

# Shared Duration of Speech Intervals Across 49 Human Languages

Lara S. Burchardt<sup>a,b</sup>, Ludger Paschen<sup>b</sup>, and Susanne Fuchs<sup>b,c</sup>

<sup>a</sup>Humboldt Universität zu Berlin

<sup>b</sup>Leibniz-Zentrum Allgemeine Sprachwissenschaft

<sup>c</sup>IMéRA and ILCB, Aix-Marseille University

June 30, 2025

## Abstract

Research on speech rhythm has predominantly focused on smaller linguistic units such as syllables and words. At these levels, rhythm is characterized by a high degree of heterogeneity between languages, suggesting that it is intrinsically distinct from nonhuman animal communication, where the underlying isochrony is frequently observed. In this study, we investigate speech rhythm based on units similar to those used in work on nonhuman animals, i.e., sound elements surrounded by silence. Interval durations are surprisingly similar between a comprehensive sample of 49 languages from all inhabited continents, with a median of  $\sim$ 2 seconds. This pattern is independent of gender, age, or language-specific factors. IOI beats are also very similar, at around 0.5 Hz, but isochronous beat models are not a good fit for human speech compared to those of other animal species. By revealing cross-linguistic regularities alongside a lack of isochrony, our results help to place human speech within the wider spectrum of rhythmic strategies observed in nature.

**Keywords:** speech rhythm | cross-species comparison | animal communication | Inter-Onset Interval

Corresponding author: Lara Sophie Burchardt, Email: lara.sophie.burchardt@hu-berlin.de

Rhythm is a fundamental component of communication, organizing the temporal structure of speech to enhance cognitive processing and enable predictions of forthcoming speech signals. Previous research on human speech production has reported strong heterogeneity in speech rhythms at the syllable and word levels, without recurrent isochronous patterns<sup>1,2,3</sup>. Isochrony has been discussed as present in the communication signals of nonhuman animals though, as highlighted by a recent review and several individual publications<sup>4,5,6,7,8,9,10,11</sup>. Strong arguments for a comparative approach in the study of rhythm in communication signals across species<sup>12</sup> warrant re-evaluating speech rhythms using a comparable data base and methodology. Thus, we investigate whether the presence or absence of similar speech rhythms is a true differentiator between human behaviors and between humans and other species, and if the differences reported are an artifact of varying definitions and measurement techniques for rhythm across disciplines.

For this study, we adopt a contemporary definition of *rhythm* as a biological signal characterized by a non-random, ordered, predictable, and recurrent alternation of events in a temporal sequence<sup>13</sup>. *Isochrony* refers to a metronome-like rhythm in which interval durations between events are constant. We consider isochrony as a basic framework for analyzing recurring patterns over time. While true isochrony is rarely observed in nature, the concept remains useful as an underlying model describing a sequence of events. Like any model, these models need to be assessed for their goodness-of-fit to determine how well an isochronous model actually matches an IOI sequence<sup>14,15</sup>.

To facilitate a meaningful comparison between species, we employ a common framework and data-driven metrics that do not rely solely on human models as reference<sup>16</sup>. We investigate communicative rhythms in spontaneous speech data from 49 primarily endangered and low-resource languages spoken across all six inhabited continents (Figure 1A). We explore the potential for consistency in speech unit intervals and underlying isochronous rhythms in humans by adopting a methodology used in nonhuman animal communication, enabling direct comparison with previous findings in that domain. Ultimately, we aim to contribute to understanding the function and evolution of rhythmic behavior across species, spanning the full range from random temporal structure to highly precise, coordinated group activity.

A common term used in studies of rhythm in human and nonhuman acoustic communication is the *syllable*<sup>17,18,19</sup>. In linguistics, a *syllable* refers to a unit of speech consisting of a sequence of sounds, though its specific characteristics vary. In biology, syllables are often defined as communicative expressions delineated by silent pauses. For human speech, an analogous temporal unit is the *Inter-Pausal Unit* (IPU), a stretch of speech uninterrupted by silent pauses, a term used in large corpus studies<sup>20</sup>. An interval from the start of an IPU to the onset of the next is referred to as an *Inter-Onset Interval* (IOI), and it forms

the basis of the rhythm analysis in this study (Figures 1B-C), analogous to approaches used in animal studies<sup>12</sup> (see Table S1 for further explanation of terminology, including pause, IPU, IOI and sequence).

## Rhythms in human communication

The idea of rhythmicity in speech has its roots in attempts to apply principles of music to speech dating back to the 18th century<sup>2</sup>. Traditional models of speech rhythm categorized languages as stress-timed (e.g., English, Dutch), syllable-timed (e.g., French, Spanish) and mora-timed (e.g., Japanese) based on the idea that either stresses, syllables or moras occur at regular intervals<sup>21,22,23</sup>. The notion of strict isochrony at the foot, syllable or mora level has since been refuted, and the consensus view today is that speech rhythm is better characterized by relative variability in segmental and prosodic timing, with languages existing on a continuum of rhythmic variation<sup>24,25</sup>. To capture the inherent variability of speech rhythms and the influence of segmental and prosodic structure, quantitative methods measuring vocalic and consonantal intervals have been proposed. The impression of isochrony appears to be primarily grounded in perception<sup>26,27,28,29</sup>. Rhythm typologies have also been proposed at the phrase level with the concept of *macro-rhythm*, a pitch-based measure that quantifies the regularity of tonal patterns<sup>30,31,32,33</sup>. Currently, it remains unclear whether human languages share similar rhythms beyond the phrase level.

Variability in rhythm within and between languages may be influenced by linguistic and extra-linguistic factors. Broad prosodic properties such as stress or tone shape the rhythmic profile of a language. Tone languages, especially those with complex tone systems like Mandarin Chinese or Cantonese, tend to rely less on phrase-level intonation than languages without grammatical tone<sup>30</sup>. Grammatical differences between languages may also indirectly affect rhythmic variation. As an IPU minimally consists of one word, languages in which words tend to be longer and more complex due to a high degree of morphological synthesis could be expected to show a floor effect and overall less variability regarding IOI duration. Tentative links between morphophonology and rhythm have also been suggested<sup>34</sup>. Furthermore, systemic differences in speech rate have been reported between males and females, as well as among speakers of different age groups<sup>35,36,37,38</sup>.

Whether languages exhibit similarities in their speech rhythm depends on the measures and units analyzed. Linguistic rhythm is sometimes quantified by intervals between stressed syllables<sup>24</sup>, while other approaches focus on ratios between consonants and vowels<sup>3</sup>. More recent work derived temporal measures in a data-driven manner from the decomposition of the amplitude envelope of the acoustic speech signal<sup>17</sup>. The study in<sup>39</sup> analyzed rhythms derived from intonation unit (IU) onsets and periodic components of

the speech envelope across six languages. An intonation unit is a fundamental prosodic unit defined by specific pitch patterns at its boundaries and prominent positions<sup>40</sup>.

The study found a high phase consistency of approximately 1 Hz between the onset of an IU and the speech envelope in five of the six languages, suggesting that intonation units might exhibit an underlying isochronous rhythm in some languages.

These varied findings suggest that the question of whether similarities in rhythm or even isochronous rhythms are present in human language depends critically on the units of analysis. Faster syllable rhythms (in the linguistic sense) appear to be specific to the respective language, reflecting each language's unique phonotactic and prosodic systems. However, larger units and slower speech rhythms may exhibit some degree of cross-linguistic, species-wide patterning, paralleling what is observed in nonhuman animal communication.

## Rhythms in nonhuman animal communication

Rhythms in animal communication have been studied for decades and the literature, especially on arthropod rhythms, is vast<sup>11</sup>, but the field of research has recently experienced renewed momentum. In addition to several methods and opinion pieces that have been published in the last years<sup>41,42,12,43</sup>, new rhythm studies have appeared for species such as zebra finches (*Taeniopygia guttata*)<sup>4</sup>, bats<sup>8,19</sup>, rock hyraxes (*Procavia capensis*)<sup>44</sup>, fish<sup>9,13</sup>, frogs<sup>6</sup>, sperm whales (*Physeter macrocephalus*)<sup>7,42</sup>, gibbons<sup>5</sup>, palm cockatoos (*Probosciger aterrimus*)<sup>45</sup> or the northern elephant seal (*Mirounga angustirostris*)<sup>46</sup> among others. Despite employing diverse methodologies, these studies consistently identify a pattern approximating isochrony to describe various sequence types. While the overarching rhythmic structure appears similar across species, differences emerge in the specific frequency rates (e.g., beats per second) and the degree to which each species adheres to an underlying isochronous rhythm, i.e., how well an isochronous model fits or how well a species keeps a beat. Through a combination of new methodological developments and efforts to find an evolutionary basis for communication rhythms, an increasing amount of knowledge about these rhythms can now be reasonably compared across species, enabling more general conclusions about their function, mechanisms, ontogeny, and evolution.

Based on the striking similarities observed across very different animal species and classes, we believe it is likely that underlying isochronous rhythms are also present in human speech. To search for possible commonalities both within humans and across nonhuman and human animals, we analyzed rhythmic parameters and further checked for underlying isochronous rhythms using the same methodology applied

to acoustic communication signals of many animal species. This approach includes the careful analysis of Inter-Onset Intervals (IOI) between and across languages, the calculation of potential underlying isochronous rhythms, and goodness-of-fit calculations evaluating how well isochronous rhythms describe IOI sequences.

## Material and Methods

### Data preparation

Our study uses data from the DoReCo corpus (v2.0)<sup>48</sup>. DoReCo is a corpus containing speech data with time-aligned transcriptions and annotations originating from language documentation collections covering a wide range of typologically diverse languages from all inhabited continents. The majority of speakers in DoReCo represent non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations<sup>49</sup>, as well as endangered, understudied or low-resource languages. In total, the corpus contains datasets from 53 languages. The core datasets in DoReCo include manually verified alignments of words and silent pauses, the latter coded based on (i) the presence of silence in the acoustic signal that cannot be attributed to articulatory processes, such as the closure phase of stop consonants, and (ii) perceptible disruptions of the speech flow. Only data from the core subset of the corpus were used for this study (see<sup>50</sup> for more details on corpus creation principles and the manual and automatic steps involved in creating the time alignments). IUPs were defined as sequences of word units uninterrupted by silent pauses (see Figure 1). IOI units were determined by grouping sequences of IUPs and silent pauses and calculating the overall duration of these units. Data were filtered to include only monological texts with personal and traditional narratives, excluding recordings of conversations or stimulus retellings. We further excluded intervals consisting entirely of filled pauses, false starts, or other content annotated with a label for unusual discourse events in the corpus. After filtering, 49 languages remained in our sample. An overview of these languages, their genealogical affiliations, and the number of Inter-Onset Intervals is provided in Table 2.

To probe the correlation between rhythm and various speaker-related and typological features (Figure 3), we employed several methods to generate the relevant data frames. Information on speaker age and gender was extracted from the metadata files included in the corpus. Language-level information on longitude, latitude, area and genealogical family were obtained from Glottolog<sup>51</sup>. Languages were manually coded as tonal or non-tonal based on available descriptions in published grammars and documentation materials. To quantify the degree of morphological synthesis, synthesis scores were calculated as the aver-

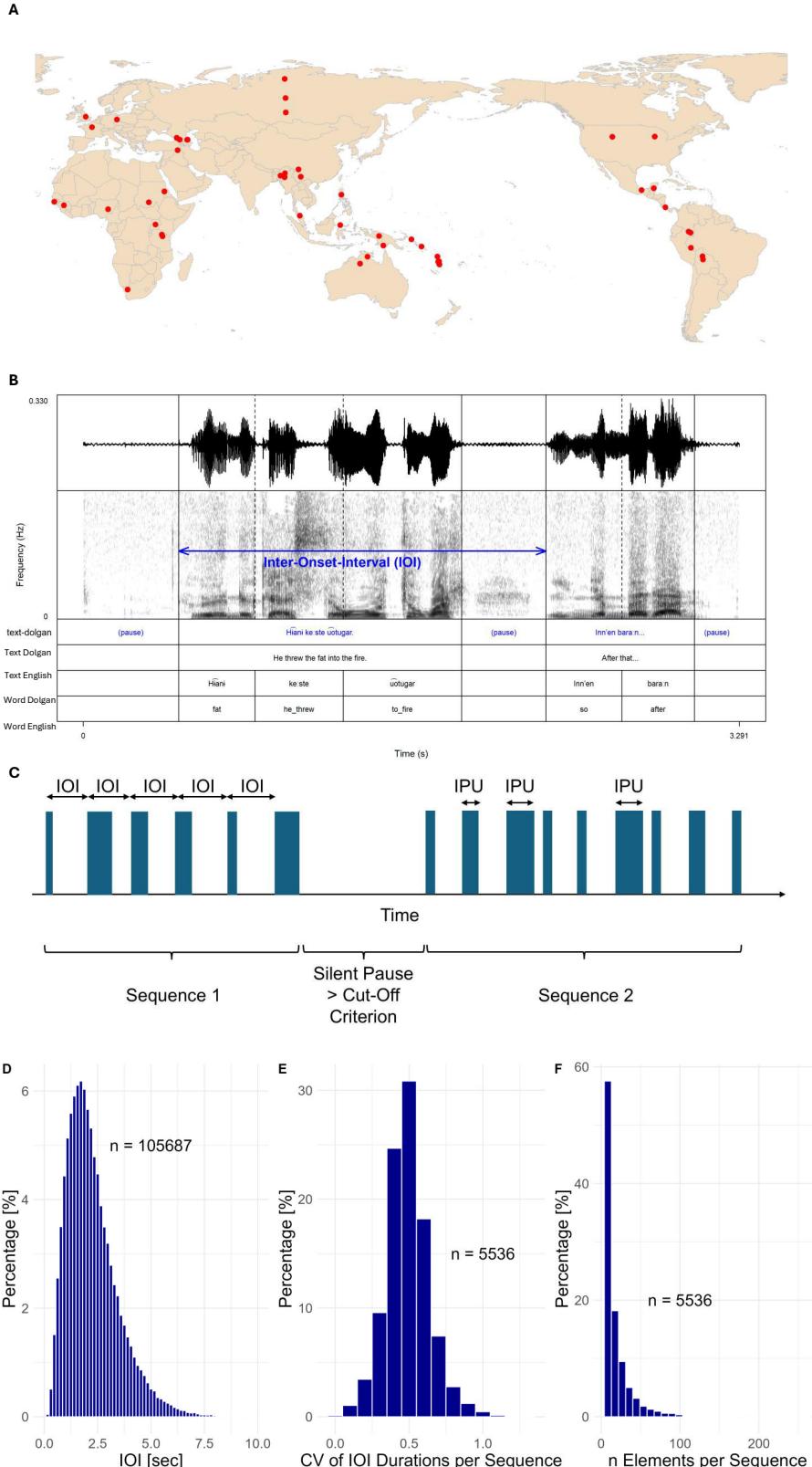


Figure 1: 1A: Map showing the approximate locations of the 49 languages included in our study. 1B: An illustration of an Inter-Onset Interval (IOI), defined as spanning an Inter-Pausal Unit (IPU) and a following speech pause, with an image of a recording from the Dolgan DoReCo dataset<sup>47</sup>. 1C: A recording with two sequences, split because of a silent pause exceeding the 3.36 seconds' threshold, with illustrations of IOIs and IPUs. 1DEF: Results for various rhythm measures based on our data.

age number of morphs per word type for 39 languages with morphological annotations. These calculations were performed on core datasets from the word-level CSV files in DoReCo 2.0. Pre-processing, analysis and plotting were performed in R<sup>52</sup> using RStudio<sup>53</sup> and Praatpicture<sup>54</sup>.

## Analysis

The data were prepared as individual files for each sequence. Sequences were automatically separated at silent breaks exceeding a certain threshold. The threshold was calculated as the 99th percentile of silent breaks across the entire dataset, which consisted of more than 100,000 IPUs and silent breaks thereafter. The data-driven duration of silent breaks used to define the start of a new sequence was 3.36 seconds of silence (Figure S1). Sequences were also separated at excluded intervals (e.g., due to disfluencies). Rhythm analysis was performed with the [shiny app RANTO](#)<sup>42,14,12</sup>. We calculated IOI beats based on the median IOI duration in a sequence transformed into a frequency using the equation in 1.

$$IOI_{beat} = \frac{1}{\widetilde{IOI}} \quad (1)$$

There can be one or multiple sequences per recording, depending on pauses and filtered elements. For this reason, the number of sequences analyzed for a given speaker varies across recordings. Consequently, the IOI beat data points are independent of each other to varying degrees, a fact that has to be taken into account when interpreting the results. Beat Precision values per sequence were calculated following<sup>14</sup> (Equation 2), with the adjustments in Equation 3<sup>15</sup>.

$$bp = \frac{|\Delta|}{\Delta_{\max}} \quad (2)$$

$$bp_t = 0.5 - |bp - 0.5| \quad (3)$$

Beat Precision indicates how well a given isochronous model fits the raw IOI sequence. It does so by measuring the deviation of each onset from the closest theoretical beat, relative to the maximum possible deviation (half a beat length). This produces a value per element between 0 and 1. These values are then summarized per sequence and transformed according to<sup>15</sup> (Equation 2), since an average value of 0.1 per sequence would indicate the same very good fit, as an average value of 0.9, just with a phase shift. After transformation, a value of 0.9 would also be 0.1. Consequently, all values range from 0 to 0.5, where 0 represents a perfect fit and 0.5 a highly variable fit across a sequence. While Beat Precision is still a

method under development, we find it to be a useful indicator within the context of this study.

## Statistics

When comparing different data groups, we calculated the effect size Cohen's  $d$ (eltha). Furthermore, we computed Pearson's correlation coefficient  $r$  for the relationship between the IOI beat and speaker age as well as morphological synthesis. The reported  $R^2$  values correspond to the square of  $r$ .

## Results

We analyzed a total of 5,536 sequences consisting of 104,628 IOIs produced by 300 speakers across 49 languages. The results show consistent interval durations across all languages, with a slightly right-skewed unimodal distribution (Figure 1D). We find a median of approximately 2 seconds, corresponding to a rhythm of 0.5 Hz. Based on the intervals, we calculated the Coefficient of Variation (CV) per sequence (Figure 1E), which reveals a considerable degree of similarity in intervals between languages. The density distributions of Inter-Onset Intervals are strikingly similar between languages and differ clearly from a randomly generated interval distribution (Figure 2A; see Methods for details on how random intervals were calculated). Median interval durations across the 49 languages are also very similar in both shape and range (Figure 2B). Median interval durations ranged from 1.3 seconds (Sümi) to 2.7 seconds (Nisvai). 63 percent of the analyzed languages show median IOI durations between 1.8 and 2.2 seconds, with linguistic families and areas being evenly distributed throughout the continuum. The highest two peaks of the density plots correspond to Sümi and Cabécar, while the curve with a low peak to the right at  $\sim 2.5$  seconds corresponds to Asimjeege Datoga (Figure 2A). Note that density plots are dependent on sample size and languages differ in sample size.

A similar pattern emerges for CV values of sequences. The distribution of sequence-level CV values is unimodal, peaking at 0.5 and showing no clear skew in either direction. CV values are largely similar across most languages, with outliers corresponding to languages with small sample sizes or few speakers (Sümi, Yali, Ruuli; see Figure S3 and Table 2).

We tested four additional parameters that may potentially account for some of the observed variations in IOI durations: speaker age and morphological synthesis (as continuous variables) and speaker gender and the distinction between tonal and non-tonal languages (as categorical variables). We found no correlation between IOI and speaker age (Figure 3A). Morphological synthesis did not show any correlation with

IOI duration either (Figure 3B). For the categorical variables, we calculated effect sizes. We found only negligible effects for male vs. female speakers (Cohen's  $d = -0.1$ , Figure 3C) and tonal vs. non-tonal languages (Cohen's  $d = -0.06$ , Figure 3D).

Since IOI distributions were unimodal with mostly strong peaks and CV values were comparable to values found in animal sequences, we tested whether an underlying isochronous rhythm might be present (Table 1) – that is, whether IOI sequences could be modeled with an isochronous beat. To this end, we calculated beat frequencies based on the median interval per sequence (1/median IOI, see Methods) as an isochronous model describing the rhythm of a particular sequence. The sequence length of successive IOI intervals lies between 4 and 266 elements per sequence, with a median of 11 elements (Figure 1F). Beat frequencies for the 5,536 sequences across 49 languages follow a normal distribution (Figure S2A), with a mean of 0.5 Hz and a standard deviation of 0.08 Hz. Rhythms ranged from an average of 0.38 Hz in Nisvai and Asimjeeg Datooga to 0.76 Hz in Sümi, and were generally similar across languages (Figure S2B). At an individual speaker level, rhythms ranged from 0.25 Hz (across five sequences) to 0.8 Hz (over three sequences). On average, there were 18 sequences per speaker (median: 10, min: 1, max: 277).

To assess how well the modeled isochronous rhythms fit the IOI sequences, we calculated Beat Precision values<sup>14,15</sup>. Beat Precision is reported as a summary value per sequence. A value of 0 indicates a perfect fit of the isochronous model to the IOI sequence, while a value of 0.5 indicates maximal deviation (i.e., strong onset jitter). Beat Precision values range from 0.0136 to 0.5, with a median of 0.40. Languages showed relatively little variation in these values (Figure S4).

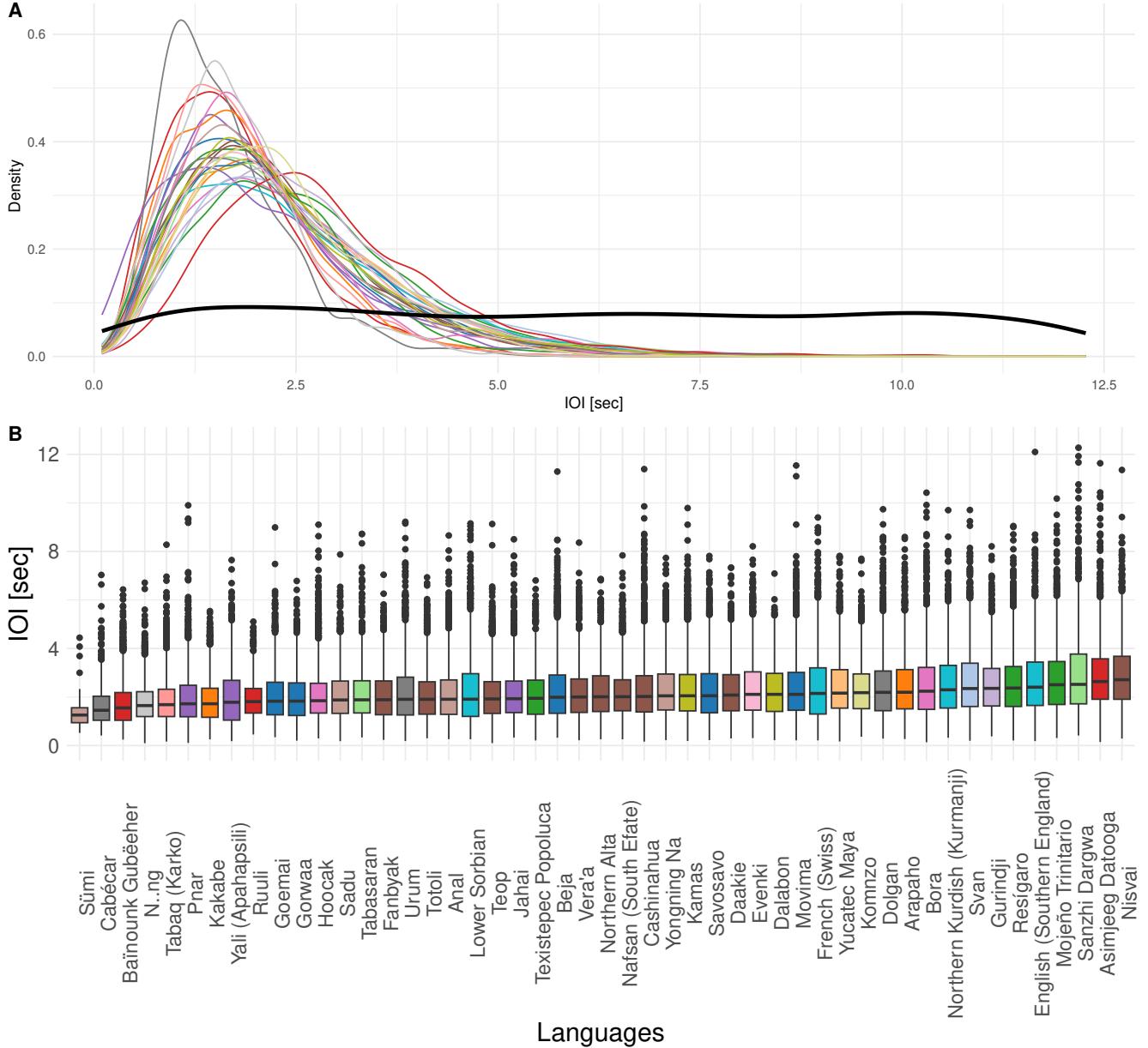


Figure 2: IOI durations are similar across languages and language families. (A) Density plots of intervals of all 49 languages. The black line represents the density distribution of a random distribution of IOIs within the same general range of durations. Colors represent language families. (B) Boxplot of IOI durations for all languages, sorted by median duration. Median IOI durations and variability look very similar for most languages.

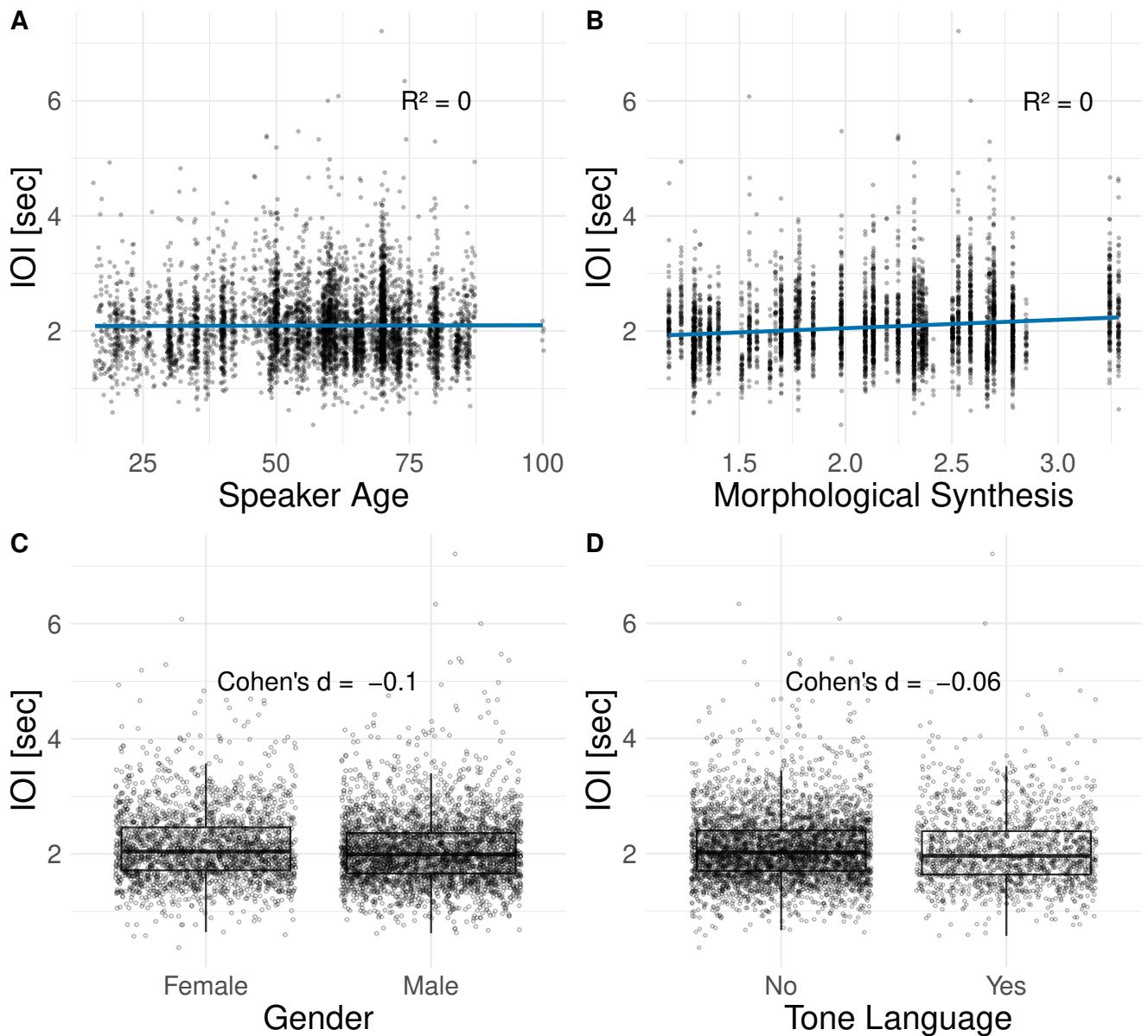


Figure 3: IOI durations are similar across age groups (A), genders (C), languages with various levels of morphological complexity (B), and between languages with and without lexical tone (D).

## Discussion

Surprisingly, speakers of languages from vastly different areas and genealogical families showed a high degree of uniformity in Inter-Onset Intervals based on Inter-Pausal Units. Across all languages, these intervals followed a slightly right-skewed unimodal distribution, with a median of 2 seconds, corresponding to a rhythm of 0.5 Hz. 63 percent of the analyzed languages showed median IOI durations between 1.8 and 2.2 seconds, with linguistic families and areas evenly distributed along this continuum (Figure 2). Median interval durations ranged from 1.3 seconds (Sümi) to 2.7 seconds (Nisvai). Neither the ‘gender’ and ‘age’ factors at the speaker level nor the typological factors ‘morphological synthesis’ and ‘tone language’ at the language level had a notable impact on IOI durations. Therefore, an underlying isochronous rhythm of approximately 0.5 Hz could be assumed for Inter-Onset Intervals in human speech, mirroring observations of isochronous rhythms in nonhuman animals. However, our Beat Precision results suggest that the strong similarities in interval durations do not translate into a consistent rhythm comparable to that of nonhuman animals. Beat Precision values in our sample range from 0.0136 to 0.5, with a median of 0.4 (Figure S4). These results indicate that while some sequences closely align with an underlying isochronous rhythm, the majority deviate from it. It is important to interpret these findings with caution, as beat precision analysis is still under active development<sup>15</sup>. Recent methodological advances further complicate direct comparisons across species. Nevertheless, a median Beat Precision of 0.4 does not support a general characterization of human speech as following an isochronous rhythm.

Although IOIs showed a high degree of uniformity across languages and speakers, some variation was still observed that warrants further investigation. This variation may be influenced by a range of linguistic and speaker-related factors. Physiological factors including body height and lung volume may be among the reasons underlying this variation.

Certain physiological features are closely tied to genetic and environmental factors; for example, populations living at high altitudes, such as in mountainous regions, generally exhibit adaptations in lung volume<sup>55</sup>.

We did not include physiological or altitude data as factors in our study as we did not have access to the relevant metadata for all speakers. Our study did not find strong effects for gender or age, despite previous reports of differences in speech rates between male and female and between younger and older individuals<sup>35,36,37,38</sup>. The most likely explanation for this is that we investigated Inter-Onset Intervals within sequences, whereas speech rate is typically measured in terms of linguistic units such as syllables

or segments. Similarly, the broad grammatical and prosodic profiles of languages, operationalized here as the degree of morphological synthesis and the presence or absence of lexical tone, did not show an effect on IOIs. As the number of sequences per speaker varied substantially, the distributions of sequences per language and speaker must be considered when interpreting the results. It is sometimes impossible to disentangle individual speaker rhythms and language rhythms when data from a limited number of speakers are available, as in the case of Kamas and Sümi. The Sümi dataset showed a fast median IOI beat at 0.76 Hz while the Kamas dataset showed a median IOI beat of 0.48 Hz, but for both datasets, only a single speaker was included in our sample.

In addition to measuring rhythm in human speech in terms of IOIs, we also calculated various related rhythm metrics. Table 1 summarizes results from recent studies on nonhuman animals using the same methodology applied to human speech data here. As indicated earlier, many more animal species are producing extremely consistent rhythmic signals, the currently known “champion” being male fireflies of the species *Pteroptyx malaccae*, which show coefficients of variation across flashing displays of as low as 0.004<sup>56,10</sup>. In our comparison, we chose to only include acoustic communication signals that were measured using the exact methodology that we applied to human data. Our findings indicate that an underlying isochronous pattern is not present in human speech when considering Beat Precision in contrast to what we hypothesized and to what is often found in animal species. This will at least in part be due to these acoustic communication signals differing immensely in production mechanisms, functions, and environmental constraints. While human speech rhythms at this level are among the slowest observed rhythms, the Coefficient of Variation as an independent measure of variability within a sequence is comparable to that of many nonhuman animals (Table 1). However, it is important to recognize that the underlying mechanisms governing rhythmic signal production are diverse and often not directly comparable across species. For example, birds use multiple inflatable air sacs and a syrinx to produce vocalizations, whereas humans rely on the lungs and larynx; whales transmit sound waves through water, a medium distinct from air. Similar to other complex behaviors, such as vocal learning<sup>57</sup>, rhythmicity might be better understood as a continuum or modular system<sup>15</sup> rather than as a binary trait.

The variability of intervals within a sequence of sound elements was similar between our human speech data and certain animal communication signals, independently of the actual beats and how well an isochronous model fits the data. Our sequence lengths were also comparable to many of the species for which data are currently available. With an average CV of 0.5, the human Coefficient of Variation compares to some echolocation call sequences of different bat species (where only the slower, consistent

Class	Species	IOI Duration [sec]	CV Mean	CV Median	npvi Mean	npvi Median	Beat [Hz] Mean or Range	Notes	Source
Mammals	Humans	2.23	0.5	0.5	67.8	60.1	0.5		This study
Amphibia	Eastern Sedgefrog						0.8	Chorus	6
Amphibia	Wallum Sedgefrog						0.8	Chorus	6
Aves	Skylark	0.18	1.3	0.7			6.9	Flightsong	14*
Osteichthyes	Brown Meagre			0.16		16.7			58
Osteichthyes	Scorpaena spp.		0.8		90.9		0.7	Kwa	58
Mammals	Cape Fur Seal	0.32	0.3		36.3		2.9	Seal Pup	59
Mammals		0.37	0.08		9.7		3.2	Adult Female Seal	59
Mammals	<i>Carollia perspicillata</i>	0.04	0.2		35.9			Isolation Calls	42
		0.06	0.8				9.8 – 43.2	Echolocation	60
Mammals	<i>Desmodus rotundus</i>	0.08	0.6				7.4 – 30.5	Echolocation	60
Mammals	<i>Glossophaga soricina</i>	0.03	0.6				21.9 – 38.9	Echolocation	60
Mammals	<i>Lonchorhina aurita</i>	0.09	0.5				4.8 – 19.0	Echolocation	60
Mammals	<i>Molossus molossus</i>	0.12	0.3				5.2 – 12.5	Echolocation	60
Mammals	<i>Myotis nigricans</i>	0.08	0.3				10.2 – 18.2	Echolocation	60
Mammals	<i>Phyllostomus hastatus</i>	0.06	0.5				10.4 – 22.5	Echolocation	60
Mammals	<i>Pteronotus parnellii</i>	0.06	0.4				10.4 – 27	Echolocation	60
Mammals	<i>Rhynchoycteris naso</i>	0.05	0.2				15.1-24.2	Echolocation	60
Mammals	<i>Saccopteryx bilineata</i>	0.08	0.2			~22.8		Isolation Calls	42
		0.12	0.3				3.7 – 13	Echolocation	60
Mammals	<i>Saccopteryx leptura</i>	0.06	0.4				13.4 – 32.4	Echolocation	60
Mammals	Sperm Whale	0.46	0.1			~5.2		Click Coda	42
Mammals	<i>Thyroptera tricolour</i>	0.009	0.2				86 – 144.4	Echolocation	60

Table 1: Rhythmic parameters calculated in this study compared to other animal species. Note that various rhythm parameters have been calculated for more species (see for example<sup>11</sup>). Here we limit ourselves strictly to studies investigating acoustic communication using the same methodology as our study. Echolocation call sequences only considered the slower and rhythm-wise more consistent search phase, the same is true for the click coda of the Sperm whale. \*The data in this table was calculated from the open-access dataset from this study.

search phase was considered<sup>60</sup>). At one end of the variability spectrum from this limited group of species (see<sup>11,10</sup> for more examples), we have the very low Coefficient of Variation in barks of adult Cape fur seal females with  $CV = 0.08^{59}$  or the click coda of a female sperm whale with  $CV = 0.14^{42}$ , which are both discussed in the literature as highly rhythmic and stereotyped and consist of only one element type repeated in sequences. At the other end of the spectrum is the complex and long flight song of skylarks ( $CV = 1.27,^{14}$ ), which is made up of several motifs and up to 300 different syllables (in the biological sense) per individual, giving rise to enormous variability in syntactic combinations and, therefore, to high variability in intervals<sup>61</sup>. Humans also have enormous variability in the size of speech chunks and their combinatorial possibilities; therefore, it should be noted that the CV in this specific skylark flight song dataset is more than double the CV for human speech. Nevertheless, the analyzed sequences of skylark flight songs were mostly considerably longer than the analyzed speech sequences, which most likely increases variability. Our study adds a new data point to the ongoing effort for increased inter-species comparability to uncover the evolutionary origins of rhythms in communication signals<sup>12</sup>, showing an example where, even though a strong unimodal distribution of IOIs is found, underlying isochronous models are not generally applicable.

The unit of our study was the Inter-Onset Interval, i.e., uninterrupted stretches of speech and a following silent pause. At this level, languages are far less heterogeneous than at smaller units such as words or syllables (in the linguistic sense). This finding aligns with previous reports that intonation units

(IU's) exhibit shared rhythms across five languages. An IU rhythm of ca. 1 Hz is reported in<sup>39</sup>, indicating average durations of ~ 1 second. This IU rhythm is twice as fast as the IOIs measured here. The crucial difference between IUs and IOIs is that the former are embedded in language-specific phonologies, forming part of the prosodic hierarchy<sup>62</sup>, and have no clear correlate in nonhuman animal vocalization. IOIs, on the other hand, are defined in terms of sequences of vocalized and pause intervals that constitute universal building blocks of both human and nonhuman communication and thus allow for meaningful cross-species comparisons. Studies on speech pauses often report a bimodal distribution for pause durations, with shorter pauses occurring after simple clauses and longer pauses after more complex phrases<sup>63,64</sup>. However, this does not translate into a bimodal distribution for IOI durations in our data (Figure 1D). We also did not find this biomodal distribution in pause durations between IPUs (Figure S1).

Our study included data from recorded monologues and no recordings of conversations. This decision was informed by two considerations. First, the DoReCo corpus primarily comprises recordings from non-WEIRD populations<sup>49</sup>, where oral storytelling plays a central role in cultural transmission. These communities often lack the writing traditions characteristic of modern Western societies. Storytelling is a fundamental aspect of human communication<sup>65</sup>, and therefore, recorded monologues can be considered a natural human vocal signal. Second, our comparison with nonhuman animal sounds is focused around self-communication (echo location, but also directed song without an immediate response). We intentionally leave out duetting and chorusing, where effects such as synchronization, coupling, and entrainment strongly influence rhythmic patterns<sup>10</sup>. Turn-taking in human conversation, on the other hand, involves a range of additional factors such as pitch, voice quality, and discourse particles<sup>66</sup>. We would expect to find a higher degree of variability of rhythms within human conversational data, but it is beyond the scope of this article to attempt a meaningful comparison with rhythms found in duetting or chorusing.

The temporal window of intervals identified in our study may serve as critical anchors for multimodal communication. Consistent IOI durations of around 2 seconds could provide an organizing framework that integrates slower body language movements<sup>67</sup> with faster speech components during complex multimodal communication. Further research is needed to explore the connection between speech rhythm and multimodality.

A key factor that may explain consistencies in Inter-Onset Intervals in nonhuman mammals and humans is respiratory rhythm. Respiration is a fundamental driver of speech production, providing the expiratory airstream for sound production, separated by brief periods of inhalation<sup>68</sup>. The temporal scope of respiratory cycles may set an upper boundary for continuous vocalization. Breathing cycles in humans

are approximately 4 seconds on average, with great flexibility. Breath groups, i.e., speech events occurring on exhalation in spontaneous speech, have been reported to range between 0.9–14 seconds or 0.3–12.6 seconds<sup>69,70</sup>, broadly matching the range of IOI durations observed in our data. There is a notable lack of cross-linguistic data on speech breathing cycles. Breath groups have been proposed as a phonological feature by which an increase in crico-thyroid muscular activity marks the onset of prosodic units<sup>71</sup>. Breath groups have been proposed as a phonological feature, with increased crico-thyroid muscular activity marking the onset of prosodic units<sup>71</sup>. Like intonation units, these can correspond to syntactic or semantic units, though mismatches are common<sup>72</sup>.

Most existing data on breathing derive from controlled laboratory recordings involving participants from WEIRD populations<sup>49</sup> performing reading tasks or constrained speaking activities. In contrast, speakers in the DoReCo corpus represent a substantially more diverse population sample, encompassing a broader range of spoken languages, age groups (the median age of our sample is 56 years), and cardiovascular health, among other factors. This suggests that physiologically constrained breathing cycles are one possible explanation for the cross-linguistic similarities in IOI durations.

Another possible explanation for the observed IOI durations is related to the capacity of working memory, which typically diminishes after 2–3 seconds<sup>73</sup>. This temporal window aligns closely with the average IOI duration of about 2 seconds found here and may reflect cognitive planning constraints or the temporal span needed for predicting upcoming speech chunks.

While links to brain oscillations—particularly in the slow theta range—are plausible, the literature does not consistently identify a slow rhythm near 0.5 Hz. Rhythms at or below 0.5 Hz are sometimes termed “infra-rhythm,” a temporal scale too slow to be integrated as rhythmic neural oscillations<sup>29</sup>.

However, speech onsets following pauses, which are the time points that were important in our analysis, may be directly linked to brain activation<sup>74</sup>. The exact relation between IOI durations and cognitive or neural functions requires more attention in future studies.

In conclusion, we report that using a comparable data basis and methodology, 49 diverse languages from all six inhabited continents show strong and unexpected similarities concerning their temporal structure, with similar durations of Inter-Onset Intervals across all analyzed languages. Despite those shared durations, no evidence for an underlying isochronous rhythm was found, as idealized isochronous models poorly fit the IOI sequences. Instead, our results showcase strong flexibility of human rhythm production abilities. While the overall variability of IOIs is similar in humans and certain animal species, there appear to be striking differences regarding the presence or absence of underlying isochrony. By revealing

cross-linguistic regularities alongside a lack of isochrony, our results help to place human speech within the wider spectrum of rhythmic strategies observed in nature.

## Data, Materials, and Software Availability

All relevant scripts and files are stored on [GitHub](#), including the scripts for data processing, plotting, and analysis. The input data are taken from the DoReCo corpus<sup>48</sup>.

## Acknowledgments

We thank Lena Henke for discussions on rhythm and working memory. This work was supported by grants from the German Research Foundation to L.S.B. (DFG BU 4375/1-1, project number: 528064681), L.P. (DFG PA 2368/1-1, project number: 496083013), S.F. (DFG FU791-9-1, project number: 502013782) as well as by the IMéRA and ILCB of Aix-Marseille University.

	Language	Glottocode	Family	Speakers	Files	Sequences	IOI Units	Source
1.	Anal	anal1239	Sino-Tibetan	8	21	225	3179	75
2.	Arapaho	arap1274	Algic	4	10	132	1024	76
3.	Asimjeeg Datooga	tsim1256	Nilotic	13	40	84	1929	77
4.	Baïnouunk Gubééher	bain1259	Atlantic-Congo	5	13	136	3190	78
5.	Beja	beja1238	Afro-Asiatic	5	55	95	3565	79
6.	Bora	bora1263	Boran	6	9	176	2618	80
7.	Cabécar	cabe1245	Chibchan	6	12	42	874	81
8.	Cashinahua	cash1254	Pano-Tacanan	3	7	127	2868	82
9.	Daakie	port1286	Austronesian	11	16	83	2235	83
10.	Dalabon	ngal1292	Guwinyguan	3	6	88	994	84
11.	Dolgan	dolg1241	Turkic	6	8	71	2199	47
12.	English (Southern England)	sout3282	Indo-European	2	4	64	1245	85
13.	Evenki	even1259	Tungusic	22	35	256	3559	86
14.	Fanbyak	orko1234	Austronesian	8	16	132	1595	87
15.	French (Swiss)	stan1290	Indo-European	8	8	129	1957	88
16.	Goemai	goem1240	Afro-Asiatic	3	9	57	1684	89
17.	Gorwaa	goro1270	Afro-Asiatic	3	5	39	1402	90
18.	Gurindji	guri1247	Pama-Nyungan	3	13	119	1644	91
19.	Hoocák	hoch1243	Siouan	9	18	155	2508	92
20.	Jahai	jeha1242	Austroasiatic	3	9	105	2327	93
21.	Kakabe	kaka1265	Mande	4	9	48	2471	94
22.	Kamas	kama1351	Uralic	1	26	277	3942	95
23.	Komnzo	komn1238	Yam	10	17	60	2265	96
24.	Lower Sorbian	lowe1385	Indo-European	4	15	157	2257	97
25.	Mojeño Trinitario	trin1278	Arawakan	5	8	134	2012	98
26.	Movima	movi1243	Isolate	5	12	172	2881	99
27.	Nafsan (South Efaté)	sout2856	Austronesian	13	30	169	2785	100
28.	Nisvai	nisv1234	Austronesian	8	12	57	2047	101
29.	Northern Alta	nort2875	Austronesian	2	3	101	1621	102
30.	Northern Kurdish	nort2641	Indo-European	2	5	34	1574	103
31.	N̩ng	nngg1234	Tuu	6	13	178	2144	104
32.	Pnar	pnar1238	Austroasiatic	4	6	48	1457	105
33.	Resígaro	resi1247	Arawakan	3	18	117	3134	106
34.	Ruuli	ruul1235	Atlantic-Congo	3	3	35	703	107
35.	Sadu	sadu1234	Sino-Tibetan	6	13	73	1719	108
36.	Sanzhi Dargwa	sanz1248	Nakh-Daghestanian	5	9	53	1156	109
37.	Savosavo	savo1255	Isolate	7	7	86	2633	110
38.	Svan	svan1243	Kartvelian	9	28	89	2174	111
39.	Sümi	sumi1235	Sino-Tibetan	1	1	9	81	112
40.	Tabaq (Karko)	kark1256	Nubian	4	16	139	3285	113
41.	Tabasaran	taba1259	Nakh-Daghestanian	2	5	64	1306	114
42.	Teop	teop1238	Austronesian	11	11	117	2517	115
43.	Texistepec Popoluca	texi1237	Mixe-Zoque	1	7	230	2716	116
44.	Totoli	toto1304	Austronesian	7	10	154	2071	117
45.	Urum	urum1249	Turkic	30	129	290	3320	118
46.	Vera'a	vera1241	Austronesian	7	8	80	2207	119
47.	Yali (Apahapsili)	apah1238	Nuclear TNG	2	3	66	1024	120
48.	Yongning Na	yong1270	Sino-Tibetan	1	6	80	1935	121
49.	Yucatec Maya	yuca1254	Mayan	6	10	104	1946	122
	Sum			300	754	5536	104628	

Table 2: Languages included in our sample with unique language identification code (“Glottocode”), top-level family as indicated in Glottolog<sup>51</sup>, number of speakers (after filtering), number of files (after splitting), number of IOI units, and references for individual datasets.

## **Author contributions**

LSB: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Software, Visualization, Writing — Original Draft Preparation, Writing — Review & Editing  
LP: Conceptualization, Data Curation, Funding Acquisition, Investigation, Writing -- Original Draft Preparation, Writing -- Review & Editing  
SF: Conceptualization, Funding Acquisition, Investigation, Visualization, Writing -- Original Draft Preparation, Writing — Review & Editing

## **Authors' ORCID**

LSB: [0000-0002-9210-7934](#)

LP: [0000-0001-7176-8576](#)

SF: [0000-0001-6751-9286](#)

## **Competing interests**

The authors declare that they have no competing interests.

## References

- [1] Francis Nolan and Hae-Sung Jeon. “Speech rhythm: a metaphor?” In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1658 (2014), p. 20130396. DOI: [10.1098/rstb.2013.0396](https://doi.org/10.1098/rstb.2013.0396).
- [2] Fred Cummins. “Rhythm and Speech”. In: *The Handbook of Speech Production*. Ed. by Melissa A. Redford. Wiley Online Library, 2015, pp. 158–177.
- [3] Franck Ramus, Marina Nespor, and Jacques Mehler. “Correlates of linguistic rhythm in the speech signal”. In: *Cognition* 73.3 (1999), pp. 265–292. DOI: [10.1016/S0010-0277\(99\)00058-X](https://doi.org/10.1016/S0010-0277(99)00058-X).
- [4] Philipp Norton and Constance Scharff. ““Bird song metronomics”: Isochronous organization of zebra finch song rhythm”. In: *Frontiers in Neuroscience* 10 (2016), p. 196177.
- [5] Chiara De Gregorio, Teresa Raimondi, Bevilacqua Valeria, Pertosa Chiara, Daria Valente, Filippo Carugati, Bandoli Francesca, Livio Favaro, Lefaux Brice, Andrea Ravignani, et al. “Isochronous singing in three crested gibbon species (*Nomascus spp.*)” In: *Current Zoology* (2023), pp. 00–00.
- [6] Alannah Filer, Lara S Burchardt, and Berndt J van Rensburg. “Assessing acoustic competition between sibling frog species using rhythm analysis”. In: *Ecology and Evolution* 11.13 (2021), pp. 8814–8830.
- [7] Taylor A Hersh, Andrea Ravignani, and Hal Whitehead. “Cetaceans are the next frontier for vocal rhythm research”. In: *Proceedings of the National Academy of Sciences* 121.25 (2024), e2313093121. DOI: [10.1111/2041-210X.14118](https://doi.org/10.1111/2041-210X.14118).
- [8] Lara S Burchardt, Philipp Norton, Oliver Behr, Constance Scharff, and Mirjam Knörnschild. “General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*”. In: *Royal Society Open Science* 6.1 (2019), p. 181076.
- [9] Marta Picciulin, Marta Bolgan, and Lara S Burchardt. “Rhythmic properties of *Sciaena umbra* calls across space and time in the Mediterranean Sea”. In: *Plos one* 19.2 (2024), e0295589.
- [10] M. D. Greenfield and B. Merker. “Coordinated rhythms in animal species, including humans: Entrainment from bushcricket chorusing to the philharmonic orchestra”. In: *Neuroscience and Biobehavioral Reviews* 153 (2023), p. 105382. DOI: [10.1016/j.neubiorev.2023.105382](https://doi.org/10.1016/j.neubiorev.2023.105382).

- [11] Gabriela-Alina Sauciuc, Jennifer N Schneider, Lara S Burchardt, Michael D Greenfield, Gerlinde Höbel, and Nao Ota. *The functional and structural diversity of rhythmic communication across species*. OSF [Preprint], Mar. 2025. DOI: [10.31219/osf.io/rqsg5\\_v1](https://doi.org/10.31219/osf.io/rqsg5_v1). URL: [osf.io/rqsg5\\_v1](https://osf.io/rqsg5_v1).
- [12] Taylor A Hersh, Andrea Ravignani, and Lara S Burchardt. “Robust rhythm reporting will advance ecological and evolutionary research”. In: *Methods in Ecology and Evolution* 14.6 (2023), pp. 1398–1407.
- [13] Lara S. Burchardt, Marta Picciulin, Eric Parmentier, and Marta Bolgan. “A primer on rhythm quantification for fish sounds: a Mediterranean case study”. In: *Royal Society Open Science* 8.9 (2021), p. 210494. DOI: [10.1098/rsos.210494](https://doi.org/10.1098/rsos.210494).
- [14] Lara S. Burchardt, Elodie F. Briefer, and Mirjam Knörnschild. “Novel ideas to further expand the applicability of rhythm analysis”. In: *Ecology and Evolution* 11.24 (2021), pp. 18229–18237. DOI: [10.1002/ece3.8417](https://doi.org/10.1002/ece3.8417). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.8417>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.8417>.
- [15] Lara S Burchardt, Marisa Hoeschele, Henkjan Honing, Edward Large, Bjorn Merker, and Jennifer N Schneider. *Chapter 1: What is rhythm and how can we study it?* OSF [Preprint], 2025. DOI: [10.31219/osf.io/69v5a\\_v1](https://doi.org/10.31219/osf.io/69v5a_v1). URL: [osf.io/69v5a\\_v1](https://osf.io/69v5a_v1).
- [16] Marisa Hoeschele, Bernhard Wagner, and Dan C Mann. “Lessons learned in animal acoustic cognition through comparisons with humans”. In: *Animal Cognition* 26.1 (2023), pp. 97–116.
- [17] Sam Tilsen and Amalia Arvaniti. “Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages”. In: *The Journal of the Acoustical Society of America* 134.1 (2013), pp. 628–639.
- [18] David Poeppel and M Florencia Assaneo. “Speech rhythms and their neural foundations”. In: *Nature Reviews Neuroscience* 21.6 (2020), pp. 322–334. DOI: [10.1038/s41583-020-0304-4](https://doi.org/10.1038/s41583-020-0304-4).
- [19] Ahana A Fernandez, Lara S Burchardt, Martina Nagy, and Mirjam Knörnschild. “Babbling in a vocal learning bat resembles human infant babbling”. In: *Science* 373.6557 (2021), pp. 923–926.
- [20] Brigitte Bigi and Daniel Hirst. “Speech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody”. In: *Proc. Speech Prosody 2012*. 2012, pp. 19–22.
- [21] Kenneth L. Pike. *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press, 1945.

- [22] David Abercrombie. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967.
- [23] Robert F. Port, Jonathan Dalby, and Michael O'Dell. "Evidence for mora timing in Japanese". In: *The Journal of the Acoustical Society of America* 81.5 (1987), pp. 1574–1585. DOI: [10.1121/1.394510](https://doi.org/10.1121/1.394510).
- [24] Rebecca M Dauer. "Stress-timing and syllable-timing reanalyzed". In: *Journal of Phonetics* 11.1 (1983), pp. 51–62.
- [25] Marina Nespor, Mohinish Shukla, and Jacques Mehler. "Stress-timed vs. syllable-timed languages". In: *The Blackwell Companion to Phonology*. Ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice. John Wiley & Sons, 2011. Chap. 48, pp. 1147–1159. DOI: [10.1002/9781444335262.wbctp0048](https://doi.org/10.1002/9781444335262.wbctp0048).
- [26] Ilse Lehiste. "Isochrony reconsidered". In: *Journal of Phonetics* 5.3 (1977), pp. 253–263.
- [27] Amalia Arvaniti. "Rhythm, timing and the timing of rhythm". In: *Phonetica* 66.1–2 (2009), pp. 46–63. DOI: [10.1159/000208930](https://doi.org/10.1159/000208930).
- [28] Tamara V. Rathcke and Rachel H. Smith. "Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies". In: *J. Acoust. Soc. Am.* 137 (2015), pp. 2834–2845. DOI: [10.1121/1.4919322](https://doi.org/10.1121/1.4919322).
- [29] Aviad Albert and Martine Grice. "Rhythm is a timescale". In: *Rhythms of Speech and Language: Culture, Cognition, and the Brain*. Ed. by Lars Meyer and Antje Strauss. Cambridge University Press, forth.
- [30] Sun-Ah Jun. "Prosodic typology: by prominence type, word prosody, and macro-rhythm". In: *Prosodic typology II*. Oxford University Press, 2014.
- [31] L. Polyanskaya, M. G. Busà, and M. Ordin. "Capturing cross-linguistic differences in macro-rhythm: the case of Italian and English". In: *Language and Speech* 63.2 (2020), pp. 242–263. DOI: [10.1177/0023830919835849](https://doi.org/10.1177/0023830919835849).
- [32] Constantijn Kaland. "Bending the string: intonation contour length as a correlate of macro-rhythm". In: *Proc. Interspeech 2022*. 2022, pp. 5233–5237. DOI: [10.21437/Interspeech.2022-185](https://doi.org/10.21437/Interspeech.2022-185).
- [33] Christine Prechtel. "A cross-linguistic comparison of lexical stress strength and macro-rhythm strength". PhD thesis. University of California, Los Angeles, 2023.

- [34] Daffyd Gibbon. “The rhythms of rhythm”. In: *Journal of the International Phonetic Association* 53.1 (2023), pp. 233–265. DOI: [10.1017/S0025100321000086](https://doi.org/10.1017/S0025100321000086).
- [35] Mary Fitzsimons, Noirin Sheahan, and Hugh Staunton. “Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies”. In: *Brain and Language* 78 (2001), pp. 94–108. DOI: [10.1006/brln.2000.2448](https://doi.org/10.1006/brln.2000.2448).
- [36] Jo Verhoeven, Guy De Pauw, Kloots, and Hanne. “Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands”. In: *Language and Speech* 47.3 (2004), pp. 297–308. DOI: [10.1177/00238309040470030401](https://doi.org/10.1177/00238309040470030401).
- [37] Svetlana Stepanova. “Speech rate as reflection of speaker’s social characteristics”. In: *Approaches to Slavic Interaction*. Ed. by Nadine Thielemann and Peter Kosta. John Benjamins, 2013, pp. 117–129. DOI: [10.1075/ds.20.09ste](https://doi.org/10.1075/ds.20.09ste).
- [38] Massimo Pettorino, Elisa Pellegrino, and Marta Maffia. “‘Young’ and ‘old’ voices: the prosodic auto-transplantation technique for speaker’s age recognition”. In: *Proc. Speech Prosody* 7. 2014, pp. 135–139. DOI: [10.21437/SPEECHPROSODY.2014-15](https://doi.org/10.21437/SPEECHPROSODY.2014-15).
- [39] Maya Inbar, Eitan Grossman, and Ayelet N Landau. “Sequences of Intonation Units form a  $\sim 1$  Hz rhythm”. In: *Scientific Reports* 10.1 (2020), p. 15846. DOI: [10.1038/s41598-020-72739-4](https://doi.org/10.1038/s41598-020-72739-4).
- [40] Ilse Lehiste. “The units of speech perception”. In: *Speech and Cortical Functioning*. Ed. by J. H. Gilbert. New York: Academic Press, 1972, pp. 187–236.
- [41] Andrea Ravignani, Simone Dalla Bella, Simone Falk, Christopher T. Kello, Florencia Noriega, and Sonja A. Kotz. “Rhythm in speech and animal vocalizations: a cross-species perspective”. In: *Annals of the New York Academy of Sciences* 1453.1 (2019), pp. 79–98. DOI: [10.1111/nyas.14166](https://doi.org/10.1111/nyas.14166). eprint: <https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.14166>. URL: <https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.14166>.
- [42] Lara S Burchardt and Mirjam Knörnschild. “Comparison of methods for rhythm analysis of complex animals’ acoustic signals”. In: *PLoS Computational Biology* 16.4 (2020), e1007755.
- [43] Marisa Hoeschele, Hugo Merchant, Yukiko Kikuchi, Yuko Hattori, and Carel ten Cate. “Searching for the origins of musicality across species”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1664 (2015), p. 20140094. DOI: [10.1098/rstb.2014.0094](https://doi.org/10.1098/rstb.2014.0094).

- [44] Vlad Demartsev, Michal Haddas-Sasson, Amiyaal Ilany, Lee Koren, and Eli Geffen. “Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success”. In: *Journal of Animal Ecology* 92.8 (2023), pp. 1520–1531.
- [45] Robert Heinsohn, Christina N Zdenek, Ross B Cunningham, John A Endler, and Naomi E Langmore. “Tool-assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music”. In: *Science advances* 3.6 (2017), e1602399.
- [46] Nicolas Mathevon, Caroline Casey, Colleen Reichmuth, and Isabelle Charrier. “Northern elephant seals memorize the rhythm and timbre of their rivals’ voices”. In: *Current Biology* 27.15 (2017), pp. 2352–2356.
- [47] Chris Lasse Däbritz, Nina Kudryakova, Eugénie Stapert, and Alexandre Arkhipov. “Dolgan DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.f09eikq3](https://doi.org/10.34847/nkl.f09eikq3). URL: <https://doreco.humanum.fr/languages/dolg1241>.
- [48] Frank Seifart, Ludger Paschen, and Matthew Stave. *Language Documentation Reference Corpus (DoReCo)*. Version 2.0. Lyon, 2024. DOI: [10.34847/NKL.7CBFQ779](https://doi.org/10.34847/NKL.7CBFQ779).
- [49] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. “The weirdest people in the world?” In: *Behavioral and Brain Sciences* 33.2–3 (2010), pp. 61–83. DOI: [10.1017/S0140525X0999152X](https://doi.org/10.1017/S0140525X0999152X).
- [50] Ludger Paschen, François Delafontaine, Christoph Draxler, Susanne Fuchs, Matthew Stave, and Frank Seifart. “Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo)”. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: European Language Resources Association, 2020, pp. 2657–2666. URL: <https://aclanthology.org/2020.lrec-1.324>.
- [51] Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank. *Glottolog v5.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2024. DOI: [10.5281/zenodo.8131084](https://doi.org/10.5281/zenodo.8131084).
- [52] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria, 2024. URL: <https://www.R-project.org/>.

- [53] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio 2024.04.2 Build 764. 2024. URL: <https://www.rstudio.com/> (visited on 10/20/2024).
- [54] Rasmus Puggaard-Rode. “praatpicture. A library for making flexible Praat Picture-style figures in R”. In: *Proceedings of the 13th International Seminar on Speech Production (ISSP 2024), Autrans*. 2024, pp. 145–148.
- [55] Tarshi Droma, Rosann G. McCullough, Robert E. McCullough, Jianguo Zhuang, Allen Cymerman, Shinfu Sun, John R. Sutton, and Lorna G. Moore. “Increased vital and total lung capacities in Tibetan compared to Han residents of Lhasa (3,658 m)”. In: *American Journal of Physical Anthropology* 86.3 (1991), pp. 341–351. DOI: [10.1002/AJPA.1330860303](https://doi.org/10.1002/AJPA.1330860303).
- [56] John Buck and Elisabeth Buck. “Mechanism of Rhythmic Synchronous Flashing of Fireflies”. In: *Science* 159.3821 (1968), pp. 1319–1327. DOI: [10.1126/science.159.3821.1319](https://doi.org/10.1126/science.159.3821.1319).
- [57] Morgan Wirthlin, Edward F. Chang, Mirjam Knörnschild, Leah A. Krubitzer, Claudio V. Mello, Cory T. Miller, Andreas R. Pfenning, Sonja C. Vernes, Ofer Tchernichovski, and Michael M. Yartsev. “A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait”. In: *Neuron* 104.1 (2019), pp. 87–99. DOI: [10.1016/j.neuron.2019.09.036](https://doi.org/10.1016/j.neuron.2019.09.036).
- [58] Lara S. Burchardt, Marta Picciulin, Eric Parmentier, and Marta Bolgan. “A primer on rhythm quantification for fish sounds: a Mediterranean case study”. In: *Royal Society Open Science* 8.9 (2021), p. 210494. DOI: [10.1098/rsos.210494](https://doi.org/10.1098/rsos.210494).
- [59] Anna N Osiecka, Jack Feary, Andrea Ravignani, and Lara S Burchardt. “Isochrony in barks of Cape fur seal (*Arctocephalus pusillus pusillus*) pups and adults”. In: *Ecology and Evolution* 13.3 (2024), e11085.
- [60] Lara Sophie Burchardt. “Rhythm in Animals’ Acoustic Signals: Novel Methods for the Analysis of Rhythm Production and Perception on the Example of Bats, Birds, and Whales”. Dissertation. Freie Universität Berlin, 2021. URL: <http://dx.doi.org/10.17169/refubium-31257>.
- [61] Elodie F. Briefer, Tomasz S. Osiejuk, Fanny Rybak, and Thierry Aubin. “Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach”. In: *Journal of Theoretical Biology* 262.1 (2010), pp. 151–164. DOI: [10.1016/j.jtbi.2009.09.020](https://doi.org/10.1016/j.jtbi.2009.09.020).
- [62] Marina Nespor and Irene Vogel. *Prosodic phonology*. Dordrecht: Foris, 1986.

- [63] Jelena Krivokapić. “Prosodic planning: Effects of phrasal length and complexity on pause duration”. In: *Journal of Phonetics* 35 (2007), pp. 162–179. DOI: [10.1016/j.wocn.2006.04.001](https://doi.org/10.1016/j.wocn.2006.04.001).
- [64] Naomi Peck and Laura Becker. “Syntactic pausing? Re-examining the associations”. In: *Linguistics Vanguard* (2024). DOI: [10.1515/lingvan-2022-0156](https://doi.org/10.1515/lingvan-2022-0156).
- [65] John D Niles. *Homo Narrans: The Poetics and Anthropology of Oral Literature*. Philadelphia, PA: University of Pennsylvania Press, 1999.
- [66] Augustín Gravano and Julia Hirschberg. “Turn-taking cues in task-oriented dialogue”. In: *Computer Speech & Language* 25.3 (2011), pp. 601–634. DOI: [10.1016/j.csl.2010.10.003](https://doi.org/10.1016/j.csl.2010.10.003).
- [67] Britta Grimme, Susanne Fuchs, Pascal Perrier, and Gregor Schöner. “Limb versus speech motor control: A conceptual review”. In: *Motor control* 15.1 (2011), pp. 5–33.
- [68] Susanne Fuchs and Amélie Rochet-Capellan. “The respiratory foundations of spoken language”. In: *Annual Review of Linguistics* 7.1 (2021), pp. 13–30.
- [69] Alison L Winkworth, Pamela J Davis, Roger D Adams, and Elizabeth Ellis. “Breathing patterns during spontaneous speech”. In: *Journal of Speech, Language, and Hearing Research* 38.1 (1995), pp. 124–144.
- [70] Yu-Tsai Wang, Jordan R Green, Ignatius SB Nip, Ray D Kent, and Jane Finley Kent. “Breath group analysis for reading and spontaneous speech in healthy adults”. In: *Folia Phoniatrica et Logopaedica* 62.6 (2010), pp. 297–302.
- [71] Philip Lieberman, Masayuki Sawashima, Katherine S Harris, and Thomas Gay. “The articulatory implementation of the breath-group and prominence: Crico-thyroid muscular activity in intonation”. In: *Language* 46 (2 1970), pp. 312–327. DOI: [10.2307/412281](https://doi.org/10.2307/412281).
- [72] Nikolaus P Himmelmann. “Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference”. In: *Language* 90.4 (2014), pp. 927–960. DOI: [10.1353/lan.2014.0105](https://doi.org/10.1353/lan.2014.0105).
- [73] Lena Henke, Eleonora Beier, Mikael Roll, and Lars Meyer. “Cognitive and neural constraints on timing and rhythm in language”. In: *Rhythms of Speech and Language: Culture, Cognition, and the Brain*. Ed. by Lars Meyer and Antje Strauss. Cambridge University Press, forth.
- [74] Nikos Chalas, Christoph Daube, Daniel S Kluger, Omid Abbasi, Robert Nitsch, and Joachim Gross. “Speech onsets and sustained speech contribute differentially to delta and theta speech tracking in auditory cortex”. In: *Cerebral Cortex* 33.10 (2023), pp. 6273–6281.

- [75] Pavel Ozerov. “Anal DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.0dbazp8m](https://doi.org/10.34847/nkl.0dbazp8m). URL: <https://doreco.huma-num.fr/languages/anal1239>.
- [76] Andrew Cowell. “Arapaho DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.36f5r1b6](https://doi.org/10.34847/nkl.36f5r1b6). URL: <https://doreco.huma-num.fr/languages/arap1274>.
- [77] Richard Griscom. “Asimjeeg Datooga DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.f77c7m72](https://doi.org/10.34847/nkl.f77c7m72). URL: <https://doreco.huma-num.fr/languages/tsim1256>.
- [78] Alexander Yao Cobbinah. “Bainounk Gubéher DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.a332abw8](https://doi.org/10.34847/nkl.a332abw8). URL: <https://doreco.huma-num.fr/languages/bain1259>.
- [79] Martine Vanhove. “Beja DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.edd011t1](https://doi.org/10.34847/nkl.edd011t1). URL: <https://doreco.huma-num.fr/languages/beja1238>.
- [80] Frank Seifart. “Bora DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.6eaf5laq](https://doi.org/10.34847/nkl.6eaf5laq). URL: <https://doreco.huma-num.fr/languages/bora1263>.
- [81] Juan Diego Quesada, Stavros Skopeteas, Carolina Pasamonik, Carolin Brokmann, and Florian Fischer. “Cabécar DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.ebc4ra22](https://doi.org/10.34847/nkl.ebc4ra22). URL: <https://doreco.huma-num.fr/languages/cabe1245>.

- [82] Sabine Reiter. “Cashinahua DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.a8f9q2f1](https://doi.org/10.34847/nkl.a8f9q2f1). URL: <https://doreco.huma-num.fr/languages/cash1254>.
- [83] Manfred Krifka. “Daakie DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.efead519](https://doi.org/10.34847/nkl.efead519). URL: <https://doreco.huma-num.fr/languages/port1286>.
- [84] Maïa Ponsonnet. “Dalabon DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.fae299ug](https://doi.org/10.34847/nkl.fae299ug). URL: <https://doreco.huma-num.fr/languages/ngal1292>.
- [85] Nils Norman Schiborr. “English (Southern England) DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.9c271u5g](https://doi.org/10.34847/nkl.9c271u5g). URL: <https://doreco.huma-num.fr/languages/sout3282>.
- [86] Olga Kazakevich and Elena Klyachko. “Evenki DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.5e0d27cu](https://doi.org/10.34847/nkl.5e0d27cu). URL: <https://doreco.huma-num.fr/languages/even1259>.
- [87] Michael Franjeh. “Fanbyak DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.02084446](https://doi.org/10.34847/nkl.02084446). URL: <https://doreco.huma-num.fr/languages/orko1234>.
- [88] Mathieu Avanzi, Marie-José Béguelin, Gilles Corminboeuf, Federica Diémoz, and Laure Anne Johnsen. “French (Swiss) DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.35201685](https://doi.org/10.34847/nkl.35201685). URL: <https://doreco.huma-num.fr/languages/stan1290>.

- [89] Birgit Hellwig. “Goemai DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.b93664ml](https://doi.org/10.34847/nkl.b93664ml). URL: <https://doreco.huma-num.fr/languages/goem1240>.
- [90] Andrew Harvey. “Gorwaa DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.a4b4ijj2](https://doi.org/10.34847/nkl.a4b4ijj2). URL: <https://doreco.huma-num.fr/languages/goro1270>.
- [91] Felicity Meakins. “Gurindji DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.ab2d89mx](https://doi.org/10.34847/nkl.ab2d89mx). URL: <https://doreco.huma-num.fr/languages/guri1247>.
- [92] Iren Hartmann. “Hoocâk DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.b57f5065](https://doi.org/10.34847/nkl.b57f5065). URL: <https://doreco.huma-num.fr/languages/hoch1243>.
- [93] Niclas Burenhult. “Jahai DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.6a71xp0p](https://doi.org/10.34847/nkl.6a71xp0p). URL: <https://doreco.huma-num.fr/languages/jeha1242>.
- [94] Alexandra Vydrina. “Kakabe DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.d5aeu9t6](https://doi.org/10.34847/nkl.d5aeu9t6). URL: <https://doreco.huma-num.fr/languages/kaka1265>.
- [95] Valentin Gusev, Tiina Klooster, Beáta Wagner-Nagy, and Alexandre Arkhipov. “Kamas DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.cdd8177b](https://doi.org/10.34847/nkl.cdd8177b). URL: <https://doreco.huma-num.fr/languages/kama1351>.

- [96] Christian Döhler. “Komnzo DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.c5e6dudv](https://doi.org/10.34847/nkl.c5e6dudv). URL: <https://doreco.huma-num.fr/languages/komn1238>.
- [97] Hauke Bartels and Marcin Szczepański. “Lower Sorbian DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.6c6e4e9k](https://doi.org/10.34847/nkl.6c6e4e9k). URL: <https://doreco.huma-num.fr/languages/lowe1385>.
- [98] Françoise Rose. “Mojeño Trinitario DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.cbc3b4xr](https://doi.org/10.34847/nkl.cbc3b4xr). URL: <https://doreco.huma-num.fr/languages/trin1278>.
- [99] Katharina Haude. “Movima DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.da42xf67](https://doi.org/10.34847/nkl.da42xf67). URL: <https://doreco.huma-num.fr/languages/movi1243>.
- [100] Nick Thieberger. “Nafsan (South Efate) DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.ba4f7601](https://doi.org/10.34847/nkl.ba4f7601). URL: <https://doreco.huma-num.fr/languages/sout2856>.
- [101] Jocelyn Aznar. “Nisvai DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.2801565f](https://doi.org/10.34847/nkl.2801565f). URL: <https://doreco.huma-num.fr/languages/nisv1234>.
- [102] Alejandro Garcia-Laguia. “Northern Alta DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. doi: [10.34847/nkl.efea0b36](https://doi.org/10.34847/nkl.efea0b36). URL: <https://doreco.huma-num.fr/languages/nort2875>.
- [103] Geoff Haig, Maria Vollmer, and Hanna Thiele. “Northern Kurdish (Kurmanji) DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen,

- and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.ca10ez5t](https://doi.org/10.34847/nkl.ca10ez5t). URL: <https://doreco.huma-num.fr/languages/nort2641>.
- [104] Tom Güldemann, Martina Ernszt, Sven Siegmund, and Alena Witzlack-Makarevich. “N|ng DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.f6c37fi0](https://doi.org/10.34847/nkl.f6c37fi0). URL: <https://doreco.huma-num.fr/languages/nngg1234>.
- [105] Hiram Ring. “Pnar DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.5ba1062k](https://doi.org/10.34847/nkl.5ba1062k). URL: <https://doreco.huma-num.fr/languages/pnar1238>.
- [106] Frank Seifart. “Resigar DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.ffb961o8](https://doi.org/10.34847/nkl.ffb961o8). URL: <https://doreco.huma-num.fr/languages/resi1247>.
- [107] Alena Witzlack-Makarevich, Saudah Namyalo, Anatol Kiriggwajjo, and Zarina Molochieva. “Ruuli DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.fde4pp1u](https://doi.org/10.34847/nkl.fde4pp1u). URL: <https://doreco.huma-num.fr/languages/ruul1235>.
- [108] Xianming Xu and Bibo Bai. “Sadu DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.3db4u59d](https://doi.org/10.34847/nkl.3db4u59d). URL: <https://doreco.huma-num.fr/languages/sadu1234>.
- [109] Diana Forker and Nils Norman Schiborr. “Sanzhi Dargwa DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.81934177](https://doi.org/10.34847/nkl.81934177). URL: <https://doreco.huma-num.fr/languages/sanz1248>.

- [110] Claudia Wegener. “Savosavo DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.b74d1b33](https://doi.org/10.34847/nkl.b74d1b33). URL: <https://doreco.huma-num.fr/languages/savo1255>.
- [111] Jost Gippert. “Svan DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.9ba054c3](https://doi.org/10.34847/nkl.9ba054c3). URL: <https://doreco.huma-num.fr/languages/svan1243>.
- [112] Amos Teo. “Sümi DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.5ad4t01p](https://doi.org/10.34847/nkl.5ad4t01p). URL: <https://doreco.huma-num.fr/languages/sumi1235>.
- [113] Birgit Hellwig, Gertrud Schneider-Blum, and Khaleel Bakheet Khaleel Ismail. “Tabaq (Karko) DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.eea8144j](https://doi.org/10.34847/nkl.eea8144j). URL: <https://doreco.huma-num.fr/languages/kark1256>.
- [114] Natalia Bogomolova, Dmitry Ganenkov, and Nils Norman Schiborr. “Tabasaran DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.ad7f97xr](https://doi.org/10.34847/nkl.ad7f97xr). URL: <https://doreco.huma-num.fr/languages/taba1259>.
- [115] Ulrike Mosel. “Teop DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.9322sdf2](https://doi.org/10.34847/nkl.9322sdf2). URL: <https://doreco.huma-num.fr/languages/teop1238>.
- [116] Søren Wichmann. “Texistepec Popoluca DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.c50ck58f](https://doi.org/10.34847/nkl.c50ck58f). URL: <https://doreco.huma-num.fr/languages/texi1237>.

- [117] Maria Bardají, Christoph Bracks, Claudia Leto, Datra Hasan, Sonja Riesberg, Winarno S. Alamudi, and Nikolaus P. Himmelmann. “Totoli DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.c8b6ei29](https://doi.org/10.34847/nkl.c8b6ei29). URL: <https://doreco.huma-num.fr/languages/toto1304>.
- [118] Stavros Skopeteas, Violeta Moisidi, Nutsa Tsetereli, Johanna Lorenz, and Stefanie Schröter. “Urum DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.ac166n10](https://doi.org/10.34847/nkl.ac166n10). URL: <https://doreco.huma-num.fr/languages/urum1249>.
- [119] Stefan Schnell. “Vera'a DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.3e2cu8c4](https://doi.org/10.34847/nkl.3e2cu8c4). URL: <https://doreco.huma-num.fr/languages/vera1241>.
- [120] Sonja Riesberg. “Yali (Apahapsili) DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.9d91nkq2](https://doi.org/10.34847/nkl.9d91nkq2). URL: <https://doreco.huma-num.fr/languages/apah1238>.
- [121] Alexis Michaud. “Yongning Na DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.abe65p95](https://doi.org/10.34847/nkl.abe65p95). URL: <https://doreco.huma-num.fr/languages/yong1270>.
- [122] Stavros Skopeteas, Amedee Colli Colli, Daniela Schellenbach, Carolin Brokmann, Florian Fischer, and Maya Gálvez Wimmelmann. “Yucatec Maya DoReCo dataset”. In: *Language Documentation Reference Corpus (DoReCo) 2.0*. Ed. by Frank Seifart, Ludger Paschen, and Matthew Stave. Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2), 2024. DOI: [10.34847/nkl.9cbb3619](https://doi.org/10.34847/nkl.9cbb3619). URL: <https://doreco.huma-num.fr/languages/yuca1254>.

# **Supporting Information for**

## **Shared Duration of Speech Intervals Across 49 Human Languages**

Lara S. Burchardt<sup>a,b</sup>, Ludger Paschen<sup>b</sup>, Susanne Fuchs<sup>b,c</sup>

a Humboldt Universität zu Berlin

b Leibniz-Zentrum Allgemeine Sprachwissenschaft

c IMéRA and ILCB, Aix-Marseille University

Corresponding Author: Lara S. Burchardt

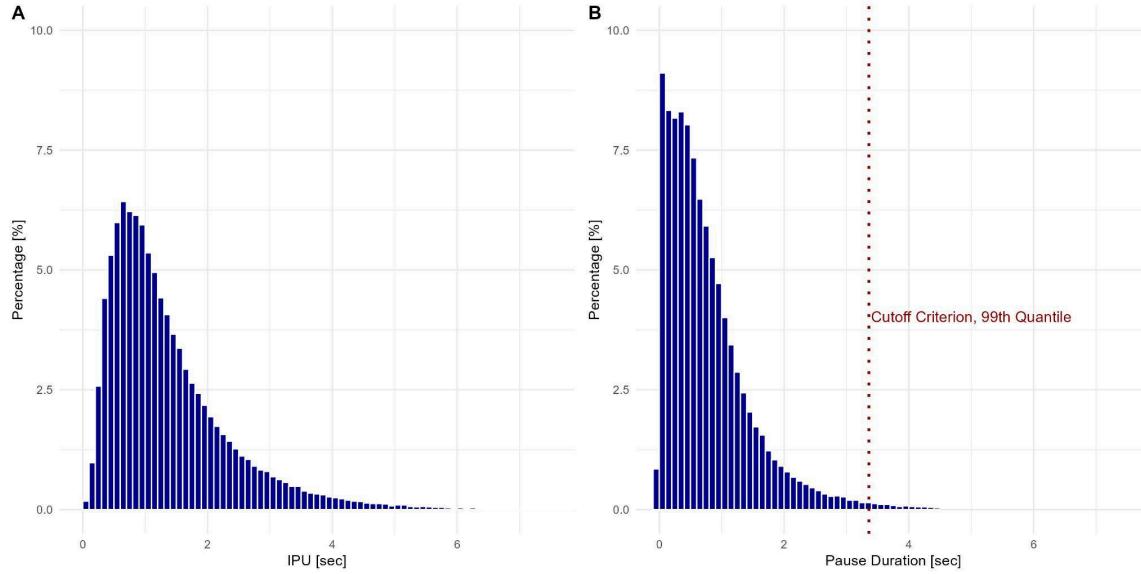
Email: lara.sophie.burchardt@hu-berlin.de

### **This PDF file includes:**

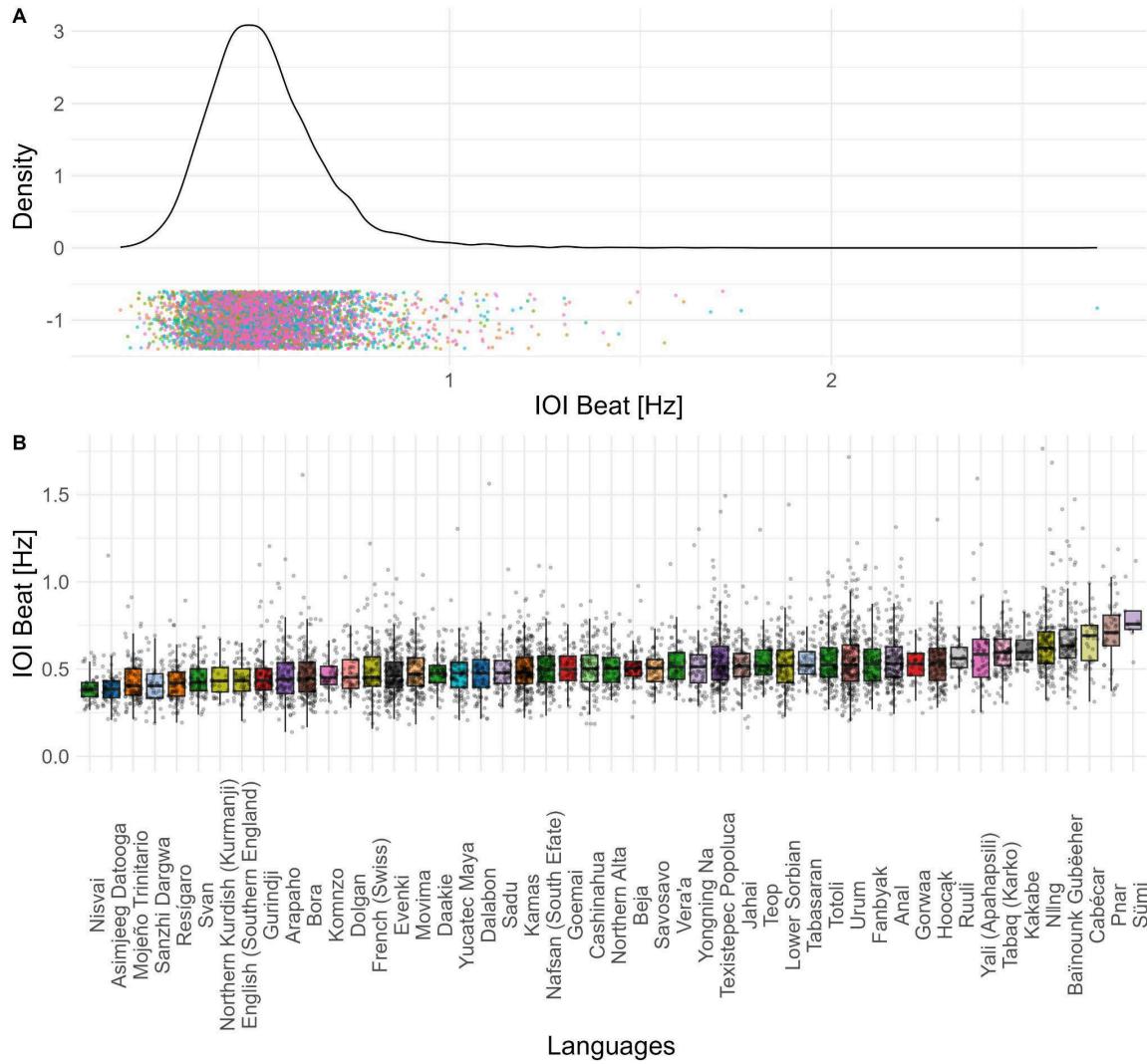
Figures S1 to S4

Table S1

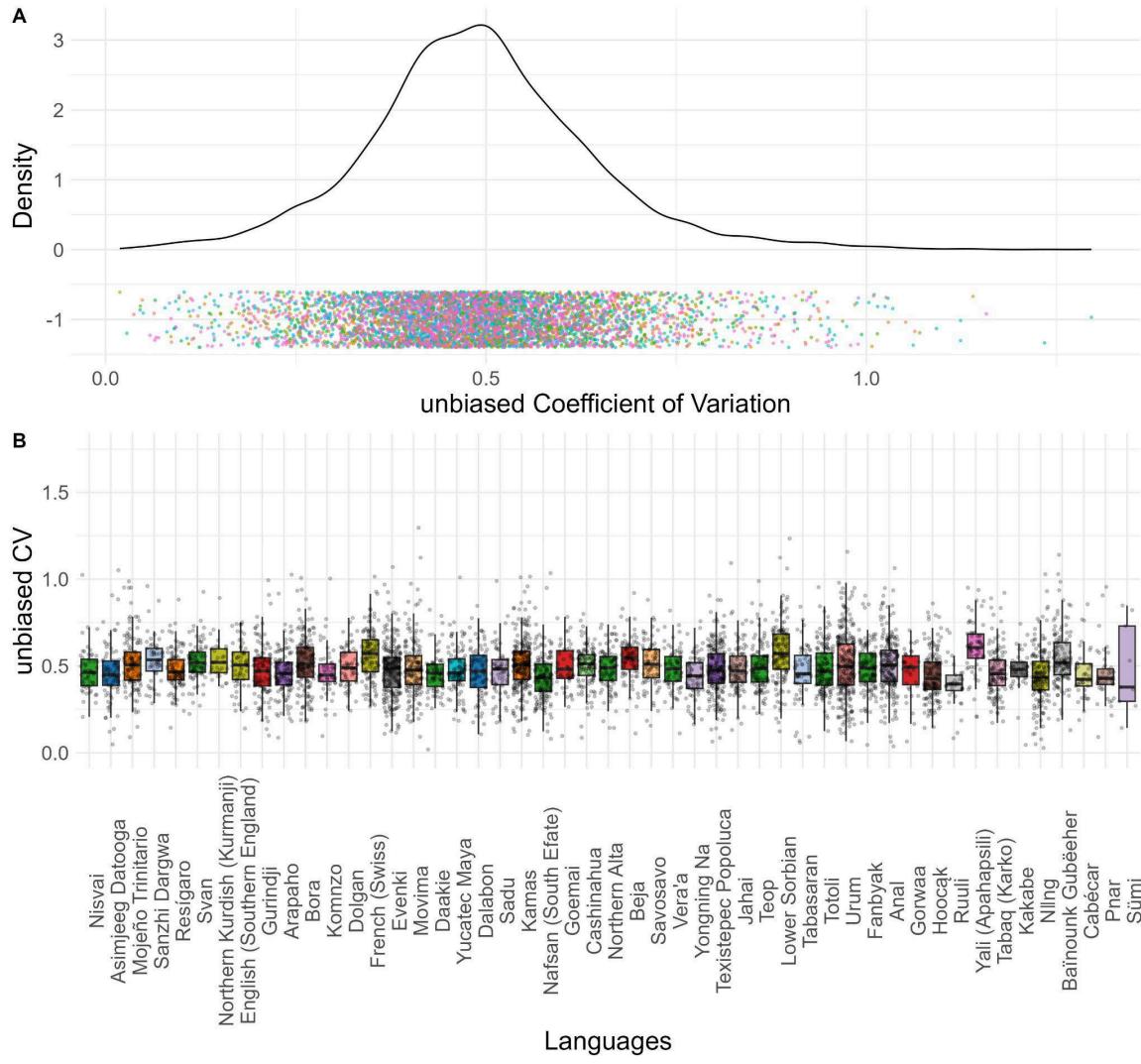
SI References



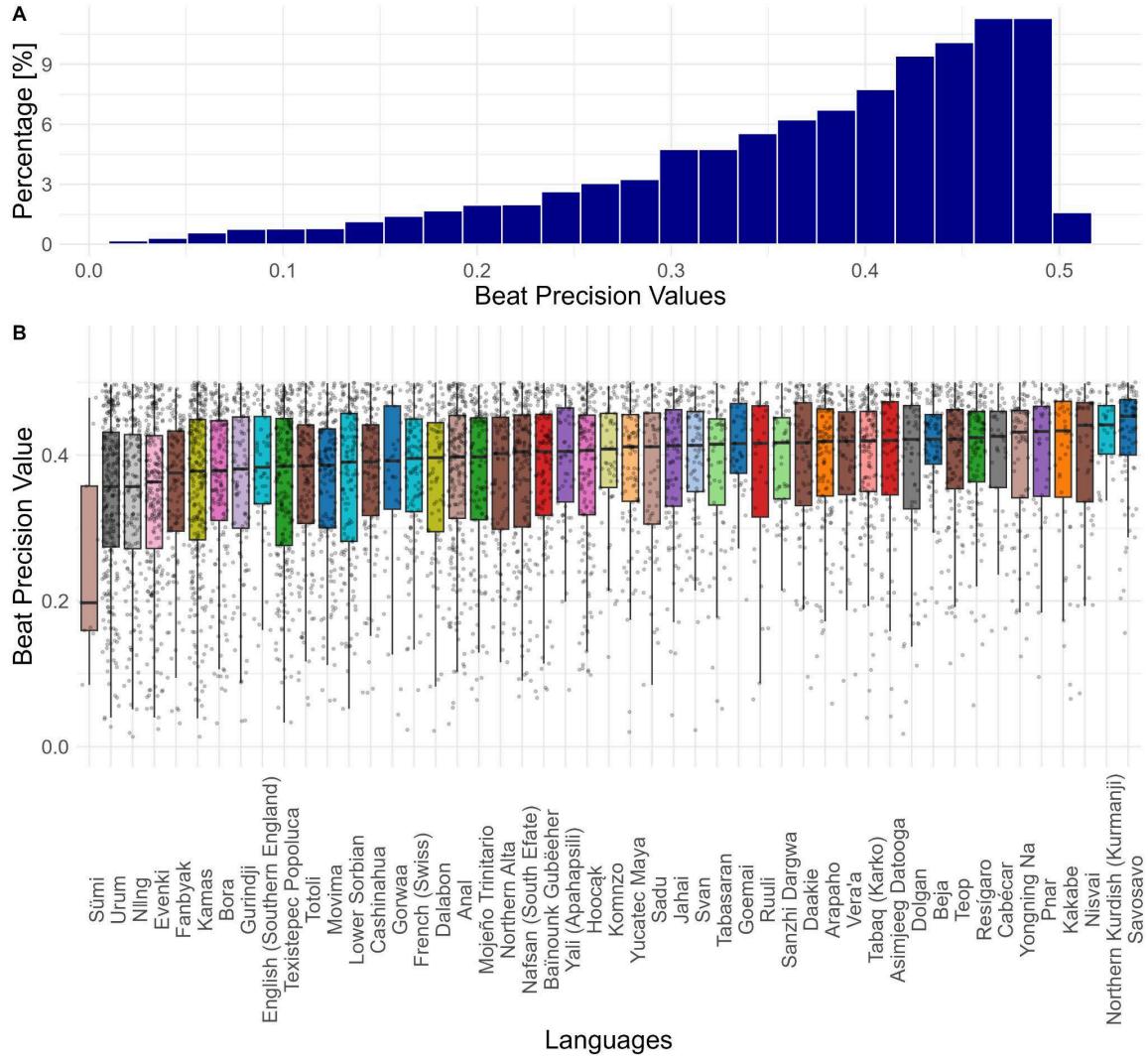
**Fig. S1.** (A) Histogram of IPU durations. Most IPUs are around 1 second long, but we see IPUs up to 6 seconds and above. Mean IPU length is 1.43 seconds. (B) Silent pause durations of the whole dataset. This is before sequences were split, based on the raw recordings. Mean Pause duration is 0.83 seconds. Illustrated also the cut-off criterion of the 99th quantile. Silent pauses above 3.3 seconds resulted in a sequence being terminated. That way, one recording could be split up into several sequences. See Fig. 1 in the main manuscript for an explanation of sequences.



**Fig. S2.** IOI Beats in Hz are similar across languages and language families based on Inter-Onset Intervals. (A) Density plots of IOI Beats of all 49 languages. Colors represent language families. (B) Boxplot of IOI Beats for all languages, sorted by median IOI Beat.



**Fig. S3.** Similar Coefficients of Variation (CV) across languages and language families based on the Coefficient of Variation per sequence. We see that CV density is slightly right-skewed, as there are many low and medium values, but a few large outliers. The order is based on the order of IOI beats per language as illustrated in Fig. S2.



**Fig. S4.** Beat Precision values overall and per language and language family. (A) Histogram of Beat Precision values. While some sequences follow an underlying isochronous model extremely well, many sequences show Beat Precision values between 0.4 and 0.5 indicating a bad fit of an underlying isochronous model with the IOI sequences. (B) Beat precision values look fairly similar between languages and language families (indicated by different colors). Only Sümi shows a clear difference, but we have a very small sample size for Sümi potentially explaining this.

**Table S1.** Glossary of key terms used in the manuscript.

<b>Pause</b>	In the context of this study, pause refers to a stretch of non-vocalization of any length occurring in spoken narratives. We make no assumptions about the function of pauses, but it is known that pauses may be inserted due to respiratory, physiological, planning, rhetorical, emphatic, and other reasons. In the DoReCo corpus, pauses are annotated based on (i) the presence of silence in the acoustic signal that could not be attributed to articulatory processes, such as the closure phase of stop consonants, and (ii) perceptual disruptions in the speech flow. Vocalized disfluencies and hesitations, although sometimes referred to as “filled pauses” (1), do not fall under the definition of pause applied here.
<b>IPU (Inter-Pausal Unit)</b>	In linguistics, a stretch of speech uninterrupted by a silent pause is referred to as an Inter-Pausal Unit or IPU (2). Some definitions of IPUs require a minimum pause length, for example, 100ms (3). However, in this study, no such threshold criterion is applied, as pause annotations in the DoReCo corpus have been manually verified (4). IPUs are defined solely in terms of silent pauses and do not refer to syntactic structure or prosodic domains. Usually, IPUs comprise one or two intonation units. IPUs flanked by longer pauses are more likely to constitute syntactic clauses than IPUs delineated by shorter pauses (5). IPUs may contain stretches of disfluent speech or be made up entirely of disfluencies; in the latter case, those IPUs were excluded from the analysis in the present study.
<b>IOI (Inter-Onset Interval)</b>	In biology and bioacoustics, IOI (Inter-Onset Interval) refers to the time interval between the onsets of successive behavioral events, such as vocalizations, movements, or neural spikes. IOI is a critical parameter in studying rhythmic patterns, temporal structure, and sequencing in animal communication and motor behaviors (6-8). In the context of this study, an IOI is defined as spanning a pair of one IPU and a following pause.
<b>Sequence</b>	A span of IOIs forms the basis of calculating speech rhythm in this study. Sequences are determined as stretches of at least 3 IOIs uninterrupted by a long pause, where “long” refers to falling outside the 99% interval of pause durations observed in the corpus (Figure S1). A single .wav file containing a speech recording may have between one and 109 IOI sequences.

## SI References

1. M. Belz, Defining filler particles: A phonetic account of the terminology, form, and grammatical classification of “filled pauses.” *Languages* **8**, 57 (2023). <https://doi.org/10.3390/languages8010057>
2. L. Paschen, S. Fuchs, F. Seifart, Final lengthening and vowel length in 25 languages. *Journal of Phonetics* **94**, 101179 (2022). <https://doi.org/10.1016/j.wocn.2022.101179>
3. H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa, Y. Den, An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech* **41**, 295–321 (1998). <https://doi.org/10.1177/002383099804100404>
4. L. Paschen, F. Delafontaine, C. Draxler, S. Fuchs, M. Stave, F. Seifart, Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari et al. (eds.), European Language Resources Association, Marseille, France, 2657–2666 (2020). <https://aclanthology.org/2020.lrec-1.324>
5. N. Peck, L. Becker, Syntactic pausing? Re-examining the associations. *Linguistics Vanguard* (2024). <https://doi.org/10.1515/lingvan-2022-0156>
6. T. A. Hersh, A. Ravignani, L. S. Burchardt, Robust rhythm reporting will advance ecological and evolutionary research. *Methods in Ecology and Evolution* **14**, 1398–1407 (2023). <https://doi.org/10.1111/2041-210X.14118>
7. M. D. Greenfield, B. Merker, Coordinated rhythms in animal species, including humans: Entrainment from bushcricket chorusing to the philharmonic orchestra. *Neuroscience and Biobehavioral Reviews* **153**, 105382 (2023). <https://doi.org/10.1016/j.neubiorev.2023.105382>
8. A. A. Rouse, A. D. Patel, S. Wainapel, M. H. Kao, Sex differences in vocal learning ability in songbirds are linked with differences in flexible rhythm pattern perception. *Animal Behaviour* **203**, 193–206 (2023). <https://doi.org/10.1016/j.anbehav.2023.05.001>