

SUPPLEMENTAL MATERIAL ACCOMPANYING, “A COMBINED
MODEL OF SENSORY AND COGNITIVE REPRESENTATIONS
UNDERLYING TONAL EXPECTATIONS IN MUSIC: FROM AUDIO
SIGNALS TO BEHAVIOR,” BY COLLINS, TILLMANN, BARRETT,
DELBÉ, AND JANATA.

CALCULATING EXPLANATORY VARIABLES FROM PERIODICITY PITCH,
CHROMA VECTOR, AND TONAL SPACE REPRESENTATIONS

For our statistical analyses of the RT data, explanatory variables were derived from each of the representational spaces (PP, CV, and TS). Each explanatory variable refers to a combination of attributes (variable classes) shown in Supplementary Table S1. These attribute values give rise to the sets of abbreviations used to label the explanatory variables.

Table S1. *Four attributes involved in calculating variables from processed audio*

Attribute	Options	Labels	x_{PP} , x_{CV} , and x_{TS}	y_{PP}	y_{CV}	z_{PP}
1. Representational space	Periodicity pitch, chroma vector, or tonal space	PP, CV, TS	PP, CV, TS, respectively	PP	CV	PP
2. Calculation type	Mean correlation (MC) or maximum value (MV)	MC, MV	MC	MV	MV	MC
3. Window comparison	Post (absolute) or pre-post (relative)	abs, rel	rel	abs	abs	abs
4. Post-target window (ms)	[0, 200] (early) or [201, 600] (late)	early, late	early	late	early	early

Note. The last four columns contain examples of attribute combinations that give rise to variables described in the *Regression results*. For instance x_{PP} arises from the combination {PP, MC, rel, early}.

Representational space

Representational space refers to whether a periodicity pitch (PP), chroma vector (CV) or tonal space (TS) representation is used as the basis for the variable.

Calculation type

Calculation type indicates whether a variable is based on the *mean correlation (MC)* of local (0.1 s time constant) and global (4 s time constant) context images within a specific time window, or instead on the *maximum value (MV)* of the mean image in a specific time window (using only the 4 s time constant constant). The rationale for using the maximum value in a mean image is that it provides a coarse estimate of the degree to which the energy distribution is concentrated within a specific region of the representational space. This information might be related to the clarity of a tonal center, in which a high MV would indicate a clearly established tonal center, while a low MV would indicate that the map is more uniformly activated at a lower amplitude. Note that one can attain a high MC independently of the shape of the distribution.

Window comparison

We use the term “relative” (*rel*) to refer to the difference between pre- and post-target values, and “absolute” (*abs*) to refer to the post-target values only. The pre-target window was 100 ms in duration, and is indicated by the black horizontal bar to the left of the vertical line in Figure 4D (value .728). In Figure 4D we show an early post-target correlation of .469 (gray horizontal bar) and a late post-target correlation of .351 (black horizontal bar to the right of the gray bar). The relative window comparison presumes that responses to target events are driven not as much by the absolute level of activation (either .469 or .351 in this case), but rather by the amount of change that the activation state of the target event represents from the activation state immediately preceding the onset of the target, that is, by a local differencing operation. In Figure 4D, the change is $-.259 \approx .469 - .728$ for the early post-target window, and $-.377 \approx .351 - .728$ for the late window. By contrast, the absolute readout implies that the absolute state of the model at each moment is the relevant parameter that listeners are tracking.

Post-target window

We used two post-target windows. An *early* window spanned 0 – 200 ms following target onset, while a *late* window spanned 201 – 600 ms. Initially, the late window was split into two sections (201 – 400 and 401 – 600 ms), but collapsed subsequently into a single window due to high correlations.

In total we considered a pool of 24 ($= 3$ representational spaces $\times 2$ calculation types $\times 2$ window comparisons $\times 2$ post-target windows) variables. As an example of the abbreviation system used in Table S2, the last variable discussed in relation to Figure 4D, taking the value -0.377 , is labeled {PP, MC, rel, late}. It comprised the mean correlation between two ($t = 0.1$ s and $t = 4$ s) leaky-integrated PP images (hence PP, MC), using values from either side of the target onset (relative or rel), and a window of 201-600 ms following the target onset (late).

The correlation structure of the matrix containing the 24 explanatory variables for all of the 303 stimuli was rank deficient, indicating the presence of highly correlated variables. These were general pairs of early/late variables. For multiple regression analyses that simultaneously estimated the variance associated with all variables in the model, we removed the *late* variable of a pair. A reduced set of 17 explanatory variables was thus created. To estimate the maximum value of R^2 that can be achieved for this RT data, a model was fitted consisting of the entire reduced set of 17 explanatory variables. For this model, $R^2 = .33$, $s = 93.10$. This model established an upper boundary for all analyses.

STEPWISE SELECTION

To address which weighted sum of variables, potentially drawn from different representational stages, was best able to model the observed RTs, we performed stepwise selection that began with the original set of 24 explanatory variables. Stepwise selection begins by comparing univariate models. A univariate model contains one explanatory variable. The univariate model that most reduces the residual sum of squares (RSS) becomes the stepwise model (if its explanatory variable is significant at the .05 level). The results of the first stage of stepwise selection are shown in Supplementary Table S2, and afford the interested reader the opportunity to see the explanatory power of each variable used in isolation. In the next stage, the least contributing variable in the model is

removed, unless it is significant. Next, models are formed by adding each remaining variable, in turn, to the stepwise model. The stepwise model is updated to include the remaining variable that most reduces the RSS (if significant at the .05 level). The process of adding and removing variables is repeated until no further changes can be made according to these rules. Tables similar in format to Supplementary Table S2 are constructed for each stage of stepwise selection, and the final table showing the addition and elimination of variables that resulted in Equation 1 is provided in Appendix B.

Table S2. *Individual fittings for the first stage of stepwise selection, in descending order of R^2*


Variable	B	$SE\ B$	b	R^2
$x_{TS} = \{TS, MC, rel, early\}^{\dagger\dagger}$	-142.66	15.46	-51.76***	.22
$\{PP, MC, abs, early\}^{\dagger}$	-392.18	43.34	-50.98***	.21
$\{TS, MC, abs, late\}^{\dagger}$	-106.16	12.35	-48.94***	.20
$\{TS, MC, rel, late\}$	-86.93	10.78	-46.48***	.18
$\{PP, MC, rel, late\}$	-259.39	33.13	-45.35***	.17
$\{TS, MV, rel, early\}^{\dagger}$	$-2.87 \cdot 10^6$	$3.73 \cdot 10^5$	-44.70***	.16
$\{TS, MC, abs, early\}^{\dagger}$	-113.03	15.31	-43.17***	.15
$\{TS, MV, rel, late\}^{\dagger}$	$-9.83 \cdot 10^5$	$1.33 \cdot 10^5$	-43.13***	.15
$z_{PP} = \{PP, MC, abs, late\}$	-248.28	34.66	-42.07***	.15
$x_{CV} = \{CV, MC, rel, early\}^{\dagger\dagger}$	-161.73	24.57	-39.10***	.13
$x_{PP} = \{PP, MC, rel, early\}^{\dagger\dagger}$	-239.28	38.31	-37.34***	.11
$\{CV, MC, rel, late\}$	-105.57	19.24	-33.24***	.09
$\{CV, MC, abs, late\}^{\dagger}$	-118.50	21.98	-32.71***	.09
$\{CV, MC, abs, early\}^{\dagger}$	-113.28	22.98	-30.13***	.07
$\{PP, MV, rel, early\}^{\dagger}$	-0.91	0.22	-24.73***	.05
$\{PP, MV, rel, late\}^{\dagger}$	-0.31	0.08	-24.48***	.05
$\{CV, MV, rel, early\}^{\dagger}$	-28.24	7.34	-23.90***	.05
$\{CV, MV, rel, late\}^{\dagger}$	-7.73	2.17	-22.22***	.04
$\{PP, MV, abs, early\}^{\dagger\dagger}$	0.06	0.03	14.17*	.02
$y_{PP} = \{PP, MV, abs, late\}$	0.04	0.03	9.69	.01
$\{TS, MV, abs, early\}^{\dagger}$	$6.37 \cdot 10^4$	$4.87 \cdot 10^4$	8.30	.01
$y_{CV} = \{CV, MV, abs, early\}^{\dagger}$	0.59	1.05	3.57	.00
$\{CV, MV, abs, late\}$	-0.51	1.14	-2.83	.00
$\{TS, MV, abs, late\}$	$-9.77 \cdot 10^3$	$5.08 \cdot 10^4$	-1.22	.00

Note. * $p < .05$. *** $p < .001$. The dotted indicates a cut-off point above which variables are significant at the .05 level. One dagger † indicates a member of the reduced set of seventeen variables; two daggers indicate significance in this model ($p < .05$).

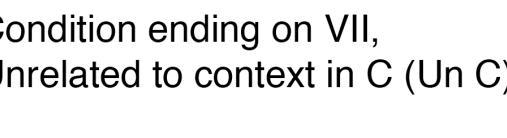
EXAMPLES OF STIMULUS MATERIALS

Three hundred and three stimuli were obtained from seven tonal priming experiments. Several stimulus examples, highlighting manipulations of interest, are illustrated in musical notation in Figures S1–S4.


Condition ending on I,
Related to context in C (Rel C)



Condition ending on VII,
Unrelated to context in C (Un C)



Condition ending on I,
Related to context in B (Rel B)



Condition ending on bII,
Unrelated to context in B (Un B)

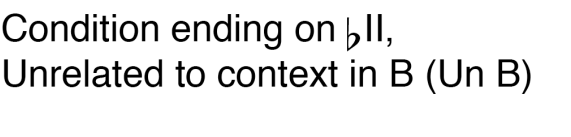


Figure S1. Stimuli from Tillmann, Janata, and Bharucha (2003). “Rel B” = stimulus in B major that ends on B major (related), “Rel C” = stimulus in C major that ends on C major, “Un B” = stimulus in B major that ends on C major (unrelated), “Un C” = stimulus in C major that ends on B major.



Figure S3. (A) Stimuli from Marmel, Tillmann, and Dowling (2008), demonstrating melodies with tonic and subdominant endings. “I” = stimulus that ends on the tonic scale degree, “IV” = stimulus that ends on the subdominant scale degree. (Adapted from Marmel et al. (2008), Figure 1A) (B) Stimuli from Marmel and Tillmann (2009), demonstrating melodies with median and leading tone endings. “III” = stimulus that ends on the median scale degree, “VII” = stimulus that ends on the leading tone; (C) Stimuli from Marmel et al. (2010), demonstrating melodies with tonic (I) and subdominant (IV) endings. The extra condition here, compared with Marmel et al. (2008), was to investigate the effect of a piano timbre versus a pure-tone timbre. Thus, the melodies shown were rendered in each of these timbres.

A ♩ = 120

B

C (♩ = 120)

Figure S4. (A) Stimuli from Tillmann et al. (2008) demonstrating a chord sequence with a tonic ending “I”, and a baseline sequence that crisscrosses the circle of fifths “I BL”; (B) From the same experiment, a chord sequence with a subdominant ending “IV”, and a paired baseline sequence “IV BL”; (C) A chord sequence with dominant ending “V” and its paired baseline sequence “V BL” (adapted from Tillmann, Janata, Birk, and Bharucha (2008), Figure 1B). Tillmann, Janata, Birk, and Bharucha (2003) used similar chord sequences, but restricted to tonic and subdominant endings and paired baseline sequences.