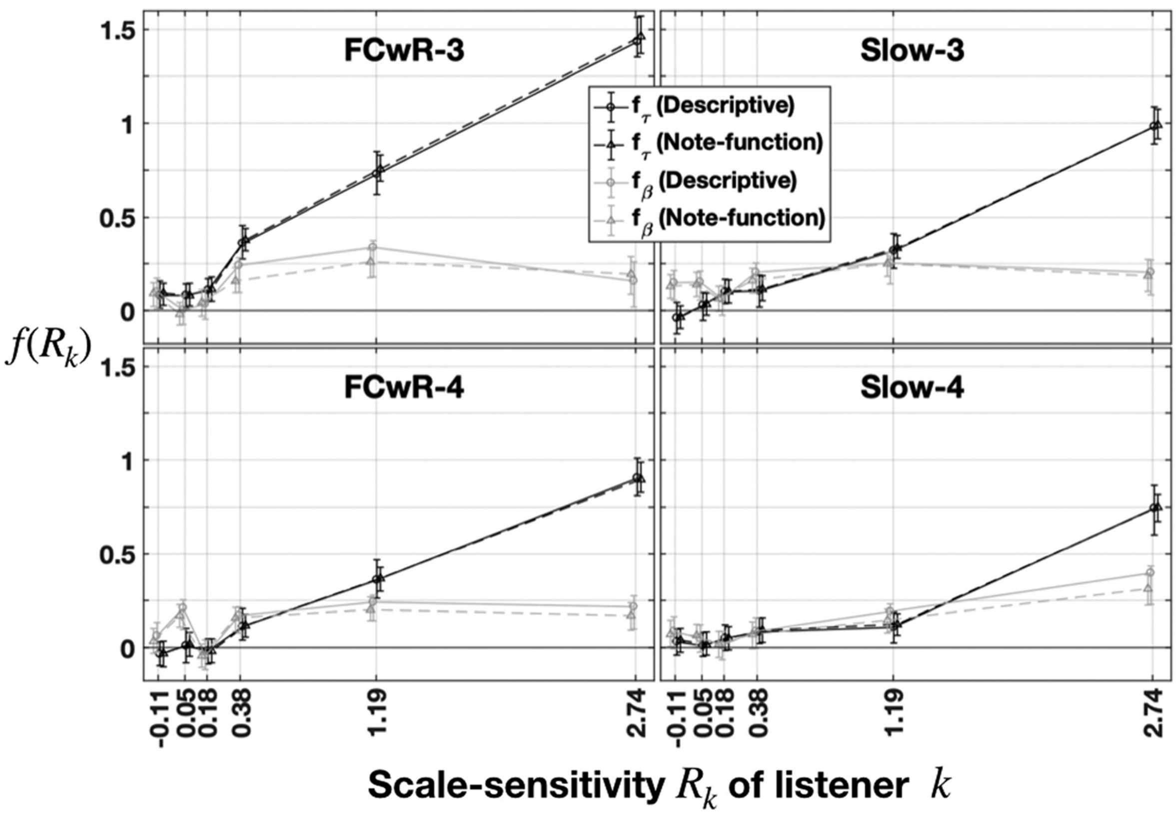

FIG. 7. The functions f_τ and f_β in the four tasks. The values of R_k along the horizontal axis are the mean values of the six sextiles of R_k observed across the 98 listeners. f_τ is plotted in black; f_β is plotted in gray. Solid (dashed) lines show the fits from the Descriptive (Note-function-biased) model. Error bars are 95% Bayesian credible intervals.

https://doi.org/10.1121/10.0001398

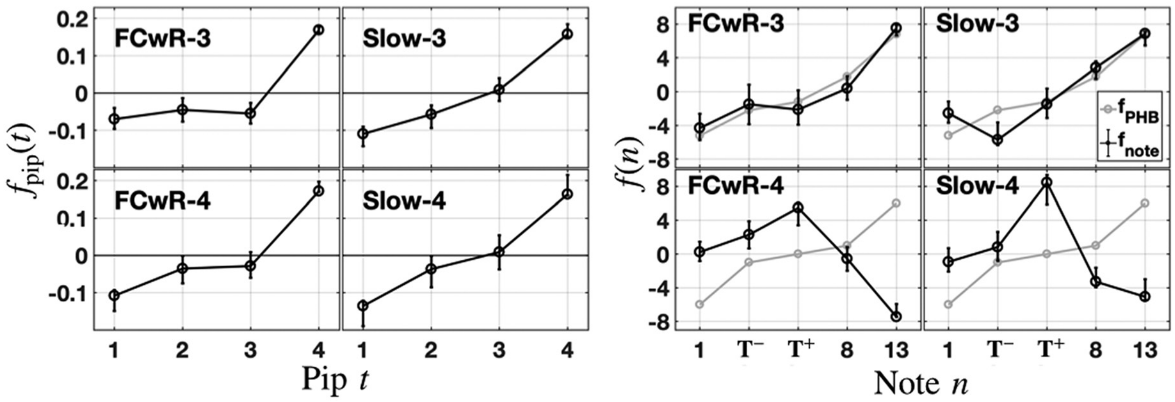

FIG. 8. The functions $f_{\rm pip}$ (left panel) and $f_{\rm note}$ (right panel) for the four tasks. The gray lines in the right panel show the form of $f_{\rm note}$ predicted under the Pitch-height-biased model (Mednicoff *et al.*, 2018). Markers show maximum likelihood estimates; error bars are 95% Bayesian credible intervals.

of Fig. 8). In each of the FCwR-3 and Slow-3 tasks, $f_{\rm note}$ is similar to f_{PH} [Eq. (7)]. This is not true for the FCwR-4 and Slow-4 tasks: $f_{\rm note}$ reaches its maximum at T^+ and descends to its minimum at note 13. It should be noted that $f_{\rm note}(T^-) \approx f_{\rm note}(T^+)$ in the FCwR-3 task. In the other three tasks, however, $f_{\rm note}(T^-) < f_{\rm note}(T^+)$.

G. Discussion of note-order effects

The note-order effects first observed by Mednicoff *et al.* (2018) were unanticipated and mysterious. The judgment required in the Slow-3 task depends only on which of the two target notes (pitch height values 4 or 5) occurs in the stimulus; the order of the notes is irrelevant. Nonetheless, the listeners' judgments were strongly influenced by shared biases that depend on the note-order of the stimulus. Mednicoff *et al.* (2018) accounted for their results in terms of the Pitch-height-biased model, which proposed that listeners' responses to a given stimulus S are influenced by a bias β_S according to Eq. (6), with $f_{\rm note}$ equal to the function f_{PH} [Eq. (7)] and $f_{\rm pip}$ similar in form to the functions plotted in Fig. 8.

The current experiment sought to broaden our understanding of these biases by probing two questions:

(1) *How do the biases depend on the target notes used in a given task?*
(2) *How do the biases depend on the temporal structure of the stimuli of a given task?*

1. Sequence-specific biases are influenced more strongly by the target notes of a stimulus than by temporal structure

The stimuli in the FCwR-task variants differ strongly in temporal structure from the stimuli in the Slow-task variants. Individual tones last 5 times as long in the Slow-task variants than they do in the FCwR-task variants. In addition, the 4-note sequence that determines a stimulus in the Slow-task variants occurs only once, but is repeated 5 times in the FCwR-task variants.

Nonetheless, as revealed by the Note-function-biased model fits, the overall pattern of biases in the FCwR-3 (FCwR-4) task is similar to that in the Slow-3 (Slow-4) task. The left plot of Fig. 8 demonstrates that the last pip seems to play a more important role in the FCwR-task variants than in the Slow-task variants. For both FCwR-3 and FCwR-4 tasks, $f_{\rm pip}$ is fairly flat across pips 1, 2, and 3, and then jumps up abruptly on pip 4. By contrast, $f_{\rm pip}$ rises more gradually for the Slow-3 and Slow-4 tasks.

Differences between the note-functions of the 3-task variants and the note-functions of the 4-task variants are strikingly clear in the right panel of Fig. 8. The 3-task note-functions assign maximally positive values to note 13 (the high tonic) whereas the 4-task note-functions assign maximally negative values. In conjunction with the fact that the functions $f_{\rm pip}$ are maximally positive at pip 4, the note-functions for both 3-tasks imply that ending on note 13 biases listeners to respond high-target (major). In contrast, in both 4-tasks, ending on note 13 biases listeners to respond low-target ("fourth").

In itself, this observation might be taken to suggest that the 4-task note-functions are negatives of the corresponding 3-task note-functions. This would imply that the effects that operate in the 3-tasks to bias listeners to respond major also operate in the 4-tasks to bias listeners to respond fourth. However, this does not appear to be true. Negating the note-functions for the FCwR-4 and Slow-4 tasks in the right panel of Fig. 8 does not convert them into the note functions of their 3-task counterparts. Furthermore, 4-task note-functions reach their maximal values at T^+ whereas both 3-task note-functions assign T^+ a value near 0. These findings imply that features of the stimuli other than the difference in pitch between the target notes are critical in determining the pattern of the sequence-specific biases in the task.

The difference in pitch between the target notes T^- and T^+ is the same in the FCwR-3, Slow-3, FCwR-4, and Slow-4 tasks; however, the sequence-specific biases are dramatically different between the two 3-task variants versus the two 4-task variants. This implies that it is not the difference in pitch between target notes that determines the biases. Plausibly, the pattern of sequence-specific biases in a given task is determined by the intervals formed between the two target notes and the context-defining notes 1, 8, and 13. Several features of the note-functions provide clues to the nature of this effect.

First, ending on note 13 (the high tonic) exerts a powerful influence on the biases in both 3- and 4-task variants; by contrast, ending on 1 (low tonic) exerts much less influence in the 3-task variants and little or no influence in the 4-task variants. Music theory suggests that the high and low tonic should play similar roles in controlling the scale-defined qualities of a tone sequence. The different roles of the high and low tonic in influencing the sequence-specific biases in both the 3- and 4-task variants thus suggest that the source of these biases may lie outside the scope of standard music theory. We discuss some possibilities in Sec. VI.

2. Sequence-specific biases may not be intrinsic to the system that is recruited to classify tone-scrambles

If sequence-specific biases are intrinsic to tone-scramble classification, then we might expect these biases to operate with increasing strength in listeners with higher levels of scale-sensitivity. However, in each of the four tasks, the $f_\beta(R_k)$ are flat across listeners k with R_k above the median. In addition, in the Slow-3 and FCwR-4 tasks, $f_\beta(R_k)$ also appears to be greater than 0 for some listeners with R_k near or below 0.

VI. GENERAL DISCUSSION

Previous research (e.g., Chubb *et al.* (2013)) has shown that ≈70% of listeners perceive little or no difference between major vs minor tone-scrambles. Moreover, a single processing resource predominates in controlling performance in a range of tone-scramble tasks that use target notes unrelated to the difference between the major vs minor scales (Dean and Chubb, 2017). This suggests that the resource recruited in these tasks confers general sensitivity to the qualities that music can achieve by establishing a tonic and selecting a scale, i.e., a distribution of intervals relative to the tonic used in the music. This led Dean and Chubb (2017) to call this resource scale-sensitivity. Plausibly, the sensitivity of a listener to scale variations in actual music is also controlled (at least in part) by his-or-her level of scale-sensitivity.

The current study shows that the temporal structure of a task's stimuli exerts substantial influence on the ease with which listeners can extract scale-defined qualities from proto-musical stimuli. For example, for $n = 3$, 4, scale-sensitivity facilitates performance with roughly twice the strength in the FRwR-n task vs the Slow-n task.

Across the four stimulus temporal-structures tested (the FR-, FRwR-, FCwR-, and Slow-task variants) the facilitation strengths vary intuitively. For $n = 3$, 4, scale-sensitivity was found to facilitate performance in the FRwR-n task most strongly, less strongly and approximately equally in the FR-n and FCwR-n tasks, and most weakly in the Slow-n task.

Stimuli in the FRwR-n task include five bursts of four notes, with each burst containing a randomly-ordered sequence of the low tonic, target note, dominant, and high tonic (notes 1, **T**, 8, and 13, respectively). Stimuli in the FR-n task also contain five sets of these notes; however, they are presented in random order as a single, unbroken stream. Thus, the stimuli in the FRwR-n task work in two ways to structure the note sequence to enhance performance: (1) they break up the stream into separate bursts, and (2) they homogenize the stream by forcing each burst to contain one each of the four notes defining the stimulus. Either or both of these features may underlie the difference between $F_{\mathrm{FRwR}-n}$ vs $F_{\mathrm{FR}-n}$ evident in Fig. 1.

Stimuli in the FCwR-n task have the same temporal structure as those in the FRwR-n task; however, each of the five bursts contains the same sequence of four notes (one each of notes 1, **T**, 8, and 13). As shown in Sec. V, performance in the FCwR-n is undermined by sequence-specific biases. Plausibly, these biases tend to cancel out in the FRwR-n stimuli to yield the difference between $F_{\mathrm{FRwR}-n}$ vs $F_{\mathrm{FCwR}-n}$ evident in Fig. 1.

The difference between $F_{\mathrm{FCwR}-n}$ vs $F_{\mathrm{Slow}-n}$ is more interesting. Stimuli in the Slow-n task contain the same information as those in the FCwR-n task: in each case, the stimulus is defined by a 4-note sequence. Moreover, the total duration of the stimuli from both tasks is roughly equal. Nonetheless, $F_{\mathrm{FCwR}-n}$ is substantially greater than $F_{\mathrm{Slow}-n}$. This shows clearly that speeding up and repeating a sequence can increase the legibility of its scale-defined qualities. Increasing the frequency of occurrence of tones in a musical sequence is known to establish a stronger perception of tonal hierarchy (Knopoff and Hutchinson, 1983; Krumhansl, 1990; Krumhansl and Kessler, 1982; Youngblood, 1958; Rosenthal and Hannon, 2016). However, increasing tone duration should enhance the perception of tonal hierarchy by an even greater magnitude (Lantz and Cuddy, 1998; Smith and Schmuckler, 2004). For example, Lantz and Cuddy (1998) found that when the total duration of tone sequences are held constant, the sequences that contain fewer tones of longer duration correspond to higher ratings of tonal stability. Our results and those of Mednicoff *et al.* (2018) regarding duration are at odds with these findings. The results of the current study may be explained by the repetition in the stimuli. Playing the same sequence of tones several times in a loop enhances the musicality of a tone sequence (Margulis and Simchy-Gross, 2016). Thus, perhaps the scale-defined qualities in our stimuli became clearer as a result of the increased musicality and the reorganization of the tones into a regularly-occurring rhythmic structure.

We might expect that using repetition in the Slow conditions (i.e., 5 repetitions of the same 4 tones) would improve performance in the Slow task; however, we are skeptical of this prediction. Based on anecdotal observations of our tasks, we predict that combining the slow speed of the tones with an overall longer stimulus length (resulting from the repetitions) would lead listeners to quickly become bored of the task and consequently not attend to the full stimulus length. Further, as demonstrated in Sec. V, listeners are susceptible to sequence-specific biases for stimuli that are defined by a 4-note sequence.

The sequence-specific biases analyzed in Sec. V remain mysterious. This analysis reveals a striking difference in the pattern of biases in the 3-task variants vs the 4-task variants. Notably, ending on note 13 (the high tonic) biases listeners to respond high-target (major) in the FCwR-3 and Slow-3 tasks; by contrast, ending on the same note biases listeners to respond low-target (fourth) in the FCwR-4 and Slow-4 tasks. Mednicoff *et al.* (2018) suggest that the bias in the 3-task may be explained by theories about the relationship between music and speech (Patel, 2005; Patel *et al.*, 2006). Specifically, intonation of speech that is spoken happily tends to end on a higher pitch (Juslin and Laukka, 2003; Swaminathan and Schellenberg, 2015; Curtis and Bharucha, 2010). Shared emotional expressiveness between music and speech would therefore suggest that a tone sequence that ends on a high note might be perceived as happier (major). Although the happy-vs-sad distinction applies most naturally to major-vs-minor stimuli, we speculate that consonant-vs-dissonant stimuli might also differ along this spectrum. Listeners prefer consonance in music (Trainor *et al.*, 2002), which is associated with harmoniousness and stability based on pitch intervals (Meyer, 2008). On the other hand, the tritone interval (which is present in the high-target 4-task stimuli) is highly dissonant and associated with unpleasantness (Meyer, 2008; Plomp and Levelt, 1965), which listeners may relate to sadness more easily than the consonant fourth interval. Thus, if a tone sequence that ends on a high note is perceived as happier, then perhaps listeners are biased to associate it with the more pleasant interval (fourth). Further research is needed to explore these ideas in depth.

ACKNOWLEDGMENTS

This work was supported by NIH Training Grant No. T32-DC010775 awarded to J.H. by the UC Irvine Center for Hearing Research. The authors also thank Melissa Huynh, Nguyen Pham, Olivia Capizzi, Nellie Kwang, Luis Zambrano, Jose Almaraz, and Alfonso Valenzuela for assisting with data collection.

APPENDIX

This appendix describes the constraints imposed on the parameters of the Descriptive model and Note-function-biased model from Sec. V in order to uniquely specify each model.

The stimulus-specific biases β_S from the Descriptive model are constrained as follows:

$$\sum_S \beta_S = 0 \quad \text{and} \quad \frac{1}{48}\sum_S \beta_S^2 = 1, \tag{A1}$$

where each sum is over all 48 stimuli S. The first constraint prevents β_S values from trading off with the threshold values η_k. The second constraint prevents β_S values from trading off with f_β, and also enables comparison of their magnitudes to those of the τ_S values (which also satisfy $(1/48)\sum_S \tau_S^2 = 1$).

Following Mednicoff *et al.* (2018), we forced the functions $f_\tau(R)$ and $f_\beta(R)$ from the Descriptive model to assign a fixed value to all R_k in a given sextile of the distribution of scale-sensitivities observed across all listeners k in the study.

The parameters of the Descriptive model are the 48 β_S values, the six values each of f_τ and f_β, and the 98 η_k values. Therefore, taking into account the two degrees of freedom sacrificed by imposing the constraints of Eq. (A1) on the β_S values, the model absorbs $48 + 12 + 98 - 2 = 156$ degrees of freedom.

The reader will note that the Note-function-biased model is under-constrained. For example, for any choice of the functions f_{note} and f_{pip} in the model, and any non-zero scalar α, if we replace f_{note} and f_{pip} with $\hat{f}_{\text{note}} = \alpha f_{\text{note}}$ and $\hat{f}_{\text{pip}} = f_{\text{pip}}/\alpha$, the new model will yield exactly the same predictions. To uniquely determine model parameters, we must specify the relative signs and amplitudes of f_β, f_{note}, and f_{pip}. We impose particular constraints to facilitate comparison of results from the FCwR-3, Slow3, FCwR-4, and Slow-4 tasks and also from Mednicoff *et al.* (2018).

To make the results from the Note-function-biased model comparable to those of the Pitch-height-biased model of Mednicoff *et al.* (2018), we constrain f_{note} to sum to 0 and also to have the same sum of squares as f_{PH} [Eq. (7)]. To ensure that the β_S values that result from Eq. (6) will satisfy Eq. (A1), f_{pip} is constrained to sum to 0, and scaled to make β_S's satisfy the right side of Eq. (A1). In addition, $f_{\text{pip}}(4)$ is constrained to be positive (which makes the f_{pip}'s from all four tasks similar in form). Finally, the sum of f_β taken across sextiles 3, 4, 5 of the R distribution is constrained to be positive (which makes the f_β's from all four tasks similar in form).

Thus, the total number of degrees of freedom absorbed by the Note-function-biased model is 115: f_{pip} uses 2 degrees of freedom; f_{note} uses 3; the η_k's use 98; and each of f_τ and f_β uses 6.

Blechner, M. J. (**1977**). "Musical skill and the categorical perception of harmonic mode," Haskins Laboratories Status Report on Speech Perception, SR-51/ **52**, 139–174.

Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., Guan, M., Silva, A. E., Gregersen, P. K., and Kowalski, E. (**2013**). "Bimodal distribution of performance in discriminating major/minor modes," J. Acoust. Soc. Am. **134**(4), 3067–3078.

Crowder, R. G. (**1984**). "Perception of the major/minor distinction: I. Historical and theoretical foundations," Psychomusicology **4**(1/2), 3–12.

Crowder, R. G. (**1985a**). "Perception of the major/minor distinction: II. Experimental investigations," Psychomusicology **5**(1/2), 3–24.

Crowder, R. G. (**1985b**). "Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations," Bull. Psychonom. Soc. **23**(4), 314–316.

Curtis, M. E., and Bharucha, J. J. (**2010**). "The minor third communicates sadness in speech, mirroring its use in music," Emotion **10**(3), 335–348.

Dean, T., and Chubb, C. (**2017**). "Scale-sensitivity: A cognitive resource basic to music perception," J. Acoust. Soc. Am. **142**(3), 1432–1440.

Deutsch, D. (**1982**). "Organizational processes in music," in *Music, Mind, and Brain* (Springer, New York), pp. 119–136.

Dowling, W. J. (**1973**). "Rhythmic groups and subjective chunks in memory for melodies," Percept. Psychophys. **14**(1), 37–40.

Gagnon, L., and Peretz, I. (**2003**). "Mode and tempo relative contributions to 'happy-sad' judgements in equitone melodies," Cognit. Emotion **17**(1), 25–40.

Gerardi, G. M., and Gerken, L. (**1995**). "The development of affective responses to modality and melodic contour," Music Percept. **12**(3), 279–290.

Halpern, A. R. (**1984**). "Perception of structure in novel music," Memory Cognit. **12**, 163–170.

Halpern, A. R., Bartlett, J. C., and Dowling, W. J. (**1998**). "Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience," Music Percept. **15**, 335–356.

Heinlein, C. P. (**1928**). "The affective characters of the major and minor modes in music," J. Comparative Psychol. **8**(2), 101–142.

Hevner, K. (**1935**). "The affective character of the major and minor mode in music," Am. J. Psychol. **47**, 103–118.

Hoel, P. G., Port, S. C., and Stone, C. J. (**1971**). *Introduction to Statistical Theory* (Houghton-Mifflin, Boston, MA).

Juslin, P. N., and Laukka, P. (**2003**). "Communication of emotions in vocal expression and music performance: Different channels, same code?," Psychol. Bull. **129**(5), 770–814.

Kastner, M. P., and Crowder, R. G. (**1990**). "Perception of the major/minor distinction: IV. Emotional connotations in young children," Music Percept. **8**(2), 189–202.

Knopoff, L., and Hutchinson, W. (**1983**). "Entropy as a measure of style: The influence of sample length," J. Music Theory **27**(1), 75–97.

Krumhansl, C. L. (**1990**). *Cognitive Foundations of Musical Pitch* (Oxford University Press, New York).

Krumhansl, C. L., and Kessler, E. J. (**1982**). "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys," Psychol. Rev. **89**(4), 334–368.

Lantz, M. E., and Cuddy, L. L. (**1998**). "Total and relative duration as cues to surface structure in music," Canadian Acoust. **26**(3), 56–57.

Leaver, A. M., and Halpern, A. R. (**2004**). "Effects of training and melodic features on mode perception," Music Percept. **22**, 117–143.

Macmillan, N. A., and Kaplan, H. L. (**1985**). "Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates," Psychol. Bull. **98**(1), 185–199.

Margulis, E. H., and Simchy-Gross, R. (**2016**). "Repetition enhances the musicality of randomly generated tone sequences," Music Percept. **33**(4), 509–514.

Mednicoff, S., Mejia, S., Rashid, J. A., and Chubb, C. (**2018**). "Many listeners cannot discriminate major vs minor tone-scrambles regardless of presentation rate," J. Acoust. Soc. Am. **144**(4), 2242–2255.

Meyer, L. B. (**2008**). *Emotion and Meaning in Music* (University of Chicago Press, Chicago, IL).

Miller, G. A. (**1956**). "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychol. Rev. **63**(2), 81–97.

Patel, A. D. (**2005**). "The relationship of music to the melody of speech and to syntactic processing disorders in aphasia," Ann. N. Y. Acad. Sci. **1060**(1), 59–70.

Patel, A. D., Iversen, J. R., and Rosenberg, J. C. (**2006**). "Comparing the rhythm and melody of speech and music: The case of British English and French," J. Acoust. Soc. Am. **119**(5), 3034–3047.

Plomp, R., and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**(4), 548–560.

Rosenthal, M. A., and Hannon, E. E. (**2016**). "Cues to perceiving tonal stability in music: The role of temporal structure," Music Percept. **33**(5), 601–612.

Smith, N. A., and Schmuckler, M. A. (**2004**). "The perception of tonal structure through the differentiation and organization of pitches," J. Exp. Psychol. **30**(2), 268–286.

Swaminathan, S., and Schellenberg, E. G. (**2015**). "Current emotion research in music psychology," Emotion Rev. **7**(2), 189–197.

Temperley, D., and Tan, D. (**2013**). "Emotional connotations of diatonic modes," Music Percept. **30**(3), 237–257.

Trainor, L. J., Tsang, C. D., and Cheung, V. H. (**2002**). "Preference for sensory consonance in 2- and 4-month-old infants," Music Percept. **20**(2), 187–194.

Wilks, S. S. (**1944**). *Mathematical Statistics* (Princeton University Press, Princeton, NJ).

Youngblood, J. E. (**1958**). "Style as information," J. Music Theory **2**(1), 24–35.