Fractal structure enables temporal prediction in music

Summer K. Rankin[a)]
Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Ross Building #826, Baltimore, Maryland 21205
srankin5@jhmi.edu

Philip W. Fink
School of Sport and Exercise, Massey University Manawatu, Private Bag 11 222, Palmerston North, New Zealand, 4442
P.Fink@massey.ac.nz

Edward W. Large
Departments of Psychology and Physics, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, Connecticut 06269-1020
edward.large@uconn.edu

Abstract: 1/*f* serial correlations and statistical self-similarity (fractal structure) have been measured in various dimensions of musical compositions. Musical performances also display 1/*f* properties in expressive tempo fluctuations, and listeners predict tempo changes when synchronizing. Here the authors show that the 1/*f* structure is sufficient for listeners to predict the onset times of upcoming musical events. These results reveal what information listeners use to anticipate events in complex, non-isochronous acoustic rhythms, and this will entail innovative models of temporal synchronization. This finding could improve therapies for Parkinson's and related disorders and inform deeper understanding of how endogenous neural rhythms anticipate events in complex, temporally structured communication signals.

PACS numbers: 43.75.Cd, 43.66.Mk, 43.66.Lj, 43.75.St [Q-JF]
Date Received: March 14, 2014 **Date Accepted:** June 30, 2014

1. Introduction

The temporal organization of acoustic signals is critical for perception and communication. Neuronal rhythms in auditory,[1] visual,[2] motor,[3] and frontal[4] cortical areas synchronize with the rhythms of communication signals. Cortical synchronization has been linked to attentional allocation,[5] scene analysis,[4] intelligibility,[6] and behavioral coordination.[7] For example, entrainment of cortical rhythms to speech may engage neurodynamic mechanisms of temporal prediction[8] to segregate incoming information[9] and organize spike timing.[10] However, key questions surround the prediction of quasi-periodic rhythms typical of music performance or conversational speech. Here we present the first empirical demonstration that fractal, or 1/*f*, temporal structure provides sufficient information for prediction of event onset times in a quasi-periodic musical rhythm.

Music is an important model system for the study of rhythmic communication because its rhythmic structure is well-understood. Musical rhythms are generally perceived to have a pulse, or basic beat, in the range of approximately 0.5–4 Hz.[11] Meter corresponds to the percept of alternating strong and weak beats; in addition to the pulse frequency, metrical structures comprise faster beat frequencies that subdivide the pulse (4–8 Hz), and slower beat frequencies (<2 Hz) that accent pulse cycles.[11] Rhythmic

[a)]Author to whom correspondence should be addressed.

patterns are also perceptually grouped into phrase structures. Like music, speech is generally hierarchical, where syllables (4–8 Hz) assemble into lexical and phrasal units at slower timescales (<4 Hz).[12] However, the fundamental timing of speech is more flexible than that of most music. Nevertheless, hierarchically nested auditory cortical rhythms[10] are entrained by both music[1,3] and speech rhythms.[13]

In musical performance, musicians may substantially vary the frequency of the underlying pulse (the tempo), to interpret compositions[14] and express emotion.[15] Interestingly, when people synchronize with musical performances that contain large tempo changes they predict the fluctuations, such that changes in the tapping rate correlate with changes in the pulse rate at zero lag.[16] Tempo fluctuations display fractal structure,[16] therefore it may be the intrinsic nature of $1/f$ serial correlations and fractal scaling that enables prediction. Conversely, it could be the information in other acoustic dimensions that renders tempo fluctuations predictable. When synchronizing with mechanical performances that lack temporal fluctuation, peoples' timing profiles resemble the metrical structure of the score.[17] Expert tempo fluctuations correlate with the musical attributes of the composition, including metrical structure, rhythmic pattern, and melody,[14] which also exhibit fractal structure.[16,18]

In this experiment we assessed the extent to which predictions rely on the intrinsic $1/f$ structure of temporal fluctuations versus the structure in other acoustic dimensions, by controlling both the structure of the fluctuations [3 temporal structure conditions: Natural ($1/f$), random (shuffled), and synthetic ($1/f$)] and the presence of musical information (4 acoustic information levels: Quarter-note, eighth-note, rhythm, and pitch).

2. Methods

2.1 Participants

Twelve volunteers (8 males, 4 females) from the Florida Atlantic University community participated in the experiment. Each participant signed an informed consent form that was approved by the Institutional Review Board at Florida Atlantic University.

2.2 Procedure

Participants tapped on a MIDI drum pad (Roland Hand-sonic HPD-15) to a sequence of MIDI events generated using a custom Max program (Cycling'74, San Francisco). The stimuli included 4 levels of musical information (quarter-note, eighth-note, rhythm, and pitch) crossed with 3 types of temporal structure (natural, synthetic, and random). In all 12 conditions (natural, random, synthetic) × (quarter-note, eighth-note, rhythm, pitch) the participants' task was to synchronize tapping at the quarter-note frequency [Fig. 1(E)], keeping pace with the changes in tempo [Fig. 1(F)]. An induction sequence of 5 clicks, with inter-onset intervals (IOIs) equal to the first IOI of the sequence, was presented to prepare the participants to tap at the correct metrical level. After the induction sequence, participants tapped to the performances for the full duration of the stimulus (2:07 min) using the index finger on their dominant hand.

2.3 Stimuli

The stimuli were based on an expert piano performance of Isaac Albeniz's *Iberia II, Triana* [Fig. 2(A); Mm. 1] recorded at the biannual Minnesota International Piano-e-competition, a highly competitive, judged piano competition.[19] In this competition, performances were recorded on a Yamaha CFIIIS concert grand piano equipped with Disklavier Pro recording technology, which collects the MIDI data via fiber optics. Jie Chen's first-prize winning performance (2004) of this piece was selected to create the stimuli because (1) the composition would be unfamiliar to the average participant, (2) the piece had a strong pulse which could easily be felt by the average listener, and (3) it was a master performance with large tempo fluctuations that exhibited fractal structure (see Table 1). Additionally, there was rhythmic activity throughout this piece at the 16th-note metrical level, which facilitated construction of the different tempo fluctuation conditions, described next.

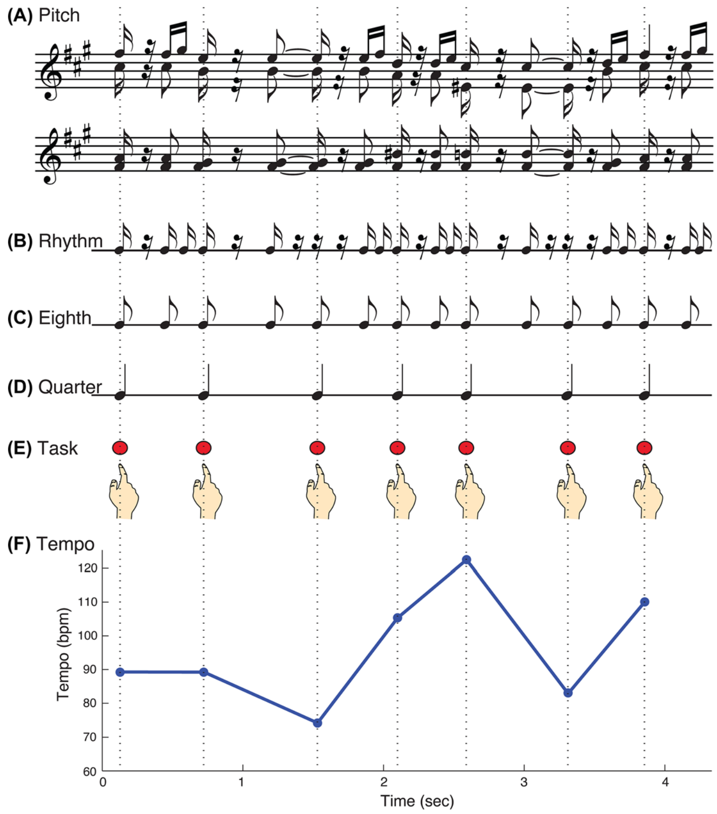

Fig. 1. (Color online) Musical information. Rhythmic and acoustic features were extracted from *Iberia II, Triana* and progressively added to the stimuli to create 4 levels of musical information. The quarter-note level (D) consisted of a sound for each beat at the quarter-note metrical level. The eighth-note level (C) consisted of a sound for each beat at the eighth-note metrical level. The rhythm level (B) consisted of a sound for each event in the musical score. The pitch level (A) contained the temporal information from the rhythm level with the addition of the pitch information for every event. Participants tapped at the quarter-note level (E) for each stimulus, keeping pace with large tempo changes (fluctuations from the natural stimulus are plotted in (F); see Sec. 2 for further information).

Mm. 1. Jie Chen's piano performance of Triana from Isaac Albeniz's *Iberia II.* (3.4 MB) [URL: http://dx.doi.org/10.1121/1.4890198.1].

2.3.1 Tempo fluctuations

The performance was matched to its written score using a custom dynamic programming algorithm.[20] Beat times were extracted from the first 2:07 min (measures 1-66) of the performance at the 16th-note metrical level. Beats without a corresponding sounded event were interpolated using local tempo, and inter-beat intervals (IBIs) were calculated by subtracting successive beat times. The 16th-note metrical level was chosen for the analysis of tempo fluctuations because there were a large number of beats ($N = 784$) with enough corresponding events to eliminate the need for excessive interpolation of beat times.

The Power Spectral Density (PSD) and Hurst's rescaled range (R/S) analyses were used to compute the fractal properties of IBI time series [Fig. 2(A) and Table 1]. The Hurst exponent, H, was calculated as the slope of the normalized range (R/S) as a function of interval length. The resulting value, H, can assume any value between 0 and 1, and gives a measure of smoothness (dimension) of a fractal object or time series. When $H = 0.5$, the points in the time series are uncorrelated and independent. When $H > 0.5$, positive statistically self-similar correlations (long-term memory) are present.[21]

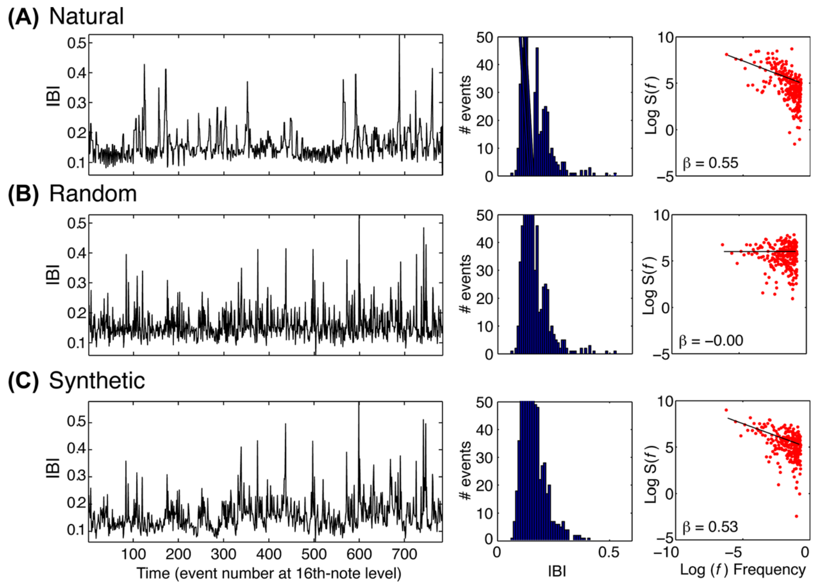

Fig. 2. (Color online) Temporal structure of stimuli. Time series (IBIs), distribution, and PSD for the 3 types of tempo fluctuation. In all 3 conditions mean IBI = 159 ± 52 ms (standard deviation). The natural (1/*f*) time series (A) was obtained by extracting beats at the 16th-note metrical level from the original performance. The random time series (B) was obtained by shuffling the natural time series (A) until $\beta = 0.00$. The synthetic time series (C) was obtained using the spectral synthesis method to essentially filter the random time series so that it would contain the same fractal structure as the natural time series.

Stimuli with 3 different types of tempo fluctuation were created for the experiment: *Natural*, *synthetic*, and *random.* We first computed the Hurst exponent and the slope of the PSD to measure the scaling and long-range correlation [see Fig. 2(A)] in the *natural* (fractal) condition. Next, we shuffled the 16th-note level natural IBIs until the PSD and R/S analyses showed that there were no long-term correlations (i.e., spectral slope = 0, $H = 0.5$) to produce the *random* time series, which had the same IBI distribution as the natural time series but was not fractal and did not correlate with the musical structure of the Albeniz composition [Fig. 2(B)]. Finally, we generated a *synthetic* fractal time series, using the spectral synthesis method,[21] that had the same Hurst exponent ($H = 0.76$; Table 1) and approximately the same IBI distribution as the natural time series [Fig. 2(C)], but did not correlate with the musical structure of the Albeniz composition. For each tempo fluctuation condition, IBIs were cumulatively summed to provide the event times.

2.3.2 Musical information

Each of the tempo conditions was presented using 4 levels of musical information (*quarter-note*, *eighth-note*, *rhythm*, and *pitch*). To create these levels we extracted

Table 1. Mean, standard deviation, number of events, H_{fGn}, and the autocorrelation at lag 1 (ac_1) for the IBIs of each fluctuation type at all 3 metrical levels (1/16 = sixteenth-note, 1/8 = eighth-note, 1/4 = quarter-note).

Fluctuation	Mean IBI 1/16	St dev IBI 1/16	Number of events 1/16	Number of events 1/8	Number of events 1/4	H_{fGn} 1/16	H_{fGn} 1/8	H_{fGn} 1/4	ac_1 1/16	ac_1 1/8	ac_1 1/4
Natural	159 ms	51.8	784	392	196	0.76	0.70	0.67	0.565	0.401	0.302
Synthetic	159 ms	59.4	784	392	196	0.76	0.77	0.78	0.383	0.451	0.469
Random	159 ms	51.8	784	392	196	0.51	0.53	0.54	−0.007	0.056	0.144

different kinds of musical structure from the Albeniz composition [Fig. 1(A)], and mapped this information onto the *natural*, *random*, and *synthetic* temporal structures described above. The *quarter-note* time series [Fig. 1(D)] was created by extracting every 4th beat from the 16th-note level time series of each of the 3 tempo fluctuation conditions. This yielded a time series of beats at the quarter-note metrical level (mean IBI = 743 ms) which reflected only changes in the quarter-note rate [Fig. 1(F)]. The quarter-note condition did not contain information about the meter, the rhythm, loudness fluctuations, or pitches. The quarter-note stimuli were delivered to the listener as a single stream of events using a woodblock sound (Mm. 2).

Mm. 2. Natural-quarter-note (1.4 MB) [URL: http://dx.doi.org/10.1121/1.4890198.2].

The *eighth-note* time series [Fig. 1(C)] was created by extracting every second beat from the 16th-note level time series. The *eighth-note* stimulus contained eighth-note level beat times that provided a metrical subdivision of the quarter-note level, but did not contain information about the rhythm, loudness fluctuations, or pitches. The eighth-note stimuli were played as a series of events (clave) with the quarter-note level beats presented simultaneously using a lower pitched woodblock sample. Participants were instructed to tap with the lower pitch (quarter-note). Thus, participants heard one sound for each beat at the eighth-note level, between each tap (Mm. 3).

Mm. 3. Natural-eighth-note (1.4 MB) [URL: http://dx.doi.org/10.1121/1.4890198.3].

The *rhythm* time series [Fig. 1(B)] was created by retaining all event times from the Albeniz composition (i.e., the rhythm), but eliminating pitch information. In other words, each of the 3 tempo fluctuation conditions (*natural*, *synthetic*, *random*) was mapped onto the rhythm from the Albeniz composition. The rhythm also did not include chord asynchronies or loudness fluctuations from the performance, because chord asynchronies are known to improve tempo tracking, and loudness fluctuations are typically highly correlated with tempo fluctuations.[14] The monotonic rhythm was presented using a clave sample, simultaneously with the quarter-note level events [Fig. 1(D)], which were presented using a lower pitched woodblock sample (Mm. 4). Participants were instructed to tap with the lower pitch (quarter-note), as in the eighth-note level.

Mm. 4. Random rhythm (1.4 MB) [URL: http://dx.doi.org/10.1121/1.4890198.4].

The *pitch* time series [Fig. 1(A)] included the same temporal information as the rhythm time series with the addition of the pitches from the Albeniz composition, which provided melodic and harmonic information. The pitch time series was presented using piano sounds, simultaneously with the quarter-note level events which were presented using a woodblock sound. Participants were instructed to tap with the woodblock sound (quarter-note level), as in the eighth-note and rhythm levels (Mm. 5).

Mm. 5. Random-pitch (1.4 MB) [URL: http://dx.doi.org/10.1121/1.4890198.5].

3. Results

Lag zero cross-correlation coefficients between the IBIs of the stimuli and participants' corresponding inter-tap intervals (ITIs), allowed us to assess the extent to which participants adjusted the length of each tapping cycle to anticipate the duration of the current cycle, which was determined by the time of the upcoming event. Thus, this measure indexed participants' ability to predict upcoming tempo changes.[22] A two-way analysis of variance revealed differences in the level of prediction as a function of temporal structure (natural, synthetic, random) and musical information (quarter-note, eighth-note, rhythm, pitch; Fig. 3). Significance was computed at $p < 0.01$ and pairwise t-tests were used for *post hoc* comparisons.

A main effect of temporal structure was observed [$F(2,22) = 250.40$, $p < 0.001$]. The correlation coefficients for the natural and synthetic fractal conditions were not significantly different ($p = 0.982$), but both were significantly greater than for the random temporal structure ($p < 0.001$). A main effect of musical information was also

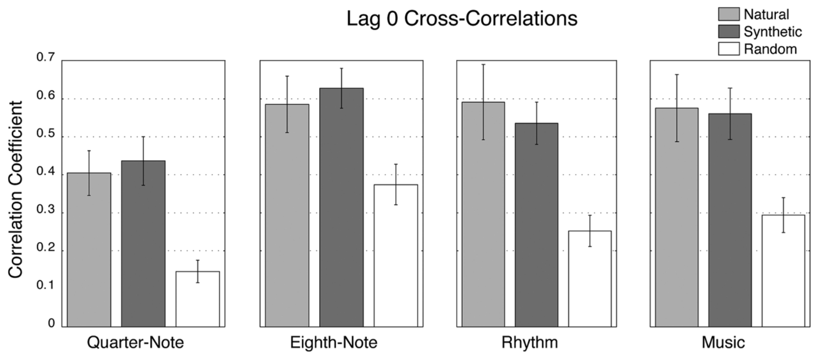

Fig. 3. Mean lag 0 cross-correlation (ITI $\times$ IBI) coefficients for each tempo condition and information level. Error bars are standard error of the mean. A main effect of temporal structure was observed where natural and synthetic fractal conditions were not significantly different ($p = 0.982$), but both were significantly greater than random ($p < 0.001$). A main effect of musical information was observed. Metrical subdivisions significantly improved prediction ($p < 0.001$). Adding rhythm or pitch information to the natural condition did not improve predictions ($p > 0.441$). The addition of rhythm or pitch information decreased prediction for the synthetic and random conditions: Synthetic eighth-note $\times$ synthetic rhythm $p < 0.001$; synthetic eighth-note $\times$ synthetic pitch $p < 0.001$; random eighth-note $\times$ random rhythm $p < 0.001$; random eighth-note vs random pitch $p = 0.002$.

observed [$F(3,33) = 40.40$, $p = 0.001$]. The addition of metrical subdivisions (eighth-note level) significantly improved prediction ($p < 0.001$). The addition of rhythm and pitch information caused a significant overall decrease in prediction beyond metrical subdivision ($p < 0.004$), and there was no difference between the rhythm and pitch conditions ($p = 0.516$). Interestingly, a significant two-way interaction was also observed [$F(6,66) = 11.74$, $p < 0.001$]. In the natural fractal condition, the prediction performance for the rhythm ($p = 0.653$) and pitch ($p = 0.441$) levels was not better than the eighth-note level (Fig. 3). However, in the synthetic and random conditions, where the tempo fluctuations were not related to the musical composition, correlations significantly deteriorated with the addition of rhythm and pitch information ($p < 0.002$), revealing that rhythm and pitch information were not irrelevant for prediction.

4. Discussion

These results demonstrate that participants use fractal temporal structure to predict tempo fluctuations, and temporal structure alone is sufficient to anticipate changes in tempo. This is true even when the tempo fluctuations are inappropriate for the music (i.e., in the synthetic condition), although predictions deteriorate slightly when inappropriate rhythm and pitch information is added, which indicates the presence of a competing source of information. Further experiments will be necessary to understand how multiple sources of information interact to enable temporal predictions. Understanding which sources of information are important for predicting upcoming sensory events is critical to understanding how attention is allocated in time.[8] Current models of dynamic attending and temporal synchronization predict that people react to changes in tempo, adjusting their frequency in response to the previous cycle length. The current work shows that such models are inadequate, and more sophisticated models will be necessary to explain anticipatory synchronization capabilities.[23,24] Moreover, further studies are needed to compare the influence of short- versus long-term memory on prediction, and to compare the predictability of fractal and non-fractal stimuli with similar positive lag-1 autocorrelations. The development of new models may have a direct application in music therapy; for example, temporal interaction is important in the (re)emergence of healthy levels of $1/f$ structure in human gait for Parkinson's disease patients. Such models will also shed light on emotional communication in music performance, as temporal fluctuations predict both reported affect[15] and real-time changes in neural activation.[15] Future models will also inform our understanding of the perception of spoken language[13] and other communication signals, more generally, because synchronization with and anticipation of irregular rhythms appears to play a critical role in rhythmic communication processes.

Acknowledgments

E.W.L. was supported by the Air Force of Scientific Research (AFOSR FA9550-07-C0095). Portions of this paper were adapted from S.K.R.'s doctoral dissertation.

References and links

1. S. Nozaradan, I. Peretz, M. Missal, and A. Mouraux, "Tagging the neuronal entrainment to beat and meter," J. Neurosci. **31**, 10234–10240 (2011).
2. P. Lakatos, G. Karmos, A. D. Mehta, I. Ulbert, and C. E. Schroeder, "Entrainment of neuronal oscillations as a mechanism of attentional selection," Science **320**, 110–113 (2008).
3. T. Fujioka, L. J. Trainor, E. Large, and B. Ross, "Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations," J. Neurosci. **32**(5), 1791–1802 (2012).
4. E. M. Zion Golumbic, N. Ding, S. Bickel, P. Lakatos, C. A. Schevon, G. M. McKhann, R. R. Goodman, R. Emerson, A. D. Mehta, J. Z. Simon, D. Poeppel, and C. E. Schroeder, "Mechanisms underlying selective neuronal tracking of attended speech at a 'Cocktail Party'," Neuron **77**(5), 980–991 (2013).
5. G. Stefanics, B. Hangya, I. Hernádi, I. Winkler, P. Lakatos, and I. Ulbert, "Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed," J. Neurosci. **30**(41), 13578–13585 (2010).
6. O. Ghitza and S. Greenberg, "On the possible role of brain rhythms in speech perception: Intelligibility of time compressed speech with periodic and aperiodic insertions of silence," Phonetica **66**, 113–126 (2009).
7. J. M. Mayville, S. L. Bressler, A. Fuchs, and J. A. S. Kelso, "Spatiotemporal reorganization of electric activity in the human brain associated with a timing transition in rhythmic auditory-motor coordination," Exp. Brain. Res. **127**, 371–381 (1999).
8. E. Large and M. Jones, "The dynamics of attending: How people track time-varying events," Psychol. Rev. **106**(1), 119–159 (1999).
9. O. Ghitza, "On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum," Front. Psychol. **3**(238) (2011).
10. P. Lakatos, A. S. Shah, K. H. Knuth, I. Ulbert, G. Karmos, and C. E. Schroeder, "An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex," J. Neurophysiol **94**, 1904–1911 (2005).
11. J. M. London, *Hearing in Time: Psychological Aspects of Musical Meter* (Oxford University Press, Oxford, 2004).
12. A.-L. Giraud and D. Poeppel, "Cortical oscillations and speech processing: Emerging computational principles," Nat. Neurosci. **15**, 511–517 (2012).
13. H. Luo and D. Poeppel, "Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex," Neuron **54**, 1001–1010 (2007).
14. C. Palmer, "Music performance," Annu. Rev. Psychol. **48**, 115–138 (1997).
15. H. Chapin, E. Large, K. Jantzen, J. Kelso, and F. Steinberg, "Dynamic emotional and neural responses to music depend on performance expression and listener experience," PLoS ONE **5**(12), e13812 (2010).
16. S. Rankin, E. Large, and P. Fink, "Fractal tempo fluctuation and pulse prediction," Music Percept. **26**(5), 401–413 (2009).
17. B. Repp, "Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper," Percept Psychophys. **61**, 529–548 (1999).
18. R. Voss and J. Clarke, "1/f noise in music and speech," Nature **258**(5533), 317–318 (1975).
19. "International Piano-e-Competition," www.piano-e-competition.com. (Last viewed July 20, 2010)
20. E. Large and S. Rankin, in *MIDI Toolbox: MATLAB Tools for Music Research*, edited by T. Eerola and P. Toiviainen (University of Jyväskylä, Kopijyvä, Jyväskylä, Finland, 2007).
21. D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics* (Cambridge University Press, Cambridge, 1997).
22. B. Repp, in *Common Mechanisms in Perception and Action: Attention and Performance XIX*, edited by W. Prinz and B. Hommel (Oxford University Press, Oxford, 2002), pp. 245–265.
23. D. Stephen, N. Stepp, J. Dixon, and M. Turvey, "Strong anticipation: Sensitivity to long-range correlations in synchronization behavior," Physica A **387**(21), 5271–5278 (2008).
24. M. van der Steen and P. E. Keller, "The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization," Front. Hum. Neurosci. **7**(253) (2013).