

## **Supplementary Information**

### **Sequences of Intonation Units form a ~1 Hz rhythm**

Maya Inbar, Eitan Grossman and Ayelet N. Landau

\* Corresponding author: Ayelet N. Landau

Email: ayelet.landau@mail.huji.ac.il

### **This Supplementary Information file includes:**

Appendix 1: Supplementary Materials and Methods

Supplementary Table S1

Appendix 2: Supplementary Analyses

Supplementary Figures S1-S2

References for Supplementary Information citations

## Appendix 1: Supplementary Materials and Methods

### *Data*

We analyzed data from six corpora of spontaneous speech. The corpora included speech transcriptions that were segmented into IUs according to the same set of criteria, and differed with respect to the degree of further transcription detail; however, these differences were at levels that did not affect our study, which required only the temporal boundaries of IUs. In general, the corpora fall into two groups. The first group consists of corpora in English, Hebrew and Russian, transcribed and segmented by professional teams, each working on their native language: the Santa Barbara Corpus of Spoken American English<sup>1</sup>, the Haifa Corpus of Spoken Hebrew<sup>2</sup>, and the Russian Multichannel Discourse corpus<sup>3</sup>. We constructed a sample from each of the three corpora according to the following protocol: we sampled randomly 10 recordings from the English and Hebrew corpora, and included all 3 recordings available as a preview to the Russian corpus. The English and Hebrew corpora include transcriptions segmented into intonation units (IUs) that are not timestamped to the audio file. Author MI manually measured the IU- onset and -offset times in the first 30 seconds of each recording, continuing until the next speaker change in each recording. Pauses between IUs were not considered as part of the neighboring IUs' duration, in deviation from the transcription guidelines of Du Bois et al.<sup>4</sup>. Spectrogram analyses in Praat 6.0.23<sup>5</sup> accompanied this procedure, including pitch and intensity contours produced by Praat's default settings. Measurements in milliseconds were entered into ELAN 4.9.4<sup>6</sup>, on separate tiers for different speakers. The Russian corpus includes the onset and offset times of IUs throughout each recording, measured in a similar fashion. Finally, in English and Hebrew, additional information was recorded for each IU, but was not further analyzed: the intonation contour attributed by the corpus constructors, the position within the turn, and a categorization to type of IU – Fragmentary, Regulatory or Substantive<sup>7</sup>. We replicated the results of the main analysis with 20 recordings from a Russian corpus that included narratives recorded by individual speakers rather than multi-speaker conversations<sup>8</sup>. In that dataset, although the sample of speakers was larger, the total recording time was less than a quarter of the length of the dataset analyzed in the current report. Nevertheless, the results were identical and we do not report them here.

The second group consists of corpora in Papuan Malay, Woi and Yali: the DoBeS Summits-PAGE Collection of Papuan Malay<sup>9</sup>, the DoBeS Woi Documentation<sup>10</sup>, and the DoBes Yali Documentation<sup>11</sup>. Papuan Malay (ISO 639-3 code: pmy) is the *lingua franca* of West Papua, the western half of the island of New Guinea governed by Indonesia. Woi (ISO 639-3 code: wbw) and Yali (ISO 639-3 code: yac) belong to different language families, have very different grammatical profiles, and are spoken in different regions in West Papua – on the coast and in the highlands, respectively. All three corpora include retellings of the plot of a silent film by a narrator to an interlocuter. More data on these languages

and corpora can be found in Himmelmann et al.<sup>12</sup>, a study that was dedicated to the human ability to identify IUs in unfamiliar languages. In this study, researchers with varying degrees of familiarity with Papuan Malay, Woi and Yali listened to recordings in these languages and in their native language and segmented all the recordings into IUs. The quantified agreement between researchers as to the correct segmentation of the entire corpus was well above chance, and was equally high for the unfamiliar languages as it was for the familiar one. Nikolaus Himmelmann kindly provided us with the consensus transcriptions obtained in this study, the joint product of the team that was verified by senior researchers highly familiar with the languages. These were already entered into ELAN<sup>6</sup> and time-aligned to the recording. We retrieved the respective audio files from the DoBeS archive. The Eastern Indonesia data in Himmelmann et al.<sup>12</sup> included an additional language that we did not attempt to analyze because it included only two recordings. Additionally, we could not retrieve the audiofile of one of the original 6 narratives from the Yali consensus set, so only 5 Yali recordings were analyzed.

Data were extracted from all the ELAN files with the aid of custom-written scripts in R 3.6.0<sup>13</sup> to data frames including IU onset and offset times, the computed duration, and the speaker of each IU. All analyses were performed in MATLAB 2016b<sup>14</sup> on a CentOS Linux 7 computer, using custom-written scripts and the FieldTrip toolbox<sup>15</sup>. Figures were produced using MATLAB 2017b<sup>16</sup> on a Windows 10 Home computer.

### *Speech envelope computation*

The amplitude envelope was computed for the audio file of each speech segment following the methods used in several studies demonstrating neural entrainment to the envelope, (e.g.<sup>17</sup>), methods that partly follow Chandrasekaran et al.<sup>18</sup>.

Speech segments that were recorded in stereo were converted to mono by averaging the channels. Recordings with a sampling rate above 20 kHz were downsampled to 20 kHz. Speech segments were band-pass filtered into 10 bands between 200 Hz and half the audio file's sampling frequency, with cut-off points designed to be equidistant on the human cochlear map. Amplitude envelopes for each band (the narrowband envelopes) were computed as absolute values of the Hilbert transform. These narrowband envelopes were downsampled to 1000 Hz and subsequently averaged, yielding the wideband envelope. The wideband envelope was smoothed using a 50 ms sliding Gaussian filter and divided by its maximal value to be on a scale of 0-1.

<b>Corpus</b>	<b>Recordings</b>	
Santa Barbara Corpus of Spoken American English <sup>1</sup>	Just_wanna_hang	Swingin_kid
	New_Yorkers_anonymous	The_mama_of_dada
	Noise_pollution	Throw_me
	Oh_you_need_a_breadbox	What_is_a_brand_inspection
	Shaggy_dog_story	You_baked
Haifa Corpus of Spoken Hebrew <sup>2</sup>	Changing_the_education_system	Korin_is_in_love
	Conversation_about_conversation_and_babies	Lack_of_tact_and_sensitivity
	Costume_for_purim	Looking_for_work
	Grenade	Luck_in_cards_and_girls
	Gym	Parenthood_and_education
Russian Multichannel Discourse corpus <sup>3</sup>	Pears04	Pears22
	Pears23	
DoBeS Summits-PAGE Collection of Papuan Malay <sup>9</sup>	PMY_pear_Boas_KONS	PMY_pear_Miry_KONS
	PMY_pear_Carl_KONS	PMY_pear_Moha_KONS
	PMY_pear_Fant_KONS	PMY_pear_Nova_KONS
	PMY_pear_Irma_KONS	PMY_pear_Sofi_KONS
	PMY_pear_Lala_KONS	PMY_pear_Suar_KONS
	PMY_pear_Laod_KONS	PMY_pear_Titi_KONS
	PMY_pear_Lia_KONS	PMY_pear_Wolt_KONS
	PMY_pear_Lisa_KONS	PMY_pear_Yanu_KONS
	PMY_pear_Maik_KONS	PMY_pear_Yul1_KONS
	PMY_pear_Maya_KONS	PMY_pear_Yul2_KONS
DoBeS Woi Documentation <sup>10</sup>	WBW_pear_Abra_KONS	WBW_pear_John_KONS
	WBW_pear_Agus_KONS	WBW_pear_Kosm_KONS
	WBW_pear_Davi_KONS	WBW_pear_Mart_KONS
	WBW_pear_Ello_KONS	WBW_pear_Oni_KONS
	WBW_pear_Feli_KONS	WBW_pear_Sofi_KONS
	WBW_pear_Heri_KONS	WBW_pear_Yuli_KONS
Yali Documentation <sup>11</sup>	YAC_pear_Edis_KONS	YAC_pear_Yust_KONS
	YAC_pear_Edo_KONS	YAC_pear_Yusu_KONS
	YAC_pear_Ibra_KONS	

Table S1. List of recordings used in the study.

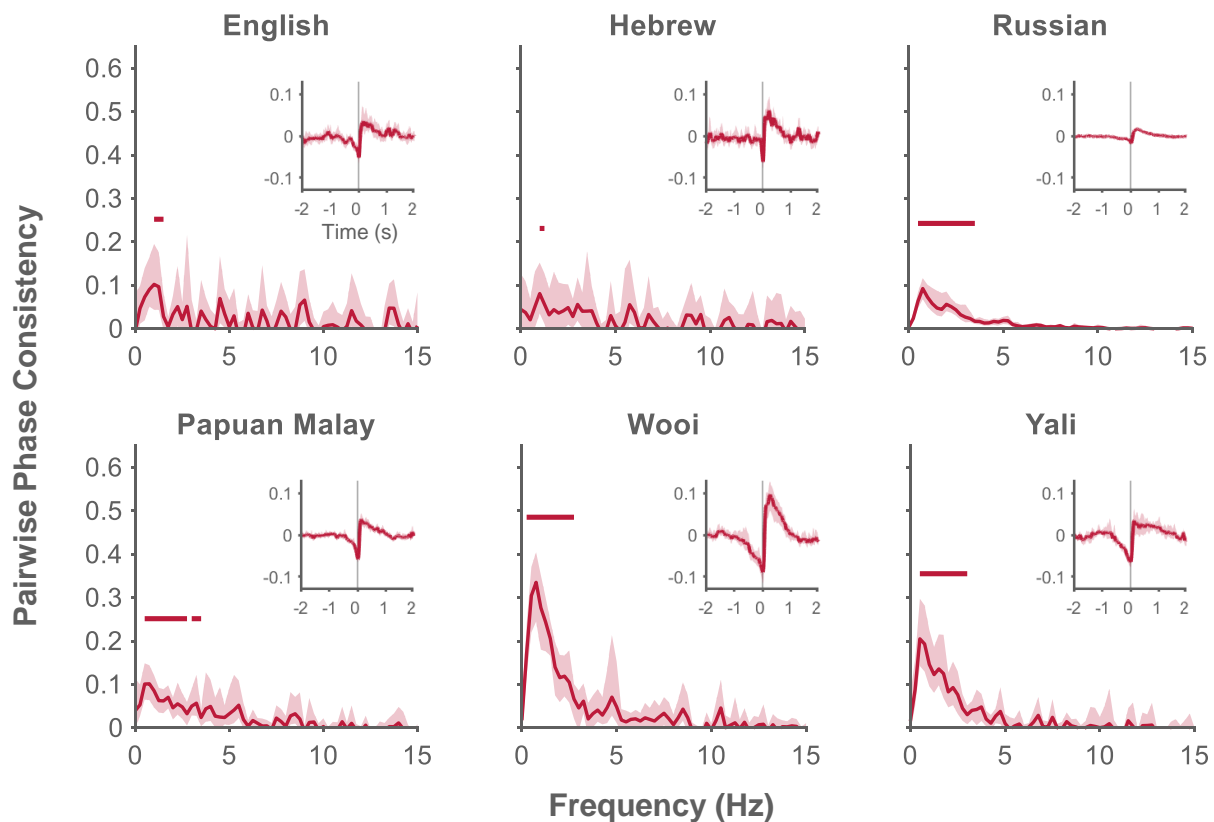
## **Appendix 2: Supplementary Analyses**

### *Larger windows for spectral decomposition in the phase consistency analysis*

In the main analysis, we extracted 2-second windows of the speech envelope centered on each IU, decomposed them and submitted the resulting phase estimations to a consistency analysis. The inclusion of as much data as possible guided this choice, as the size of the window influences the amount of data submitted to the analysis. With 2-second windows centered on IU onsets, only the first and last seconds of each recording are excluded from the analysis. This corresponded to 5% of the data in our smallest dataset, the Hebrew dataset, whose recordings were also generally shorter. In comparison, 4-second windows would lead to the exclusion of the first and last two seconds of each recording, corresponding to 12% of the Hebrew data. However, since larger analysis windows allow for phase estimations in a finer frequency resolution and based on more cycles, we repeated the analysis using 4-second windows of the speech envelope around IU onsets (Figure S1).

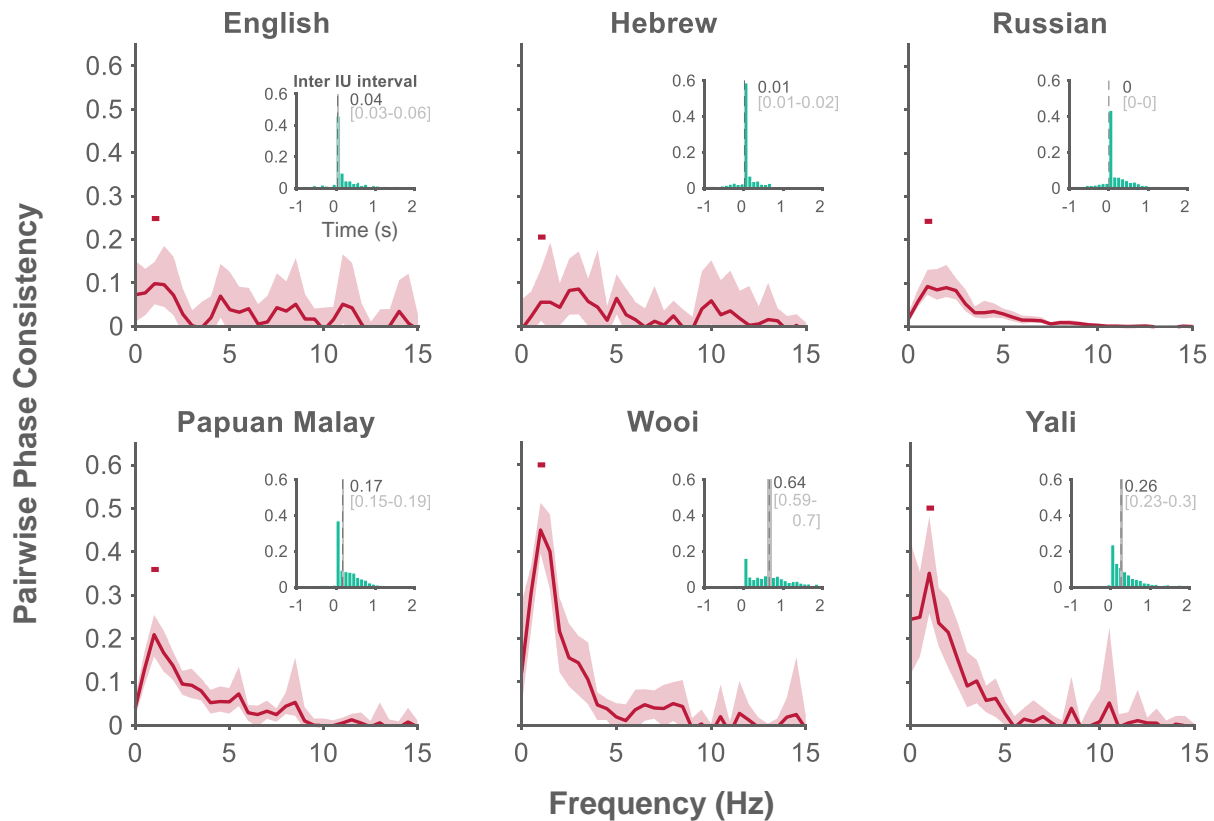
IU onsets still appear at significantly consistent phases of the low-frequency components of the speech envelope. The significant ranges using the same statistical criteria (i.e. while still ensuring that FDR < 1%) now exclude the higher frequencies (English: 1-1.25 Hz; Hebrew: 1 Hz; Russian: 0.5-3.25 Hz; Papuan Malay: 0.5-3.25 Hz; Wooroi: 0.25-2.5 Hz; and Yali: 0.5-2.75 Hz, all  $p$ 's < 0.001). The highest phase consistency is measured at 1 Hz for English and Hebrew, at the neighboring frequency bin 0.75 Hz for Papuan Malay, Russian and Wooroi, and at 0.5 Hz for Yali.

The insets in Figure S1 present, for each language, the grand average of the speech envelope segments that were submitted to spectral decomposition. Each single envelope window was demeaned, the windows around IUs produced by the same speaker were averaged, and the grand average was then calculated across all the within-speaker average time courses. The grand average time courses clearly show that the envelope captures IU onsets. Although it is the individual segments that are submitted to the spectral decomposition and not the average of segments, this representation allows to appreciate the invariant component across the windows. This component gives rise to the measured phase consistency.



**Figure S1. Larger analysis windows do not qualitatively change the characterization of the temporal structure of Intonation Units** (item corresponding to Figure 2). Phase-consistency analysis results include the average of phase consistency spectra across speakers for each language. Shaded regions denote bootstrapped 95% confidence intervals of the averages. Significance is denoted by a horizontal line above the spectra, after correction for multiple comparisons across neighboring frequency bins using an FDR procedure. Inset: Average envelope segments time-locked to IU onsets (depicted by vertical gray line). Shaded regions denote bootstrapped 95% confidence intervals of the averages.

*Addressing a question of transient effects*



**Figure S2. Removal of IUs following long pauses does not qualitatively change the characterization of the temporal structure of Intonation Units** (item corresponding to Figure 2). Phase-consistency analysis results include the average of phase consistency spectra across speakers for each language. Shaded regions denote bootstrapped 95% confidence intervals of the averages. Significance was assessed for the 1 Hz frequency component of each spectrum and is denoted by a horizontal line above the peak. Inset: Probability distribution of all inter-IU-interval durations within each language corpus, calculated for 50 ms bins and pooled across speakers. Overlaid are the medians (dashed line; dark gray) and the bootstrapped 95% confidence intervals of the medians (light gray).

## References

1. Du Bois, J. W. *et al.* Santa Barbara Corpus of Spoken American English, Parts 1-4. <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus> (2005).
2. Maschler, Y. *et al.* The Haifa Corpus of Spoken Hebrew. [http://weblx2.haifa.ac.il/~corpus/corpus\\_website/](http://weblx2.haifa.ac.il/~corpus/corpus_website/) (2017).
3. Kibrik, A. A. *et al.* Russian Multichannel Discourse. <https://multidiscourse.ru/main/?en=1> (2018).
4. Du Bois, J. W., Cumming, S., Schuetze-Coburn, S. & Paolino, D. *Discourse Transcription. Santa Barbara Papers in Linguistics* vol. 4 (University of California, Santa Barbara, 1992).
5. Boersma, P. & Weenink, D. Praat: doing phonetics by computer. (2016).
6. Wittenburg, P., Brugman, H., Russel, A., Klassman, A. & Sloetjes, H. ELAN: a professional framework for multimodality research. (2006).
7. Chafe, W. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing.* (University of Chicago Press, 1994).
8. Kibrik, A. A. & Podlesskaya, V. I. Spokencorpora. Rasskazy o snovidenijax i drugie korpusa zvučaščeje reči (Prosodically annotated corpus of spoken Russian). <http://spokencorpora.ru/> (2014).
9. Himmelmann, N. P. & Riesberg, S. The DoBeS Summits-PAGE Collection of Papuan Malay 2012-2016. <https://hdl.handle.net/1839/00-0000-0000-0019-FF78-5> (2016).
10. Kirihió, J. K. *et al.* The DobeS Wooi Documentation 2009-2015. <https://hdl.handle.net/1839/00-0000-0000-0014-C76C-1> (2015).
11. Riesberg, S., Walianggen, K. & Zöllner, S. The DobeS Yali Documentation 2012-2016. <https://hdl.handle.net/1839/00-0000-0000-0017-EA2D-D> (2016).
12. Himmelmann, N. P., Sandler, M., Strunk, J. & Unterladstetter, V. On the universality of intonational phrases – a cross-linguistic interrater study. *Phonology* **35**, (2018).
13. R Core Team. R: A language and environment for statistical computing. (2019).
14. MATLAB version 9.1.0.441655. (2016).
15. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* (2011) doi:10.1155/2011/156869.
16. MATLAB version 9.3.0.713579. (2017).
17. Gross, J. *et al.* Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* **11**, e1001752 (2013).
18. Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A. & Ghazanfar, A. A. The natural statistics of audiovisual speech. *PLoS Comput. Biol.* **5**, e1000436 (2009).