Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Modeling Response Times in Tonal Priming Experiments

Tom Collins,*[1] Barbara Tillmann,#[2] Charles Delbé,#[2] Frederick S. Barrett,*[1] Petr Janata*[1]

*Janata Lab, Center for Mind and Brain, University of California, Davis, USA

#Universite de Lyon, and Centre National de la Recherche Scientifique, France

[1]{tecollins, fsbarrett, pjanata}@ucdavis.edu, [2]{btillmann, charles.delbe}@olfac.univ-lyon1.fr

ABSTRACT

In tonal priming experiments, participants make speeded judgments about target events in short excerpts of music, such as indicating whether a final target tone or chord is mistuned. By manipulating the tonal function of target events, it is possible to investigate how easily targets are processed and integrated into the tonal context. We investigate the psychological relevance of attributes of processed audio signals, by relating those attributes to response times for over three hundred tonal priming stimuli, gathered from seven reported experiments. To address whether adding a long-term, "cognitive," representation of tonal hierarchy improves the ability to model response times, Leman's "sensory" periodicity pitch (PP) model is compared with a "cognitive" model (projection of PP output to a tonal space (TS) representing learned knowledge about tonal hierarchies), which incorporates pitch probability distributions and key distance relationships. Results revealed that variables calculated from the TS model contributed more to explaining variation in response times than variables from PP, suggesting that a cognitive model of tonal hierarchy leads to an improvement over a purely sensory model. According to stepwise selection, however, a combination of sensory and cognitive attributes accounts better for response times than either variable category in isolation. Despite the relative success of the TS representation, not all response time trends were simulated adequately. The addition of attributes based on transition probabilities may lead to further improvements.

Background

The tonal priming paradigm is one behavioral method by which inferences can be made about listeners' sensitivities to changes in tonal function. Participants make speeded judgments about the final chord or tone (called the target) of short excerpts of music (see Fig. 1A) consisting of several chords or melody notes (called the prime or context). Example tasks include tuning, dissonance, and timbre discrimination, with faster response times interpreted as reflecting more strongly expected events. Manipulating the prime and/or target makes it possible to investigate how easily listeners process different tonal functions (Tillman, Janata, Birk, & Bharucha, 2003; 2008). Other methods for investigating tonal expectancy include eliciting ratings of fit (Krumhansl, 1990), and more recently behavioral approaches have been complemented by neuroimaging techniques (Tillmann et al., 2003a).

Leman's (2000) periodicity pitch (PP) model and Bharucha's (1987) MUSACT are two well-known computational models for simulating observed trends in response times. While these models emulate some experimental observations, MUSACT is limited to symbolic input, and while the PP model works with acoustic input, it is

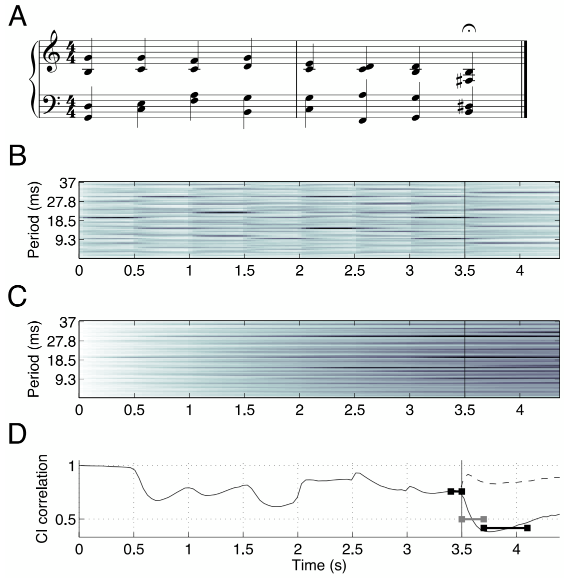

Figure 1: (A) Staff notation for a stimulus from Tillmann, Janata, & Bharucha (2003); (B) Integrated periodicity pitch images at successive time points (also called a context image) using the 0.1 s time constant. White for low to black for high relative intensity; (C) Context image (CI) for the same stimulus, using the 2 s time constant; (D) Plot against time of the correlation coefficient between a column vector from the local (0.1 s) context image and the corresponding column vector from the global (2 s) context image. The dashed line indicates the correlation coefficient for the stimulus ending on a C major chord, as opposed to the B major chord. Horizontal bars are the windows across which mean correlations are calculated, as explained in the text. Vertical bars in these figures indicate the position of the target chord.

limited to only an echoic, "sensory" musical memory. The latter limitation raises the question of whether adding a long-term, "cognitive," representation of tonal hierarchy improves the ability to model response times. A PP image can be thought of as a power spectrum of the audio signal over time, but taking into account the structure of a listener's auditory periphery. Leman (2000) integrated PP images using two time constants, for relatively local and global memories of the signal. A PP image integrated with a 0.1 s constant appears less leaky and more temporally local (Figure 1B) than a PP image integrated with a 2 s constant (Figure 1C). The correlation coefficient between corresponding column vectors of Figs. 1B and C is plotted in Fig. 1D. Leman (2000) proposed the correlation value at a time sample just after the target event's onset as a sensory explanation of probe tone profiles, challenging Krumhansl's (1990) cognitive account.

Aims

Our aim is to establish the psychological relevance of attributes of processed audio signals, by regressing response times for over three hundred musical priming stimuli on those attributes. First, Leman's (2000) sensory periodicity pitch (PP) model is compared with Janata's (2007) cognitive model, the latter incorporating pitch probability distributions and key distance relationships (TS). Second, we investigate whether linear combinations of sensory and cognitive attributes are useful predictors for observed response times.

Method

We considered different attributes of processed audio as explanatory variables for response times to priming stimuli. A third of the attributes were calculated using Leman's (2000) periodicity pitch (PP) model (www.ipem.ugent.be/toolbox), and the remaining two thirds were calculated by extending the PP model, creating a tonal space (TS) representation and also experimenting with chroma vectors (CV). The TS representation involves projecting integrated PP images to the surface of a toroidal (donut-shaped) self-organizing map (Kohonen, 1995), which encapsulates pitch probability distributions and key distance relationships (Janata, 2007).

For each of the three spaces (PP, CV, TS), correlation was calculated as described in relation to Fig. 1D. As well as defining the additional spaces CV and TS, we extended Leman's (2000) work, which used mean post-target correlations, in three ways: (1) alternative early and late post-target time windows were considered (0-200 and 201-600 ms, respectively), indicated by gray and black horizontal bars to the target's right in Fig. 1D; (2) some explanatory variables used the difference between mean post-target correlation and mean pre-target correlation (indicated by the black horizontal bar to the target's left in Fig. 1D); (3) we defined variables using the maximum value in a single image over a time window, as well as the mean correlation between two images.

Mean response times for 303 stimuli from seven reported experiments (Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003; Marmel & Tillmann, 2009; Marmel, Tillmann, & Delbé, 2010; Marmel, Tillmann, & Dowling, 2008; Tillmann et al., 2003a; 2003b; 2008) were regressed on sixteen variables calculated from attributes of the corresponding processed audio. Analyses of variance (ANOVA) were conducted. Here we focus on one that grouped the explanatory variables by space (PP, CV, TS) to see which group explained the greatest proportion of variation in response time data. A linear regression model was also built using stepwise selection, investigating whether a weighted sum of audio attributes might be related to observed response times. Stepwise selection adds/eliminates the most/least significant variables to/from a model, until no further changes can be made at $p = .05$. All models contained indicator variables for the experiments to allow for fixed between-experiment differences in response times.

Results

In the ANOVA grouped by representation (PP, CV, TS), the TS attributes contributed more to explaining response times ($p < .001$, Cohen's $f^2 = 0.09$) than the PP attributes ($p < .01, f^2 = 0.07$), which contributed more in turn than the CV attributes ($p < .05, f^2 = 0.05$).

The stepwise model consisted of four attributes beyond experiment indicators ($F(10, 292) = 55.36$, $p <$ 1E-16). For the stepwise model, $r^2 = .65$, meaning that it accounts for 65% of variation in response times. This should be contrasted with $r^2 = .44$ for a model consisting of experiment indicators only. A variable from the TS representation entered the stepwise model first (i.e., was most significant), but importantly the model contained both TS and PP attributes (two each).

Conclusions

The tonal space (TS) representation defined by Janata (2007) appears to be a more useful source of predictors of response times to tonal priming stimuli than either the periodicity pitch (PP) model (Leman, 2000) or the chroma vector (CV) representation. This conclusion is supported by the result of the ANOVA grouped by representation, and by the first addition to the stepwise model being a TS variable. Our findings shed new light on the sensory-cognitive debate concerning theories of tonal expectancy: the stepwise model contains sensory (PP) and cognitive (TS) variables, suggesting that both sensory and cognitive variables are necessary to account for listeners' tonal expectations in music.

Although our stepwise model simulates the majority of response time trends observed in the seven experiments we examined, there remains room for improvement. In particular, the simulated response times for some tonic targets are longer than observed. Future work—focused on modeling transition probabilities for chord and note events extracted from the stimulus audio—will attempt to refine simulated response times for identified problematic tonic targets.

Acknowledgment
This work was supported by NSF grant #1025310.

SELECTED REFERENCES

Bharucha, J.J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5*, 1-30.

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D.A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 159-171.

Janata, P. (2007). Navigating tonal space. *Computing in Musicology, 15*, Tonal Theory for the Digital Age, 39-50.

Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.

Krumhansl, C.L. (1990a). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17*, 481-509.

Marmel, F., & Tillmann, B.. (2009). Tonal priming beyond tonics. *Music Perception, 26*, 211-221.

Marmel, F., Tillmann, B., & Delbé, C.. (2010). Priming in melody perception: Tracking down the strength of cognitive expectations. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1016-1028.

Marmel, F., Tillmann, B., & Dowling, W.J. (2008). Tonal expectations influence pitch perception. *Perception and Psychophysics, 70*, 841-852.

Tillmann, B., Janata, P., & Bharucha, J.J. (2003a). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research, 16*, 145-161.

Tillmann, B., Janata, P., Birk, J., & Bharucha, J.J. (2003b). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 470-482.

Tillmann, B., Janata, P., Birk, J., and Bharucha, J.J. (2008). Tonal centers and expectancy: Facilitation or inhibition of chords at the top of the harmonic hierarchy. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1031-1043.