OXFORD

Tutorial/Introduction

Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations

Andrea Ravignani[1,2,*,#] and Philipp Norton[3,#]

[1]Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium, [2]Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525, The Netherlands and [3]Department of Animal Behavior, Freie Universität Berlin, Takustraße 6, 14195 Berlin, Germany

*Corresponding author: andrea.ravignani@gmail.com
#The authors contributed equally to this study.

Abstract

Research on the evolution of human speech and phonology benefits from the comparative approach: structural, spectral, and temporal features can be extracted and compared across species in an attempt to reconstruct the evolutionary history of human speech. Here we focus on analytical tools to measure and compare temporal structure in human speech and animal vocalizations. We introduce the reader to a range of statistical methods usable, on the one hand, to quantify rhythmic complexity in single vocalizations, and on the other hand, to compare rhythmic structure between multiple vocalizations. These methods include: time series analysis, distributional measures, variability metrics, Fourier transform, auto- and cross-correlation, phase portraits, and circular statistics. Using computer-generated data, we apply a range of techniques, walking the reader through the necessary software and its functions. We describe which techniques are most appropriate to test particular hypotheses on rhythmic structure, and provide possible interpretations of the tests. These techniques can be equally well applied to find rhythmic structure in gesture, movement, and any other behavior developing over time, when the research focus lies on its temporal structure. This introduction to quantitative techniques for rhythm and timing analysis will hopefully spur additional comparative research, and will produce comparable results across all disciplines working on the evolution of speech, ultimately advancing the field.

Key words: evolution of speech; animal communication; vocal learning; timing; rhythm.

1. Introduction

Research on the evolution of speech has greatly benefitted from the comparative approach (Fitch 2000; Yip 2006; Rauschecker and Scott 2009; Fedurek and Slocombe 2011; Scharff and Petri 2011; de Boer 2012; Rauschecker 2012; Ghazanfar 2013; Lameira et al. 2014; Lameira et al. 2015; Fehér 2016; Filippi 2016; Gustison and Bergman, manuscript under review). In fact, relating human speech and animal vocal production learning can inform the evolution of speech both by comparing the signals produced across species, and by inverse inference, comparing the neurocognitive machinery used to generate the signals (Bolhuis et al.

2010; Fitch 2014; Bowling and Fitch 2015, and Collier et al. 2014).

Three main avenues seem particularly relevant to the comparative study of the evolution of speech (Fig. 1).

Positional regularities: how the building blocks of speech or language, each taken holistically, are organized and related to each other (Fitch 2014; Kershenbaum et al. 2014; ten Cate 2016).

Spectral characteristics: how different frequencies and their intensities relate to one another, giving rise to tone in speech, liaison, vowel quality, harmonicity, etc. (Fitch 2000; Yip 2006; Spierings and ten Cate 2016).
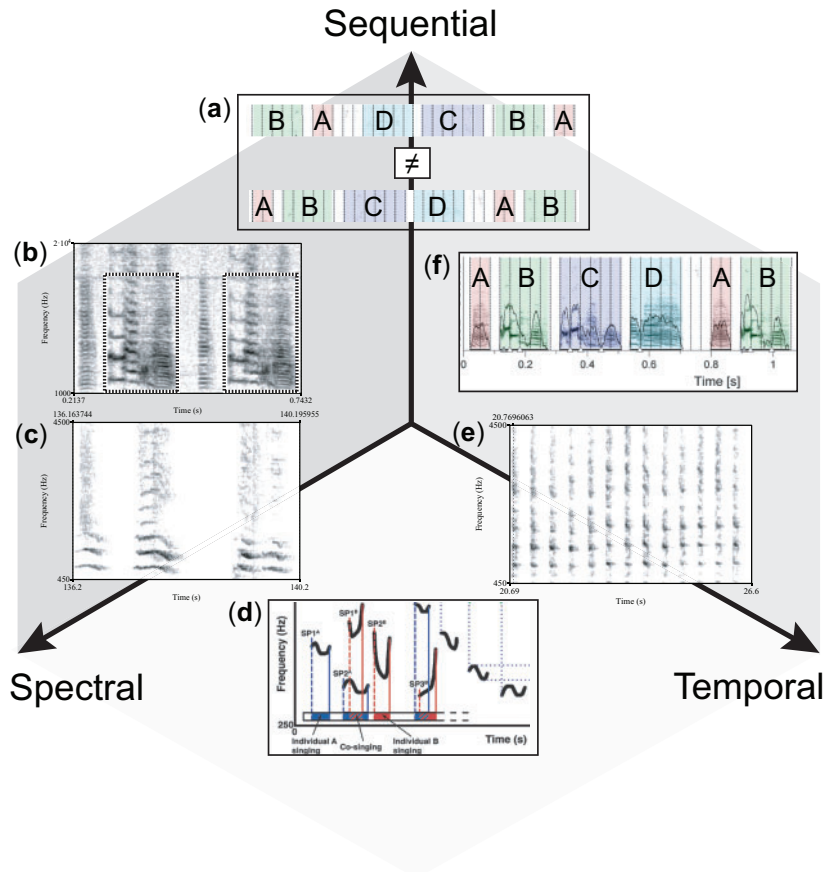

Figure 1. Three approaches to the analysis of animal vocalizations, including human speech, laying emphasis on sequential, spectral, or temporal features of the signal. The three-dimensional space presents a common multicomponent approach to speech and animal vocalizations, showing that these components constitute complementary dimensions, rather than a mutually exclusive classification. Depending on the species and scientific question, one-dimension might be particularly relevant. Research focus on structural, spectral, and temporal information is exemplified along the three main axes, with many cases of vocalizations falling between two categories. Counter-clockwise, the figure shows structures where a possible research emphasis is laid on: (a) *Sequential:* an abstract representation of a series of song notes, and its mirror permutation, containing the same elements and having the same duration but different sequential properties; (b) *Spectro-sequential:* spectrogram of a chunk of zebra finch song (*Taeniopygia guttata*), where complex spectral information is combined into a sequential repetitive structure (dashed boxes); (c) *Spectral:* spectrogram of a harbor seal pup call (*Phoca vitulina*), showing harmonic features, with less emphasis on structural/positional regularities; (d) *Spectro-temporal:* schematic representation of a spectrogram showing a singing lemur vocal duet (*Indri indri*), where spectral and temporal information might be interacting (Gamba et al. 2016); (e) *Temporal:* spectrogram of a California sea lion (*Zalophus californianus*)'s bark, showing a clear rhythmic, isochronous pattern: barks contain little spectral information and variability, but are produced with remarkable regularity, like a metronome; (f) *Tempo-sequential:* notes composing a zebra finch song, which in turn possesses an underlying rhythmic structure (Norton and Scharff 2016). Notice that these examples concern the research focus, rather than the nature of the signal: for instance, panel (e) is taken as a prototype of temporal signal, but it also has a clear spectral structure. This article focuses on useful methods to analyze cases (d), (e), and (f), that is, when emphasis is put on analysing the temporal structure of vocalizations. Panel (d) reproduced and modified from Gamba et al. (2016) and panels (a) and (f) from Norton and Scharff (2016), both published open access under the CC BY 4.0. Other panels generated using Praat (Boersma and Weenink, 2013).

Table 1. Definition of crucial concepts in comparative rhythm research

Name (abbreviation)	Definition
Rhythm	Structure over time; durations marked by (acoustic) events (McAuley 2010)
Isochronous	A series of events repeating at a constant rate
Synchronous	A series of events individually occurring at the same time as events from another series
Time series	Sequence of events occurring over time, sampled at regular time intervals
IOI	Time elapsed between the beginning of one event (i.e. onset) and the beginning of the next event
Meter	Hierarchical organization of timed events based on spectral properties (stress, e.g. loudness alternation, pitch variation)
Syllable-timed	Language in which all syllables (both accented and unaccented) are roughly isochronous
Stress-timed	Language with an isochronous occurrence of stressed syllables
Mora-timed	Language in which all moras (syllables and some combinations thereof) are isochronous
Pulse	Isochronous grid most suitable to a given temporal pattern
Iambic	Metrical form alternating a weak (unstressed) syllable with a strong (stressed) syllable
Trochaic	Metrical form alternating a strong syllable with a weak syllable

Temporal structure: how the speech/vocal signal is organized and develops over time (Ramus et al. 1999; Goswami and Leong 2013; Bekius et al. 2016; Benichov et al. 2016; Hannon et al. 2016; Jadoul et al. 2016). Timing, together with some spectral characteristics, contributes to rhythm (see Table 1 for a definition of rhythm and other major concepts presented in this article).

Rhythmic properties of human speech (Ramus, et al. 1999; Grabe and Low 2002; Tilsen 2009; Lehiste 1977) and animal vocal communication (Saar and Mitra 2008; Norton and Scharff 2016; Ravignani et al. 2016) are often investigated. However, lack of an interdisciplinary statistical approach to rhythmicity of vocalizations across species hinders comparative and comparable research.

Here we introduce the reader to a broad range of quantitative methods useful for rhythm research within and across species. The article is structured as follows. A first distinction between methods concerns measuring either rhythmic *complexity of one pattern*, that is, how regular and predictable a sequence of events is (Section 2), or *rhythmic relationships* between multiple patterns, that is, similarities in the temporal structure of two or more sequences of events (Section 3).

In Section 2, we characterize inter-onset intervals (IOIs), used to measure temporal information, and show how rhythmicity can be quantified using: distributional measures, such as histograms (subsection 2.2), Kolmogorov–Smirnov D (subsection 2.3), and normalized pairwise variability index (nPVI)(subsection 2.4); autocorrelation (subsection 2.5); beat-finding algorithms, such as Fourier transform (subsection 2.6) and pulse generate-and-test (GAT)(subsection 2.7); time

series analysis (subsection 2.10). We also present two powerful techniques to visualize regularities in rhythmic patterns: phase portraits (subsection 2.8) and recurrence plots (subsection 2.9).

Section 3 extends Section 2 by focusing on measures of temporal similarity between two patterns, namely: cross-correlation (subsection 3.1); multidimensional time series analysis, testing how one pattern can be linearly estimated (subsection 3.2) or causally predicted (subsection 3.3) from the other; statistical analysis of circular data, including visualization with rose plots (subsection 3.4).

Section 4 describes some research areas where the techniques we present may be particularly useful. Section 5 provides some overall conclusions. Section 6 describes how to access the data and scripts used in this article. To get a better understanding of the methods presented in this article, we encourage the reader to use the provided code (Supplementary data) to experiment with different patterns. New versions of the code will be uploaded at http://userpage.fu-berlin.de/phno/mrc/.

2. Within-pattern analytical techniques: quantifying rhythmicity

2.1 Overview
The temporal structure of vocalizations can be analyzed on the raw audio signal, or more commonly on time series extracted from the signal, like the amplitude envelope. Often however, one is interested in the timing of certain discrete events, for example, syllable onsets, amplitude peaks, or any other behavior developing over time. Some of the methods described here apply to
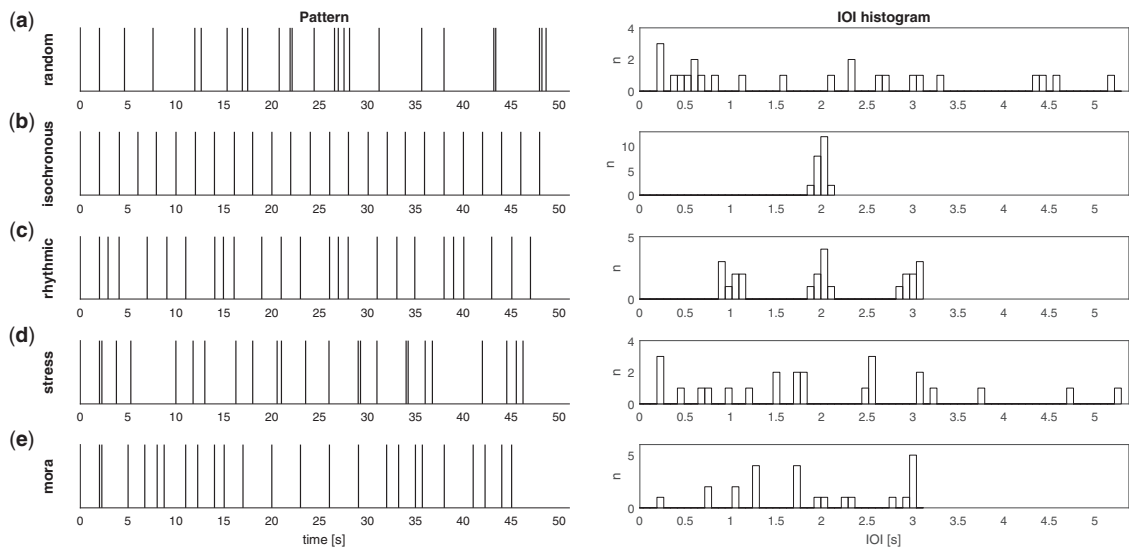
Figure 2. Five different example patterns that differ in their rhythmic structure (a–e, left column), as well as histograms of their inter-onset intervals (IOI, right column). (a) *Random sequence*: The time points of events of the first sequence are pseudorandomly drawn from a uniform distribution. (b) *Isochronous sequence*: The second sequence is isochronous, that is, IOIs of successive events are equal for all events (in this case 2 s). This sequence with added noise is taken to loosely represent the hypothesized syllable-timed structure of some languages (Grabe and Low 2002). (c) *Rhythmic sequence*: The third pattern contains four repetitions of a sub-pattern consisting of six events (IOIs: 1, 1, 3, 2, 2 and 3 s). (d–e) The final two sequences are loosely based on concepts of timing division in language: stress timing and mora timing. (d) The *stress sequence* consists of clusters of four events each. The duration of all clusters is 8 s and the pattern of IOIs within a cluster are different each time (e.g. [0.5,1.5,3.75,2.25], [2,1,3.5,1.5], …). (e) The *mora sequence* contains clusters of 3 s duration each, that either contain one or two events (e.g. [1,2], [3], [2.25,0.75], …). Gaussian noise with a SD of 0.04 s (isochronous and rhythmic sequence) or 0.02 s (stress and mora sequence) was added to the timing of the events.

patterns represented by their IOIs, while others—like autocorrelation and Fourier transform—are more readily applied to a time series, that is, a sequence of values sampled at equally spaced points in time (see Supplementary data for details).

To demonstrate the methods described in this section, we generated five sequences that differ in their rhythmic structure, each consisting of twenty-four events (Fig. 2, left column). To simplify notation, we demonstrate techniques for events with IOIs in the range of 1–6 s, though all presented techniques can be used at any time scale, depending on the observed behavior.

2.2 IOI distribution: which intervals occur in the pattern, and how often?

A histogram of all IOIs in a sequence provides a first visualization of its rhythmic structure (Fig. 2, right column). A normal distribution suggests a roughly isochronous pattern, where a lower spread indicates a more isochronous pattern (Fig. 2b). The existence of a few distinct IOI categories may appear as a multimodal distribution (Fig. 2c). Uniformly distributed IOIs might hint at a lack of such categories or a rhythmic structure in

general (Fig. 2a). However, when many durational categories exist, a uniform distribution might also appear, concealing higher order structure (Fig. 2d). A particular limitation of histograms and statistical measures of durational distributions is their lack of sensitivity to structure: for instance, any permutation of the order of IOIs will change the structure while leaving distributional measures unaltered.

2.3 Kolmogorov–Smirnov D: does interval distribution differ from any hypothesized distribution?

The one-sample Kolmogorov–Smirnov test (K-S test, Table 2) can be used to evaluate how different an observed IOI distribution is from a particular, specified distribution, for example, a normal distribution in the isochronous case, or a uniform distribution in the case of no actual structure (Lilliefors 1967). The test uses the statistic D, measuring the maximum distance between two cumulative distribution functions. This measure can be used to compare, for example, different languages based on the normality of the intervals between syllable nuclei (Jadoul et al. 2016). When compared to a normal distribution, the isochronous pattern in Fig. 2b has a D

of 0.085, about half of the 0.14 of the mora pattern (Fig. 2e). In order to statistically test an IOI distribution for normality, a modification of the K-S test by Lilliefors can be used instead (Lilliefors 1967). Of the five example patterns in Fig. 2, only the random and rhythmic patterns are significantly non-normal according to the Lilliefors test (P = 0.035 and 0.024, respectively; $n = 24$).

2.4 The nPVI: how variable are adjacent intervals?

The nPVI is a measure originally developed to quantify temporal variability in speech rhythm (Table 2; Grabe and Low 2002; Toussaint 2013). It is computed as follows. For each *pair of adjacent IOIs*, one calculates their difference and divides it by their average. The nPVI equals the average of all these ratios, multiplied by 100, namely:

$$nPVI = \frac{100}{n-1}\left[\left|\frac{IOI_2 - IOI_1}{\frac{IOI_2 + IOI_1}{2}}\right| + \ldots + \left|\frac{IOI_n - IOI_{n-1}}{\frac{IOI_n + IOI_{n-1}}{2}}\right|\right]$$

(1)

The nPVI equals zero for a perfectly isochronous sequence, and it increases with an increasing *alternation* in onset timing. Some have suggested nPVI captures cross-linguistic rhythmic classes, with syllable-timed languages exhibiting low nPVI, stress-timed languages having high nPVI, and mora-timed languages lying between the two extremes (Grabe and Low 2002). The nPVI values of our computer-generated example patterns align with this classification: the isochronous pattern, representing syllable-timing, has a low nPVI of 3.3, the stress pattern a high value of 92.74 (close to the random pattern with 94.54), while the mora pattern lies in between with 52.84. Notice however that nPVI has little explanatory power beyond simple, zeroth-order distributional statistics, such as Kolmogorov–Smirnov D (Jadoul et al. 2016; Ravignani, in press).

2.5 Autocorrelation: how similar is a pattern to itself at different points in time?

Autocorrelation is a technique that can help reveal higher order structure within a pattern, specifically repeating temporal sub-patterns (Hamilton 1994; Ravignani and Sonnweber 2015). Autocorrelation correlates a time series (Fig. 3b) with a copy of itself at different time lags (Fig. 3c–f, red dotted line). It is calculated by having one copy of the signal slide stepwise across the other; at each step the products of all points in the two signals are calculated and added together. The result of the process is a function of these

sums at the different time lags between the two signals (Fig. 3h). The autocorrelation starts at a time lag of zero, where the two copies overlap perfectly (Fig. 3c). At zero lag, the autocorrelation function is normalized to 1 (Fig. 3h). Since the signal is nonzero only at the time points of the events, the function will be zero at the lags at which none of the events overlap (Fig. 3d,h). When one or more events in the two signals overlap a peak appears in the function (Fig. 3e,h). The more the events overlap at a certain lag, the higher the peak in the function (Fig. 3f,h), suggesting that a sub-pattern might repeat after a duration that equals this lag. Note that unless the IOIs are exactly equal in duration, the events in the raw time series will not overlap. Real-world patterns, however, are rarely isochronous and any rhythmic structure usually contains some amount of noise or jitter. To deal with noise one can convolute the time series with a narrow normal probability density function (npdf) prior to autocorrelation, effectively replacing every single time point event with narrow Gauss curves (Fig. 3a,b). Several nearby, though not simultaneously overlapping, events can thereby partially overlap over a range of lags and together contribute to a single peak in the autocorrelation function (e.g. Fig. 3e).

Figure 2b shows an isochronous pattern (IOI ≈ 2 s) of twenty-four events with a small amount of Gaussian noise added. At a lag of 2 s, a peak close to 1 appears in the autocorrelation function, as events 2–24 of the signal overlap with events 1–23 of the lagged signal (Fig. 3g). The height of the following peaks—regularly occurring at n × 2 s—gradually decreases as the lagged signal moves beyond the original signal and fewer events overlap.

The 'rhythmic sequence' contains three repetitions of a random sub-pattern consisting of six events with IOIs equaling 1, 2, or 3 s (total duration = 12 s; Fig. 2c). Looking at the autocorrelation plot of this pattern, the first peak appears at a lag of 1 s, where some of the events start to overlap, and all following peaks are at lags 2,3,4,...s (Fig. 3h). This suggests 1 s is the basic time unit of this pattern, that is, all IOIs are multiples of 1 s. The largest peak is located at 12 s lag, because 12 s is the total duration of the repeating sub-pattern. At this point all events of sub-patterns 2, 3 and 4 of the signal overlap with sub-patterns 1, 2 and 3 of the lagged signal. Note that a higher peak could potentially occur at a lower lag, if more than three-quarters of events were to overlap. Relatively high peaks are likely to appear at harmonics (i.e. multiples) of the 12 s lag, namely at 24 and 36 s where the following repetitions of the sub-pattern overlap. In the 'mora sequence' events occur every 3 s as well as at varying time points in between

Table 2. Analytical techniques to unveil temporal structure in one pattern, including their specific function

Technique (References)	Function	Advantage	Disadvantage
Kolmogorov–Smirnov D (Jadoul et al. 2016)	One number describing the distribution of a sequence of IOI	Common and well-studied statistical tool	Not an absolute number, instead the relative distance from a specific hypothesized distribution
nPVI (Grabe and Low 2002; Jadoul et al. 2016)	One number describing temporal variability of a sequence of IOI, taking only information about adjacent intervals into account	Summarizes temporal regularity with one number	Not very robust to different speakers and replications
Autocorrelation (Ravignani and Sonnweber 2015; Hamilton 1994)	Probes the existence of repeating sub-patterns within a pattern	Few assumptions required	Provides only little information on the pattern (difficult to map autocorrelations to necessary/ sufficient conditions on pattern structure)
Fourier analysis (Saar and Mitra 2008; Norton and Scharff 2016)	Decomposes signal into sum of isochronous pulses	Common, fast method in signal analysis	May return several best-fitting pulses
GAT pulse matching (Norton and Scharff 2016)	Finds a metronomic pulse best fitting the pattern	Aimed and proficient at finding the slowest pulse in the pattern	Computationally intensive; less used than Fourier analysis
Phase portrait (Rothenberg et al. 2014; Wagner 2007; Ravignani, in press)	Enables temporal regularities to be visualized as geometric regularities	Easy to plot, straightforward interpretation	Not an analytical, rather a visualization technique
Recurrence plots (Thiel et al. 2004; Coco and Dale 2014)	Visualizes temporal regularities among nonadjacent elements	Provides a quick glance at higher order regularities	Sensitive to the initial parameters (threshold for considering two IOIs similar), which can produce false positives/ negatives
ARMA(Hamilton 1994; Jadoul et al.2016)	Tests whether each IOI can be expressed as a linear combination of previous IOI	Most common method to analyze time series; based on much theoretical work; many statistical packages available; using Akaike sets increases robustness	Only captures linear relations

(Fig. 2e). Accordingly, the autocorrelation function shows peaks at 3, 6 s, etc., as well as some noise stemming from less-regularly occurring events (Fig. 3i).

2.6 Fourier transform: decomposing the pattern into a sum of regular, clock-like oscillators

The Fourier transform is used to express any signal as a sum of sine waves of different frequencies and phases. In acoustics, one application is to decompose a sound wave into its constituent frequencies. When applied to the time series of a rhythmic pattern, the Fourier transform finds periodicities in the timing of the events. A particularly important periodicity is the pulse: the slowest isochronous sequence to which most events align. The Fourier transform can be visualized in a power spectrum, where the *x*-axis denotes the frequencies of the waves and the *y*-axis their magnitude, that is, how much they contribute to the signal (Fig. 4).

In the 'isochronous sequence', the first and largest peak in the power spectrum is located at 0.5 Hz corresponding to the pulse of the pattern, that is, the frequency at which the events regularly occur, namely every 2 s (Fig. 4a). If a signal is a sine wave, the power spectrum will consist of only this peak. However, often the signal is a sequence of narrow spikes, hence additional waves are needed to reconstruct the signal. This explains the higher frequency waves in the power spectrum in Fig. 4a.

In the stress and mora patterns, the maximum peaks in the power spectra are located at 4 Hz (Fig. 4c,d) because all IOIs are integer multiples of 0.25 s, so a 4-Hz pulse coincides with all events.
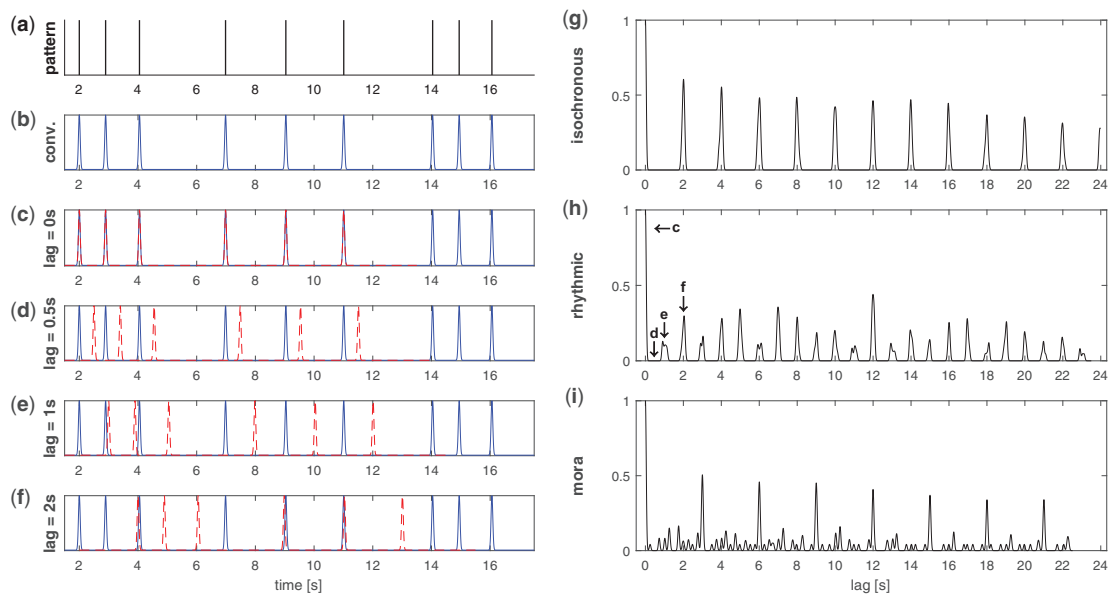
Figure 3. Visualization of the autocorrelation process (a–f) and the resulting autocorrelation function for three of the example patterns (g–i). (a) The first nine events of the rhythmic sequence depicted in Fig. 2c. (b) The same events converted to a time series and convoluted with a narrow npdf. (c) At zero lag, the stationary copy of the time series (blue, continuous line) and the time-lagged copy (red, dotted line) overlap perfectly. For the sake of the example the time-lagged copy is only shown for the first six events (i.e. the first repetition of the sub-pattern). (d) At a time lag of 0.5 s none of the events overlap, that is, across the whole time range at least one of the two signals is zero. (e) At 1 s lag, events 1 and 2 of the lagged copy partially overlap with events 2 and 3 of the stationary signal. (f) At a lag of 2 s, three of the events overlap, resulting in a higher peak ('f' in h). The peak is also narrower than the one at 1 s lag, indicating that the events overlap more closely. (g) Autocorrelation function of the isochronous sequence, depicted in Fig. 2b. (h) Autocorrelation function of the rhythmic sequence, seen in Fig. 2c. The letters c–f point to the lags that are depicted in (c–f). (i) Autocorrelation function of the mora sequence (see Fig. 2e).

2.7 Pulse GAT: which isochronous grid fits the pattern best?

A more direct, albeit computationally intensive, approach to finding the pulse of a sequence is 'generate-and-test' (GAT). First, a low frequency, isochronous grid or pulse is created in the form of regularly spaced timestamps (Fig. 5a). The deviation of each event in the pattern to its nearest grid element is measured and the root-mean-squared deviation of the whole pattern, multiplied by the grid frequency (frmsd) is calculated. The grid is then shifted forward in time in small steps in order to find the best fit (i.e. lowest frmsd) for the grid of that particular frequency (Fig. 5b). Next, the grid frequency is increased in small steps and the minimal frmsd is again calculated for each step (Fig. 5c,d). If a rhythmic sequence has an underlying isochronous regularity, its frequency can be determined by finding the grid frequency with the overall lowest frmsd value. This GAT approach was recently used to find isochrony in zebra finch song (Norton and Scharff 2016).

In many cases, the GAT frequency with the lowest frmsd corresponds to the frequency of highest power in the Fourier transform. The GAT method, however, is specifically aimed at finding the slowest pulse that still fits all events in the sequence, also called the *tatum* (Bilmes 1993). In our computer-generated 'rhythmic sequence', for example, the tatum has a frequency of 1 Hz, because all IOIs (1, 2, and 3 s) are integer multiples of 1 s (Fig. 5f). However, the frequency of highest power in the Fourier transform is located at twice the frequency, 2 Hz (Fig. 4b). While a pulse of 2 Hz covers all events as well, the slower pulse of 1 Hz is enough to explain the underlying rhythmic structure. An advantage of the GAT method is that it provides the frmsd as a measure for goodness of fit. This can be used to statistically compare different patterns in terms of their pulse fidelity, albeit only to pulses of similar frequency, since the frmsd is frequency-dependent.

2.8 Phase portraits: how to visualize repeating rhythmic patters

Unlike IOI histograms, phase portraits visualize the durations of all IOIs of a pattern, while taking their first-order sequential structure (i.e. adjacent IOIs) into account. Pairs of adjacent IOIs hereby serve as *x*- and *y*-
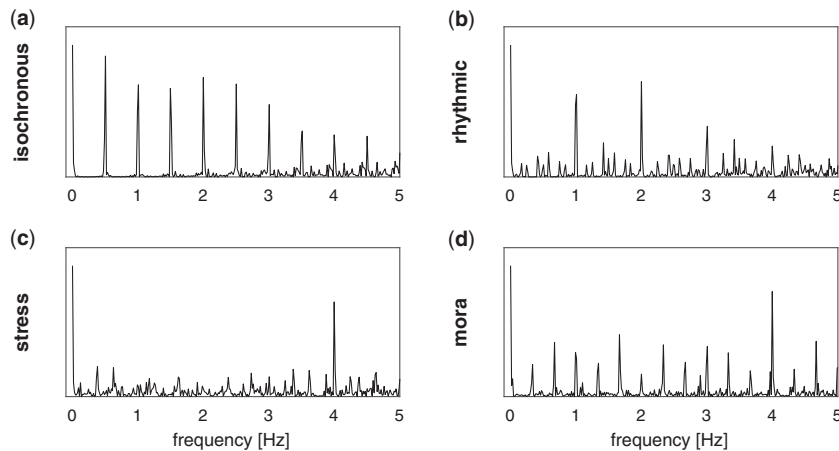
Figure 4. Power spectra of four of the example patterns. The power spectrum is the result of a Fourier transform, which decomposes a signal into waves of different frequencies. It shows in what magnitude (*y*-axis) the different frequencies (*x*-axis) contribute to the signal.

coordinates, respectively. A line connects coordinates of neighboring pairs. Together, these lines form a continuous trajectory on a two-dimensional plane. In the 'rhythmic sequence' (Fig. 2c), for example, the first two IOIs are both 1 s long, hence a dot is plotted at coordinates (1,1) (Fig. 6a). The next dot is plotted at coordinates (1,3), corresponding to the second and third IOI, which are 1 and 3 s, respectively. The trajectory connects these coordinates and moves from (1,1) to (1,3) to (3,2), and so on until it reaches (1,1) again at the end of the first sub-pattern (Fig. 6b–d). As the pattern repeats, the trajectory traces the same path three more times, albeit with slight deviations due to the noise in the timing of events (Fig. 6e).

Geometrical regularities in phase portraits correspond to rhythmic regularities in the acoustic patterns (Ravignani, in press; Ravignani et al. 2016). A repeated rhythmic pattern corresponds to similar superimposed polygons. A cyclical permutation of a rhythmic pattern (e.g. from 1,3,2,2 to 2,2,1,3) will produce the same polygon from a different starting point. Reversing the pattern (e.g. from 1,2,3 to 3,2,1) will result in the same, though rotated, polygon. Patterns that are palindromic on any phase (the 'rhythmic sequence', for example, has a palindromic cyclical permutation: 1,3,2,2,3,1) appear as polygons symmetrical with respect to the diagonal (Fig. 6h). An isochronous pattern or sub-pattern appears as a relatively dense cloud of nearby points (Fig. 6g). Some other structural regularities, like the fact that sets of four IOIs in the 'stress pattern' add up to the same duration, might go undetected and resemble a 'random pattern' (Fig. 6f,i).

2.9 Recurrence plots: visualizing pairwise similarities among all intervals in a pattern

Recurrence plots are another descriptive method for visualizing rhythmic structure, with the potential to reveal repeating sub-patterns at a glance. These plots show when a time series revisits the same regions of phase space (Dale et al. 2011; Coco and Dale 2014). Recurrence plots can be used to visualize any time series or—as shown in Fig. 7—the sequence of IOIs. As such, the recurrence plot is a two-dimensional matrix in which the similarity between any two IOIs is color-coded. The color code can be a gradient or, as shown here, monochromatic, where a black square represents similarity between two IOIs above a particular threshold. The IOI indices are noted on both axes of a recurrence plot in sequential order. Looking at the bottom row from left to right, the position of black squares indicates the sequential positions of IOIs that are similar to the first IOI, the second row to the second IOI, and so on.

The plot for the 'isochronous sequence' is all black, indicating that all IOIs are similar to each other (Fig. 7b). The repetition of the sub-pattern (6 IOIs) of the 'rhythmic sequence' is plotted as repeating patterns in the horizontal and vertical directions. As the sub-pattern is palindromic, the plot is symmetrical with respect to both diagonals (Fig. 7c). Larger black patches, as seen in the 'mora sequence' (Fig. 7e), indicate that a number of neighboring events share a similar IOI. Like phase portraits, recurrence plots visualize sequential structure, but fail to capture higher level structures, such as the equal cluster duration in the 'stress pattern' (Fig. 7d).
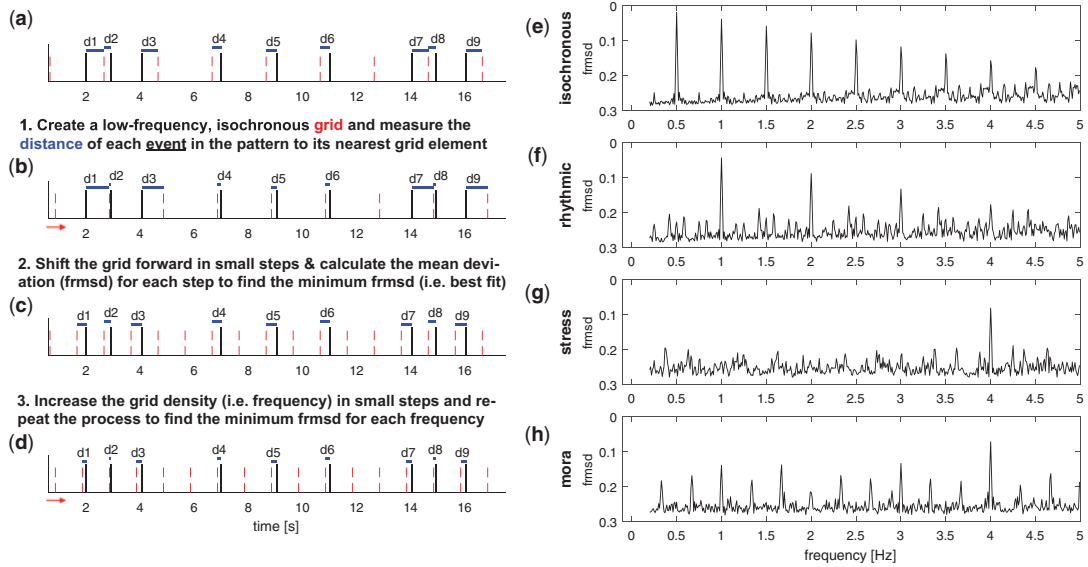
Figure 5. Visualization of the pulse GAT method (a–d), and the result of its application to four of the example patterns (e–h). Note that the *y*-axis in (e–h) is inverted. (a–d) Depicted are the first nine events of the rhythmic sequence (black vertical lines) and the deviations (d1–d9, blue lines) to their nearest element of the isochronous grid (red dotted lines). Example grid frequencies are 0.5 Hz (a and b) and 1 Hz (c and d). (e) The isochronous sequence, where all events occur roughly 2 s apart, best fits a pulse of 0.5 Hz. (f) In the rhythmic sequence all IOIs are multiples of one (1, 2 and 3), so a pulse of 1 Hz has the lowest frmsd (see also d). (g and h) In the mora and stress sequences all IOIs are multiples of 0.25 s (0.25, 0.5, 0.75, . . .), so a pulse of 4 Hz fits best to both.
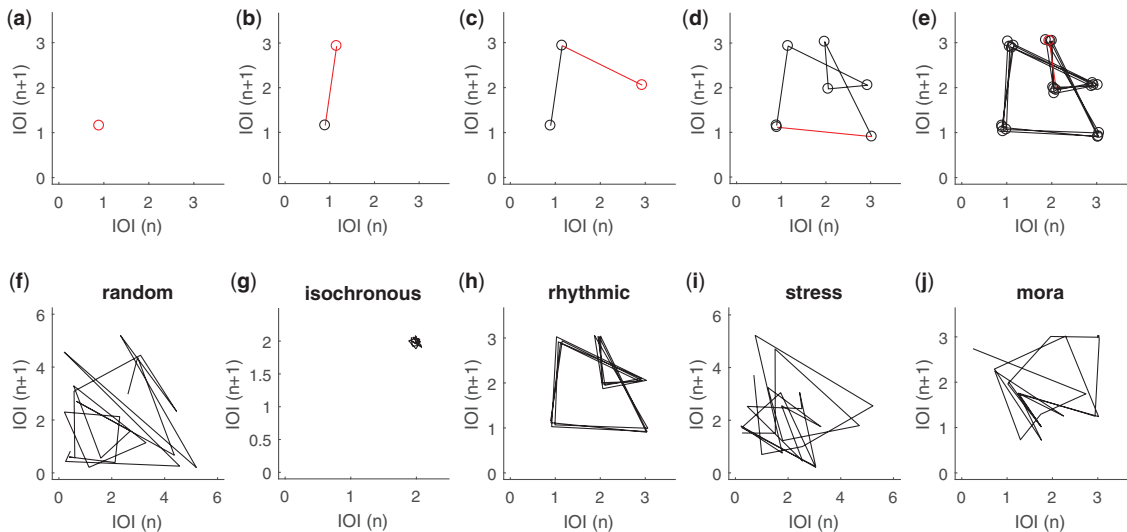

Figure 6. Step-by-step visualization of the phase portrait based on the rhythmic sequence (a–e) and phase portraits of the five example patterns (f–j). Pairs of adjacent IOIs serve as *x*- and *y*-coordinates, respectively and a line connects coordinates of neighboring pairs. (f) The random pattern shows no discernable geometrical structure. (g) The isochronous sequence appears as a dense cloud of nearby points. (h) Repetitions of a sub-pattern as in the rhythmic sequence appear as superimposed polygons, slightly shifted due to the jitter introduced by the Gaussian noise. (i and j) Higher order regularities like the constant cluster duration of the stress sequence do not appear in the phase portrait.
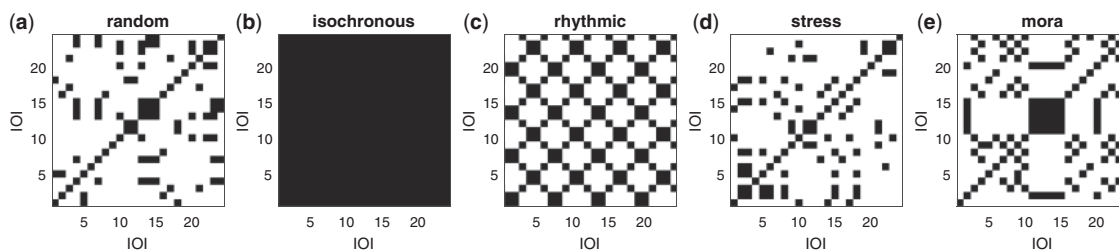
Figure 7. Recurrence plots of the five example patterns (a–e). Each square reflects the similarity of two IOIs. Black squares represent pairs of IOIs whose difference is below a certain threshold (here: 0.3 s). Recurrence plots are always symmetrical on the diagonal. (a) The random sequence shows no regular structure. (b) The plot for the isochronous sequence is all black, as all IOIs are similar to each other. (c) The repeating sup-patterns of the rhythmic sequence appear as repeating patterns in both dimensions. (d) Higher order regularities of the stress pattern do not appear in the recurrence plot. (e) Repeating IOIs of similar duration (3 s, in the middle of the mora pattern) appear as larger black patches.

2.10 Time series and regressions: can the structure in a pattern be described by a linear equation?

Time series analysis denotes a broad range of statistical methods used to extract information from data points ordered in time. Time series analysis encompasses some of the techniques described above, namely autocorrelation, cross-correlation, and Fourier analysis. In addition, autoregressive moving average models (ARMA) are promising time series techniques for speech rhythm (Jadoul et al. 2016), until now commonly employed in ecology, neuroscience, demography, and finance (Hamilton 1994). While autocorrelation probes the existence of some linear relationship between intervals in a pattern, ARMA is used to test for, model, and quantify this linear relationship. In brief, an ARMA model is a parametric, linear description of a time series: one IOI is expressed as a linear combination (i.e. weighted sum) of previous IOI values, possibly previous values of another time series (e.g. intensity of previous units of speech/vocalization), and random noise. In principle, a number of metrical stress patterns in human speech could be captured by the equation:

$$IOI_t = a + b\ IOI_{t-1} + c\ IOI_{t-2} + e_t, \qquad (2)$$

where *a*, *b*, and *c* are the (empirically estimated) model parameters, and e_t is the random error. For instance, if $IOI_1 = 1$, $a = 4$, $b = -1$, and $c = 0$, the equation becomes

$$IOI_t = 4 - IOI_{t-1} \qquad (3)$$

and it describes both iambic and trochaic meters, namely an alternation of one short and one long interval, corresponding to a strong–weak, or weak–strong, alternation in stress. (The reader can try this herself by plugging $IOI_{t-1} = 1$ in Equation (3) and iterating, i.e.

the resulting IOI_t becomes the next IOI_{t-1}.) Instead, if $c = -1$, then

$$IOI_t = 4 - IOI_{t-1} - IOI_{t-2} \qquad (4)$$

corresponding to the dactyl meter, that is, one long interval followed by two short ones. ARMA models are extremely powerful but, apart from few exceptions (Gamba et al. 2016; Jadoul et al. 2016), still scarcely used in human phonology and animal communication.

3. Between-pattern analytical techniques: comparing rhythms

3.1 Cross-correlation: are patterns linearly related?

The process of cross-correlation is similar to the autocorrelation described above, except that it correlates two different signals instead of two copies of the same signal. This makes it a useful tool to find common rhythmic properties in multiple sequences.

The cross-correlation function of two isochronous sequences (Table 3), one with noise added ('noisy-isochronous') and one without ('isochronous'), is similar to the autocorrelation function, except that the function is smaller than one at lag zero, as the events do not perfectly overlap (Fig. 8b, middle column). Cross-correlating two different 'mora sequences' reveals a common rhythmic property: a subset of events occurs regularly every 3 s in both sequences (Fig. 8d).

3.2 Multidimensional time series: how are patterns linearly related?

The ARMA models presented above for one pattern can be used to model the linear relationship between two patterns' IOIs. To test for temporal structure between two individuals vocalizing synchronously, performing

Table 3. Analytical techniques, including their specific function, to compare temporal structure between two or more patterns

Technique (References)	Function	Advantage	Disadvantage
Cross correlation (Ravignani and Sonnweber 2015)	Common sub-patterns between two patterns	Few statistical assumptions required	Provides neither sufficient nor necessary conditions for two patterns to be structurally similar
ARMA (Hamilton 1994)	IOIs in one pattern can be expressed as a linear combination of IOIs from another pattern	Common method for multidimensional time series	Only captures linear relations between patterns
Granger causality (Hamilton 1994; Seth 2010; Fuhrmann et al. 2014; Ravignani and Sonnweber 2015)	Information from one pattern's IOIs helps better predict the other pattern	Useful test, powerful inference, easy to implement	Not a real test for logical causality, rather testing the added value of one time series in predicting another series
Circular statistics (Fisher 1995; Berens 2009)	Testing hypotheses on periodically occurring events	Necessary for some data types; using classical statistics would be a mistake	A number of assumptions, concerning periodicity of events generating the data, are required

turn-taking, etc., the ARMA equation for one individual can be enriched by past IOI terms of the other individual (e.g. IOI'_{t-1} and IOI'_{t-2}), leading to the equation:

$$IOI_t = a + b\ IOI_{t-1} + c\ IOI_{t-2} + b'\ IOI'_{t-1} + c'\ IOI'_{t-2} + e_t. \quad (5)$$

Statistical estimation of parameters a, b, c, b', and c' allows in turn to draw inference about the relative contribution of the two individuals to each other's timing.

3.3 Granger causality: can one pattern structure help predict the other?

Based on ARMA modeling, one can test whether one time series significantly affects the other, using a test for Granger causality (Hamilton 1994). This technique tests whether future values of a target time series are better predicted by considering values of another time series, rather than using past values of the target series alone. In terms of equation (5), Granger causality corresponds to testing whether $b'\ IOI'_{t-1} + c'\ IOI'_{t-2}$ contributes to the statistical prediction of IOI(t). Granger causality has been successfully applied in neuroscience and economics, and it is being increasingly employed to analyze production of rhythmic patterns in humans and other animals (Seth 2010; Fuhrmann et al. 2014; Gamba et al. 2016).

3.4 Circular statistics: how does a pattern relate statistically to a fixed-pace clock?

Circular statistics (or directional statistics) offer a set of techniques that are useful when dealing with data from a circular distribution, like compass direction and time of day (Fisher 1995). In such distributions both ends of the scale have the same value, that is, they 'wrap around' and can be mapped to a circle. On a linear scale, for example, the mean of 355° and 5° would be 180°, which makes little sense when calculating, for example, compass directions. The circular mean on the other hand is 360°, which is identical to 0°.

Circular statistics can be used to compare any sequence against a known isochronous pattern. In such cases the distance of all events in the sequence to their nearest event in the isochronous pattern can be expressed as a circular distribution. An event that exactly coincides with an isochronous event is defined to have an angular direction of 0°. Events that lag behind the nearest isochronous event have positive angles >0° (i.e. positive delay), and events that precede have negative angles (equivalently, positive angles of <360°).

The distribution of events can be visualized on a socalled rose plot, which is the circular equivalent of a linear histogram (Fig. 8, right column). An arrow can be added to mark the circular mean vector, which indicates the mean direction and has a length between 0 (events are uniformly distributed) and 1 (events concentrate on a single direction).
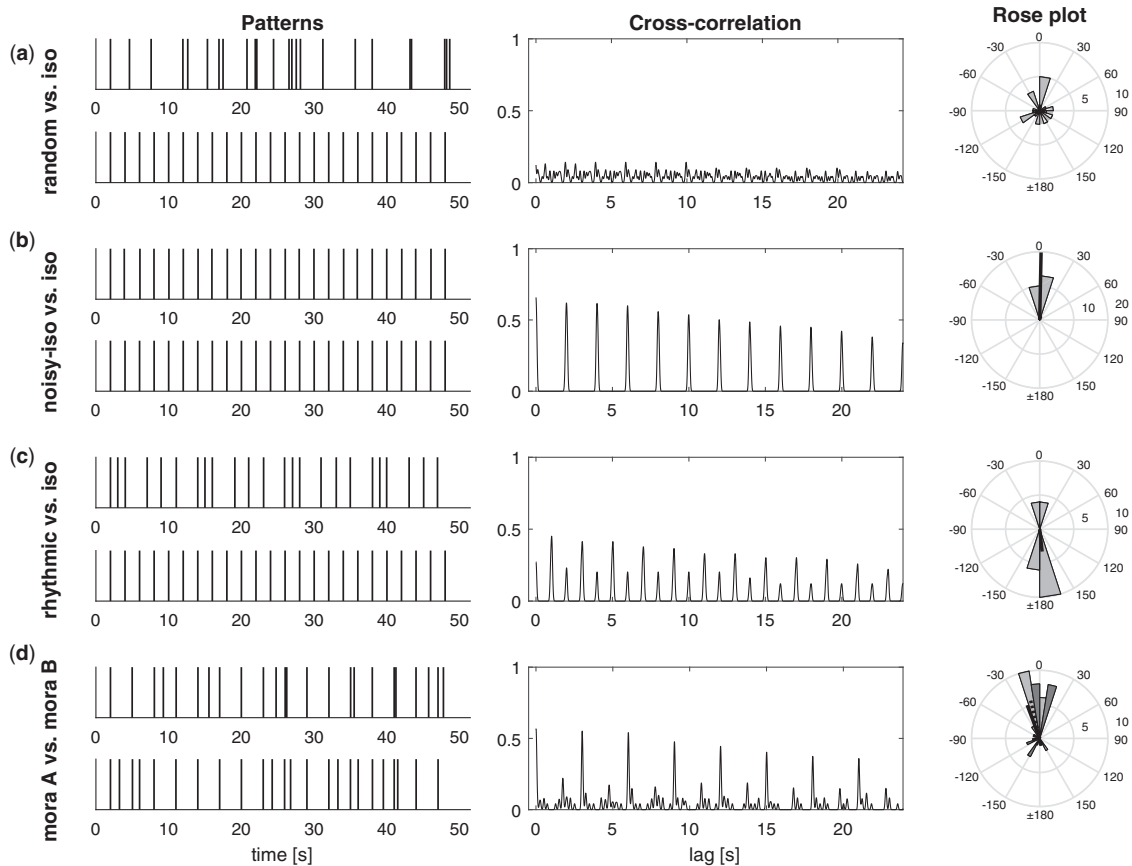
Figure 8. Cross-correlation functions (middle) and rose plots (right) for four sets of two patterns each, with different rhythmic structure (a–d). Rose plots show the mean vector as a thick black line. The mean vector can have a length between 0 (uniform distribution) and the radius of the plot (all vectors have the same direction) (a) Comparison of a random pattern and an isochronous pattern (without added noise). (b) Comparison of two isochronous patterns, one with added noise (top, 'noisy-iso'), the other without (bottom, 'iso'). (c) Comparison of a 'rhythmic' pattern and an isochronous pattern (without added noise). (d) Comparison of two different mora patterns (mora A, top and mora B, bottom). In the rose plot for the mora patterns each of the twenty sections is split between values from mora A (light gray, dotted mean vector) and mora B (dark gray, solid mean vector). Both patterns are compared against an isochronous sequence with IOI = 3 s.

Several significance tests exist to evaluate the circularity of a distribution (Zar 2010). The Rayleigh z-test can determine whether the sample data comes from a uniform distribution; For instance, the deviations of the random sequence from the isochronous pattern are uniformly distributed (Fig. 8a, P = 82.7, z = 216, n = 24). The z-test assumes that the data is unimodally distributed and sampled from a von Mises distribution, the circular equivalent of a normal distribution. Kuiper's test is the circular analogue of the Kolmogorov–Smirnov test and can be used to test the difference between two distributions, for example, the sample data against a von Mises distribution. The bimodal distribution of the rhythmic pattern (Fig. 8c) for example differs significantly from a von Mises distribution (Kuiper's test, P < 0.005, V = 336,

n = 24). If an expected mean direction is known beforehand (e.g. 0° in synchronization experiments), the more precise V-test can be used to test against uniformity. If directions are not uniformly distributed and the mean direction does not significantly differ from 0°, the patterns are synchronous (V-test, P < 0.001, V = 23.8, n = 24; Fig. 8b, isochronous pattern).

4. Where can these methods be fruitfully used?

Here we suggest some avenues of research where to apply the methods described above.

First, the field of birdsong bioacoustics should put more emphasis on temporal and rhythmic, not only

spectral or positional, properties of the signal. Building on recent findings, songs from different birdsong species could be compared to temporal structures hypothesized *a priori* or across species (Benichov et al. 2016; Norton and Scharff 2016; Spierings and ten Cate 2016, Janney et al. 2016), ultimately constructing 'rhythmic phylogenies'. Fourier analysis, GAT pulse matching and circular statistics might be particularly suitable to uncover isochronous structure in birdsong and human speech.

Second, the comparative study of vocal complexity has traditionally encompassed species capable of vocal production learning, putting relatively more emphasis on animals' abilities to modify spectral, rather than temporal, properties of the vocalizations (e.g. Kershenbaum et al. 2014). The methods presented here could be used to analyze rhythmic properties of acoustic behavior in animals with limited vocal learning, which however have potential for temporal flexibility, for instance, primates' chorusing (Fedurek et al. 2013; Gamba et al. 2016), sea lions barks (Schusterman 1977, Ravignani et al. 2016), ape drumming (Ravignani et al. 2013; Dufour et al. 2015), etc. Phase portraits, autocorrelation, circular statistics, and recurrence plots could be used, among others, to uncover the possibly flexible temporal structures in these acoustic behaviors.

Third, conversational turn-taking across human languages and animal species is a topic gaining increasing scientific interest (Levinson 2016; Vernes 2016). This area is, however, missing an advanced and unified quantitative analytical framework. We suggest that the methods described here can be used for exactly such purpose, enabling comparisons of temporal structure in communicative interaction across human languages, modalities, and species. Time series analysis, in particular ARMA models

and Granger causality, might be suitable to investigate turn-taking.

Fourth, the methods we present make almost no top-down assumptions about the structure present in the signal. Hence, they can be employed to investigate phonology, individual timing, and rhythm in modalities and domains other than speech (e.g. Dale et al. 2011; Fan et al. 2016), most notably in sign languages (de Vos et al. 2015). Distributional indexes and all other structural measures in Table 2 could be used to quantify temporal structure in a given modality or domain.

Fifth, and finally, the statistical methods presented can be used to test for cross-modal influences in multimodal communication (e.g. Tilsen 2009). Temporal interdependencies between modalities could arise, for instance, in the co-evolution of human rhythmic vocalizations and coordinated movement (Laland, et al. 2016; Ravignani and Cook 2016) or the simultaneous song and 'dance' display of some bird species (Dalziell et al. 2013; Ullrich et al. 2016). Levenshtein distance (Post and Toussaint 2011; Ravignani et al. 2016), and other dyadic techniques in Table 3 could be used to relate temporal structure between modalities.

5. Conclusions

Quantitative cross-species comparison of rhythmic features can inform the evolution of human speech (Fitch 2000; Yip 2006; Petkov and Jarvis 2012; Spierings and ten Cate 2014, 2016; Lameira et al. 2015; Norton and Scharff 2016; Ravignani et al. 2016), be it its origin, mechanisms, or function.

Here we present a number of techniques to analyze rhythmic temporal structure in animal vocalizations,
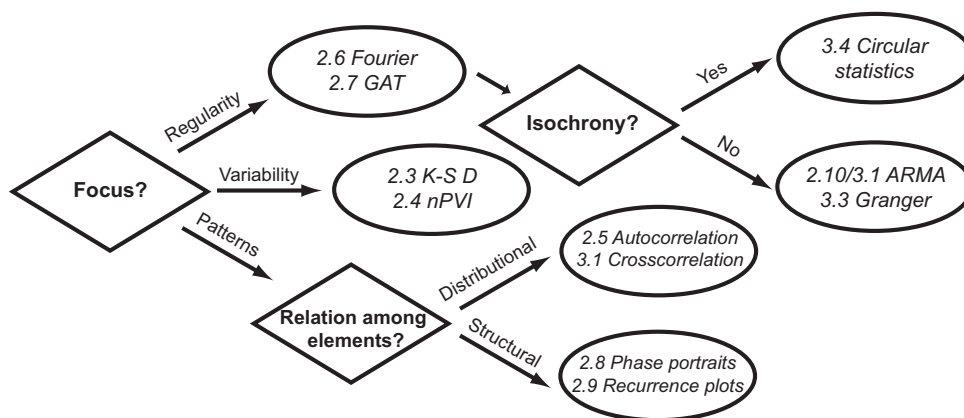

Figure 9. Flux diagram to select the most appropriate technique for rhythm analysis. Answers (regular font) to specific questions (in bold) guide the reader toward the appropriate technique and corresponding subsection (in italics).

Table 4. Some available software to perform comparative speech/vocal production analyses of rhythmic features. Software (columns) aimed at specific measures (rows), specifying packages (in square brackets), and functions (in parentheses). All software discussed here is free except for Matlab (paid software, free packages). Note that the Octave implementation of the Kolmogorov–Smirnov test differs from that of Matlab and R (R core team 2013)

Technique	Matlab	Octave	R
Histogram	histogram() (since version 2014b), hist()	hist()	hist()
Kolmogorov–Smirnov D	kstest()	kolmogorov_smirnov_ test() [statistics]	ks.test()
Auto-/Cross-correlation	xcorr()	xcorr() [signal]	acf()
Fast Fourier transform	fft()	fft()	fft(); periodogram() [TSA]
Rose plots	polarhistogram() (since version 2016b), rose()	rose()	rose.diag() [ggplot2]
Circular statistics	[CircStats] (Berens 2009).	[CircStats]	[circular]
ARMA	arima() [Econometrics]; [GCCA toolbox] (Seth 2010)	arma_rnd()	arima() [TSA]
Recurrence plots	plotRecurrence()	plotRecurrence()	[crqa] (Coco and Dale 2014)

including human speech. Figure 9 provides an aid to select the appropriate technique in the behavioral analysis of timing. This figure guides researchers in their initial approach to rhythm analysis, and presents an overview of all techniques and methods described here.

Theoretically, comparative research should avoid treating rhythm as a monolithic entity (2015). This theoretical advance can only be achieved via improvement of the methodological tools to discern and test for specific properties of rhythm. Figure 9 and the techniques it refers to provide a roadmap to testing hypotheses on specific rhythmic characteristics in the data.

Finally, this article focuses on temporal structure in rhythm. Timing is the main but not the only component underlying speech rhythm. Similarly, rhythm in animal phonology might derive from an interaction between temporal and spectral characteristics of the signal. New statistical techniques for comparative analyses should also be developed to incorporate spectral information like intonation, stress, etc. (Ramus et al. 1999).

6. Data accessibility

All sequences, figures, and statistical results were generated in Mathworks Matlab R2012b. All code used in this article can be freely downloaded as Supplementary data. It can be run in Matlab as well as the free and open-source alternative GNU Octave. The Supplementary data accompanying this article contains a detailed explanation of the code and instructions on its usage. Table 4 describes useful software (columns), packages (in square brackets), and functions, with their intended use.

Supplementary data

Supplementary data is available at *Journal of Language Evolution* online.

Conflict of interest statement. None declared.

References

Bekius, A. et al. (2016) 'The Beat to Read: A Cross-Lingual Link Between Rhythmic Regularity Perception and Reading Skill', *Frontiers in Human Neuroscience*, 10: 425.

Benichov, J. I. et al. (2016) 'Finding the Beat: From Socially Coordinated Vocalizations in Songbirds to Rhythmic Entrainment in Humans', *Frontiers in Human Neuroscience*, 10: 255.

Berens, P. (2009) 'CircStat: A Matlab Toolbox for Circular Statistics', *Journal of Statistical Software*, 31/10: 1–21.

Bilmes, J.A. (1993) 'Techniques to Foster Drum Machine Expressivity', *International Computer Music Conference*, 93: 276–83.

Boersma, P., and Weenink, D. (2013) *PRAAT: Doing Phonetics by Computer, version 5.3.49*. Amsterdam: Universiteit van Amsterdam. Retrieved from www.praat.org.

Bolhuis, J. J. et al. (2010) 'Twitter Evolution: Converging Mechanisms in Birdsong and Human Speech', *Nature Reviews Neuroscience*, 11/11: 747–59.

Bowling, D. L., and Fitch, W. T. (2015) 'Do Animal Communication Systems Have Phonemes?', *Trends in Cognitive Sciences*, 19/10: 555–7.

Coco, M. I., and Dale, R. (2014) 'Cross-Recurrence Quantification Analysis of Categorical and Continuous Time Series: An R Package', *Frontiers in Quantitative Psychology and Measurement*, 5: 1–14.

Collier, K. et al. (2014) 'Language Evolution: Syntax before Phonology?', *Proceedings of the Royal Society B: Biological Sciences*, 281/1788: 1–7.

Dale, R. et al. (2011) 'Nominal Cross Recurrence as a Generalized Lag Sequential Analysis for Behavioral Streams', *International Journal of Bifurcation and Chaos*, 21/04: 1153–61.

Dalziell, A. H. et al. (2013) 'Dance Choreography is Coordinated with Song Repertoire in a Complex Avian Display', *Current Biology*, 23/12: 1132–5.

de Boer, B. (2012) 'Air Sacs and Vocal Fold Vibration: Implications for Evolution of Speech', *Theoria Et Historia Scientiarum*, 9: 13–28.

de Vos, C. et al. (2015) 'Turn-Timing in Signed Conversations: Coordinating Stroke-to-stroke Turn Boundaries', *Frontiers in Psychology*, 6/268: 10–3389

Dufour, V. et al. (2015) 'Chimpanzee Drumming: A Spontaneous Performance with Characteristics of Human Musical Drumming', *Scientific Reports*, 5: 1–6.

Fan, P. F. et al. (2016) 'Rhythmic Displays of Female Gibbons Offer Insight into the Origin of Dance', *Scientific Reports*, 6: 1–8.

Fedurek, P., and Slocombe, K. E. (2011) 'Primate Vocal Communication: A Useful Tool for Understanding Human Speech and Language Evolution?', *Human Biology*, 83/2: 153–73.

——— et al. (2013) 'The Acoustic Structure of Chimpanzee Pant-Hooting Facilitates Chorusing', *Behavioral Ecology and Sociobiology*, 67/11: 1781–9.

Fehér, O. (2016) 'Atypical Birdsong and Artificial Languages Provide Insights into How Communication Systems are Shaped by Learning, Use, and Transmission', *Psychonomic Bulletin & Review*, 1–9.

Filippi, P. (2016) 'Emotional and Interactional Prosody across Animal Communication Systems: A Comparative Approach to the Emergence of Language', *Frontiers in Psychology*, 7: 1393.

Fisher, N. I. (1995) *Statistical Analysis of Circular Data*. Cambridge, UK: Cambridge University Press.

Fitch, W. T. (2000) 'The Evolution of Speech: A Comparative Review', *Trends in Cognitive Sciences*, 4/7: 258–67.

——— (2009). 'The biology and evolution of rhythm: unraveling a paradox,' in P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (eds), *Language and Music as Cognitive Systems*, Oxford, UK: Oxford University Press, pp. 73-95.

Fitch, W. T. (2014) 'Toward a Computational Framework for Cognitive Biology: Unifying Approaches from Cognitive Neuroscience and Comparative Cognition', *Physics of Life Review*, 11: 329–64.

Fitch, W. T. (2015) 'The Biology and Evolution of Musical Rhythm: An Update' in *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*, Cambridge US: MIT Press, pp. 293–324.

Fuhrmann, D. et al. (2014) 'Synchrony and Motor Mimicking in Chimpanzee Observational Learning', *Scientific Reports*, 4: 1–7.

Gamba, M. et al. (2016) 'The Indris Have Got Rhythm! Timing and Pitch Variation of a Primate Song Examined Between Sexes and Age Classes', *Frontiers in Neuroscience*, 10: 249.

Ghazanfar, A. A. (2013) ' Multisensory Vocal Communication in Primates and the Evolution of Rhythmic Speech', *Behavioral Ecology and Sociobiology*, 67/9: 1441–8.

Goswami, U., and Leong, V. (2013) 'Speech Rhythm and Temporal Structure: Converging Perspectives', *Laboratory Phonology*, 4/1: 67–92.

Grabe, E., and Low, E. L. (2002) 'Durational Variability in Speech and the Rhythm Class Hypothesis', *Papers in Laboratory Phonology*, 7: 515–46.

Gustison, M. L., and Bergman, T. 'The Spectro-Temporal Properties of Gelada Calls Resemble Human Speech', *Journal of Language Evolution*. Manuscript under review.

Hamilton, J. D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.

Hannon, E. E. et al. (2016) 'Exaggeration of Language-Specific Rhythms in English and French Children's Songs', *Frontiers in Psychology*, 7: 939. doi: 10.3389/fpsyg.2016.00939

Jadoul, Y., Ravignani, A., Thompson, B., Filippi, P. and de Boer, B. (2016) 'Seeking Temporal Predictability in Speech: Comparing Statistical Approaches on 18 World Languages'. *Frontiers in Human Neuroscience* 10/586: 1–15.

Janney, E. et al. (2016) 'Temporal Regularity Increases with Repertoire Complexity in the Australian Pied Butcherbird's Song', *Royal Society Open Science*, 3: 160357.

Kershenbaum, A. et al. (2014) 'Acoustic Sequences in Non-human animals: A Tutorial Review and Prospectus', *Biological Reviews*.

Lehiste, I. (1977) 'Isochrony Reconsidered', *Journal of Phonetics*, 5/3: 253–63.

Laland, K. et al. (2016) 'The Evolution of Dance', *Current Biology*, 26/1: R5–9.

Lameira, A. R. et al. (2014) 'Primate Feedstock for the Evolution of Consonants', *Trends in Cognitive Sciences*, 18/2: 60–2.

——— et al. (2015) 'Speech-like Rhythm in a Voiced and Voiceless Orangutan Call', *PloS One*, 10/1: e116136.

Levinson, S. C. (2016) 'Turn-taking in Human Communication–Origins and Implications for Language Processing', *Trends in Cognitive Sciences*, 20/1: 6–14.

Lilliefors, H. W. (1967) 'On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown', *Journal of the American Statistical Association*, 62/318: 399–402.

McAuley, J. D. (2010) 'Tempo and rhythm,' *Springer Handbook of Auditory Research, Vol.36: Music Perception*, eds M. R. Jones, R. R. Fay, and A. N. Popper New York, NY:Springer.

Norton, P., and Scharff, C. (2016) 'Bird Song Metronomics': Isochronous Organization of Zebra Finch Song Rhythm', *Frontiers in Neuroscience*, 10: 309.

Petkov, C. I., and Jarvis, E. (2012) 'Birds, Primates, and Spoken Language Origins: Behavioral Phenotypes and Neurobiological Substrates', *Frontiers in Evolutionary Neuroscience*, 4: 12

Post, O., and Toussaint, G. (2011) 'The Edit Distance as a Measure of Perceived Rhythmic Similarity', *Empirical Musicology Review*, 6/3: 164–79.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ramus, F. et al. (1999) 'Correlates of Linguistic Rhythm in the Speech Signal', *Cognition*, 73/3: 265–92.

Rauschecker, J. P. (2012) 'Ventral and Dorsal Streams in the Evolution of Speech and Language', *Frontiers in Evolutionary Neuroscience*, 4: 7

——, and Scott, S. K. (2009) 'Maps and Streams in the Auditory Cortex: Nonhuman Primates Illuminate Human Speech Processing', *Nature Neuroscience*, 12/6: 718–24.

Ravignani, A. et al. (2013) 'Primate Drum Kit: A System for Studying Acoustic Pattern Production by Non-human Primates Using Acceleration and Strain Sensors', *Sensors*, 13/8: 9790–820.

Ravignani, A., and Sonnweber, R. (2015) 'Measuring Teaching Through Hormones and Time Series Analysis: Towards a Comparative Framework', *Behavioral and Brain Sciences*, 38: 40–41.

——, and Cook, P. (2016) 'The Evolutionary Biology of Dance Without Frills', *Current Biology*, 26: 19.

—— et al. (2016) 'What Pinnipeds Have to Say about Human Speech, Music, and the Evolution of Rhythm', *Frontiers in Neuroscience*, 10: 274.

—— (2017, in press). Visualizing and Interpreting Rhythmic Patterns Using Phase Space Plots. *Music Perception*.

—— et al. (2016) 'Musical Evolution in the Lab Exhibits Rhythmic Universals', *Nature Human Behaviour*, 1/0007: 1–7.

Rothenberg, D. et al. (2014) 'Investigation of Musicality in Birdsong', *Hearing Research*, 308: 71–83

Saar, S., and Mitra, P. P. (2008) 'A Technique for Characterizing the Development of Rhythms in Bird Song', *PLoS ONE*, 3: e1461.

Scharff, C., and Petri, J. (2011) 'Evo-devo, Deep Homology and FoxP2: Implications for the Evolution of Speech and Language', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366/1574: 2124–40.

Schusterman, R. J. (1977) 'Temporal Patterning in Sea Lion Barking (Zalophus californianus)', *Behavioral Biology*, 20/3: 404–8.

Seth, A. K. (2010) 'A MATLAB Toolbox for Granger Causal Connectivity Analysis', *Journal of Neuroscience Methods*, 186/2: 262–73.

Spierings, M., and ten Cate, C. (2014) 'Zebra Finches are Sensitive to Prosodic Features of Human Speech', *Proceedings of the Royal Society of London B: Biological Sciences*, 281/1787: 20140480.

——, and —— (2016) 'Zebra Finches as a Model Species to Understand the Roots of Rhythm', *Frontiers in Neuroscience*, 10: 345.

ten Cate, C. (2016) 'Assessing the Uniqueness of Language: Animal Grammatical Abilities Take Center Stage', *Psychonomic Bulletin & Review*, 1–6.

Thiel, M. et al. (2004) 'How Much Information is Contained in a Recurrence Plot?', *Physics Letters A*, 330/5: 343–9.

Tilsen, S. (2009) 'Multitimescale Dynamical Interactions Between Speech Rhythm and Gesture', *Cognitive Science*, 33/5: 839–79.

Toussaint, G. T. (2013) *The Geometry of Musical Rhythm: What Makes a 'Good' Rhythm Good?*. Boca Raton, FL, USA: CRC Press.

Ullrich, R. et al. (2016) 'Waltzing Taeniopygia: Integration of Courtship Song and Dance in the Domesticated Australian Zebra Finch', *Animal Behaviour*, 112: 285–300.

Vernes, S. C. (2016) 'What Bats Have to Say about Speech and Language', *Psychonomic Bulletin & Review*, 1–7.

Wagner, P. (2007). Visualizing levels of rhythmic organization. In Trouvain, Jürgen and Barry, William J (eds) *XVIth International Congress of the Phonetic Sciences*, Saarbrücken, 6–10 August 2007: pp. 1113–1117.

Yip, M. J. (2006) 'The Search for Phonology in Other Species', *Trends in Cognitive Sciences*, 10/10: 442–6.

Zar, J. H. (2010) *Biostatistical Analysis, 5th edn*. Upper Saddle River, NJ: Prentice Hall, Inc.