Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Short-Term Periodicity of Prosodic Phrasing: Corpus-Based Evidence

Sabrina Stehwien[1] *& Lars Meyer*[2,3]

[1]Fraunhofer Institute for Integrated Circuits, Erlangen, DE †
[2]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, DE
[3]Clinic for Phoniatrics and Pedaudiology, University Hospital Münster, DE

`sabrina.stehwien@iis.fraunhofer.de, lmeyer@cbs.mpg.de`

Abstract

Speech is perceived as a sequence of meaningful units of various lengths, from phones to phrases. Prosody is one of the means by which these are segmented: Prosodic boundaries subdivide utterances into prosodic phrases. In this corpus study, we study prosodic boundaries from a neurolinguistic perspective. To be perceived correctly, prosodic phrases must obey neurobiological constraints. In particular, electrophysiological processing has been argued to operate periodically, with one electrophysiological processing cycle being devoted to the processing of exactly one prosodic phrase. We thus hypothesized that prosodic phrases as such should show periodicity. We assess the DIRNDL corpus of German radio news, which has been annotated for intonational and intermediate phrases. We find that sequences of 2–5 intermediate phrases are periodic at 0.8–1.6 Hertz within their superordinate intonation phrase. Across utterances, the duration of intermediate phrases alternates with the duration of superordinate intonation phrases, indicating a dependence of prosodic time scales. While the determinants of periodicity are unknown, the results are compatible with an association between periodic electrophysiological processing mechanisms and the rhythm of prosody. This contributes to closing the gap between the the neurobiology of language and linguistic description.

Index Terms: prosodic phrasing, rhythm, electrophysiology

1. Introduction

Speech is a temporal succession of phonological units of various lengths, from phones to phrases. The topmost suprasegmental level is prosody: patterns of rhythm, stress, and tune that convey the phrasing of an utterance; that is, multi-word units termed intonation phrases [1, 2, 3, 4]. In the literature, there is no universally accepted definition of intonation phrases, partly due to cross-linguistic variability [5]. It is generally agreed that they are delineated by prosodic phrase boundaries, which are multi-dimensional combinations of a falling or rising pitch contour, often followed by a pause and a pitch and energy reset [6]. The widely used Tones and Break Indices (ToBI) annotation system further defines longer intonation phrases (IPs) that consist of shorter intermediate phrases (ips). IP boundaries are perceived as more salient than ip boundaries; furthermore, ip boundaries are often not marked by pauses. In ToBI, ip and IP boundaries correspond to the two strongest break indices, that is 3 and 4, respectively [7, 8].

Prosodic phrases are a core object of psycholinguistics [9] because prosodic phrasing guides the formation of macroscopic syntactic and semantic units [10, 11, 12, 13, 14, 15, 16, 17]. IPs delineate syntactic units, whereby prosodic boundaries scale with the size of the unit [18, 19]. It is often proposed that the sampling of the words contained within an IP allows the listener to decode a single piece of information [20, 21, 22].

Prosodic phrasing is also an object of neuroscientific research on speech processing. In particular, the *timing* of phrasing has recently entered focus. This reflects evidence for a role of rhythmic neuronal activity, so-called neural oscillations, in speech processing. Neural oscillations are rhythmic fluctuations in electrophysiological potentials; each cycle provides a time window for the processing of one piece of information [23, 24, 25]. During speech processing, oscillations were found to synchronize with acoustic rhythms on the time scales of phonemes, syllables, and prosody [26, 27].

One major assumption of this neuroscientific framework is that prosody is rhythmic enough to be processed by rhythmic electrophysiological activity. But is this really the case? Our corpus analysis tests this assumption, addressing a gap between the linguistic description of prosodic phrases and the electrophysiological mechanisms of the human auditory system. We follow two hypotheses: First, phrase duration should match the duration of those electrophysiological cycles that have been related to prosody processing. Second, phrase onsets should be regularly spaced, such that phrases are periodic. The quantification of rhythmicity in speech corpora is an emerging field of linguistic research [28, 29, 30] and periodicity of prosody is a recent hypothesis in linguistics [28]. Indeed, the onsets of pairs of intonation units that are analogous to ToBI's ip were found to associate with a consistent phase angle of the speech envelope at a frequency of $\sim$1 Hertz [31, 32, 20]. This suggests that acoustic amplitude modulations at prosodic boundaries might be regular enough to synchronize to.[1]

Building on this prior investigation of prosodic amplitude modulations, we here investigate the periodicity of prosodic phrase boundaries by taking into account the multiple acoustic dimensions that trigger human perception [6]. We make use of a corpus of German radio news that has been manually annotated with ToBI [39]. Prosodic events in the ToBI framework are considered multi-dimensional in the sense that annotators take several perceptual cues into account. Using autocorrelation analysis, we find evidence of short-term periodicity of prosodic phrasing. Within more that 60 percent of IPs, we observed periodic series of 2 to 5 subsequent subordinate ips. Furthermore, we find that periodicity ranges within a narrow band from roughly 0.8 to 1.6 Hertz, critically depending on the duration of the superordinate IP. We discuss implications for phonological theory, the neurobiology of language, and the interplay between neurobiological constraints and speech as such.

---

†Work carried out while at MPI CBS, Leipzig

[1] By analogy, syllable frequency across languages is confined to a narrow band between 4 and 8 Hertz [33, 34, 29], consistent with the previously proposed role of theta-band synchronization in syllable processing [35, 36, 37, 38].

10.21437/SpeechProsody.2022-141

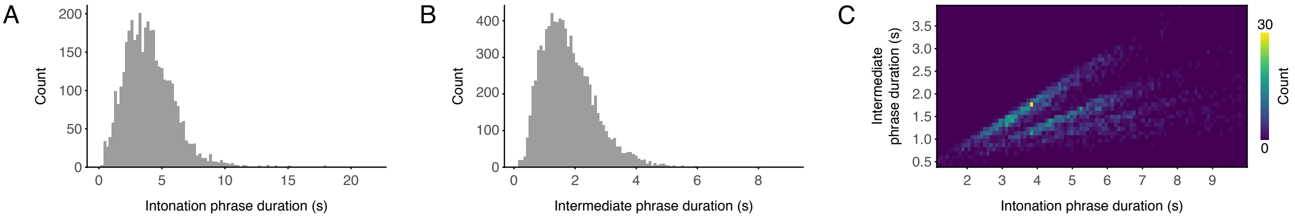

Figure 1: *Distributions of IP and ip durations. (A) Histogram of IP durations. (B) Histogram of ip duration. (C) Joint histogram of ip duration (ordinate) as a function of IP duration (abscissa); colors indicate counts within 100 × 40 equispaced bins.*

Table 1: *Variance of pairwise differences in intermediate phrase length with increasing distance between each pair of phrases (counted in number of intermediate phrases). Distance of one means that the phrases are neighbors.*

	Distance between ips (number of ips)										
Measure	1	2	3	4	5	6	7	8	9	10	11
Number of ip pairs	4,975	2,311	892	349	141	63	32	14	5	2	0
Average variance within bin	1.23	1.13	1.20	1.46	1.30	1.53	1.73	0.75	0.15	–	–

2. Speech Data

The DIRNDL corpus is a database of German radio news broadcasts [39, 40]. It consists of speech recordings of male and female professional speakers, with time-aligned transcriptions. Prosody, including IP and ip on– and offsets, was manually labeled by experts for a 5-hour subset of the corpus according to the Stuttgart German adaptation of the ToBI system [41]. The dataset contains 3,947 IPs and 9,041 ips. DIRNDL was collected over the course of a few days, and thus the news items are repeated, but not always read by the same speaker. Since the data is not annotated for speaker ID, we account for speaker differences using recording ID in our statistical analyses. Data was pre-processed using Python; statistical analyses were performed in R [42].

3. Data Analysis & Results

As first step to assessing the regularity of IPs and ips, we calculated durations and variances. Median IP duration was 3.8 s with a variance of 3.53. Median ip duration was 1.65 s and variance was 0.69; ip duration was thus less varied than IP duration, hinting at temporal regularity (Figure 1A/B).

We also explored the relationship between the durations of IPs and ips, based on the visual impression of a harmonic relationship (Figure 1C). We employed linear mixed-effects model comparison implemented in the lme4 package [43]. The baseline model included random intercepts only to account for recording-specific (and thus speaker-specific) differences. The dependent measure was ip duration. Through Analysis of Variance (ANOVA), this model was compared to a model that also included the duration of the superordinate IP as a fixed effect [44]. Fit improved ($\chi^2(1) = 277.61, p < 0.001$), indicating that ip duration depends on the duration of the superordinate IP. A confound of this could be speech rate differences, which was expected to affect both ip and IP duration. To test for this, we made two additional models that assessed the effect of speech rate (phones per second) on ip and IP duration; both included recording ID as random intercept. The models outperformed baseline models that contained only the random intercept ($\chi^2(1)_{ip} = 23.12, p < 0.001$; $\chi^2(1)_{IP} = 23.11, p < 0.001$). Standardized coefficients suggested equal effects across models (($\beta_{ip} = -0.03$; $\beta_{IP} = -0.03$); yet, IP duration in the above model was a stronger predictor of ip duration ($\beta = 0.09$). Hence, ip and IP duration scale with speech rate, but ip duration still scales with IP duration. In other words: At identical speech rates, long IPs contain long ips, but short IPs contain short ips.

We thus decided to continue periodicity analyses within IP duration bins of 1 s width. The lack of long IPs in DIRNDL had us consider IPs of up to 9 s. As an initial step, we compared the variance across ips within IP (total number of IPs = 2,664; average number of ips per IP = 2.87) to the variance across the corpus. Average variance is 0.39—smaller than across the corpus. Yet, visual inspection did not show equispacing of ips either (Figure 2A). To test this, we added the linear ip index as fixed effect. This led to further model improvement ($\chi^2(1) = 430.54, p < 0.001$), suggesting that ip duration depends on ip position within the IP. To assess this, we calculated variance across pairs of increasing distance within IP duration bin (e.g., ips 1–2 versus 1–3 versus 1–4, etc.; see Table 1). Variance was lower for nearby ips (Table 1). Hence, ip periodicity is more likely within shorter stretches of speech.

To assess periodicity of series instead of pairs, we then ran autocorrelation analyses on ip series within IP duration bin. Because of the increase of duration variance with distance, we ran separate analyses for series of incrementally increasing ip count. Each series was converted into a binary time series, where 1 marks an ip offset. A sampling rate of 100 Hz was chosen, that is, each one or zero marks a 10-ms frame. To avoid aliasing, the maximum lag matched the next ip offset minus one. To determine significant lags, we performed a permutation test within individual ip series. We created 1,000 permutations of each series to obtain 1,000 random autocorrelation functions. A lag was considered significant when the observed r was above the 97.5th percentile of the random autocorrelations at this lag, corresponding to a corrected $p < 0.05$ [45]. We removed all series for which the only significant lag was the first one; a 10-ms period would likely not be meaningful. A lack of long series limited the maximum ip count to 7. This analysis yielded a significant lag in 4,474 out of 7,346 ip series (i.e., 61 % of ip series), pointing to substantial periodicity of ip series.

Visual inspection suggested that lags were more normally

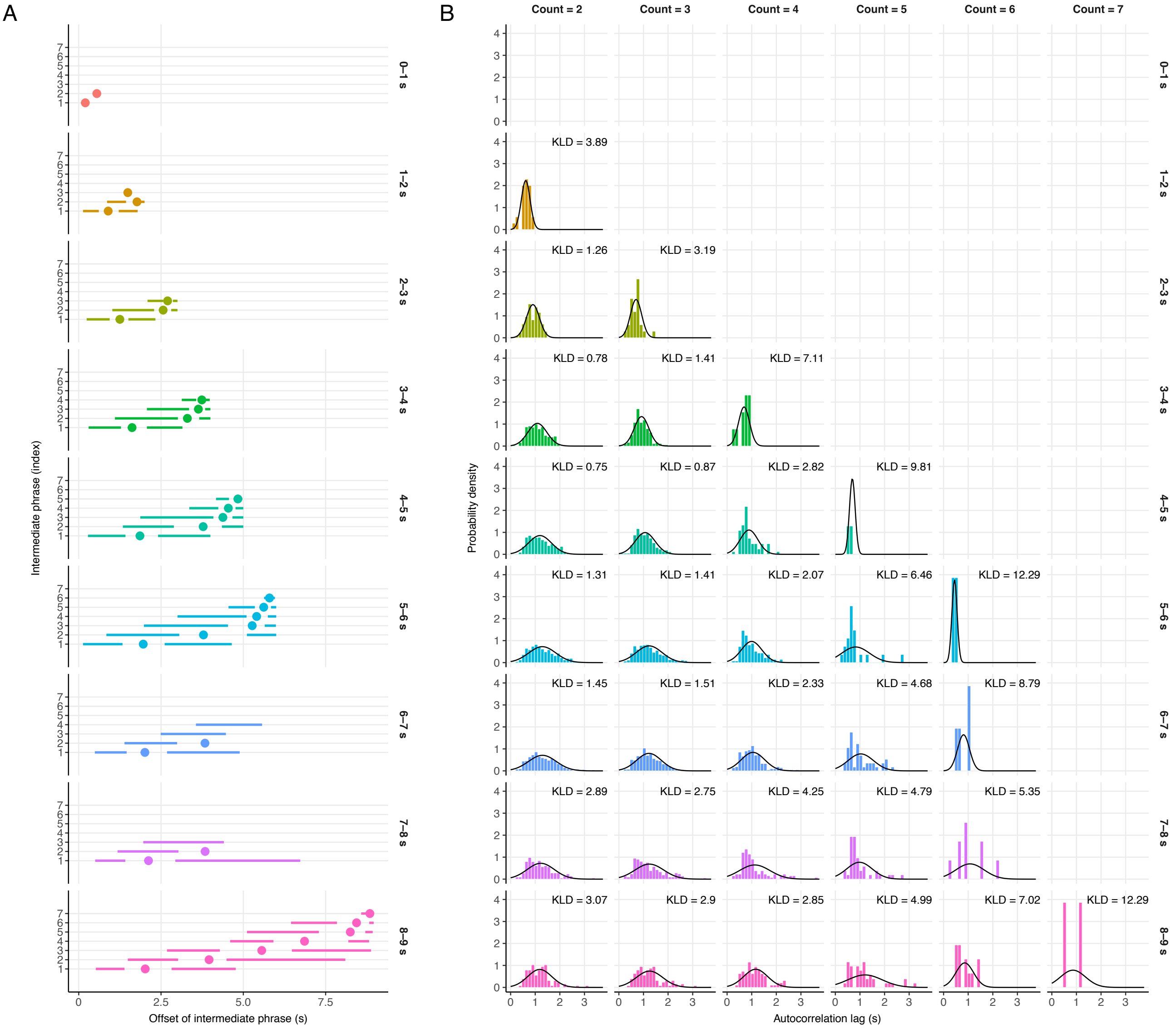

Figure 2: *Autocorrelation results. (A) Boxplots of ip offsets, binned into 1-second IP duration bins (major ordinate); minor ordinate: ip index. (B) Probability density of lags (minor abscissa) as function of ip count (major abscissa) and bin (major ordinate); black lines: normal probability density functions generated from mean and standard deviation of data; KLD = Kullback–Leibler Divergence.*

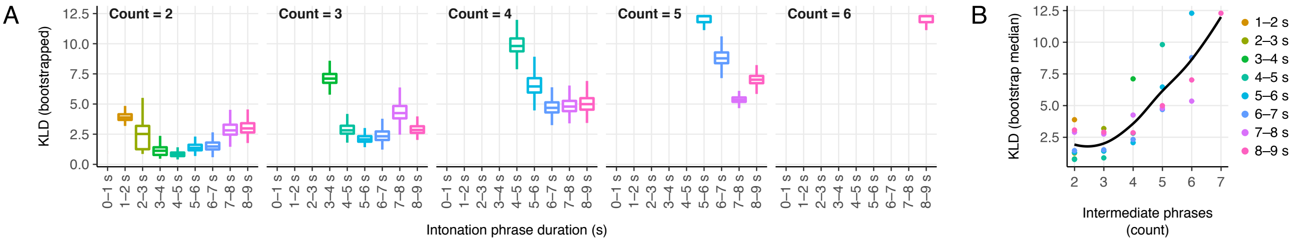

Figure 3: *Dependence of ip periodicity on ip count. (A) Kullback–Leibler Divergence (KLD) as function of ip count (major abscissa) and IP duration bin; ordinate: KLD. (B) Correlation of median KLD with number of subsequent ips, partialing out bin.*

distributed for short as compared to long ip series (Figure 2B). To test this, we calculated, within duration bin and ip count, the Kullback-Leibler Divergence (KLD) between the lag distribution and a normal distribution. A small KLD would indicate a periodicity of ip series. The baseline normal distribution consisted of 10,000 random data points organized into 100 bins, based on the mean and standard deviation of the lag distribution. We bootstrapped a median KLD from 10,000 iterations of this procedure (Figure 3A). Median KLDs were then correlated with ip count. To account for covariance between ip count and IP duration bin (i.e., large counts are only present in long IPs), we partialed out IP duration bin. Because of the skewness of the KLD distribution, we employed Kendall's Partial Rank Correlation. This was significant ($r_\tau = 0.62, p < 0.001$; Figure 3B), indicating that periodicity was high within short series of 2–5 ips. This indicates short-term periodicity of ips.

Table 2: *Median autocorrelation lags by IP duration bin and ip count within ip series, converted from samples to milliseconds.*

	Count of ips within series					
IP bin (s)	2	3	4	5	6	7
1–2	610					
2–3	930	730				
3–4	1,050	860	760			
4–5	1,140	1,000	785	735		
5–6	1,180	1,140	900	650	445	
6–7	1,200	1,165	990	885	830	
7–8	1,140	1,100	935	795	930	
8–9	1,140	1,170	1,090	1,040	750	860

4. Discussion

We found that series of 2–5 ips are periodic within their superordinate IP. The exact period depends on the duration of the superordinate IP—longer IPs contain longer ips. This entails short-term periodicity of prosodic phrasing within a specific range.

Previous work has reported that amplitude modulations, which make up one type of cue to prosodic boundaries, are periodic at 1 Hertz [20, 31, 32]. Our study extends this research in several ways: First, we consider ToBI-annotated boundaries, i.e. events determined by several perceptual cues. Second, we report periodicity beyond the previously reported time window of 2 seconds [20], containing 2 ips on average (Figure 2B). Finally, we show that ip period exhibits a range rather than a fixed value. While our study does not aim to assess the underlying mechanisms of these effects, it opens several interesting questions for further research. In the following, we discuss the implications of the above findings.

Why does periodicity stop after a few seconds? One possible reason is breathing: IP production requires exhalation; exhalation frequency fluctuates with metabolic demands [46]. IP offsets accompany, but do not always consistently coincide with the end of a breathing cycle [18, 47, 48]. Hence, cognitive constraints are a more plausible explanation. Auditory memories deteriorate after 2.4–2.7 seconds. Hence, information integration across the constituent ips of an IP must by then occur [31, 49, 47, 50, 51, 52, 53]. Our result is consistent with the idea that the brain samples the constituent ips of an IP with the help of electrophysiological rhythms. Furthermore, the currently observed ip periods of 610–1,200 milliseconds (Table 2) translate to a frequency range from 1.64–0.83 Hertz, consistent with evidence for an oscillatory synchronization with the speech envelope below 4 Hertz [54, 55, 56]. Note that a simpler alternative explanation could also be annotator error: Intervening ip boundaries can easily be misclassified as pitch accents, full IP boundaries, or even missed by human annotators [57, 58, 59]. An annotator error that disrupts an otherwise longer periodic ip series would shorten the periodic series found in our study.

One curious aspect of our results is the discrepancy between the median *duration* of ips across the corpus (1.65 seconds) and the median *periods* of ip series (0.6–1.2 s; Figures 1B/C and 3 and Table 2). At first glance, this could mean that short ips are more likely to be periodic. However, shorter ips are more frequent than longer ips (i.e., the histogram in Figure 1A is skewed). Hence, autocorrelated series of long ips are less probable, shortening the median autocorrelation lag.

A difference to previous studies is the use of annotations instead of acoustics. Prior work [20] assessed the amplitude envelope, a critical cue for boundary perception [6]. The envelope dominates the cochlear output, which is forwarded to the cortex for the cognitive inference of boundaries [60, 61]. Yet, boundaries are a multi-dimensional construct [6] and automatic detection of boundaries based on one acoustic dimension alone is inferior to multi-dimensional detection [62]. ToBI's formal definition of boundaries takes multi-dimensionality into account [7] and human annotators tend to agree on boundary locations [63, 64]. While one could argue that human annotators are not immune to cognitive constraints either, we consider it unlikely that annotation labor as such evoked periodicity of ips in DIRNDL. Nevertheless, future work should specify the link between the periodicity of the multiple dimensions of prosody and the periodicity of boundaries as such.

Generalization of our results would require replications on further languages, registers, and annotation systems. First, there may be cross-linguistic variability in periodicity, in particular when typological boundaries are transversed [20]. Second, idiosyncrasies of radio news [65, 66, 67] could lead to an overestimation of periodicity; hence, conversational and spontaneous speech should be assessed, which are characterized by short utterances, frequent turn-taking, and disfluencies such as hesitations, pauses, and repairs [68]. Finally, ToBI could have biased the current analyses [69], which should thus be replicated on corpora annotated with other standards [20, 31].

5. Conclusion

In this study, we aimed to find evidence of periodicity in prosody that could reflect neural oscillatory processing of speech. Our corpus analysis showed that the prosodic phrases in a German radio news corpus are periodic within a specific frequency range, confirming our assumption that prosody may be rhythmic enough to conform with current neuroscientific frameworks of speech processing. While this investigation was limited to one corpus and one annotation standard (ToBI) we believe that it motivates further research on the rhythmicity of speech corpora to help bridge the gap between phonological theory and neurolinguistic research.

6. Acknowledgements

This work was supported by the award of the Max Planck Research Group *Language Cycles* to Lars Meyer. We thank Dafydd Gibbon and Sascha Griffiths for helpful comments.

7. References

[1] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[2] E. Selkirk, “Phonology and syntax: The relation between sound and structure,” 1984.

[3] M. Nespor and I. Vogel, *Prosodic phonology: with a new foreword*. Walter de Gruyter, 2007, vol. 28.

[4] C. Féry, *Intonation and Prosodic Structure*, ser. Key Topics in Phonology, 2017.

[5] S. A. Jun, “Prosodic Typology,” in *Prosodic Typology: The Phonology of Intonation and Phrasing*, 2010.

[6] A. Brugos, M. Breen, N. Veilleux, J. Barnes, and S. Shattuck-Hufnagel, “Cue-based annotation and analysis of prosodic boundary events,” in *Proceedings of Speech Prosody*, 2018, pp. 245–249.

[7] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. B. Pierrehumbert, and J. Hirschberg, “ToBI: A standard for labelling English prosody,” in *Proceedings of the 2nd International Conference on Spoken Language Processing*, 1992, pp. 867–870.

[8] J. Pierrehumbert, “The phonology and phonetics of english intonation,” Ph.D. dissertation, Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy, 1980.

[9] L. Frazier, K. Carlson, and C. Clifton, “Prosodic phrasing is central to language comprehension,” *Trends in Cognitive Sciences*, vol. 10, no. 6, pp. 244–249, 2006.

[10] H. Kreiner and Z. Eviatar, “The missing link in the embodiment of syntax: Prosody,” *Brain and Language*, vol. 137, pp. 91–102, 2014.

[11] M. Wagner and D. G. Watson, “Experimental and theoretical advances in prosody: A review,” *Language and Cognitive Processes*, 2010.

[12] A. Cutler, D. Dahan, and W. Van Donselaar, “Prosody in the Comprehension of Spoken Language: A Literature Review,” 1997.

[13] I. I. Lehiste, J. P. Olive, L. A. Streeter, and I. I. Lehiste, “Role of duration in disambiguating syntactically ambiguous sentences,” *The Journal of the Acoustical Society of America*, vol. 60, no. 5, pp. 1199–1202, 1976.

[14] P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, “The use of prosody in syntactic disambiguation,” *Journal of Acoustic Society of America*, vol. 90, no. December 1991, pp. 2296–2970, 2015.

[15] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, “Segmental Durations In The Vicinity Of Prosodic Phrase Boundaries,” *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1707–1717, 1992.

[16] J. Snedeker and E. Casserly, “Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities,” *Language and Cognitive Processes*, vol. 25, no. 7-9, pp. 1234–1264, 2010.

[17] K. Steinhauer, K. Alter, and A. D. Friederici, “Brain potentials indicate immediate use of prosodic cues in natural speech processing,” *Nature Neuroscience*, vol. 2, no. 2, pp. 191–196, 1999.

[18] F. Grosjean, L. Grosjean, and H. Lane, “The patterns of silence: Performance structures in sentence production,” *Cognitive Psychology*, 1979.

[19] H. Truckenbrodt, “On the relation between syntactic phrases and phonological phrases,” *Linguistic Inquiry*, vol. 30, no. 2, pp. 219–255, 1999.

[20] M. Inbar, E. Grossman, and A. Landau, “Sequences of intonation units form a 1 Hz rhythm,” *Sci Rep.*, vol. Sep 28;10(1):15846, 2020.

[21] A. Wahl, “Intonation unit boundaries and the storage of bigrams,” *Review of Cognitive Linguistics*, vol. 13, no. 1, pp. 191–219, 2015.

[22] S. R. Speer and K. Ito, “Prosody in first language acquisition - Acquiring intonation as a tool to organize information in conversation,” *Linguistics and Language Compass*, 2009.

[23] G. Hickok, H. Farahbod, and K. Saberi, “The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation,” *Psychological science*, vol. 26, no. 7, pp. 1006–1013, 2015.

[24] M. J. Henry and J. Obleser, “Frequency modulation entrains slow neural oscillations and optimizes human listening behavior,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 49, pp. 20 095–20 100, 2012.

[25] C. E. Schroeder and P. Lakatos, “Low-frequency neuronal oscillations as instruments of sensory selection,” *Trends in Neurosciences*, vol. 32, no. 1, pp. 9–18, 2009.

[26] A.-L. Giraud and D. Poeppel, “Cortical oscillations and speech processing: emerging computational principles and operations,” *Nature Neuroscience*, vol. 15, no. 4, pp. 511–7, 2012.

[27] L. Meyer, “The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms,” *European Journal of Neuroscience*, vol. 48, no. 7, pp. 2609–2621, 2018.

[28] D. Gibbon, “The Future of Prosody: It’s about Time,” in *Proceedings of Speech Prosody*, 2018, pp. 1–9.

[29] S. Tilsen and A. Arvaniti, “Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages,” *The Journal of the Acoustical Society of America*, vol. 134, pp. 628–, 2013.

[30] A. Arvaniti, “The usefulness of metrics in the quantification of speech rhythm,” *Journal of Phonetics*, vol. 40, no. 3, pp. 351–373, 2012.

[31] W. L. Chafe, “The flow of thought and the flow of language,” in *Discourse and Syntax*. Brill, 1979, pp. 159–181.

[32] J. W. Du Bois, “Discourse transcription,” *Santa Barbara papers in linguistics*, vol. 4, pp. 1–225, 1992.

[33] N. Ding, A. D. Patel, L. Chen, H. Butler, C. Luo, and D. Poeppel, “Temporal modulations in speech and music,” *Neuroscience and biobehavioral reviews*, vol. 81(Pt B), pp. 181–187, 2017.

[34] C. Coupé, Y. Oh, D. Dediu, and F. Pellegrino, “Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche,” *Science Advances*, vol. 5, no. 9, p. eaaw2594, 2019.

[35] H. Luo and D. Poeppel, “Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex,” *Neuron*, vol. 54, no. 6, pp. 1001–1010, 2007.

[36] J. E. Peelle, J. Gross, and M. H. Davis, “Phase-locked responses to speech in human auditory cortex are enhanced during comprehension,” *Cerebral Cortex*, vol. 23, no. 6, pp. 1378–1387, 2013.

[37] K. B. Doelling, L. H. Arnal, O. Ghitza, and D. Poeppel, “Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing,” *NeuroImage*, vol. 85, pp. 761–768, 2014.

[38] M. F. Howard and D. Poeppel, “The neuromagnetic response to spoken sentences: Co-modulation of theta band amplitude and phase,” *NeuroImage*, vol. 60, no. 4, pp. 2118–2127, 2012.

[39] K. Eckart, A. Riester, and K. Schweitzer, “A Discourse Information Radio News Database for Linguistic Analysis,” in *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*, C. Chiarcos, S. Nordhoff, and S. Hellmann, Eds., 2012, pp. 65–75.

[40] A. Björkelund, K. Eckart, A. Riester, N. Schauffler, and K. Schweitzer, “The extended DIRNDL corpus as a resource for automatic coreference and bridging resolution,” in *In Proceedings of LREC*, 2014, pp. 3222–3228.

[41] J. Mayer, “Transcribing German intonation - the Stuttgart system,” University of Stuttgart, Tech. Rep., 1995.

[42] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017.

[43] D. Bates, M. Maechler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, vol. 67(1), pp. 1–48, 2015, doi:10.18637/jss.v067.i01.

[44] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[45] E. Maris and R. Oostenveld, “Nonparametric statistical testing of EEG- and MEG-data,” *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007.

[46] A. B. Tort, J. Brankačk, and A. Draguhn, “Respiration-Entrained Brain Rhythms Are Global but Often Overlooked,” pp. 186–197, 2018.

[47] P. A. White, “The three-second ”subjective present”: A critical review and a new proposal,” *Psychological Bulletin*, vol. 143, no. 7, pp. 735–756, 2017.

[48] J. Pierrehumbert, “The perception of fundamental frequency declination,” *The Journal of the Acoustical Society of America*, vol. 66, no. 2, pp. 363–369, 1979.

[49] H. Simpson and F. M. del Prado Martín, “Memory capacity limits in processing of natural connected speech: The psychological reality of intonation units,” in *37th annual conference of the Cognitive Science Society, Pasadena, CA.*, 2015, pp. 2206–2211.

[50] L. Henke and L. Meyer, “Endogenous oscillations time-constrain linguistic segmentation: Cycling the garden path,” *Cerebral Cortex*, 2021.

[51] S. R. Speer, R. G. Crowder, and L. M. Thomas, “Prosodic Structure and Sentence Recognition,” *Journal of Memory and Language*, 1993.

[52] A. Wingfield and D. L. Byrnes, “Decay of information in Short-term memory,” *Science*, vol. 176, no. 4035, pp. 690–692, 1972.

[53] M. Roll, M. Lindgren, K. Alter, and M. Horne, “Time-driven effects on parsing during reading,” *Brain and Language*, vol. 121, no. 3, pp. 267–272, 2012.

[54] M. Bourguignon, X. De Tiège, M. O. De Beeck, N. Ligot, P. Paquier, P. Van Bogaert, S. Goldman, R. Hari, and V. Jousmäki, “The pace of prosodic phrasing couples the listener’s cortex to the reader’s voice,” *Human Brain Mapping*, vol. 34, no. 2, pp. 314–326, 2013.

[55] J. M. Rimmele, D. Poeppel, and O. Ghitza, “Acoustically driven cortical δ oscillations underpin prosodic chunking,” *Eneuro*, vol. 8, no. 4, 2021.

[56] V. J. Boucher, A. C. Gilbert, and B. Jemel, “The role of low-frequency neural oscillations in speech processing: Revisiting delta entrainment,” *Journal of Cognitive Neuroscience*, vol. 31, no. 8, pp. 1205–1215, 2018.

[57] A. K. Syrdal and J. McGory, “Inter-transcriber reliability of ToBI prosodic labeling,” *6th International Conference on Spoken Language Processing, ICSLP 2000*, no. Icslp, 2000.

[58] M. Grice, M. Reyelt, R. Benzmüller, J. Mayer, and A. Batliner, “Consistency in transcription and labelling of German intonation with GToBI,” in *Proceeding of the Fourth International Conference on Spoken Language Processing. ICSLP’96*, vol. 3, 1996, pp. 1716–1719.

[59] M. Grice, D. R. Ladd, and A. Arvaniti, “On the place of phrase accents in intonational phonology,” *Phonology*, pp. 143–185, 2000.

[60] O. Ghitza, A.-L. L. Giraud, D. Poeppel, L. M. Miller, and U. C. Davis, “Neuronal oscillations and speech perception: Critical-band temporal envelopes are the essence,” *Frontiers in Human Neuroscience*, vol. 6, no. JAN, pp. 1–4, 2013.

[61] J. Obleser, B. Herrmann, and M. J. Henry, “Neural oscillations in speech: Don’t be enslaved by the envelope,” *Frontiers in Human Neuroscience*, vol. 6, no. AUGUST, pp. 2008–2011, 2012.

[62] A. Rosenberg, *Automatic detection and classification of prosodic events.* Columbia University, 2009.

[63] M. Breen, L. C. Dilley, J. Kraemer, and E. Gibson, “Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch),” *Corpus Linguistics and Linguistic Theory*, no. 2, 2012.

[64] J. Pitrelli, M. Beckman, and J. Hirschberg, “Evaluation of prosodic transcription labeling reliability in the ToBI framework,” in *Proceedings of the 3rd Workshop on Spoken Language Processing*, 1994, pp. 123–126.

[65] D. Bolinger, “The network tone of voice,” *Journal of broadcasting*, vol. 26, pp. 7726–728, 1982.

[66] ——, *Intonation and its uses*, 1989.

[67] C. Cotter, “Prosodic aspects of the broadcast news register,” in *Proceedings of the 19th Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Semantic Typology and Semantic Universals*, 1993, pp. 90–100.

[68] P. A. Heeman and J. F. Allen, “Speech repairs, intonational phrases, and discourse markers: Modeling speakers’ utterances in spoken dialogue,” *Computational Linguistics*, vol. 25, no. 4, pp. 527–571, 1999.

[69] C. W. Wightman, “ToBI or not ToBI?” in *Proceedings of Speech Prosody*, 2002, pp. 25–29.