Synchronization in human musical rhythms and mutually interacting complex systems

Holger Hennig[1]

Department of Physics, Harvard University, Cambridge, MA 02138

Though the music produced by an ensemble is influenced by multiple factors, including musical genre, musician skill, and individual interpretation, rhythmic synchronization is at the foundation of musical interaction. Here, we study the statistical nature of the mutual interaction between two humans synchronizing rhythms. We find that the interbeat intervals of both laypeople and professional musicians exhibit scale-free (power law) cross-correlations. Surprisingly, the next beat to be played by one person is dependent on the entire history of the other person's interbeat intervals on timescales up to several minutes. To understand this finding, we propose a general stochastic model for mutually interacting complex systems, which suggests a physiologically motivated explanation for the occurrence of scale-free cross-correlations. We show that the observed long-term memory phenomenon in rhythmic synchronization can be imitated by fractal coupling of separately recorded or synthesized audio tracks and thus applied in electronic music. Though this study provides an understanding of fundamental characteristics of timing and synchronization at the interbrain level, the mutually interacting complex systems model may also be applied to study the dynamics of other complex systems where scale-free cross-correlations have been observed, including econophysics, physiological time series, and collective behavior of animal flocks.

time series analysis | long-range cross-correlations | anticorrelations | musical coupling | interbrain synchronization

In his book *Musicophilia*, neurologist Oliver Sacks writes: "In all societies, a primary function of music is collective and communal, to bring and bind people together. People sing together and dance together in every culture, and one can imagine them having done so around the first fires, a hundred thousand years ago" (1). Sacks adds, "In such a situation, there seems to be an actual binding of nervous systems accomplished by rhythm" (2). These thoughts lead to the question: Is there any underlying and quantifiable structure to the subjective experience of "musical binding"? Here, we examine the statistical nature of musical binding (also referred to as musical coupling) when two humans play rhythms in synchrony.

Every beat a single (noninteracting) layperson or musician plays is accompanied by small temporal deviations from the exact beat pattern, i.e., even a trained musician will hit a drum beat slightly ahead or behind the metronome (with a SD of typically 5–15 ms). Interestingly, these deviations are statistically dependent and exhibit long-range correlations (LRC) (3, 4). Listeners significantly prefer music mirroring long-range correlated temporal deviations over uncorrelated (white noise) fluctuations (5, 6). LRC are also inherent in the reproduction of both spatial and temporal intervals of single subjects (4, 7–9) and in musical compositions, such as pitch fluctuations (a simple example of pitch fluctuations is a melody) (10, 11) and note lengths (12). The observation of power law correlations in fluctuations of pitch and note length in compositions reflects a hierarchical, self-similar structure in these compositions.

In this article, we examine rhythmic synchronization, which is at the foundation of musical interaction, from orchestral play to audience hand-clapping (13). More specifically, we show that the interbeat intervals (IBIs) of two subjects synchronizing musical rhythms exhibit long-range cross-correlations (LRCCs), which

appears to be a general phenomenon given that these LRCC were found both in professional musicians and in laypeople.

The observation of LRCCs may point to characteristics of criticality in the dynamics of the considered complex system. LRCCs are characterized by a power law decay of the cross-correlation function and indicate that the two time series of IBIs form a self-similar (fractal) structure. Here, self-similarity implies that trends in the IBIs are likely to repeat on different timescales, i.e., patterns of IBI fluctuations of one musician tend to reproduce in a statistically similar way at a later time—even in the other musician's play. A variety of complex systems exhibit LRCCs; examples include price fluctuations of the New York Stock Exchange (where the LRCCs become more pronounced during economic crises) (14–16), heartbeat and EEG fluctuations (15, 17), particles in a Lorentz channel (18), the binding affinity of proteins to DNA (15), schools of fish (19), and the collective response of starling flocks (20, 21). The origin of collective dynamics and LRCCs based on local interactions often appears elusive (20), and is the focus of current research (19, 21). Of particular interest are the rules of interactions of the individuals in a crowd (22, 23) and transitions to synchronized behavior (16, 24). We introduce a stochastic model for mutually interacting complex systems (MICS) that generates LRCCs and provides a physiologically motivated explanation for the surprising presence of long-term memory in the cross-correlations of musical performances.

Interbrain synchronization has received growing attention recently, including studies of interpersonal synchronization (see ref. 4 for an overview), coordination of speech rhythm (25), social interactions (26), cortical phase synchronization while playing guitar in duets (27, 28), and improvisation in classical music performances (29).

Significance

Though the statistical properties of musical compositions have been widely studied, little is known about the statistical nature of musical interaction—a foundation of musical communication. The goal of this study was to uncover the general statistical properties underlying musical interaction by observing two individuals synchronizing rhythms. We found that the interbeat intervals between individuals exhibit scale-free cross-correlations, i.e., the next beat played by an individual is dependent on the entire history (up to several minutes) of their partner's interbeat intervals. To explain this surprising observation, we introduce a general stochastic model that can also be used to study synchronization phenomena in econophysics and physiology. The scaling laws found in musical interaction are directly applicable to audio production.

Notably, the differences between the beats of two musicians are on the order of only a few milliseconds, not much larger than the typical duration of a single action potential (~1 ms). The neurophysical mechanisms of timing in the millisecond range are still widely open (30, 31). EEG oscillatory patterns are associated with error prediction during music performance (32). Fine motor skills, such as finger-tapping rhythm and rate, are used to establish an early diagnosis of Huntington disease (33). The neurological capacity to synchronize with a beat may offer therapeutic applications for Parkinson disease, but the mechanisms are unknown to date (34). This study offers a statistical framework that may help to understand these mechanisms.

Experimental Setup and Methods

The performances were recorded at the Harvard University Studio for Electroacoustic Composition (see *SI Text* for details) on a Studiologic SL 880 keyboard yielding 57 time series of Musical Instrument Digital Interface (MIDI) recordings (Fig. 1, *Top*). Each recording typically lasted 6–8 min and contained ~1,000 beats per subject. The temporal occurrences $t_1, \ldots, t_n$ of the beats were extracted from the MIDI recordings and the interbeat intervals read $I_n = t_n - t_{n-1}$ with $t_0 = 0$. However, the results presented here apply not only to MIDI but also to acoustic recordings. The subjects were asked to press a key with their index finger according to the following. Task type Ia: Two subjects played beats in synchrony with one finger each. Task type Ib: Sequential recordings were made, where subject B synchronized with prior recorded beats of subject A. Sequential recordings are widely used in professional studio recordings, where typically the drummer is recorded first, followed by layers of other instruments. Task type II: One subject played beats in synchrony with one finger from each hand. Task type III: One subject played beats with one finger (finger-tapping). Finger-tapping of single subjects is well studied in literature (4) and serves as a baseline, but our focus is on synchronization between subjects. In addition to periodic tapping, a 4/4 rhythm {1, 2.5, 3, 4}, where the second beat is replaced by an offbeat, was used in tasks I–III.

For analysis of the cross-correlations between two nonstationary time series (here, between two sequences of interbeat intervals; Fig. 1*A*), a modified version of detrended cross-correlation analysis (DCCA) was used (17) (*Materials and Methods*). We added global detrending as an initial step before DCCA, which has been shown crucial in analyzing slowly varying nonstationary signals (35). DCCA calculates the detrended covariance $F(s)$ in windows of size s. LRCC are present if $F(s)$ asymptotically follows a power law $F(s) \sim s^\delta$ with $0.5 < \delta < 1.5$. In contrast, $\delta = 0.5$ indicates absence of LRCC.

A time series is considered long-range correlated if its power spectral density (PSD) asymptotically decays in a power law, $p(f) \sim 1/f^\beta$ for small frequencies f and $0 < \beta < 2$. The limits $\beta = 0$ ($\beta = 2$) indicate white noise (Brownian motion), whereas $-2 < \beta < 0$ indicates anticorrelations. Throughout this article, we measure the power spectral frequency f in units of the Nyquist frequency ($f_{\text{Nyquist}} = 1/2$ Hz), which is half the sampling rate of the time series. A method tailored for studying long-range correlations in slowly varying nonstationary time series is detrended fluctuation analysis (DFA) (36, 37) (*Materials and Methods*). The DFA method consists of (*i*) local detrending of the signal in windows of size s with a polynomial of degree k (we used $k = 2$) and (*ii*) calculation of the variance $F_{\text{DFA}}(s)$ of detrended segments of length s. For fractal scaling, we obtain $F_{\text{DFA}}(s) \sim s^\alpha$ with DFA exponent α (which in our case is the Hurst exponent). The DFA exponent quantifies the degree of persistence of memory in a given time series and is related to the PSD exponent β via $\beta = 2\alpha - 1$.

Many time series of physical, biological, physiological, and social systems are nonstationary and exhibit long-range correlations. The PSD is commonly analyzed to gain insight into the dynamics of natural systems. However, for nonstationary time series the
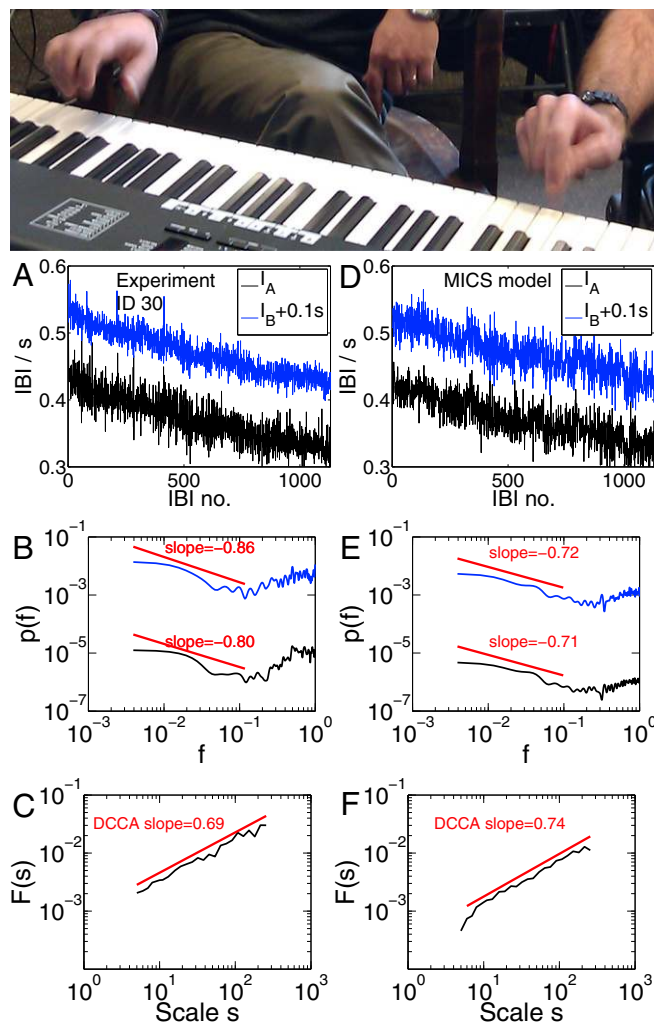
Fig. 1. (*Top*) Two professional musicians A and B synchronizing their beats: comparison of experiments (*A–C*) with MICS model (*D–F*). (*A*) The IBIs of 1,134 beats of musician A (black curve) and B (blue curve, offset by 0.1 s for clarity) exhibit slowly varying trends and a tempo increase from 133 to 182 beats per minute. (*B* and *E*) The gPSD of time series I_A, I_B shows LRC asymptotically for small f and anticorrelations for large f separated by a vertex of the curve at $f \sim 0.1 f_{\text{Nyquist}}$ (7). (*C*) Evidence of LRCC between I_A and I_B, DCCA exponent is $\delta = 0.69$. (*D–F*) The MICS model for $\beta_A = \beta_B = 0.85$, $n = 1,133$ predicts $\delta = 0.74$, in excellent agreement with the experimental data. A global trend extracted from A was added to the curve for illustration. Other parameters as in Fig. 2 *A* and *B*.

PSD method (also known as periodogram) fails: global trends manifest as spurious LRC in the PSD. We therefore propose global PSD (gPSD), which is an extension of the PSD method by including prior global detrending with a polynomial of degree $k = 1 \ldots 5$ (*Materials and Methods*). We found good agreement between DFA and gPSD in our data set within the margin of error.

Note that in addition to the numerical error of the least-squares fitting procedure, the algorithms for fractal analysis have internal error bars analyzed in detail by Pilgram and Kaplan (38). For example, DFA results for the scaling exponent α are expected to have a SD of ~±0.05 for data sets that are of the same size as ours. Because DCCA is strongly related to DFA, the internal errors of DCCA and DFA are expected to be of a comparable size.

Results

A representative example of the findings from a recording of two professional musicians, A and B, playing periodic beats in

synchrony (task type Ia) is shown in Fig. 1. Evidence for LRCCs between I_A and I_B on timescales up to the total recording time is reported in Fig. 1C with DCCA exponent $\delta = 0.69 \pm 0.05$. The two subjects are rhythmically bound together on a timescale up to several minutes, and the generation of the next beat interval of one subject depends on all previous beat intervals of both subjects in a scale-free manner. LRCCs were found in all performances of both laypeople and professionals, when two subjects were synchronizing simple rhythms (Fig. 2C). Thus, rhythmic interaction can be seen as a scale-free process. In contrast, when a single subject is synchronizing his left and right hands (task type II), no significant LRCCs were observed (Fig. 2C), suggesting that the interaction of two complex systems is a necessary prerequisite for rhythmic binding. However, further research may clarify whether different types of two-handed play in single individuals can similarly lead to LRCCs between the two hands.

What do the correlations between interbeat intervals look like in sequential recordings (task type Ib), where the musicians are recorded one after another? Surprisingly, we also found scale-free cross-correlations in sequential recordings (green circles in Fig. 2C), which is reproduced in the MICS model. Therefore, even though a wide class of pop and rock songs are currently recorded sequentially in recording studios, they do potentially contain LRCCs (and thus exhibit scale-free coupling) between individual tracks.

We identify two distinct regions in the PSD of the interbeat intervals separated by a vertex of the curve at a characteristic frequency $f_c \sim 0.1 f_{\text{Nyquist}}$ (Fig. 1B and *SI Text*): (*i*) The small frequency region asymptotically exhibits long-range correlations. This region covers long periods of time up to the total recording

time. (*ii*) The high-frequency region exhibits short-range anti-correlations. This region translates to short timescales. These two regions were first described in single-subjects' finger-tapping without a metronome (7). Because these two regions are observed in the entire data set (i.e., in all 57 recorded time series across all tasks), this finding suggests that these regions are persistent when musicians interact. Fig. 1E shows that the MICS model reproduces both regions and f_c for interacting complex systems.

The two subjects potentially perceive the deviations $d_n = t_{A,n} - t_{B,n}$ between their beats. The DFA exponent $\alpha = 0.72$ for the time series d_n indicates long-range correlations in the deviations (averaging over the entire data set, we find $\overline{\alpha} = 0.73 \pm 0.11$). In Fig. 3A the average deviation is $\overline{d_n} = -30.6$ ms, which may indicate that musician A led and B followed. An analysis how the scaling exponents differ in cases when no obvious leader–follower relation is observed can be found in *SI Text*.

MICS Model

What is the origin of the scale-free cross-correlations? In algorithms generating LRCCs, such as a two-component autoregressive fractionally integrated moving average (ARFIMA) process (35), the n'th element of the time series is essentially generated by a stochastic term plus a weighted sum over all previous elements (see Eq. **4** in *Materials and Methods*). This weighted sum is also inherent in other well-known long-memory processes, such as fractional Brownian motion. Although these statistical processes are widely applied in finance, climate analysis, etc., a weighted sum over all previous elements does not appear to be (physiologically) meaningful in many cases, including our case, because it requires explicit memory for these elements. In fact, it is highly unlikely that each subject has declarative memory for hundreds of interbeat intervals played in the preceding minutes. Hence, the occurrence of LRCCs when musicians synchronize rhythms is surprising. Moreover, our experimental results deviate strongly from the analytical prediction based on two-component ARFIMA (dashed line in Fig. 2C), where $\delta_{\text{ARFIMA}} = (\alpha_A + \alpha_B)/2$ with DFA (Hurst) exponents α_A, α_B (35). In the following, a model for MICS will be introduced (see ref. 39 for a related model based on an activation-threshold mechanism), where LRCCs emerge dynamically from a local interaction.

Gilden et al. (7) presented a model in which the generation of temporal intervals by a single person is composed of two parts: an internal clock and a motor program associated with moving a finger or limb. The delay of the motor program is given by Gaussian white noise ξ_n. The internal clock generates beat intervals C_n where the PSD consists of $1/f$ noise (7) [which as a first step we generalize to $1/f^\beta$ noise with $0 < \beta < 2$ to account for recent studies, where a range of power law exponents were found (3, 4)].

The following observation is built into the MICS model: When two subjects A and B are synchronizing a rhythm, each person attempts to (partly) compensate for the deviations $d_n = t_{A,n} - t_{B,n}$ perceived between the two n'th beats when generating the n + 1'th beat. We propose the following model for MICS:

$$
\begin{aligned}
I_{A,n} &= \sigma_A C_{A,n} + T + \xi_{A,n} - \xi_{A,n-1} - W_A d_{n-1} \\
I_{B,n} &= \sigma_B C_{B,n} + T + \xi_{B,n} - \xi_{B,n-1} + W_B d_{n-1},
\end{aligned} \tag{1}
$$

where $C_{A,n}$ and $C_{B,n}$ are Gaussian-distributed $1/f^\beta$ noise time series with exponents $0 < \beta_{A,B} < 2$, and T is the mean beat interval. We set $d_0 = 0$. The coupling strengths $0 < W_{A,B} < 2$ describe the rate of compensation of a deviation in the generation of the next beat. In the limit $W_A = W_B = 0$ and $\beta_A = \beta_B = 1$, the MICS model reduces to the Gilden model (7). The MICS model diverges for $W_A + W_B \geq 2$, i.e., when subjects are overcompensating.

Anticorrelations on short timescales arise from the term $-\xi_{n-1}$: a long interbeat interval is likely followed by a short one and vice versa to maintain a given tempo (7), which offers an explanation for the two regions in the PSD of the interbeat intervals in Fig. 1B. Because the delay of the motor program has no
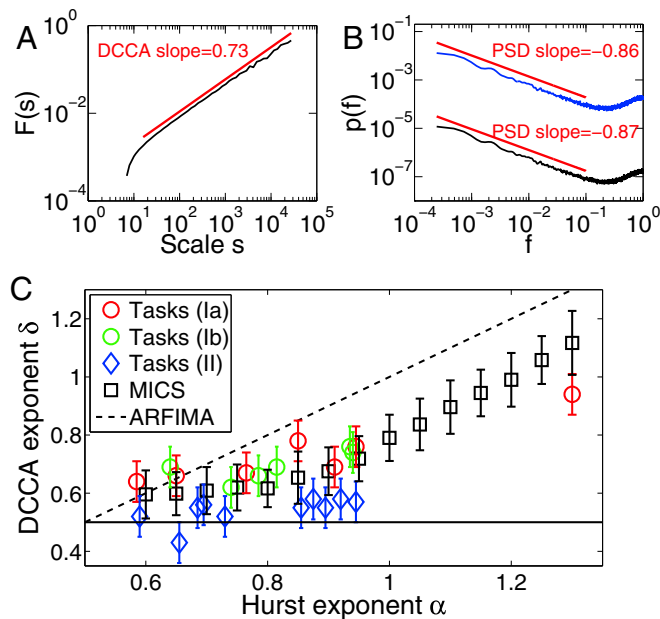
Fig. 2. (*A*) Evidence of scale-free cross-correlations in the MICS model. (*B*) The PSD of I_A (and I_B) shows two regions: LRC asymptotically for small f with exponent $\beta(I_A) = 0.86 \sim \max(\beta_A, \beta_B)$ and anticorrelations for large f. Other parameters (*A* and *B*): $n = 2^{17}$, $\beta_A = \beta_B = 0.85$, coupling $W_A = W_B = 0.5$, and $\sigma_A = \sigma_B = 6$. (*C*) Excellent agreement is found between the predicted δ of the MICS model and tasks Ia (simultaneous recordings, marked by red circles) and Ib (sequential recordings; green circles). For each DFA (Hurst) exponent $\alpha_A = \alpha_B \equiv \alpha$, 100 time series of length $n = 2,048$ were generated with the MICS model, and δ was calculated for all power law estimates with Pearson correlation coefficient $R > 0.95$ (for $\alpha \geq 0.9$ all realizations fulfilled this criterion). In contrast, the ARFIMA model (dashed line) deviates strongly for $\alpha > 0.75$ from the experiments (tasks Ia and Ib). Recordings of single subjects (tasks II) are consistent with absence of LRCC, i.e., $\delta = 0.5$ (black line).
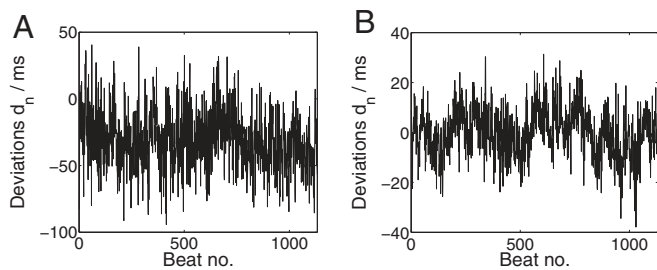
Fig. 3. The time series of deviations d_n (Eq. **2**) between the beats of two synchronizing individuals contain LRC both (*A*) in the experiment and (*B*) in the MICS model. (*A*) As a representative example, we show the d_n obtained from the interbeat intervals $I_{A,n}$ and $I_{B,n}$ of the two professional musicians in Fig. 1*A*. Average deviation is $\overline{d_n} = -30.6$ ms. The DFA exponent reads $\alpha = 0.72$, indicating long-range correlations. (*B*) Deviations d_n simulated with the MICS model (Eq. **1**) with DFA exponent $\alpha = 0.89$. Other parameters are $n = 1{,}133$ data points, $\beta_A = \beta_B = 0.85$, $\sigma_A = \sigma_B = 6$, $W_A = W_B = 0.5$.

long-term memory, its effect decays exponentially over time; hence, the anticorrelations are only seen on short timescales, whereas the long-time behavior is given by the long-range correlated clock noise. The relative strength of clock noise over motor noise is given by σ_A and σ_B and can be extracted from the slope of the anticorrelations in I_A and I_B for high f: the larger σ_A the smaller the anticorrelations in I_A (7). One would expect the anticorrelations to decrease when every other beat is neglected. Indeed, by selecting all odd number terms $\{I_{2n-1}\}$ (or the even number terms), the short-range anticorrelations in the PSD are significantly reduced (both in the MICS model and in the data) and can even vanish if clock noise dominates (i.e., for large σ_A, σ_B). LRCCs are significantly reduced when selecting only every other element of the time series.

A comparison of the MICS model (Fig. 1 *D*–*F*) with our experiments (Fig. 1 *A*–*C*) shows excellent agreement. The vertex at the characteristic frequency f_c in the PSD is reproduced by the MICS model (Fig. 1 *B* and *E*).

How do scale-free cross-correlations emerge in the MICS model? The deviations d_n, which the musicians perceive and adapt to, can be written as a sum over all previous interbeat intervals

$$d_n = t_{A,n} - t_{B,n} = \sum_{j=1}^{n} (I_{A,j} - I_{B,j}),\qquad[2]$$

thus involving all previous elements of the time series of IBIs of both musicians. Therefore, the MICS model suggests that scale-free coupling of the two subjects emerges mainly through the adaptation to deviations between their beats.

What are predictions of the MICS model? (P1) Emergence of LRCCs (Fig. 2*A*): The predicted DCCA exponents δ, which are not a parameter of the model, are in excellent agreement with the experiments (Fig. 2*C*). (P2) Asymptotically, the DFA scaling exponents $\alpha_{A,B}$ of the interbeat intervals are determined by the "clock" with the strongest persistence: $\alpha_A = \alpha_B = [\max(\beta_A, \beta_B) + 1]/2$. This result is valid for long time series of length $N \gtrsim 10^5$ (Fig. 2*B*). Surprisingly, even when turning off, say, clock A (i.e., $\beta_A = 0$), the long-time behavior of both I_A and I_B is asymptotically given by the exponent of the long-range correlated clock B (and vice versa) for large N. Thus, the musician with the higher scaling exponent determines the partner's long-term memory in the IBIs. However, in experiments the exponents can differ significantly in shorter time series of length $N \sim 1{,}000$, which can be seen by comparing the PSD exponents in Figs. 1*E* and 2*B*. (P3) Sequential recordings ($W_A = 0$ and $0 < W_B < 2$): The MICS model predicts LRCC for sequential recordings in agreement with the experimental findings (green circles in Fig. 2*C*). (P4) The

DFA exponent of the time series of deviations d_n is given by $\alpha = [\max(\beta_A, \beta_B) + 1]/2$ in the limit $N \to \infty$. However, the DFA exponents of the time series of deviations can significantly differ for finite lengths $N \sim 1{,}000$ beats. A sample sequence is shown in Fig. 3*B*.

A possible extension of the MICS model is to consider variable coupling strengths $W = W(d_n)$. Because larger deviations are likely to be perceived more distinctly, one possible scenario is to introduce couplings W that increase with d_n. For example, W may increase when large deviations such as glitches are perceived. In cases where the coupling strengths W_A or W_B depend on d_n, Eq. **1** becomes nonlinear in d_n.

Application: Musical Coupling

The observation of scale-free (fractal) musical coupling can be used to couple the interbeat intervals of two or more audio tracks, i.e., to imitate the generic interaction between musicians. Additional temporal deviations from the exact beat pattern, which, e.g., are introduced by intention by a musician to interpret a musical piece, are not modeled by the procedure below. Though the interbeat intervals are modified, all other characteristics, such as pitch, timbre, and loudness, remain unchanged.

Introducing musical coupling in two or more sequences is referred to as "group-humanizing." More than two audio tracks can be group-humanized by having each additional track responding to, e.g., the average of all other tracks' deviations. In addition, it is possible to group-humanize sequences by having a computer respond to or follow a musician in a "humanized" manner. For example, a musician can play a MIDI instrument and the computer instruments adapt to his unique input.

The procedure to introduce musical coupling in two audio tracks A and B is demonstrated using an instrumental version of the song "Billie Jean" by Michael Jackson. This song was chosen because drum and bass tracks consist of a simple rhythmic and melodic pattern that is repeated continuously throughout the entire song; this leads to a steady beat in drum and bass, which is well suited to demonstrate their generic mutual interaction. For simplicity, we merge all instruments into two tracks: track A includes all drum and keyboard sounds, and track B includes the bass. The interbeat intervals of tracks A and B read $I_{A,t} = X_t + T$ and $I_{B,t} = Y_t + T$, where T is the average interbeat interval given by the tempo (here, $T = 256$ ms, which corresponds to 234 beats per minute in the eighth notes). If the time series X_t (blue curve in Fig. 4, *Upper*) and Y_t (black curve) are long-range cross-correlated, we obtain musical coupling between drum and bass tracks. We modified a MIDI version of the song Billie Jean following three different procedures (audio samples can be found in *SI Text*). (*i*) Musical coupling (also called group-humanizing): X_t and Y_t contain LRCCs using a two-component ARMIFA process; (*ii*) "humanizing" the tracks separately: X_t and Y_t each contain $1/f^\beta$ noise, but are statistically independent; and (*iii*) randomizing: X_t and Y_t consist of Gaussian white noise.

For all three procedures, a SD of 10 ms was chosen for X_t and Y_t. Parameters are $\delta = 0.9$ for musical coupling and $\beta = 0.9$ for humanizing. The time series of deviations X_t and Y_t for musical coupling are shown in Fig. 4. The measured DCCA exponent reads $\delta = 0.93$ (in agreement with the analytical value 0.9 within margins of error) showing LRCC. For separately humanized sequences, however, we expect absence of LRCC. Indeed, the detrended covariance of X_t and Y_t oscillates around zero (*SI Text*), i.e., no LRCCs are found (35). Other processes that generate LRCC could also be used to induce musical coupling, including the MICS model (Eq. **1**). Randomizing is implemented in professional audio editing software, whereas a group-humanizer program for coupling MIDI tracks is available from www.nld.ds.mpg.de/~holgerh/download_software.html.

Our procedure of introducing musical coupling of a pair of sequences ($I_{A,t}, I_{B,t}$) reduces to humanizing the sequences I_A and I_B
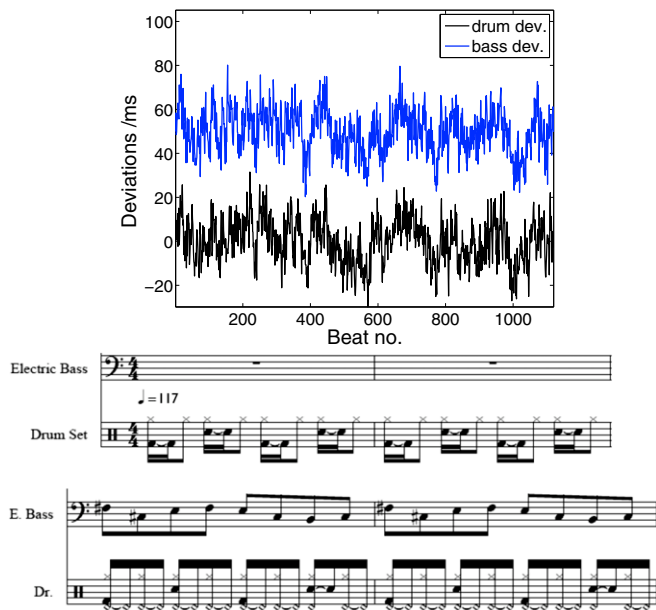
Fig. 4. (Upper) Scale-free coupling of two audio tracks: The deviations from their respective positions (e.g., given by a metronome) are shown in the drum track (upper blue curve, offset by 50 ms for clarity) and bass track (lower black curve) to introduce musical coupling. When an instrument is silent on a beat, the corresponding deviation is skipped. The time series each of length $n = 1,120$ were generated with a two-component ARFIMA process with Hurst exponents $\alpha_A = \alpha_B = 0.9$ and coupling constant $W = 0.5$. (Lower) Excerpt of the first four bars of the song "Billie Jean" by Michael Jackson. Because there is a drum sound on every beat, all 1,120 deviations are added to the drum track, whereas in the first two bars the bass pauses.

independently in the limit of vanishing coupling ($W_A = W_B = 0$). Therefore, humanizing the tracks independently can be seen as a special case of musical coupling. Because humanizing is preferred by listeners over randomizing (3, 5), we expect that the coupled/humanized samples (nos. 1 and 2) will be preferred over the randomized sample (no. 3). Audio sample 1 imitates the interplay of two musicians, whereas sample 2 only reflects the statistics of the play of single musicians that do not interact with each other.

Conclusion

We showed that rhythmic synchronization between individuals (both musicians and laypeople) exhibits long-term memory of the partner's interbeat intervals up to several minutes for which an explanation is suggested in a physiologically motivated stochastic model for MICS. The MICS model suggests that scale-free coupling of the two subjects emerges mainly through the adaptation to deviations between their beats.

The MICS model may be applicable to other mutually interacting complex systems, such as synchronization phenomena and interdependencies in finance, heartbeats, EEG signals (15, 17), or bird flocks (20), where the origin of scale-free cross-correlations and resulting collective behavior is often elusive. On the methods side, to avoid artifacts in the PSD of slowly varying nonstationary time series, our analysis suggests prior global detrending. We demonstrated that the observed memory phenomenon in rhythmic synchronization can be imitated by fractal coupling of audio tracks and thus applied in electronic music. Though this study complements our understanding of timing and synchronization at the neural networks level, we hope that this work further stimulates the interdisciplinary study of neuronal correlates of timing and synchronization at the interbrain level, e.g., based on combined audio, EEG, or fMRI measurements.

Ethics Statement. This study was reviewed by the Harvard University Committee on the Use of Human Subjects in Research, which determined that it meets the criteria for exemption to Institutional Review Board approval. Informed consent was obtained from the healthy adult subjects.

Materials and Methods

Globally Detrended Power Spectral Density. To analyze the PSD of slowly varying nonstationary time series, we propose the gPSD method, which is a modification of the PSD method and involves prior global detrending; it extends previous approaches where linear global detrending (39) was used by including higher order polynomials of degree k. The gPSD method consists of the following steps, given a time series X_n of length N: (i) Global detrending of X_n with a polynomial of degree $k = 1\ldots5$. (ii) Divide X_n in m segments of length N/m. (iii) Calculating the power spectral density $p_r(f) = \left|\text{FFT}(X_n^{(r)})\right|^2$ in each (sliding of shifted) window of size N/m, where FFT denotes the fast Fourier transform, including zero padding, and r labels the different windows. (iv) Averaging the power spectra $p(f) = \langle p_r(f)\rangle$ and calculating the power spectral exponent of $p(f)$ by means of a least-squares fit. If not noted otherwise, we used $k = 2$, $m = 8$ and shifted the windows by $N/(2m)$ thus obtaining $r = 1\ldots2m-1$ power spectra for each time series. The fitting range for gPSD is $[f_{\min}, f_c]$, where $f_{\min} = 1/(2N)$ is the smallest possible power spectral frequency, and the position of the characteristic frequency f_c in the PSD was determined automatically.

Detrended Fluctuation Analysis. A well-known method tailored for studying slowly varying nonstationary time series is DFA (for details, see refs. 36 and 37). The DFA fitting range used here reads $[s_{\min}, N/4]$, where $s_{\min}$ was determined automatically by minimizing the error of a least-squares fit (quantified by Pearson's correlation coefficient). However, in this work we are not only interested in the DFA exponent itself but also in the shape of the PSD, which exhibits two distinct regions separated by a vertex at the characteristic frequency f_c. DFA shows a much more smooth transition between the two regions; therefore, gPSD is not only a valuable independent measure for long-range correlations, but also vital to characterize the scaling behavior in both regions.

Detrended Cross-Correlations Analysis. DCCA has been developed to analyze the cross-correlations between two nonstationary time series (17). As suggested in ref. 35, we added global detrending with a polynomial of degree k as initial step before DCCA. Global detrending proved to be a crucial step to calculate the DCCA exponent of the nonstationary time series in our data. Without global detrending much larger DCCA exponents are obtained, i.e., spurious LRCC are detected that reflect global trends. Given two time series X_n, X'_n, where $n = 1\ldots N$. The DCCA method including prior global detrending consists of the following steps: (i) Global detrending: fitting a polynomial of degree k to X_n and a polynomial to X'_n, where typically $k = 1\ldots5$. We used $k = 3$ and carefully checked that the obtained DCCA scaling exponents did not change significantly with k. (ii) Integrating the time series $R_n = \sum_{i=1}^{n} X_n$ and $R'_n = \sum_{i=1}^{n} X'_n$. (iii) Dividing the series into windows of size s. Least-squares fit $\bar{R}_n$ and $\bar{R}_n$ for both time series in each window. (iv) Calculating the detrended covariance

$$F_{\text{DCCA}}(s) = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} (R_k - \bar{R}_k)(R'_k - \bar{R}'_k), \qquad [3]$$

where N_s is the number of windows of size s. For fractal scaling, $F_{\text{DCCA}}(s) \propto s^\delta$ with $0.5 < \delta < 1.5$. Absence of LRCC is indicated by fluctuations of $F_{\text{DCCA}}(s)$ around zero. If, however, the detrended covariance $F_{\text{DCCA}}(s)$ changes signs and fluctuates around zero as a function of the timescale s, LRCCs are absent. When setting $X_n = X'_n$ and without global detrending (step 1), DFA is obtained. In this respect, DFA can be seen as a detrended autocorrelation analysis.

Two-Component ARFIMA. A two-component ARFIMA process has been proposed that generates two time series $x_{1,2}$, which exhibit LRCC (15, 17). The process is defined by

$$X_t = \sum_{n=1}^{\infty} w_n(\alpha_A - 0.5)x_{t-n}$$
$$Y_t = \sum_{n=1}^{\infty} w_n(\alpha_B - 0.5)y_{t-n}$$
$$x_t = [WX_t + (1-W)Y_t] + \xi_{t,A}$$
$$y_t = [(1-W)X_t + WY_t] + \xi_{t,B}, \qquad [4]$$

with Hurst exponents $0.5 < \alpha_{A,B} < 1$, weights $w_n(d) = d\Gamma(n - d)/(\Gamma(1 - d)\Gamma(n + 1))$, Gaussian white noise $\xi_{t,A}$, and $\xi_{t,B}$ and gamma function Γ. The

coupling constant W ranges from 0.5 (maximum coupling between x_t and y_t) to 1 (no coupling). It has been shown analytically that the cross-correlation exponent is given by $\delta_{\text{ARFIMA}} = (\alpha_A + \alpha_B)/2$ (dashed line in Fig. 2C). An example of two coupled time series generated with a two-component ARFIMA process is shown in Fig. 4.

1. Sacks O (2008) *Musicophilia: Tales of Music and the Brain* (Vintage Books, New York).
2. Sacks O (2006) The power of music. *Brain* 129(Pt 10):2528–2532.
3. Hennig H, et al. (2011) The nature and perception of fluctuations in human musical rhythms. *PLoS ONE* 6(10):e26457.
4. Repp BH, Su YH (2013) Sensorimotor synchronization: A review of recent research (2006-2012). *Psychon Bull Rev* 20(3):403–452.
5. Hennig H, Fleischmann R, Geisel T (2012) Musical rhythms: The science of being slightly off. *Phys Today* 65:64–65.
6. Hennig H, Fleischmann R, Geisel T (2012) Immer haarscharf daneben [Just a hair off]. *Spektrum Wiss* 9:16–20. German.
7. Gilden DL, Thornton T, Mallon MW (1995) 1/f noise in human cognition. *Science* 267(5205):1837–1839.
8. Roberts S, Eykholt R, Thaut MH (2000) Analysis of correlations and search for evidence of deterministic chaos in rhythmic motor control by the human brain. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 62(2 Pt B):2597–2607.
9. Rankin SK, Large EW, Fink PW (2009) Fractal tempo fluctuation and pulse prediction. *Music Percept* 26:401–413.
10. Voss R, Clarke J (1975) 1/f noise in music and speech. *Nature* 258:317–318.
11. Voss R, Clarke J (1978) 1/f noise in music: Music from 1/f noise. *J Acoust Soc Am* 63:258–263.
12. Levitin DJ, Chordia P, Menon V (2012) Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proc Natl Acad Sci USA* 109(10):3716–3720.
13. Néda Z, Ravasz E, Brechet Y, Vicsek T, Barabási AL (2000) The sound of many hands clapping. *Nature* 403(6772):849–850.
14. Podobnik B, Horvatic D, Petersen AM, Stanley HE (2009) Cross-correlations between volume change and price change. *Proc Natl Acad Sci USA* 106(52):22079–22084.
15. Podobnik B, Wang D, Horvatić D, Grosse I, Stanley HE (2010) Time-lag cross-correlations in collective phenomena. *Europhys Lett* 90:68001.
16. Preis T, Schneider JJ, Stanley HE (2011) Switching processes in financial markets. *Proc Natl Acad Sci USA* 108(19):7674–7678.
17. Podobnik B, Stanley HE (2008) Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys Rev Lett* 100(8):084102.
18. Karlis AK, Diakonos FK, Petri C, Schmelcher P (2012) Criticality and strong intermittency in the Lorentz channel. *Phys Rev Lett* 109(11):110601.
19. Herbert-Read JE, et al. (2011) Inferring the rules of interaction of shoaling fish. *Proc Natl Acad Sci USA* 108(46):18726–18731.
20. Cavagna A, et al. (2010) Scale-free correlations in starling flocks. *Proc Natl Acad Sci USA* 107(26):11865–11870.
21. Bialek W, et al. (2012) Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci USA* 109(13):4786–4791.
22. Gallup AC, et al. (2012) Visual attention and the acquisition of information in human crowds. *Proc Natl Acad Sci USA* 109(19):7245–7250.
23. Moussaïd M, Helbing D, Theraulaz G (2011) How simple rules determine pedestrian behavior and crowd disasters. *Proc Natl Acad Sci USA* 108(17):6884–6888.
24. Zhao L, et al. (2011) Herd behavior in a complex adaptive system. *Proc Natl Acad Sci USA* 108(37):15058–15063.
25. Kawasaki M, Yamada Y, Ushiku Y, Miyauchi E, Yamaguchi Y (2013) Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci Rep* 3:1692.
26. Dumas G, Nadel J, Soussignan R, Martinerie J, Garnero L (2010) Inter-brain synchronization during social interaction. *PLoS ONE* 5(8):e12166.
27. Lindenberger U, Li SC, Gruber W, Müller V (2009) Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neurosci* 10:22.
28. Sänger J, Müller V, Lindenberger U (2012) Intra- and interbrain synchronization and network properties when playing guitar in duets. *Front Hum Neurosci* 6:312.
29. Dolan D, Sloboda J, Jensen HJ, Crüts B, Feygelson E (2013) The improvisatory approach to classical music performance: An empirical investigation into its characteristics and impact. *Music Performance Research* 6:1–38.
30. Buhusi CV, Meck WH (2005) What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* 6(10):755–765.
31. Buonomano DV, Karmarkar UR (2002) How do we tell time? *Neuroscientist* 8(1):42–51.
32. Ruiz MH, Strübing F, Jabusch HC, Altenmüller E (2011) EEG oscillatory patterns are associated with error prediction during music performance and are altered in musician's dystonia. *Neuroimage* 55(4):1791–1803.
33. Walker FOF (2007) Huntington's disease. *Lancet* 369(9557):218–228.
34. Hui E, Chui BT, Woo J (2009) Effects of dance on physical and psychological well-being in older persons. *Arch Gerontol Geriatr* 49(1):e45–e50.
35. Podobnik B, et al. (2009) Quantifying cross-correlations using local and global detrending approaches. *Eur Phys J B* 71:243–250.
36. Peng CK, Havlin S, Stanley HE, Goldberger AL (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5(1):82–87.
37. Kantelhardt J, Koscielny-Bunde E, Rego H, Havlin S, Bunde A (2001) Detecting long-range correlations with detrended fluctuation analysis. *Physica A* 295:441–454.
38. Pilgram B, Kaplan D (1998) A comparison of estimators for 1/f noise. *Physica D* 114:108.
39. Torre K, Wagenmakers EJ (2009) Theories and models for 1/f(beta) noise in human movement science. *Hum Mov Sci* 28(3):297–318.

Supporting Information

Hennig 10.1073/pnas.1324142111

SI Text

Experimental Setup. In total, 57 Musical Instrument Digital Interface (MIDI) recordings of subjects playing musical rhythms in synchrony were made. A MIDI file can be seen as a text file containing a set of instructions such as choice of the instrument, loudness, pitch, and start/end times of the beats. Compared with acoustic recordings, the use of MIDI simplifies data postprocessing, because it eliminates the "beat finder" as an additional step and source of errors. Thirteen subjects took part in the study (eight male, five female) with average age 32 ± 11 y. The experiments were conducted at the Harvard University Studio for Electroacoustic Composition on a Studiologic SL 880 keyboard with baud rate 19.2 kHz, i.e., a temporal resolution of 0.05 ms. The musical expertise was determined through a self-assessment on a scale from 1 (absolute layperson) to 6 (professional musician), with average 3.3 ± 2.0, covering the whole range from professional musicians to laypersons. The subjects sat in an armchair, and the forearms could be rested on the armrests of the chairs. To exclude visual information, subjects were asked to keep their eyes closed during the experiments. Playing as synchronously as possible was stressed for the synchronization tasks. The subjects were asked to repeatedly play a simple rhythm with their index finger on the keyboard in a comfortable tempo that the subjects were free to choose. The average interbeat interval in the entire data set is 454.9 s with SD 87.9 s.

Average detrended fluctuation analysis (DFA) exponents $\overline{\alpha}$ of the interbeat intervals did not significantly vary across the three different tasks. We found $\overline{\alpha} = 0.84 \pm 0.19$ (SD) for two subjects synchronizing their beats (tasks Ia and Ib) and $\overline{\alpha} = 0.78 \pm 0.13$ (SD) for one subject synchronizing a finger of both hands (task II), and $\overline{\alpha} = 0.81 \pm 0.12$ (SD) for finger-tapping (task III).

Leader–Follower: Structure of Temporal Deviations Between Musicians. Will the detrended cross-correlation analysis (DCCA) and DFA exponents differ, when one subject is leading and the other subject is following, compared with the cases where no obvious leader–follower behavior is observed? The average deviation between two subjects synchronizing their beats is used an indicator, whether one subject followed the other. We assume that no obvious leader–follower relation is observed when the average deviation is small ($-10 < \overline{d_n} < 10$ ms).

Fig. S1 shows the resulting three cases where subject B followed A ($\overline{d_n} < -10$ ms), subject A followed B ($\overline{d_n} > 10$ ms), and the case where no obvious leader–follower relation is observed ($-10 < \overline{d_n} < 10$ ms). We find that the DFA exponent of the time series of deviations $\alpha(d_n)$ is significantly lower when no lead is taken (two-sample t test, $P = 0.005$), whereas the DCCA exponent δ showed no dependence on the leader–follower behavior (two-sample t test, $P = 0.47$). However, the present data set is rather heterogeneous and it would be very interesting to thoroughly test these hypotheses in a systematic study, e.g., where one musician is instructed to follow the other compared with where musicians avoid following each other.

DCCA Analysis for Uncorrelated Time Series. The DCCA method is used to quantify the degree of long-range cross-correlations (LRCCs) between two time series. What does DCCA return for pairs of uncorrelated time series? Here, we show a representative example for a DCCA analysis for two uncorrelated time series but where each time series has long-range correlations itself. We analyze the cross-correlations of the two time series consisting of Gaussian $1/f^\beta$ noise with $\beta = 0.9$ that were used to create audio

example no. 2, i.e., where two audio tracks are humanized separately. In this case, we expect that LRCCs are absent, which we test in the following with the DCCA method. There are two possible indicators of absence of LRCCs: (*i*) The DCCA exponent is equal to 0.5 (within margins of errors) or (*ii*) the fluctuation function $F(s)$ oscillates around zero (1).

Fig. S2 *Upper* shows the fluctuation function ($F(s)$) for the two uncorrelated time series in Fig. S2 *Lower*, each consisting of $1/f^\beta$ noise with $\beta = 0.9$. The fluctuation function $F(s)$ fluctuates around zero on a wide range of scales s, thus correctly finding that the two time series are not long-range cross-correlated. Note that because $F(s) < 0$ for many s, the fluctuation function is shown on a lin-log scale, whereas it is commonly plotted on a log-log scale. This is a representative example obtained when analyzing two $1/f^\beta$ noise time series that do not exhibit LRCC. We conclude that DCCA (with prior global detrending; see *Materials and Methods* in main text) is capable of correctly discerning these cases from long-range cross-correlated time series.

Power Spectral Density With vs. Without Metronome. In this article, subject A was synchronizing rhythms with subject B. What is the difference when subject A is synchronizing beats with a metronome, i.e., when subject B is replaced by a metronome? Interestingly, there is a fundamental difference between settings where individuals are provided with a metronome click (e.g., over headphones) while playing and where no metronome is present (also referred to as self-paced play) that manifests in the power spectral density (PSD) of the interbeat intervals. The following short overview aims to qualitatively and quantitatively point out the differences in the PSD exponents between the settings where the metronome is on and off (a comparison of tapping with vs. without metronome can also be found in ref. 2). All frequencies are measured in units of the Nyquist frequency $f_{\text{Nyquist}} = 1/2$ Hz, which is half the sampling rate of the time series.

For self-paced play, the PSD of the interbeat intervals exhibits two distinct regions that may resemble the shape of a hockey stick (3). Long-range correlations are found asymptotically for small frequencies in the PSD. This region relates to correlations over long timescales of up to several minutes (as long as the subject does not frequently lose rhythm). However, for high frequencies in the PSD, anticorrelations are found. Gilden et al. (3) presented a model where the origin of the anticorrelations is attributed to the fact that if one interbeat interval is smaller than average, the following interval is likely larger than average if the subject keeps an overall constant tempo. An example of the PSD of a time series of interbeat intervals of length 10^5 generated with the Gilden model is shown in Fig. S3*A*. The vertex of the PSD at a characteristic frequency $f_c = 0.1 f_{\text{Nyquist}}$ separates the two regions.

In contrast, a different situation is observed in presence of a metronome: for play of both complex musical rhythms (4) and finger-tapping (5), long-range correlations were found in the time series of deviations of the beats from the metronome clicks. The deviations from the metronome clicks are defined as $d_n = t_n - M_n$, where t_n is the temporal occurrence (e.g., the onset) of the n'th beat, $M_n = nT$ is the temporal occurrence of the n'th metronome click, and T is the time period between two consecutive metronome clicks. The interbeat intervals read

$$I_n = t_n - t_{n-1} = d_n - d_{n-1} + T. \qquad \text{[S1]}$$

Hence, the interbeat intervals are the discrete derivative of the deviations (except for a constant). Given a time series x_n where

the PSD asymptotically decays in a power law with exponent β, let the time series $dx_n = x_n - x_{n-1}$ denote the discrete derivative of x_n. Then the PSD of the derivative time series dx_n asymptotically follows a power law with exponent $\beta - 2$. Note that when x_n exhibits long-range correlations with exponent $0 < \beta(x_n) < 2$, the derivative dx_n exhibits long-range anticorrelations with $-2 < \beta(dx_n) < 0$. To illustrate this case, a time series x_n of length 10^7 consisting of $1/f^\beta$ noise with $\beta = 1.0$ was generated. In Fig. S3*B* the PSD of the derivative dx_n follows a power law $1/f^\beta$ with measured exponent $\beta(dx_n) = -1.00$. For short time series $n < 1,000$, a finite size effect is observed where the anticorrelations in dx_n are still observed, but can be significantly reduced.

When subjects are synchronizing beats with a metronome, the time-series deviations exhibit long-range correlations with PSD exponents reported in the range $\beta(d_n) = [0.2, 1.3]$ (4). Hence, we expect the PSD exponents for the time series of interbeat intervals in the range $\beta(I_n) = \beta(d_n) - 2 = [-1.8, -0.7]$. Thus, the interbeat intervals are long-range anticorrelated for settings where a metronome is present. In contrast, for self-paced play as in the present study (i.e., in absence of a metronome), the interbeat intervals are long-range correlated on timescales of up to several minutes.

Characteristic Frequency. The numerical value of the characteristic frequency f_c (separating short-range from long-range correlations in the interbeat intervals; Fig. S3*A*) depends on how the frequencies in the PSD are normalized. In Gilden et al. (3), the frequencies are given by $f_k = (k - 1)/(2mT_0)$, where $k = 1 \ldots m$ and T_0 is the target duration, i.e., the average time between consecutive beats. The characteristic frequency is proportional to the inverse of T_0, which can also be seen in figure 1B of ref. 3, where f_c increases with $1/T_0$. In this case, the inverse of f_c indicates, how much time (in real time) passes until the anticorrelations between neighboring beats have decayed; the latter depends on how fast a rhythm is played. In contrast, in this article we set $T_0 \equiv 1$, such that f_c remains the same for experiments with different tempi and the power spectral densities can be directly compared. When setting $T_0 = 1$, then f_c is a constant and is related to the number of beats after which the anticorrelations between neighboring beats have decayed. For $m \to \infty$, the maximum frequency in ref. 3 reads $1/(2T_0)$. Thus, when setting $T_0 = 1$, the maximum frequency in the PSD is 1/2 Hz, i.e., the Nyquist frequency.

1. Podobnik B, et al. (2009) Quantifying cross-correlations using local and global detrending approaches. *Eur Phys J B* 71(2):243–250.
2. Torre K, Wagenmakers EJ (2009) Theories and models for 1/f(beta) noise in human movement science. *Hum Mov Sci* 28(3):297–318.
3. Gilden DL, Thornton T, Mallon MW (1995) 1/f noise in human cognition. *Science* 267(5205): 1837–1839.
4. Hennig H, et al. (2011) The nature and perception of fluctuations in human musical rhythms. *PLoS ONE* 6(10):e26457.
5. Repp BH, Su YH (2013) Sensorimotor synchronization: A review of recent research (2006-2012). *Psychon Bull Rev* 20(3):403–452.
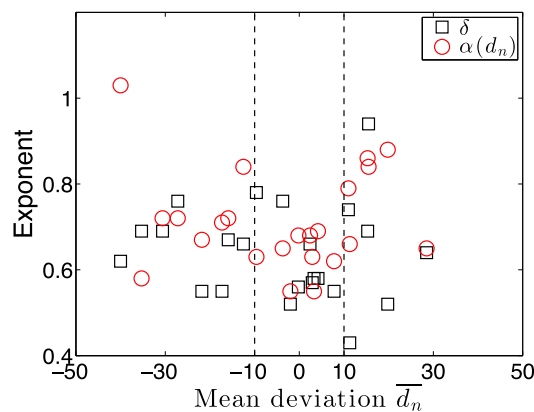
Fig. S1. When the average deviation $\overline{d_n}$ between the beats of two subjects is small ($\overline{d_n} < 10$ ms), the DFA exponent $\alpha(d_n)$ is significantly lower (two-sample t test, $P = 0.005$), and the data are consistent with the hypothesis that the DCCA exponent δ is independent from $\overline{d_n}$ (two-sample t test, $P = 0.47$). The vertical lines at ± 10 ms are a guide to the eye separating large from small average deviations.
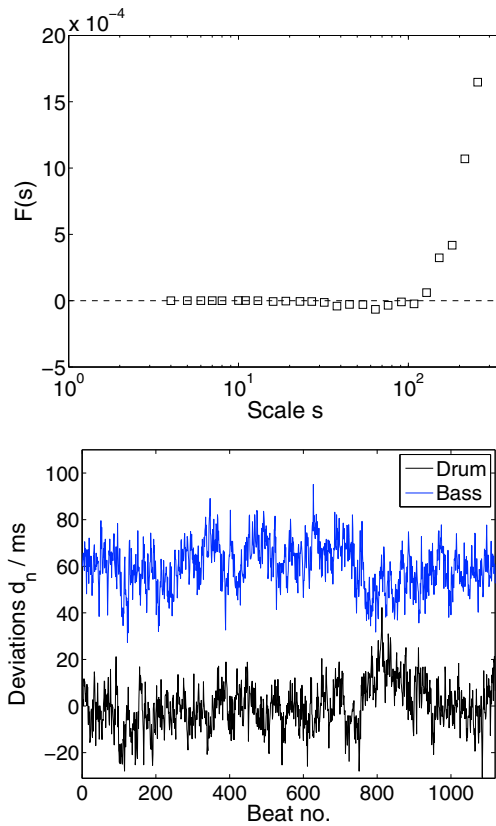
Fig. S2. (*Lower*) Shown are the two $1/f^{\beta}$ noise time series (with $\beta = 0.9$) that were used to separately humanize the two sets of tracks (referred to as drum and bass tracks) in the audio example. The blue curve is offset by 60 ms to visually separate the two curves. The time series are independent from each other, i.e., there are no LRCCs. (*Upper*) The fluctuation function $F(s)$ obtained from the DCCA analysis fluctuates around zero, indicating absence of LRCCs.
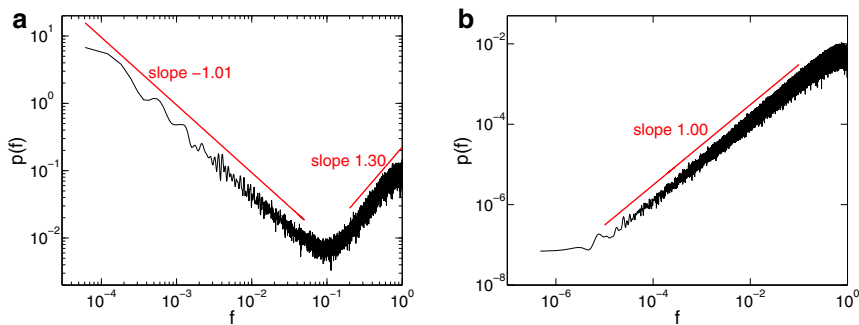
Fig. S3. Illustration of the PSD of the interbeat intervals when humans are playing or synchronizing rhythms (*A*) without and (*B*) with a metronome. (*A*) Illustration of the case where rhythms are played in absence of a metronome: The PSD of the interbeat intervals exhibits long-range correlations (asymptotically for low frequencies with PSD exponent $\beta = 1.01$) and anticorrelations for high frequencies. The characteristic frequency separating the two regions is observed at $0.1f_{\text{Nyquist}}$. The time series of interbeat intervals was calculated with the Gilden model for $\beta = 1.0$ and relative strength of clock noise over motor noise $\sigma = 0.5$, i.e., for rather dominant motor noise (which only manifests on short timescales, but does not affect the long-term behavior) (3). (*B*) Illustration of the case where rhythms are played while synchronizing beats with a metronome: The PSD of the interbeat intervals dx_n exhibits long-range anticorrelations. The PSD was calculated with the global PSD (gPSD) method (with $k = 0$ for detrending and $m = 8$ for averaging; see the description of gPSD in *Materials and Methods* for details).

Audio File S1. The audio example 1 demonstrates musical interaction. The instrumental version of the song "Billie Jean" was modified such that it includes musical coupling between drum and bass tracks, thus imitating the generic interaction of a drum and a bass player. More specifically, the interbeat intervals of drum and bass tracks contain long-range cross-correlations (SD of the deviations added to the interbeat intervals is 10 ms, DCCA exponent is $\delta = 0.9$). See *Application: Musical Coupling* in the main text for details.

[Audio File S1](#)

Audio File S2. The audio example 2 demonstrates humanized music, where there is no interaction between the tracks. The instrumental version of the song "Billie Jean" was modified such that the interbeat intervals of drum and bass tracks contain $1/f^{\beta}$ noise with $\beta = 0.9$. However, drum and bass tracks are statistically independent and do not interact (SD of the deviations added to the interbeat intervals is 10 ms). See *Application: Musical Coupling* in the main text for details.

[Audio File S2](#)

Audio File S3. The audio example 3 demonstrates randomized music using white noise. The instrumental version of the song "Billie Jean" was modified by adding uncorrelated deviations (Gaussian white noise) to the interbeat intervals of drum and bass tracks, as currently implemented in professional audio software. A jerking sounding rhythm reflects the lack of correlations in the deviations (SD of the deviations added to the interbeat intervals is 10 ms). See *Application: Musical Coupling* in the main text for details.

[Audio File S3](#)