Coordinative patterns underlying cross-linguistic rhythmic differences

Leonardo Lancia, Georgy Krasovitsky, Franziska Stuntebeck

To cite this version:

Leonardo Lancia, Georgy Krasovitsky, Franziska Stuntebeck. Coordinative patterns underlying cross-linguistic rhythmic differences. Journal of Phonetics, 2019, 72, pp.66-80. 10.1016/j.wocn.2018.08.004 . hal-02069888

HAL Id: hal-02069888
https://hal.science/hal-02069888

Submitted on 21 Oct 2021

Version of Record: https://www.sciencedirect.com/science/article/pii/S0095447018300469
Manuscript_3f8f238a182d1260287f169bcf04a367

Coordinative patterns underlying cross-linguistic rhythmic differences

Leonardo Lancia,

Laboratoire de phonétique et phonologie, CNRS/Sorbonne Nouvelle, Paris, France

leonardo.lancia@blri.fr

Georgy Krasovitsky,

University of Oxford, Oxford, UK

georgy.krasovitsky@wolfson.ox.ac.uk

Franziska Stuntebeck,

Osnabrück University, Osnabrück, Germany

franziska.stuntebeck@gmail.com

ABSTRACT

We propose a new approach to characterize cross-linguistic differences in the rhythmic structure of speech utterances by studying the degree of coordination between the production of syllables and the production of prosodic prominence at the level of the word. With this approach we compare languages traditionally considered as stress-timed (English and German) and syllable-timed (French and Italian), as well as a language that on the basis of phonological considerations does not seem to belong to either class (Polish). We analyzed recorded narrations elicited with the Pear Story technique from 26 speakers (on average 5 per language) under two elicitation conditions. The results suggest that processes underlying the production of syllables and those underlying the production of prosodic prominence are more tightly coordinated in the Germanic languages analyzed than in Romance languages. The status of Polish on the other hand is more ambiguous because while its results differ from Romance languages, the differences depend on the condition of elicitation. Overall our results suggest that the coordination between syllable production and prominence production is a pertinent dimension for the discrimination of the rhythmic characteristics of languages.

Keywords: speech rhythm, cross-linguistic comparisons, nonlinear dynamics, Generalized Synchronization, Recurrence Analysis.

1. INTRODUCTION

Some languages are perceived as rhythmically more similar than others, meaning that consistent rhythmic similarity judgments can be elicited by exposing adult and infant speakers or non-human primates to speech signals recorded from speakers of different languages and deprived of their segmental content and their F0 modulations (Ramus and Mehler, 1999; Ramus, Dupoux and Mehler, 2003, Ramus, Hauser, Miller, Morris and Mehler, 2000; Tincoff, Hauser, Tsao, Spaepen, Ramus, Mehler, 2005). Although these results provide insight into the role of acoustic energy modulations and/or their temporal characteristics in the perception of rhythmic similarity, it remains unclear which properties of the acoustic signal allow speakers to make these distinctions. A number of recent studies suggest that the perception of the rhythmic features of speech utterances may in part depend on the way syllables and prosodic prominence are coordinated in speech production (see Goswami and Leong, 2013 for a review). This leads us to explore the coordinative dimension of speech rhythm in the present paper. To this end, we propose a new approach to quantifying the degree of coordination between the production of syllables and the production of prosodic prominence. The method is based on the application of concepts from Recurrence Analysis, a technique originally introduced to characterize the behavior of non-linear dynamical systems through the analysis of their recurrent patterns. We apply the proposed methods to recordings of online and offline narrations produced by speakers of five languages (German, English, French, Italian, and Polish) whose rhythmic similarity has been assessed at the perceptual level in available works (e.g. Ramus, Dupoux and Mehler, 2003). In the remainder of the introduction, we summarize previous attempts to detect phonetic correlates of speech rhythm and set out the motivation for our new approach. In Section 2, we detail the proposed method, while Section 3 describes the data collected and the analyses conducted. In Section 4, we discuss the results with respect to current theories of speech rhythm.

1.1 Phonetic and phonological approaches to cross-linguistic rhythmic differences

Early studies on cross-linguistic rhythmic differences adopted a perspective strongly anchored in phonetics and based on the so called isochrony hypothesis (Pike, 1945; Abercrombie, 1967), which in its initial version distinguished between languages showing syllabic isochrony (i.e., languages with small variability in syllable duration) and those showing foot isochrony (i.e. languages with small variability in foot duration)[1]. Empirical evidence against isochrony principles (cf. Auer and Uhmann, 1988, Lehiste, 1977; den Os, 1983; Bertinetto, 1989 for reviews) led researchers to attempt to identify the sources of cross-linguistic rhythmic differences at a more abstract level of description (e.g. Bertinetto, 1989; Dauer, 1983; 1987; Donegan and Stampe, 1983; Gil, 1986; Pulgram, 1970). This led to the idea originally proposed by Dauer (1983, 1987) that rhythmic differences between languages result from a number of structural factors, such as the complexity of the syllabic structure, the presence of phonological vowel reduction, the presence of lexical stress, and the phonetic realization of prosodic prominence. These factors can enhance the grouping of syllables into larger units defined by prosodic prominence patterns above the syllables (i.e. stress or accent). In languages that display many of these features, prosodic prominence above the syllable is easily perceived and it provides the basis for grouping the units composing an utterance. Auer (1993) further elaborates this proposal by integrating it with the idea, already present in the works of Pulgram (1970), Holm (1987) and Kuzæmenko (1987), that in some languages (termed *word-based languages*) phonological processes tend to make reference to the phonological word, while in others (termed *syllable-based languages*) they tend to have a syllabic domain. From a theoretical perspective, the terminological shift from an opposition between syllable and foot timed languages to one between syllable and word/foot-based languages entails a move from a mainly phonetic conception of speech rhythm to a proper prosodic typology, in which the

---

[1] A third class of languages characterized by mora isochrony was added to account for languages such as Japanese that, according to Ladefoged (1975), maintain isochrony for subsyllabic rhythmic units.

notions of stress-based and syllable-based languages do not individuate language classes but prototypical structural models emerging from the coexistence of interdependent aspects of the phonological systems. As a consequence, specific languages are not sorted into classes but are expected to tend toward one prototypical pattern or the other.

1.2 Inconsistency between perception and production data

As far as distinctions between prototypical languages are concerned, results of several perceptual discrimination experiments are strongly supportive of such a conception of cross-linguistic rhythmic differences (Ramus, Dupoux and Mehler, 2003, Ramus et al., 2000; Tincoff et al., 2005). Discrimination measures consistently grouped Romance languages, typically considered syllable-based languages, and Germanic languages, considered as based on higher prosodic constructs. Perceptual data are less clear when it comes to the languages considered as intermediate between the two extremes of the continuum between these prototypical language types. Ramus et al. (2003) also report that Polish, which presents a great variety of syllable types and high syllabic complexity (like a prototypical stress/word-based language) but no vowel reduction at normal speech rate (like a prototypical syllable-based language), was perceptually distinguished from both English and Spanish, which are located close to the two extremes of the continuum. On the other hand, Catalan, which has a relatively simple syllabic structure, but also displays reduction of unstressed vowels, was distinguished from English but not from Spanish.

Based on the premise that the expression of the typological continuum should be reflected in the durational properties of the speech signal, a number of studies have therefore adopted various criteria to normalize measures of vowel and consonant durations in order to capture the underlying tendencies toward prototypical rhythmic behaviors. The results of this line of research have been questioned as it seems that the proposed metrics strongly depend on a variety of factors such as the segmental content of the utterances and stylistic, tempo, and speaker-dependent features (Arvaniti, 2012; Barry, Andreeva and Koreman, 2009; Cummins, 2002; Engstrand and Krull, 1999; Loukina et al, 2011; Steiner, 2003). For example, Arvaniti and Rodriquez (2012) suggest that there are reasons to believe that Polish utterances used in the experiments conducted by Ramus et al. (2003) display a higher speech rate compared to utterances from both stress and syllable based languages. This may explain why Polish utterances are perceived as different from utterances of both stress and syllable-based languages.

1.3 Revisiting the link between rhythmic typology and its phonetic expression

Although potential effects of the rhythmic organization of a language on the durational properties of speech utterances are not excluded by theory, the assumptions behind phonological approaches to rhythmic typology are not formulated in terms of durational properties (e.g. by stipulating different isochrony rules for different language types) but in terms of grouping (Dauer, 1983) or in terms of the domain of application of the phonological processes (e.g. the syllable vs. the phonological word, Auer, 1993). These assumptions have been translated into the language of dynamical systems by O'Dell and Nieminen (1999) who propose that the production of syllables and the production of feet are regulated by two different internal clocks behaving as linear oscillators. The slower oscillator, governing the production of feet, and the faster one, governing the production of syllables are linked by a coupling function, determining the dependence of the syllabic oscillator on the behavior of the foot oscillator. Due to inter-oscillator coupling, the duration of the inter-stress interval (I) depends linearly on the number of syllables (n) in a foot: $I=a+bn$, with slope b and intercept a. The ratio $r=a/b$ indicates the asymmetry of the coupling function, determining the dominance of one oscillator toward the other. When $r > 1$ the foot dominates the syllable (as it is expected in stress/word based languages). If n and I are known for a number of produced feet, a and b, and thus r, can be easily estimated by linear regression. In this conception of speech rhythm the notion of isochrony is retained at the theoretical level, although isochronous syllables or feet are not expected to be observed in experimental data because such behaviors are predicted to occur only when the model reaches equilibrium (i.e. when the relative phase describing the difference between the relative positions of the two oscillators in their respective cycles is constant). Moreover, conceiving speech as governed by the cascaded interactions between oscillatory processes has proved to be a productive idea to model temporal features of speech production in several studies (Saltzman, Nam, Goldstein, and Byrd, 2006; Saltzman, Nam, Krivokapic and Goldstein, 2008; Goldstein, Nam, Saltzman and Chitoran, 2009; Tilsen, 2009). However, currently available cross-linguistic comparisons that adopt this approach in conjunction with statistical analyses of the significance of the results are limited to few pairs of languages (e.g. for Brazilian vs. European Portuguese: Barbosa, Viana and Trancoso, 2009; for English vs. Polish: Malisz, 2013) and do not allow to evaluate the pertinence of this model for linguistic typology. On more methodological grounds, it should be noted that although highly simplified models of human behavior provide proofs of concept useful to understand the basic principles underlying its functioning, using their predictions to interpret observed data from a variety of conditions and tasks is often a hazardous practice. It is likely that the simplicity required to formulate interpretable models, does not permit taking into account the underlying complexities. For example, it has been proposed that in order to be tested against actual speech data, the model should be augmented in such a way that accounts for the effects of higher levels of prosodic organization (Barbosa and Madureira, 1999). Moreover, the interpretation of the *r* coefficient as an index of the dominance of the foot level over the syllable level has been questioned by Windmann, Simko and Wagner (2014), who use simulated rhythmic patterns to show how variation in the value of that coefficient could be well explained by differences in the frequency of stressed syllables that are subject to word-final lengthening.

Despite its potential drawbacks, the work by O’Dell and Nieminen (1999) has the merit of providing an original way to bridge the gap between the hypotheses about the structural factors affecting the rhythmic organization of languages and observable features of the speech signal. This is achieved by translating the theoretical assumptions into constraints on the functioning of the dynamical systems underlying speech production and not into features of the observed acoustic signals.

A similar strategy can be adopted while maintaining a higher degree of generality, without postulating a specific model for the underlying dynamics. We can hypothesize that if in some languages phonological processes depend more than in others on prosodic domains above the syllable (i.e. the foot or the phonological word), this may result into a stronger dependency between the physical processes activated to produce prosodic prominence and those underlying the production of syllables. Recent research on the perception of the rhythmic organization of speech utterances suggests that listeners pay attention to the way syllables and prominence are coordinated (see Goswami and Leong, 2013 for a review). For example, it is possible to induce the perception of trochaic or of iambic rhythms in pure sinewave stimuli by manipulating the coordination between the modulations of the acoustic energy occurring on the time scales of syllables (henceforth ***syllAM***) and the modulations of the acoustic energy occurring the time scales of word-level prosodic prominence (henceforth ***stressAM***). On the production side, Leong, Kalashnikova, Burnham and Goswami (2017) could show that, by extracting via band-pass filtering ***syllAM*** and ***stressAM*** from actual speech signals and measuring the degree of coordination between these two AM signals, it is possible to separate motherese speech from speech addressed to adults. More precisely, the degree of coordination between ***syllAM*** and ***stressAM*** is higher in motherese speech than in adult-directed speech. This observation is particularly pertinent in the context of rhythmic typology theories based on grouping, such as those of Dauer (1987) and Auer (1993), because motherese speech is usually perceived as highly rhythmical, with frequent stresses and shorter sentences (Fernald et al. 1989, Garnica 1977, Jusczyk, Cutler, & Redanz, 1993). A related approach was adopted for studying rhythmic typology by Tilsen and Arvaniti (2013) who analyzed the properties of energy modulations at the two time scales of syllables and word/foot level prominence in several languages (English, German, Greek, Italian, Korean, and Spanish). For example, the variability of the oscillatory frequency of the ***stressAM*** extracted from the acoustic signals was considered as indicative of the rhythmical stability of utterances at the level of the prosodic prominence. Another index used to characterize a potential dominance of suprasyllabic rhythms was the ratio between the spectral energy extracted from the amplitude envelope signal in the frequency bands close to the frequency of the ***syllAM*** signal and in those close to the frequency of the ***stressAM*** signal. A strongly rhythmic utterance at the level of prosodic prominence would be expected to show more energy in the frequency bands close to the frequency of the ***stressAM*** signal. However, neither measure could separate languages consistently with the postulated groupings. It should be noted that the analyses proposed by Tilsen and Arvaniti (2013) do not permit to characterize the coordination between the processes underlying syllable and stress production and that such an analysis of cross-linguistic speech data is still lacking.

1.4 Measuring the coordination between the production of syllables and the production of prosodic prominence in German, English, Polish, French and Italian

In the present paper, we follow Goswami and Leong (2013) and Tilsen and Arvnaiti (2013) in examining the relationship between ***stressAM*** and ***syllAM***. We study these, however, by means of an original approach, based on methods developed to investigate Generalized Synchronization (GS; Rulkov, Sushchik, Tsimring, and Abarbanel, 1995) between dynamical systems. A dynamical system is a law of change through which the current value of one or more variables (the current state of the system) depend on their past values and on some external influence. These laws of change can be used to model the behavior of observable processes unfolding over time. Whenever the states of two dynamical systems are related, the systems are in GS and this relation can be detected by looking at the points in time when the systems repeat or nearly repeat their behavior. Because when two states of one system observed at different points in time are similar, the states of the other system observed at the same points in time will be similar too. In analyzing two time-series it is assumed that these are affected by the behavior of two distinct underlying processes (each potentially involving different quantities) and that these processes can be modeled through two different (potentially multidimensional) dynamical systems. An important advantage this approach is that it permits relating the behavior of time-series with no surface similarity or time-series with very different frequencies.

If we assume that ***syllAM*** and ***stressAM*** depend respectively on processes responsible for syllable and prosodic prominence production, by analyzing their mutual dependency, we can infer how much the production of syllables and the production of prosodic prominence affect each other. On the basis of the hypotheses that in stress-based languages the production of syllables enhances the correlates of prosodic prominence or that phonological processes show a stronger tendency to be defined with respect to the phonological word, we expect that in stress-based languages syllables and prosodic prominence should be better coordinated than in syllable-based languages.

To test this hypothesis, we analyzed acoustic recordings of speakers of syllable-based languages (French and Italian), stress-based languages (German and English) and of a language considered as intermediate between the two types (Polish). Speakers were asked to describe the events of the Pear Story video (Chafe, 1980), which has been conceived to elicit narrations across cultures and languages. In order to provide different conditions of elicitation, the participants were asked to tell the story twice: first online, during the projection of the video, and once more offline, without the aid of the video.

In addition to measuring the coordination between ***syllAM*** and ***stressAM*** we also estimated the average speech rate by measuring the average frequency of oscillation of the ***syllAM*** signals (related to the number of syllables per seconds). This additional analysis was motivated by the potential effects of speech tempo on the observed coordinative patterns. Indeed, although our hypothesis does not permit formulating explicit predictions regarding the effects between coordinative properties of speech patterns and speech rate, such effects are commonly observed in speech repetition tasks (e.g.Tuller and Kelso, 1991; Rochet-Capellan and Schwartz, 2007; Lancia and Rosenbaum, 2018) and are related to attested distributional biases of phonological patterns across the world’s languages (see, e.g., MacNeilage, Davis, Kinney and Matyear, 2000).

2. METHOD

In order to measure GS from ***syllAM*** and ***stressAM*** time-series extracted from speech signals, we adopted a variant of Joint Recurrence Analysis (see Marwan, Romano, Thiel and Kurths, 2007 for an introduction) tailored by Lancia and Rosenbaum (2018) to the analysis of speech signals. In section 2.1 we will introduce the Joint Recurrence Analysis framework, and in section 2.2 we will briefly motivate and discuss the variant used in this paper.

2.1 Joint Recurrence Analysis

Under the Joint Recurrence Analysis approach, the detection of GS between two given dynamical systems requires the identification of all time points where the trajectory of each system repeats its behavior. Given the trajectories of two dynamical systems (see for example panels a and b of Fig. 1), the repetitions in their behaviors are determined by analyzing their recurrence plots (RPs). Given an m-dimensional trajectory X of length N represented by a sequence of vectors of length m ($X_i = \{x_{i,1,}, \dots, x_{i,m}\}$, with $i = 1, .. N$), its recurrence plot ($RP(X)$) is obtained in two steps. In the first step we compare each vector X_i to any other vector X_j (with $i, j \ = 1, \dots, N$) through a given distance function and we arrange the obtained values in a distance matrix $D_{i,j} = \|X_i, X_j\|$ (with $i, j \ = 1, \dots, N$ and $\|\cdot,\cdot\|$ representing the distance between X_i and X_j) so that the value stored at location (i, j) corresponds to the distance observed between the coordinates of the trajectory at times *i* and *j* (see panels c and d in Fig.1, where dark regions correspond to higher distances). The second step consists in transforming the distance matrix into a recurrence plot matrix containing ones at locations corresponding to values in the distance matrix lower than a given threshold ε and zeroes elsewhere:

$$RP_{i,j} = \Theta\left(\varepsilon - D_{i,j}\right), \qquad (1)$$

Where $\Theta(\cdot)$ is the Heaviside step function (equal to one if its argument is positive and to zero otherwise) and $i, j \ = 1, \dots, N$. $RP_{i,j} = 1$ indicates that at times *i* and *j* the distance between the positions of the system observed at those points in time is smaller than ε and that the two corresponding states of the trajectory can be considered equivalent. The visualization of an RP features black dots at coordinates containing ones in the matrix (see panels e and f in Fig.1), therefore each dot in an RP corresponds to a recurrent state (or to a recurrence) of the system underlying the observed trajectory. A continuous line with (locally) positive slope indicates that a whole portion of the multidimensional time-series is repeated. Depending on the application, other criteria to locate recurrences can be adopted. For example, instead of fixing the similarity threshold ε, one can adopt a predetermined value for the number of recurrences occurring in the RP and use the value of ε that produces the desired number of recurrences.

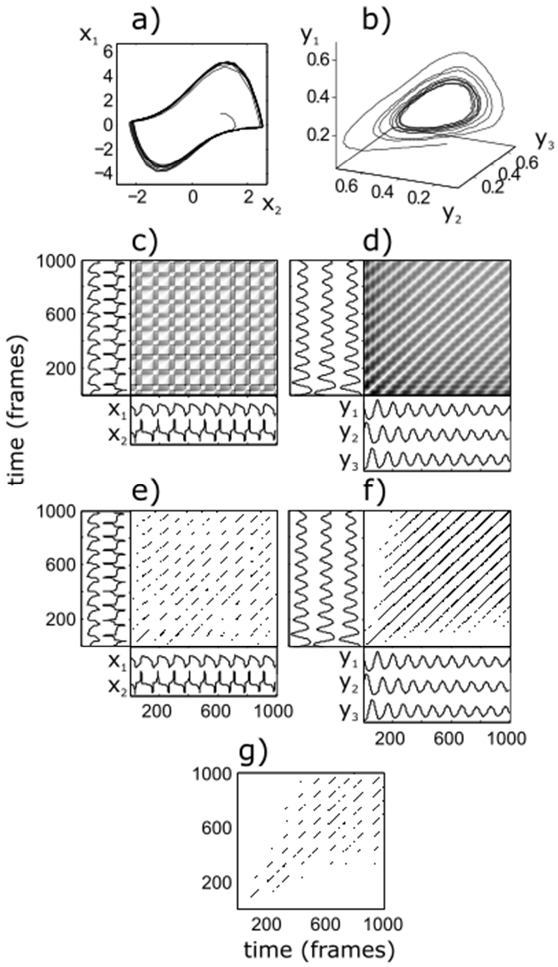

Figure 1 Recurrence and joint recurrence plots obtained from two coupled dynamical systems. First row: trajectories of a two-dimensional Van der Pol system[2] (panel a) coupled to a three-dimensional Lotka-Volterra system[3] (panel b). Second row: distance matrices of the two systems in panels a (panel c) and b (panel d). Distance values are coded with shades of grey (clear areas: small distances). The evolution over time of the systems' coordinates are replicated in parallel to the sides of the distance matrices. Third row: RPs obtained from the two distance matrices in panels c (panel e) and d (panel f). Panel g: JRP obtained from the RPs in panels e and f.

---

[2] $\dot{x_1} = x_2, \dot{x}_2 = \mu x_1^2 x_2 - x_2 + c y_1$. Where $\mu = 2$ and the coupling constant $c = 1.8$ modulates the effects of the first dimension of the Lotka-Volterra system (cf. note 2) on the Van der Pol system.

[3] $\dot{y}_1 = y_1(1 - y_1) - \gamma y_2 + \eta + \xi,\ \dot{y}_2 = y_2(1 - y_2) - \gamma y_3 + \eta,\ \dot{y}_3 = y_3(1 - y_3) - \gamma y_1 + \eta$. Where $\gamma = 2.4, \eta = 10^{-5}$and ξ is a noise term extracted from an uniform distribution with zero mean and standard deviation equal to 0.05.

Given two multidimensional trajectories X and Y of length N and their RPs ($RP(X)\ and\ RP(Y)$), a joint recurrence plot (JRP) can be computed. The JRP displays recurrences only at locations containing recurrences in both RPs (see panel g in Fig.1). We can then quantify for each point in time the probability that the corresponding state of one time series recurs given the probability that the corresponding state of the other time series recurs:

$$CR_i(X \mid Y) = \frac{\sum_{j=1}^{N} JRP_{i,j}(X,Y)}{\sum_{j=1}^{N} RP_{i,j}(Y)} \qquad (2)$$

From the two time-varying conditional probabilities of recurrence $CR_i(X \mid Y)$ and $CR_i(Y \mid X)$ we can compute an average conditional probability of recurrence that quantifies the degree of coordination between the trajectories under study and that can be used as a coordination index (CI).

$$CI(X,Y) = \frac{1}{N}\sum_{i=1}^{N} \frac{\left(CR_i(Y \mid X) + CR_i(X \mid Y)\right)}{2} \qquad (3)$$

The distribution of values obtained when using nonparametric methods to analyze time-series is affected by their statistics, which change with the shape and with the probability of occurrence of the patterns they contain. This has the practical consequence that the CI values obtained from the application of the proposed approach to the analysis of the coordination between ***syllAM*** and ***stressAM*** signals may be affected by features as their complexity or by their rate of change. These characteristics in turn may depend on structural features (e.g. the probability of complex syllables or the number of syllables per foot) paralinguistic factors (e.g. emotional state, speech rate, style), and speaker specific features (i.e. the tendency to mouth or mumble). In order to determine the presence of significant coordination between two time-series, it is necessary to build a null distribution of coordination index values that would be obtained by analyzing pairs of uncoordinated time-series showing the same features as those of the time-series observed. Significant coordination is then assessed by comparing the coordination index obtained from the analysis of the observed time series to the null distribution. These surrogate pairs of time series are constructed by repeatedly submitting one or the other observed time series to the twin surrogate construction method (Romano et al. 2009).

2.2 Measuring Generalized Synchronization of stressAM and syllAM with Joint Recurrence Analysis

Although observed processes often depend on multiple quantities, it is rarely possible to measure more than a few, if any, of the time-varying quantities reflecting the behavior of the process under investigation. For example, we may assume that the unidimensional time-series representing ***syllAM*** of the speech signal depends on the functioning of several physiological and cognitive processes activated to produce speech syllables and that modeling these interactions requires a multidimensional system. In such cases, existing studies generally adopt methods such as time-delay embedding permitting reconstruction of a multidimensional time-series representing the behavior of the dynamical system underlying the behavior of the observable unidimensional time-series. The first dimension of a time-series reconstructed via time-delay embedding corresponds to the observed time-series; the second dimension corresponds to a time-delayed copy of the first dimension with lag τ; the third dimension corresponds to another copy, lagged by 2τ. Likewise, the Mth dimension will correspond to a copy lagged by $(M\text{-}1)\tau$ (See Marwan et al., 2007, for a review of methods to find the optimal values of the number of dimensions M and of the lag τ). Once a multivariate time-series is obtained from each observed time-series, these can be submitted to Joint Recurrence Analysis to study the coordination between the underlying systems.

The application of time delay embedding to speech signals proves highly problematic because this technique relies on the assumption that the underlying time-series are stationary (i.e. the parameters governing their behaviors are constant over time). This is not generally the case for time-series generated by intentional or goal-oriented behavior whose rate of change can be modulated over time. In the analysis of a system that lacks stationarity in the temporal dimension, a single value for the lag parameter τ governing time-delay embedding will be appropriate to embed one portion of the time-series but inappropriate for modelling a different portion (because τ should increase as the system slows down). The use of locally erroneous values for the embedding parameters will introduce artefacts into the RPs and ultimately bias the analysis (Marwan, 2010; Lancia, Fuchs and Tiede, 2014, Lancia, Voigt and Krasovitskiy, 2016). Such artefactual recurrences yield groups of connected dots of variable thickness and line structures with negative slopes. In Lancia, Voigt and Krasovitskiy (2016) and in Lancia and Rosenbaum (2018), we show that this issue can be addressed by building a recurrence plot from a one-dimensional time-series without applying time-delay embedding.

Although the obtained RP will contain a high number of artefactual recurrences due to the lack of an appropriate reconstruction procedure, these can be removed through the application of common image processing algorithms. The modified approach proved to be robust to temporal nonstationarity systematically injected in the behaviour of simulated dynamical systems and it has been successfully used in the analysis of electroglottographic data (Lancia, Voigt and Krasovitskiy, 2016) and of articulatory kinematics (Lancia and Rosenbaum, 2018). The image processing steps adopted are illustrated in detail in Lancia and Rosenbaum (2018), and in the Supplemental Material of that paper a detailed guide to the implementation is provided. However, for the sake of completeness, we provide a summarized description in section 2.2.1. The different steps are illustrated in Figure 2, where they are applied to a small portion of ***syllAM*** signal obtained from one of the German speakers participating in our experiment.

Another potential issue can be encountered when two RPs are combined in a JRP because recurrences produced by two time-series may be related although not identically located. This may occur if the effect of one of the two underlying systems over the other is delayed by a fixed or variable amount of time (i.e. if a system takes some time to react to changes in the behavior of the other). In section 2.2.2 we summarize the workaround adopted by Lancia and Rosenbaum (2018) who expand the line structures in the two RPs before building the JRP and thus obtain a tolerance in the location of the recurrences. The processing steps used to build the JRP are illustrated in Figure 3.

2.2.1 Building the recurrence plots

In order to build a recurrence plot from an observed time-series, we first compute a distance matrix without submitting the time-series to time-delay-embedding (see Fig 2, panel a). From this distance matrix we compute a RP (see Fig 2, panel b) by using a very large similarity threshold ε, corresponding to a very lax similarity criterion (for the analyses presented here ε this was determined as the smallest value resulting in a number of dots equal to 1/3 of the total number of locations in each RP). The obtained RP is then submitted to a cleaning procedure permitting to reduce the number of the artefactual recurrences.

The first step of the cleaning procedure is applied separately to each group of connected dots present in the RP. The aim is to derive for each group of connected dots the longest continuous line with constant thickness equal to one dot and with locally positive or null slope passing through the connected dots. To do that we first submit each group of connected dots to morphological skeletonization, yielding a line of constant thickness equal to one dot that preserves the shape of the initial group of connected dots (see Fig 2, panel c). Portions of line structures with locally positive slope approximate the locations of recurrences in the obtained RP. However, given the lax similarity criterion adopted in building the RP, only a portion of the dots individuated in this way will correspond to true recurrences. Therefore the locations of the recurrences are refined in the second part of the algorithm. True recurrences are expected to occur close to monotonically increasing line structures. Therefore we first locate notable points (i.e. endpoints, junctions and turning points) and then we remove portions of line structures connecting two notable points that define a straight line of negative slope.

The regions were recurrences are expected to occur are obtained by substituting each remaining line structure with a monotonically increasing band with constant thickness δ (where δ is a free parameter; see Fig 2, panel d). To refine the shape of the regions of interest, a new distance matrix is obtained by setting to Infinite the values of the initial distance matrix that are not located inside the obtained bands

(Fig 2, panel e) and a new RP is obtained from the modified distance matrix using a similarity threshold ε equal to 1/10 of the standard deviation of the time-series under study (Fig 2, panel f). The obtained RP contains monotonically increasing line structures with variable thickness resulting from a still lax similarity criterion. The last cleaning step is aimed at obtaining line structures of constant thickness equal to one dot. The final RP (see Fig 2, panel g) is obtained by applying Dynamic Time Warping (DTW) separately to each region of the original distance matrix corresponding to a group of connected dots in the RP. The DTW algorithm is an iterative procedure that, starting from the top-right corner of the region of distance matrix considered, traces a line arriving to the bottom-left corner, by choosing at each step the location corresponding to the lower distance values (see Fig. 2, panel h for an illustration).

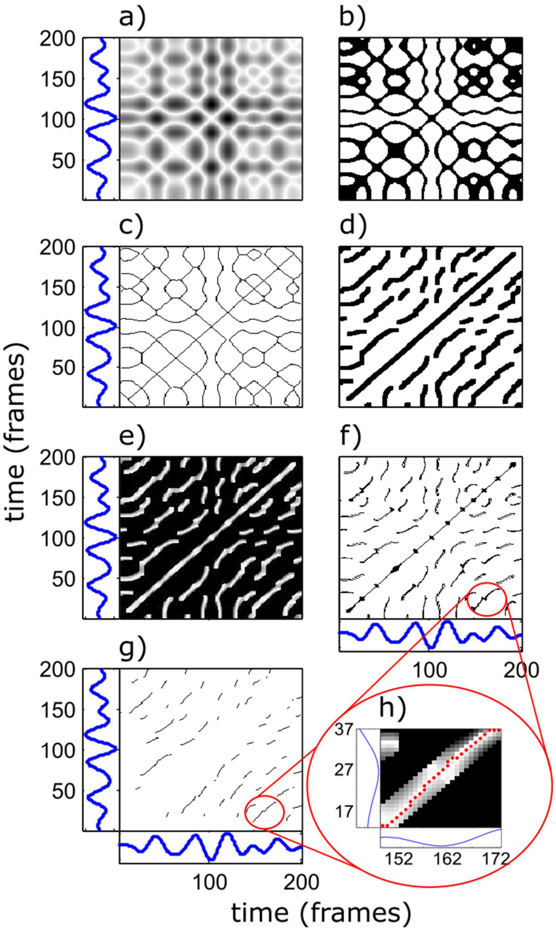

Figure 2 Recurrence plot cleaning steps. Panels contains RPs or distance matrices obtained from the same time-series corresponding to a portion of ***syllAM*** modulation from one German speaker. This is replicated in parallel to the left sides of the panels in the left column and to the lower sides of the panels at the bottom of the two columns. Panel (a): distance matrix. Panel (b): RP obtained from the distance matrix in panel a) by using a threshold allowing for a number of recurrences equal to the 30% of the number of locations in the RP. Panel (c): RP obtained by submitting the RP in panel (b) to skeletonization. Panel (d): RP obtained by expanding the line structures in panel c). Panel e): distance matrix obtained by penalizing the locations of the distance matrix in panel (a) that are not included in the bands contained in the RP in panel b). Panel f) RP obtained from the distance matrix in panel (e) by applying a similarity threshold equal to 10% of the standard deviation of the observed time-series. Panel (g): RP obtained by individuating groups of connected dots in the RP in panel (f) and by

submitting the corresponding portions of the distance matrix in panel (e) to Dynamic Time Warping. Panel (h): example of the application of the DTW algorithm to one such portion of distance matrix. The red dots are the output of the DTW algorithm and represent the best mapping between the portions of time-series corresponding to the selected portion of distance matrix (here replicated in parallel to its sides).

2.2.2 Building the JRP

Once two clean RPs are obtained (as those in panels a and b of Fig 3), they are used to build a JRP. In order to allow some tolerance in the temporal location of recurrences, the JRP is derived from RPs obtained by substituting each line structure in the clean RPs with a band of constant thickness equal to the free parameter δ_1 (see Fig 3, panels c and d). Following this modification, the recurrences of states observed at time *i* in the two time-series will be considered as joint recurrences (and appear in the JRP) if their distance in time is at most δ_1. Since the JRP obtained in this way presents line-structures of variable thickness, a new JRP is built by submitting each group of connected dots in the obtained JRP to the tracking algorithm proposed by Marwan and Kurths (2002). The algorithm is designed to derive a monotonically increasing line of constant thickness equal to one which, at each horizontal location, is equidistant from the upper and lower margins of the original group of connected dots[4]. The value of a coordination index (CI) quantifying the degree of coordination between the systems driving the

---

[4] Essentially, the procedure is an iterative one which is initialized by placing the first dot of the line which is to be tracked at coordinates corresponding to that of the dot that is closer to the lower left corner of the smallest rectangle including the considered connected dots. At each iteration an initially square window with side length equal to one is placed with its lower left corner on the last tracked dot. The lengths of window’s sides are increased by one until a new dot is found in the region of the JRP included in the window. The coordinates of the new tracked point are then determined by computing the center of mass of the dots included in the window.

observed trajectories is estimated by applying the formulas introduced in the preceding section to the obtained RPs and JRP.

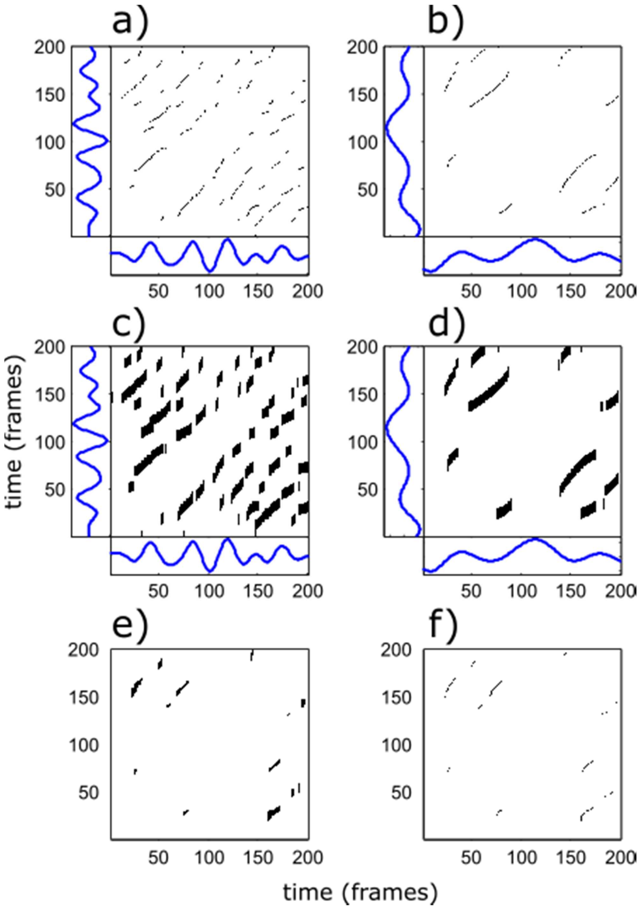

Figure 3 Construction of the JRP from two time-series corresponding to portions of ***syllAM*** and ***stressAM*** simultaneously extracted from the speech of the same German speaker. Panel a): RP from the ***syllAM*** time-series (replicated in parallel to the left and lower sides of the RP). Panel b): RP from the ***stressAM*** time-series. Panels c) and d): RPs obtained by expanding the line structures in the RPs in panels a) and b). Panel e): JRP obtained by the RPs in panels c) and d). Panel f) JRP obtained by submitting the JRP in panel e) to the skeletonization algorithm by Marwan and Kutrhs (2002).

Note that the whole procedure is based on three main free parameters: δ, δ_1 and ε. δ determines the width of the bands defining the regions in the RP where recurrences can be located. Assuming that the DTW algorithm is capable of correctly mapping between the portions of time series individuated by a group of connected dots, reducing the value of δ has the main function of speeding up the computations (by reducing the search region of the DTW algorithm). However a too small value for this parameter reduces the precision of the temporal alignment and with it the sensitivity of the obtained CI. ε determines the tolerance of the algorithm to differences between states of the observed time-series; δ_1 determines the tolerance of the algorithm to delays in the occurrence of recurrences in the two time-series under study. These two parameters determine the sensitivity of the analysis to variability in the temporal or in the amplitude dimensions. This in in turn affects the sensitivity of the analysis to weak coupling strengths. In summary fine tuning the parameters has the main effect of making the analysis more sensitive to weak coupling strengths[5].

3. EXPERIMENT

3.1. Speakers and data collection

Between 4 and 9 speakers of each language under investigation were recorded in a soundproof room (4 French, 5 Italian, 9 German, 6 Polish and 6 English). The Pear Story video clip (of approx. 5 min) was played back to the participants via a computer monitor. The speakers wore open headphones (Philips SHP 9000), enabling them to hear the video soundtrack but also their own voice. Each speaker was asked to tell the story twice: first during the projection of the video (online condition), and once more once the video ended (offline condition). While in the offline condition speakers were free to plan and produce their utterances, in the online condition they had to coordinate their speech with the ongoing

---

[5] This being the case a criterion for the choice of the parameters' settings of quite general applicability is to choose the values that better help distinguishing between the categories under study (Schinkel, Dimigen and Marwan, 2008). For example when analysing the differences in the coordination strength across languages and styles, the most appropriate set of parameters values would be the ones that maximizes the differences in coordination strength due to the language and the style (however see Eroglu, Marwan, Prasad, and Kurths, 2014 or Yang, Ren,Hu and Li, 2015 for alternative approaches).

events of the video. This constraint limited the opportunity for planning long sequences in the online condition.

3.2. Analyses

3.2.1. Pre-processing

First, we extracted a time-varying amplitude envelope from each recorded audio signal. To this end, in order to emphasize the contribution of vowels to the amplitude envelope, the audio signal was first band-pass filtered with a IIR filter (cut-off freqs.: 700 and 1750 Hz; see Tilsen and Johnson, 2008). The Hilbert envelope of the obtained signal was band-pass filtered with IIR filters with cut-off thresholds set at 0.8 and 2.7 Hz in order to capture ***stressAM*** and at 2.7 and 7Hz to extract ***syllAM***. The choice of these cut-off frequencies was based on the results obtained by Tilsen and Arvaniti (2013) which show that in data from English, Korean, German, Italian, Greek and Spanish the rates of syllabic and supra-syllabic oscillations never exceed 2.7Hz and 7Hz respectively, regardless of stylistic differences in the analyzed speech data (i.e., read or spontaneous speech).

A voice activity detection algorithm (Lee and Hasegawa-Johnson, 2007) was applied to each acoustic recording to detect silences. Then, we manually identified hesitations in each speech chunk separated by pauses. Before the application of Recurrence Analysis, portions of ***syllAM*** and ***stressAM*** signals corresponding to silences, to hesitations and to portions of speech chunks preceding hesitations were removed (and the remaining portions concatenated).

In order to obtain several measurement points from each recording, each time-series was split into portions of equal length (16 secs), amplitude-normalized (so to have zero mean and unit standard) and downsampled to 160Hz.

3.2.2. Coordination analysis

From each portion we obtained one RP from the ***stressAM*** time-series and another from the ***syllAM*** time-series. the RPs were cleaned and combined into a JRP following the procedures described in Sections 2.2.1 and 2.2.2 and a coordination index (CI) value was computed though the application of the formulas in section 2.1. The length of the portions of time-series considered was chosen in such a way to obtained at least 3 CI values for each processed recording. The parameters of the analysis were set as follows: $\delta = 5$ for ***syllAM*** and $\delta = 10$ for ***stressAM***, $\delta_1 = 10$ and $\varepsilon = 0.1$.

For each pair of ***syllAM*** and ***stressAM*** signals extracted from the same portion of speech signal lasting 16 sec, we generated 30 surrogate ***syllAM*** time-series by repeatedly applying the twin surrogate method (Romano et al. 2009) and obtained 30 CI values by measuring their coordination with the observed ***stressAM*** time-series. These surrogate CI values were averaged in order to have an average surrogate CI value for each portion of speech signal. The comparison between the CI values obtained with the original time-series and the average CI values obtained with the surrogate time-series reveals if differences between languages reflect different coupling strengths or are due to statistical characteristics of the time-series compared.

The values obtained (N=550) were submitted to a linear mixed model in order to test for the triple interaction between language (a categorical factor with German as reference level), test condition (online vs. offline, reference: online) and data type (original vs. surrogate, reference: original). Speaker identity and recording were considered as random factors and speaker-specific random slopes relative to the effect of the test condition were also included in the model.

3.2.3. Speech rate analysis

In order to detect speech rate differences between languages and conditions, we submitted the ***syllAM*** signals to the Hilbert transform and extracted the time-varying instantaneous frequency. If we assume that the cycles of ***syllAM*** activity correspond to syllabic cycles, the instantaneous frequency is an estimate of speech rate roughly interpretable in terms of syllables per seconds. Instantaneous frequency time-series were split into portions of 16s matching those used in the computation of the CI values and the instantaneous frequency values in each portion were averaged in order to obtain an estimate of average speech rate in each portion. Average speech rate values (N=550) were submitted to a linear mixed model in order to test for the interaction between language (a categorical factor with German as reference level) and test condition (online vs. offline, reference: online). Also in this model the random effect structure included two intercepts for speaker identity and recording and a speaker specific slope for the effect of the test condition.

3.3. Results

3.3.1. Coordination analysis

Fig. 4 presents the values of the coordination index obtained from surrogate time-series and original time-series in both online and offline conditions. The results of the mixed model regression are summarized in Table 1.

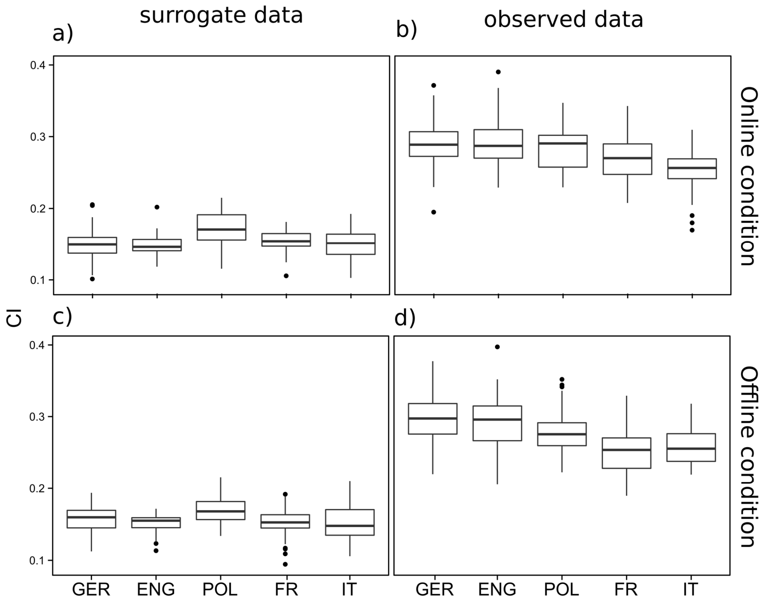

Figure 4 Coordination index over languages. Panels group data with respect to the elicitation condition (top row: online, bottom row: offline) and to the kind of data (left column: surrogate data right column: observed data).

The triple interaction was excluded from the model as it was neither significant nor improved the model fit (as established by comparing the residuals obtained with and without the interaction through Chi-square tests). The intercept of the model (row a of Table 1) represents CI values obtained from original data produced by German speakers in the offline condition. While the intercept is significantly positive, the effect of surrogate data is significantly negative (row f). This proves that there is a significant degree of coordination between ***syllAM*** and ***stressAM*** time series from German speakers. In the original data, no significant difference was observed between English and German or between Polish and German (rows b and c). However, French and Italian display CI index values significantly smaller than the intercept (rows d and e). At the same time, the interaction between the effect of language and of data type is significant and positive for both Italian and French (rows l and m), indicating that the differences between German and Italian CI values and between German and French CI values are smaller in surrogate data than in original data. This is interpreted as evidence that the differences in CI values between the two Romance languages and German are not entirely due to the statistics of the languages considered but are genuine differences in the degree of coordination between ***stressAM*** and ***syllAM***.

Because Polish was not significantly different from German in the observed data, the significantly positive interaction between data type and Polish language (row i) suggests that in surrogate data Polish displays CI values which are higher than those observed in German. The effects of the languages do not change significantly in the offline condition except for French. Indeed the effect of the interaction between the French language and the offline condition is significantly negative (row p), which is also true for the simple effect of the French language. This indicates that the difference in CI values between French and German increases in the offline condition.

Table 1. Summary of the results from mixed model regression on the coordination index values. First column: row indexes; second column: effects; third column: effects' estimates; fourth column: standard errors; fifth column: t values, sixth column: p values. The following abbreviation are used in this as well as in the following tables: Ger: German; Pol: Polish; Eng: English; It: Italian; Fr: French; Orig: original data; Surr: surrogate data. Significant effects are indicated by bold typeface.

	Fixed effects:	Estimate	Std. Error	t value	p value
a	**(Intercept)**	**2.901e-01**	**3.819e-03**	**75.955**	**<2e-16**
b	Lang (Eng)	2.581e-03	6.337e-03	0.407	0.685734
c	Lang (Pol)	-6.265e-03	6.647e-03	-0.942	0.350852
d	**Lang (Fr)**	**-2.404e-02**	**6.364e-03**	**-3.778**	**0.000519**
e	**Lang (It)**	**-3.351e-02**	**6.158e-03**	**-5.442**	**3.17e-06**
f	**Data type (Surr)**	**-1.402e-01**	**2.689e-03**	**-52.126**	**<2e-16**
g	Condition (Online)	4.778e-03	3.509e-03	1.361	0.186597
h	Lang (Eng) * Surr	-3.344e-03	4.150e-03	-0.806	0.420619
i	**Lang (Pol) * Surr**	**2.886e-02**	**4.356e-03**	**6.625**	**5.21e-11**
l	**Lang (Fr) * Surr**	**3.378e-02**	**3.760e-03**	**8.985**	**<2e-16**
m	**Lang (It) * Surr**	**3.429e-02**	**3.922e-03**	**8.743**	**<2e-16**
n	Lang (Eng) * Online	-4.272e-03	5.430e-03	-0.787	0.441147
o	Lang(Pol) * Online	-7.727e-03	5.846e-03	-1.322	0.200650
p	**Lang (Fr) * Online**	**-1.782e-02**	**5.399e-03**	**-3.300**	**0.004845**
q	Lang(It) * Online	-5.832e-03	5.417e-03	-1.077	0.296131
r	Surr * Online	2.879e-03	2.695e-03	1.068	0.285618

Post-hoc analyses confirm that:

- Differences between the Germanic languages or between the Romance languages never reach significance, regardless of the elicitation condition (see Appendix, Table A1 and Table A2).

- Every Romance languages displays smaller CI values than every Germanic language regardless of the elicitation condition (see Appendix, Table A3).
- CI values from Polish are never significantly different from those collected from speakers of the Germanic languages (see Appendix, Table A5) but, depending on the elicitation condition, may differ from those obtained from speakers of the Romance languages (see Appendix, Table A6).

The reported differences are not observed when comparing surrogate data (see Appendix, Tables A4, A7), proving that they are genuinely due to different degrees of coordination between the processes underlying the production of prosodic prominence and the production of syllables.

3.3.2. Speech rate analysis

Figure 5 shows the boxplot obtained from the average frequencies of the syllabic components computed in the portions of chunks considered, and grouped by language and elicitation condition.

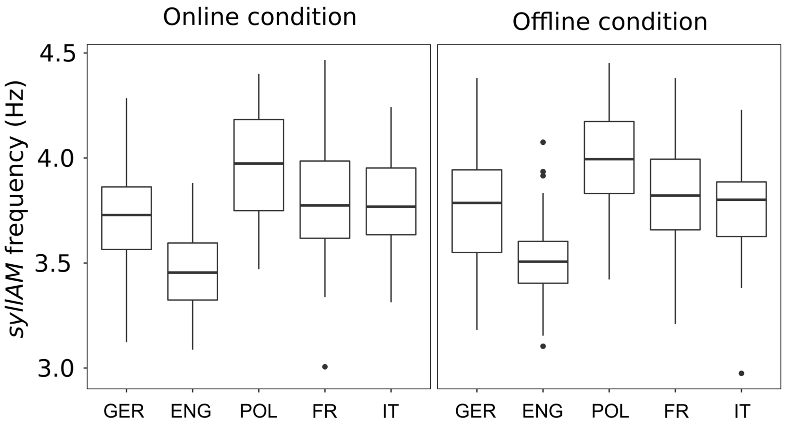

Figure 5 Box plots obtained from average frequency values (in Hz) computed in the portions of ***syllAM*** signals submitted to Recurrence Analysis. Data are separated with respect to the elicitation conditions.

While Polish seems to be faster than German, English seems to be slower. This is confirmed by the results of the mixed model regression (summarized in Table 2). Indeed the effect of English language (row b) is significant and negative, while that of Polish (row c) is significant and positive.

Table 9. Summary of results from the mixed model regression on ***syllAM*** *frequency values. Data are organized as in Table 2.*

	Fixed effects:	Estimate	Std. Error	t value	p value
a	**(Intercept)**	**3.724517**	**0.045194**	**82.411**	**<2e-16**
b	**Lang (Eng)**	**-0.27812**	**0.076032**	**-3.658**	**0.000933**
c	**Lang (Pol)**	**0.264455**	**0.078844**	**3.354**	**0.002006**
d	Lang (Fr)	0.104504	0.076266	1.37	0.181005
e	Lang (It)	0.07981	0.074213	1.075	0.291362
f	Condition (Online)	0.040376	0.0336	1.202	0.250562
g	Lang (Eng) * Cond. (Online)	0.026915	0.056986	0.472	0.643038
h	Lang (Pol) * Cond. (Online)	0.003266	0.061909	0.053	0.958527
i	Lang (Fr) * Cond. (Online)	-0.025353	0.055346	-0.458	0.655749
l	Lang (Ita) * Cond. (Online)	-0.097532	0.056466	-1.727	0.105894

4. DISCUSSION

The results presented in the preceding section clearly show that the coordinative aspects of speech production are of relevance for rhythmic typology. Indeed, the fact that syllable production and prosodic prominence production are better coordinated in Germanic languages than in Romance languages is consistent with the hypothesis formulated by Dauer (1983; 1987), and further elaborated by Auer (1993), that the syllables of some languages (e.g. Germanic languages) are better grouped in larger units such as metrical feet or phonological words than the syllables of other languages (e.g. Romance languages). Polish phonology allows complex consonant clusters but does not allow vowel reduction. Therefore structural considerations would predict a behavior which is midway between that of stress-based and that of syllable-based languages. Consistently with this prediction, our results indicate that, although Polish tends to cluster with languages considered as stress-based, significant differences between Polish and Romance languages are observed only in Polish speech produced in the online condition. Polish utterances seem to be produced with a relatively fast speech rate in comparison to other languages. This is consistent with the fact that, when comparing surrogate data, CI values obtained from Polish speakers were higher than those from speakers of the other analyzed languages. Indeed ***syllAM*** and ***stressAM*** signals characterized by high numbers of oscillations per seconds (corresponding to fast speech) will present more recurrences than ***syllAM*** and ***stressAM*** signals with lower frequencies and therefore it is more likely to observe joint recurrences just by chance.

If we assume that, as proposed by Arvaniti (2012) and supported by our data, Polish utterances display a higher syllable rate than utterances from both Romance and Germanic languages, the discrepancy between our results and those obtained from perceptual experiments showing that Polish is perceived as rhythmically different from both Romance and Germanic languages (Ramus, Dupoux and Mehler, 2003), may be explained by the effect of speech rate on the perceptual judgments.

Although our results support the idea that rhythmic differences between the Romance and the Germanic languages are related to the cohesion between syllables that are grouped together at higher levels of prosodic organization, many related questions are left unanswered. Additional work is required to better characterize the nature of the supra-syllabic grouping units (e.g. feet or phonological words) across languages. Another direction for further research is related to the location of the regions of the speech stream where the different levels of processing are more (or less) coordinated. For example, we may ask if the coordination between the syllabic and the supra syllabic levels changes close to pitch accents or to the boundaries of major prosodic constituents. Because we cannot exclude that other levels of organization are relevant for cross-linguistic rhythmic distinctions, many potential links (e.g. those between F0 and amplitude modulations) remain unexplored. Finally and most importantly, our results do not reveal the level of processing at which the features that characterize speech rhythm are determined. In other words, in this paper we leave unanswered the question if speech rhythm is defined internally by some abstract pattern generator or if it emerges from the interplay between local constraints regulating the interactions between the phonetic processes. Under the first interpretation, the production of syllables and the production of prosodic prominence are coordinated in a top down fashion as in the model proposed by O'Dell and Nieminen (1999), where speech rhythm is generated by the interaction between a syllabic oscillator and foot oscillator. In this model, the function of each oscillator is that of a clock regulating the advancement in time of the activity at a level of the prosodic structure. Since the oscillators are coupled, they work jointly as a central pattern generator that coordinates, in a top-down fashion, the timing between speech production processes triggered at different levels of the prosodic hierarchy.

Under the second interpretation, rhythmic patterns are not explicitly controlled during speech production but emerge from low-level constraints regulating the interactions between the many heterogeneous processes underlying speech production. These constraints are determined both by material and linguistic factors. Indeed, in the execution of a specific speech task (e.g. in the production of a given consonant) speakers manipulate the material constraints operating on the motor system by adjusting the mutual dependencies between the articulators. As a result of these task-specific manipulations, the articulators jointly behave in a way that is functional to the achievement of the current linguistic goal (see for example Kelso, Tuller, Vatikiotis-Bateson and Fowler, 1984; Saltzman and Munhall, 1989). The interdependency between simultaneous events (as the motion of different articulators involved in the production of the same sound) and between events that follow each other (as the motion of articulators supporting the achievement of different sounds) favors the emergence of synchrony and repetition: two core ingredients of rhythmic patterns. A number of studies explored therefore how rhythmic patterns emerge from the interactions between the different local constraints affecting the behavior of the many components of the sensory motor system during speech production (e.g.: Tuller and Kelso, 1991; Cummins and Port, 1998; Saltzman et al., 2008; Tilsen, 2009; to cite just a few). This line of research has been strongly inspired by pattern formation phenomena observed in many domains and disciplines and formalized through theories of self-organization (Haken, 1977; 1983; Prigogine, 1978). In this theoretical framework, biological rhythms are characterized by at least two different levels of organization. Indeed, while the interactions between the different organs, tissues, molecules etc. occurring at fine-grained (i.e. microscopic) levels of organization require a high number of variables to be described, this behavior can be predicted by that of few (collective) variables that capture the relations between the microscopic variables. Once a low-dimensional collective behavior emerges from the interaction between many interacting quantities, it constrains the interactions occurring at the microscopic levels in a way that favors its persistence. In this way activity is coherently structured across levels of organization and the overall complexity of the system is reduced.

On the basis of these considerations, one may ask whether cross-linguistic rhythmic differences reflect different configurations of a low-dimensional central pattern generator explicitly regulating speech rhythm or if they emerge from the interactions between local constraints provided by the phonologies of different languages. The stability of a particular collective behavior is determined by the constraints from which it emerges. Under the assumption that rhythmic patterns capture collective behaviors of the sensorimotor system, some rhythmic patterns are expected to be more stable in some languages than in others and therefore selected during language evolution by some languages but not by others.

Such an evolutionary pattern becomes even more plausible in the light of the potential role of rhythmic regularities in speech perception. On the one hand, it is reasonable to hypothesize that regularities in the speech signal help building predictions concerning future input and that these predictions permit reducing the attention paid to predictable portions of the speech stream (Cummins, 2016). On the other hand, it has been shown that low-level auditory processing is tuned to the typical time scale of syllable production, that cortical oscillations at that time scale are coordinated with syllabic amplitude modulations and that the strength of this coordination is related to speech intelligibility (see Räsänen, Doyle and Frank, 2018 for a recent review). Evidence collected in these studies motivated theoretical models in which the coordination between the amplitude modulation of the speech signal and the cortical rhythms of the listener favors chunking processes permitting the extraction of the segmental content (e.g.: Ghitza, 2011; Giraud and Poeppel, 2012). In the light of the potential role played by the rhythmic properties of the speech signal in the extraction of linguistic information, it does not seem unreasonable to assume that perceptual constraints constitute an additional source of pressure toward the selection of a stable rhythmic pattern during language evolution.

On more methodological grounds, the approach developed in this paper diverges dramatically from previous attempts to investigating cross-linguistic differences in speech rhythm since we do not rely on durational properties of the speech signal, whether durations of phonetic events or frequencies of amplitude modulations. Instead, we address the coordination between the processes that underlie amplitude modulations related to the production of syllables and prosodic prominence by studying how well the two kinds of processes repeat together behaviors displayed in the past. To this end, we adapted the tools provided by Recurrence Analysis to investigate strongly nonstationary processes, as are expected to be those involved in intentional behavior (cf. Lancia and Rosenbaum, 2018). By detecting the repetitions in the processes underlying the production of stress and the production of syllables via Recurrence Analysis we capture a core characteristic of rhythmic behavior, which is the recurrence of some events or classes of events. By measuring the degree of coordination between the processes under study, we capture another key dimension of rhythmic behavior, which is that of grouping of events in superordinate events. In this respect our approach is similar to that recently adopted by Leong et al. (2017) who measured the coordination of amplitude modulation components by estimating the variability of their generalized phase difference values. The generalized phase difference captures the relative advancement of two oscillatory signals in their own cycles by taking into account the ratio between the oscillatory frequencies of the two signals. The variability of the generalized phase difference between two oscillatory trajectories is expected to decrease as the strength of their coordination increases. Due to the flexibility of Recurrence Analysis, the approach proposed in this paper is expected to be more extensive. Indeed, it is not restricted to the analysis of narrow-band oscillatory processes (those for which instantaneous phase values can be computed, see Huang, Wu, Long, Arnold, Chen, and Blank, 2009) and therefore of more general applicability (thus permitting for example analyses of the coordination between amplitude modulations and articulatory trajectories or f0 movements); it does not require prior knowledge of the ratio between the oscillatory frequencies of the signals under study; it naturally handles cases where the oscillatory frequencies vary over time and it does not make assumptions on the distribution of the obtained values.

5. CONCLUSION

Since the first works showing that speech rhythm is constrained by preferred patterns of coordination between the realizations of phonological categories (Cummins and Port, 1998; Tajima, Zawaydeh and Kitahara, 1999; Tajima and Port, 2003), the coordinative bases of cross-linguistic differences in speech rhythms have mostly been studied as produced by specific theoretical models of coordination between syllabic and feet oscillators. This was in part due to the lack of methods that, as the one proposed in this paper, permit addressing the coordination between different phonetic processes with little knowledge about the processes themselves. We analyzed the degree of coordination between the modulations of the acoustic energy occurring at the time scale characteristic of the production of syllables and those occurring at the time scale characteristic of the production of prosodic prominence. The comparisons of data from different languages support the idea that speech rhythms are behavioral patterns that result from the coordination between the processes underlying speech production constrained by the phonological structure. This is in line with a conception of phonology in general, and of prosody in particular, as coordinative devices permitting the organic functioning of the sensorimotor processes within and across speakers involved in speech communication.

ACKNOWLEDGMENTS

Leonardo Lancia’s work, carried out within the Labex BLRI (ANR-11-LABX-0036) and EFL (ANR-10-LABX-0083), has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the program "Investissements d'Avenir".

REFERENCES

Abercombie, D. (1967). *Elements of general phonetics*. Aldine, Chicago, 1967.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351-373.

Arvaniti, A., & Ross, T. (2012). Rhythm classes and speech perception. In O. Niebuhr (Ed.), *Understanding prosody: The role of context, function and communication* (pp. 13-75). Walter de Gruyter.

Arvaniti, A., & Rodriquez, T. (2013). The role of rhythm class, speaking rate, and F0 in language discrimination. *Laboratory Phonology*, *4*(1), 7-38.

Auer, P. (1993). Is a rhythm-based typology possible. *A study of the role of prosody in phonological typology (KontRI Working Paper No. 21)*. Hamburg: Universität Hamburg.

Auer, P., & Uhmann, S. (1988). Silben-und akzentzählende Sprachen: Literaturüberblick und Diskussion. *Zeitschrift für Sprachwissenschaft*, *7*(2), 214-259.

Barbosa, P. A., & Madureira, S. (1999). Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. In *Proceedings of the XIVth international congress of phonetic sciences*, Vol. 1, 297-300.

Barbosa, P. A., Viana, M. C., & Trancoso, I. (2009). Cross-variety Rhythm Typology in Portuguese. In *Tenth Annual Conference of the International Speech Communication Association*.

Barbosa, P. N. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication, 49*(9), 725-742.

Barry, W., Andreeva, B., & Koreman, J. (2009). Do rhythm measures reflect perceived rhythm?. *Phonetica*, *66*(1-2), 78-94.

Bertinetto, P. M. (1989). Reflections on the dichotomy ‘stress’ vs.‘syllable-timing’. *Revue de phonétique appliquée, 91*(93), 99-130.

Chafe, W. L. (1980). The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. Norwood, New Jersey: Ablex.

Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. In *Speech Prosody 2002, International Conference*.

Cummins, F. (2016). Rhythm and Speech in M. A. Redford, (Ed.) *The handbook of speech production* (158-177). New York, USA: Wiley.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145-171.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*(1), 51-62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th international congress of phonetic sciences* (Vol. 5, pp. 447-450).

den Os, E. A. (1983). Stress-timed and syllable-timed languages. *Pogress Report Institute of Phonetics University of Utrecth*, *8,* 12-23. DOI:10.1162/opmi_a_0000

Donegan, P. J., & Stampe, D. (1983). Rhythm and the holistic organization of language structure. In J. F. Richardson, M. Marks, and A. Chukerman (Eds.), *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax* (pp. 337-353). Chicago, Chicago Linguistic Society.

Engstrand, O., & Krull, D. (2003). Rhythmic intentions or rhythmic consequences? Cross-language observations of casual speech. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 2789-2792).

Eroglu, D., Marwan, N., Prasad, S., & Kurths, J. (2014). Finding recurrence networks' threshold adaptively for a specific time series. *Nonlinear Processes in Geophysics*.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of child language*, *16*(3), 477-501.

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C., E., Snow & C., A., Ferguson (Eds.), *Talking to Children* (pp. 63-88). Cambridge: Cambridge University Press.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in psychology*, *2*.

Gibbon, D., Bachan, J., & Demenko, G. (2007). Syllable timing patterns in Polish: results from annotation mining. In *Eighth Annual Conference of the International Speech Communication Association*, 994-997.

Gil, D. (1986). A prosodic typology of language. *Folia Linguistica*, *20*(1-2), 165-232.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, *15*(4), 511-517.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in phonetics and speech science,* 239-250.

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives. *Laboratory Phonology, 4*(1), 67-92.

Haken, H. (1977). Synergetics: An Introduction. Berlin: Springer-Verlag.

Haken, H. (1983). Synergetics: An introduction. Nonequilibrium phase transitions and self-organization in physics, chemistry and biology. Berlin: Springer-Verlag.

Holm, C. (1987). Le cadre des changements phonétiques dans les langues Romancees. Mot et 'syntagme phonétique'. In: H. Andersen & K. Koerner (Eds.), *Historical Linguistics* (pp 231-243). Amsterdam: Benjamins.

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X., & Blank, K. (2009). On instantaneous frequency. *Advances in adaptive data analysis, 1*(02), 177-229.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child development*, *64*(3), 675-687.

Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance, 10*(6), 812.

Kuzæmenko, Y. K. (1987). The change of the phonological type of a language. In *Proceedings of the 11th International Congress of Phonetic Sciences (August 1-7, 1987, Tallinn)* (Vol. 5, pp. 429-432).

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lancia, L. & Rosenbaum, B. (2018). Coupling relations underlying the production of speech articulator movements and their invariance to speech rate. *Biological Cybernetics*, in press, DOI: 10.1007/s00422-018-0749-y.

Lancia, L., Fuchs, S., & Tiede, M. (2014). Application of concepts from cross-recurrence analysis in speech production: An overview and comparison with other nonlinear methods. *Journal of Speech, Language, and Hearing Research, 57(3)*, 718-733.

Lancia, L., Voigt, D. & Krasovitskiy, G. (2016). Characterization of laryngealization as irregular vocal fold vibrations and interaction with prosodic prominence. *Journal of Phonetics*, *54*, 80-97.

Lee, B., & Hasegawa-Johnson, M. (2007). Minimum mean squared error a posteriori estimation of high variance vehicular noise. *Biennial on DSP for In-Vehicle and Mobile Systems*.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*(3), 253–263.

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind*. Advance Publication.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech a. *The Journal of the Acoustical Society of America, 129*(5), 3258-3270.

MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child development, 71*(1), 153-163.

Malisz, Z. (2013). *Speech rhythm variability in Polish and English—a study of variability in rhythmic levels* (PhD dissertation, Adam Mickiewicz University in Poznań).

Marwan, N. (2011). How to avoid potential pitfalls in recurrence plot based data analysis. *International Journal of Bifurcation and Chaos*, 21(04), 1003-1017.

Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A, 302*(5), 299-307.

Marwan, N., Romanceo, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237-329.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems* (Vol. 191977). New York: New York, USA: Wiley.

O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In *Proceedings of the XIVth international congress of phonetic sciences* (Vol. 2, pp. 1075-1078). Berkeley: University of California.

Pike, K. L. (1945). The intonation of American English. Ann Arbor, Michigan: University of Michigan Press.

Pikovsky, A., Rosenblum, M., & Kurths, J. (2003). *Synchronization: a universal concept in nonlinear sciences*. Cambridge, UK: Cambridge university press.

Prigogine, I. (1978). Time, structure, and fluctuations. *Science, 201*(4358), 777-785.

Pulgram, E. (1970). *Syllable, word, nexus, cursus.* The Hague: Mouton.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America, 105*(1), 512-521.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 337-342. Barcelona.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349-351.

Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition, 171*, 130-150.

Romanceo, M. C., Thiel, M., Kurths, J., Mergenthaler, K., & Engbert, R. (2009). Hypothesis test for synchronization: Twin surrogates revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 19*(1), 015108.

Rulkov, N. F., Sushchik, M. M., Tsimring, L. S., & Abarbanel, H. D. (1995). Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E, 51*(2), 980.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology, 1*(4), 333-382.

Saltzman, E., Nam, H., Goldstein, L., & Byrd, D. (2006). The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In *Motor control and learning* (pp. 63-73). Springer US.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody* (pp. 175-184).

Schinkel, S., Dimigen, O., & Marwan, N. (2008). Selection of recurrence threshold for signal detection. *The european physical journal special topics*, *164*(1), 45-53.

Steiner, I. (2003) A refined acoustic analysis of speech rhythm. In *38th Linguistics Colloquium,* Piliscsaba, Hungary, 2003.

Tajima, K., & Port, R. F. (2003). Speech rhythm in English and Japanese. *Phonetic interpretation: Papers in laboratory phonology VI*, 317-334.

Tajima, K., Zawaydeh, B., & Kitahara, M. (1999). A comparative study of speech rhythm in Arabic, English and Japanese. *Proceedings of the XIV ICPhS, San Francisco, USA.*

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive science, 33*(5), 839-879.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America, 134*(1), 628-639.

Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America, 124*(2), EL34-EL39.

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. *Developmental science, 8*(1), 26-35.

Tuller, B., & Kelso, J. S. (1991). The production and perception of syllable structure. *Journal of Speech, Language, and Hearing Research, 34*(3), 501-508.

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology, 4*(1), 93-118.

Windmann, A., Simko, J., & Wagner, P. (2014). Probing theories of speech timing using optimization modeling. *Proceedings of Speech Prosody 7*.

Yang, D., Ren, W. X., Hu, Y. D., & Li, D. (2015). Selection of optimal threshold to construct recurrence plot for structural operational vibration measurements. *Journal of Sound and Vibration, 349*, 361-374.

APPENDIX

Table A1. Post-hoc comparison between observed data from Germanic languages. First column: comparison; second column: estimated difference, third column: standard error, fourth column t value, fifth column: p values.

Compared data	Estimate	Std. Err.	t value	p value
Ger,Orig,Online-Eng,Orig,Online	-2.581e-03	6.337e-03	-0.407	1.000
Ger,Orig,Online-Ger,Orig,Offline	-4.778e-03	3.509e-03	-1.361	0.992
Ger,Orig,Online-Eng,Orig,Offline	-3.086e-03	6.416e-03	-0.481	1.000
Eng,Orig,Online-Ger,Orig,Offline	-2.196e-03	6.489e-03	-0.338	1.000
Eng,Orig,Online-Eng,Orig,Offline	-5.052e-04	4.560e-03	-0.111	1.000
Ger,Orig,Offline-Eng,Orig,Offline	1.691e-03	6.284e-03	0.269	1.000

Table A2. Post-hoc comparison between observed data from Romance languages. Data are organized as in Table A1.

Compared data	Estimate	Std. Err.	t value	p value
Fr,Orig,Online-It,Orig,Online	9.468e-03	7.076e-03	1.338	0.993
Fr,Orig,Offline-It,Orig,Offline	-2.515e-03	7.024e-03	-0.358	1.000
Fr,Orig,Online-Fr,Orig,Offline	1.304e-02	4.524e-03	2.882	0.244
Fr,Orig,Online-It,Orig,Offline	1.052e-02	7.441e-03	1.414	0.988
It,Orig,Online-Fr,Orig,Offline	3.570e-03	6.904e-03	0.517	1.000
It,Orig,Online-It,Orig,Offline	1.055e-03	4.546e-03	0.232	1.000

Table A3. Post-hoc comparison between observed data from Germanic and from Romance languages. Data are organized as in Table A1.

Compared data	Estimate	Std. Err.	t value	p value
Ger,Orig,Online-Fr,Orig,Online	2.404e-02	6.364e-03	3.778	0.0236

Ger,Orig,Online-It,Orig,Online	3.351e-02	6.158e-03	5.442	<0.01
Ger,Orig,Online-Fr,Orig,Offline	3.708e-02	6.194e-03	5.987	<0.01
Ger,Orig,Online-It,Orig,Offline	3.457e-02	6.537e-03	5.287	<0.01
Eng,Orig,Online-Fr,Orig,Online	2.663e-02	7.227e-03	3.684	0.0312
Eng,Orig,Online-It,Orig,Online	3.609e-02	7.050e-03	5.120	<0.01
Eng,Orig,Online-Fr,Orig,Offline	3.966e-02	7.093e-03	5.592	<0.01
Eng,Orig,Online-It,Orig,Offline	3.715e-02	7.399e-03	5.021	<0.01
Fr,Orig,Online-Ger,Orig,Offline	-2.882e-02	6.536e-03	-4.410	<0.01
Fr,Orig,Online-Eng,Orig,Offline	-2.713e-02	7.331e-03	-3.701	0.0298
It,Orig,Online-Ger,Orig,Offline	-3.829e-02	6.276e-03	-6.101	<0.01
It,Orig,Online-Eng,Orig,Offline	-3.660e-02	7.103e-03	-5.153	<0.01
Ger,Orig,Offline-Fr,Orig,Offline	4.186e-02	6.078e-03	6.886	<0.01
Ger,Orig,Offline-It,Orig,Offline	3.934e-02	6.369e-03	6.177	<0.01
Eng,Orig,Offline-Fr,Orig,Offline	4.017e-02	6.942e-03	5.786	<0.01
Eng,Orig,Offline-It,Orig,Offline	3.765e-02	7.201e-03	5.228	<0.01

Table A4. Post-hoc comparison between surrogate data from Germanic and from Romance languages. Data are organized as in Table A1.

Compared data	Estimate	Std. Err.	t value	p value
Ger,Surr,Online-Fr,Surr,Online	-9.735e-03	6.364e-03	-1.530	0.9758
Ger,Surr,Online-It,Surr,Online	-7.772e-04	6.158e-03	-0.126	1.0000
Ger,Surr,Online-Fr,Surr,Offline	4.232e-04	6.194e-03	0.068	1.0000
Ger,Surr,Online-It,Surr,Offline	-2.602e-03	6.537e-03	-0.398	1.0000
Eng,Surr,Online-Fr,Surr,Online	-1.050e-02	7.227e-03	-1.453	0.9849
Eng,Surr,Online-It,Surr,Online	-1.540e-03	7.050e-03	-0.218	1.000
Eng,Surr,Online-Fr,Surr,Offline	-3.391e-04	7.093e-03	-0.048	1.0000
Eng,Surr,Online-It,Surr,Offline	-3.364e-03	7.399e-03	-0.455	1.0000
Fr,Surr,Online-Ger,Surr,Offline	2.078e-03	6.536e-03	0.318	1.0000
Fr,Surr,Online-Eng,Surr,Offline	7.113e-03	7.331e-03	0.970	0.9998
It,Surr,Online-Ger,Surr,Offline	-6.879e-03	6.276e-03	-1.096	0.9992
It,Surr,Online-Eng,Surr,Offline	-1.845e-03	7.103e-03	-0.260	1.0000

Ger,Surr,Offline-Fr,Surr,Offline	8.080e-03	6.078e-03	1.329	0.9939
Ger,Surr,Offline-It,Surr,Offline	5.055e-03	6.369e-03	0.794	1.0000
Eng,Surr,Offline-Fr,Surr,Offline	3.045e-03	6.942e-03	0.439	1.0000
Eng,Surr,Offline-It,Surr,Offline	2.019e-05	7.201e-03	0.003	1.0000

Table A5. Post-hoc comparison between original data from Polish and from Germanic languages. Data are organized as in Table A1.

Compared data	Estimate	Std. Err.	t value	p value
Ger,Orig,Online-Pol,Orig,Online	6.265e-03	6.647e-03	0.942	0.100
Ger,Orig,Online-Pol,Orig,Offline	9.214e-03	6.538e-03	1.409	0.989
Eng,Orig,Online-Pol,Orig,Online	8.846e-03	7.479e-03	1.183	0.998
Eng,Orig,Online-Pol,Orig,Offline	1.180e-02	7.397e-03	1.595	0.966
Pol,Orig,Online-Ger,Orig,Offline	-1.104e-02	6.795e-03	-1.625	0.959
Pol,Orig,Online-Eng,Orig,Offline	-9.351e-03	7.564e-03	-1.236	0.997
Ger,Orig,Offline-Pol,Orig,Offline	1.399e-02	6.411e-03	2.182	0.703
Eng,Orig,Offline-Pol,Orig,Offline	1.230e-02	7.236e-03	1.700	0.942

Table A6. Post-hoc comparison between original data from Polish and from Romance languages. Data are organized as in Table A1.

Compared data	Estimate	Std. Err.	t value	p value
Pol,Orig,Online-Fr,Orig,Online	1.778e-02	7.501e-03	2.370	0.574
Pol,Orig,Online-It,Orig,Online	2.725e-02	7.331e-03	3.717	0.029
Pol,Orig,Online-Pol,Orig,Offline	2.950e-03	5.049e-03	0.584	1.000
Pol,Orig,Online-Fr,Orig,Offline	3.082e-02	7.374e-03	4.179	<0.01
Pol,Orig,Online-It,Orig,Offline	2.830e-02	7.669e-03	3.690	0.029
Pol,Orig,Offline-Fr,Orig,Offline	2.787e-02	7.057e-03	3.949	0.014
Pol,Orig,Offline-It,Orig,Offline	2.535e-02	7.313e-03	3.467	0.059
Fr,Orig,Online-Pol,Orig,Offline	-1.483e-02	7.437e-03	-1.994	0.823
It,Orig,Online-Pol,Orig,Offline	-2.430e-02	7.213e-03	-3.368	0.076

Table A7. Post-hoc comparison between surrogate data from Polish and from Romance languages. Data are organized as in Table A1.

Compared data	Estimate	Std. Err.	t value	p value
Pol,Surr,Online-It,Surr,Online	2.182e-02	7.331e-03	2.976	0.200
Pol,Surr,Offline-Fr,Surr,Offline	2.295e-02	7.057e-03	3.251	0.103
Pol,Surr,Online-It,Surr,Offline	1.999e-02	7.669e-03	2.607	0.405
Pol,Surr,Offline-Fr,Surr,Offline	2.295e-02	7.057e-03	3.251	0.103