A BIMODAL APPROACH TO STUDY THE EFFECTS OF RHYTHMIC PRIMING

Contreras Roa, Leonardo[1], Mairano, Paolo[2], Moreau, Caroline[3], Basirat, Anahita[4]

[1]CORPUS (EA 4295), Université de Picardie Jules Verne
[2]STL (UMR 8163), Université de Lille, France
[3]LilNCog (U1171-2), Université de Lille, CHU Lille, Inserm, France
[4]SCALab (UMR 9193), Université de Lille, CNRS, CHU Lille, France
[1]l.contreras.roa@u-picardie.fr, [2]paolo.mairano@univ-lille.fr, [3]caroline.moreau@chru-lille.fr, [4]anahita.basirat@univ-lille.fr

ABSTRACT

We present a speech rhythm analysis workflow based on two different but complementary methodologies: one at a segmental/syllabic level, and another one at a suprasegmental/prosodic level. This bimodal methodology is devised to study the effects of regular and irregular rhythmic priming — i.e. rhythmic stimuli presented to speakers before elocution — on the speech rhythm of people with Parkinson's disease. Preliminary results show that both segmental and suprasegmental rhythm metrics vary in different priming conditions. We propose that this methodology can be used to have a better grasp of speech rhythm by peering into the potential interaction or correlation between vocalic-consonantal interval distributions and pitch movement variations. Future implementations of the workflow are possible in other speech analysis fields if models are shown to be effective.

Keywords: speech rhythm, speech production, rhythmic priming, prosody, Parkinson's disease.

1. INTRODUCTION

1.1. Studies on speech rhythm

The concept of speech rhythm is a complex one, and despite the existence of a conspicuous amount of work in this domain, there is still no real comprehensive theoretical framework to fully account for speech rhythm [1]. Acoustic measures based on vocalic and consonantal durations have been proposed, which are supposed to reflect phonological characteristics thought to be related to speech rhythm (e.g., the presence of vowel reduction or of complex consonant clusters in a given language), notably ΔC, ΔV, $\%V$ [2] and PVI (Pairwise Variability Index, [3]). Such formulas have then been variously improved to account for the effects of speech rate (VarcoV, VarcoC, cf. [4]) or applied to different segments (V-to-V, cf. [5]) or to prosodic parameters other than duration such as intensity [6] and F0 [7]. While some researchers remain sceptical of these measures [8] or more generally of language rhythmicity [9], they have been used in a large number of studies characterizing the rhythm of different languages, as well as to study the effects of the rhythmic properties of the L1 on L2 pronunciation acquisition [10] and to study the characteristics of pathological speech in the context of dysarthria [11, 12].

While rhythm metrics give an account of speech rhythm that is based on segmental (or syllabic) durations, other approaches to speech rhythm have focused on suprasegmental aspects, notably the metrical structure of utterances and accents or prominence. In this context, the literature distinguishes hierarchical linguistic units responsible for creating speech rhythm: generally, the syllable, the accentual phrase (henceforth AP, which in French is marked by a primary accent on the right), the intonational phrase (henceforth IP, cf. [13]) and, according to certain approaches, an intermediate phrase (henceforth ip, cf. [14]). The literature about rhythm metrics (that we will call segmental / syllabic rhythm) and about the metrical structure and prominence (that we will call suprasegmental rhythm) belong to different frameworks, and various authors have advocated the need to develop a combined multi-level or multi-parametric model of speech rhythm [15, 12]. In this contribution, we propose a pipeline based on this type of bimodal approach to study dysprosody in Parkinson's disease (PD).

1.2. The case of dysprosody in Parkinson's disease: is rhythmic priming beneficial?

PD is the second most common neurodegenerative disease. While it is mainly known for its motor effects on the upper/lower limbs, such as tremor and walking impairment, various studies have pointed out early impairments affecting spoken communication in individuals with Parkinson's disease (PwPD). Prosody is one of the aspects that is most often reported to be affected in PD. This is referred to as dysprosody and consists of alterations at various levels such as monotony of pitch [16], reduced intonation [17], and issues in prosodic phrasing [18] and in the temporal organization of speech [19].

There is a growing body of evidence suggesting that exposure to rhythmic auditory patterns may modulate subsequent speech and language processing, at both the reception (e.g. [20]) and production (e.g. [21]) levels. This effect is known as rhythmic priming effect and thought to be based on shared cognitive representations and mechanisms for speech and music. Importantly, the transfer from musical rhythm to language seems promising in clinical settings (for a review, see [22]). Given this emerging literature on rhythmic priming effects, it seems relevant to examine whether exposure to auditory rhythm impacts speech production at prosodic levels, and improve prosody production in PwPD suffering from dysprosody.

1.3. Aim of this study

In this contribution, we present a methodological protocol for studying the effects of rhythmic priming on speech. We applied it to investigate whether non-linguistic rhythmic priming has any effect on French-speaking individuals with Parkinson's disease (PwPD), since previous research found that rhythmic priming modulates speech production (see section 1.2). The goal of this contribution is to document the data processing protocol and to reflect upon its advantages and potential uses in other areas of speech analysis. In the following sections, we give full details of the procedure that we followed and the set of Praat and R scripts that we have created as part of our pipeline are provided on GitHub at the link specified below as part of this contribution. We also present preliminary results for 12 speakers to illustrate the type of results that can be obtained.

2. A BIMODAL APPROACH TO STUDY THE EFFECTS OF RHYTHMIC PRIMING

2.1. Task

For the purpose of the current study, our pipeline was designed to analyze read speech under different priming conditions (see below). However, this pipeline can also be used to analyze speech rhythm obtained using other tasks and under other conditions.

Recordings were obtained using a sentence read-aloud task under three conditions:

1. after listening to a regular 2-second percussive rhythmic prime coinciding with the accentual structure of each sentence to read,
2. after listening to an irregular (arrhythmia) 2-second percussive prime incongruous with the accentual structure of the sentences to read,
3. after listening to two seconds of silence (no prime).

Sentences had a fixed accentual structure: four groupings of three syllables (i.e. accentual phrases) stress on their respective final syllables: $xxX_1xxX_2xxX_3xxX_4$ (e.g. *Le poisson$_1$ se nourrit$_2$ de nombreux$_3$ végétaux$_4$*). We aimed at testing (1) whether rhythmic condition impacts production in terms of speech rhythm, and (2) whether regular rhythmic priming favours a more natural performance in terms of speech rhythm among PwPD compared to irregular rhythmic priming.

2.2. The analysis protocol

The analysis is devised to combine rhythm metrics traditionally used in phonetic studies and increasingly used in the study of patients with dysarthria, along with automatic gradual prominence detection methods -as opposed to strictly categorical detection methods- based on acoustic prosodic parameters (see section 1). In this sense, our approach is bimodal, as it considers two sets of parameters that are thought to contribute to speech rhythm: traditional rhythm metrics are taken to reflect a syllabic or intersegmental level of analysis, while acoustic parameters of prominence are taken as a rhythmic measure at the accentual level. Rhythm metrics and acoustic parameters of prominence stem from different theoretical frameworks (the former from speech rhythm typology, the latter from studies on the metrical structure of speech), and - although a multilevel approach to speech rhythm has been often advocated - they have seldom been combined within the same study. We consider this to be one of the originalities of our contribution. The data analysis workflow is illustrated in Figure 1 and encompasses 4 main stages: (1) automatic speech annotation and segmentation, (2) manual check, (3) automatic prominence detection, (4) rhythm metrics calculation.

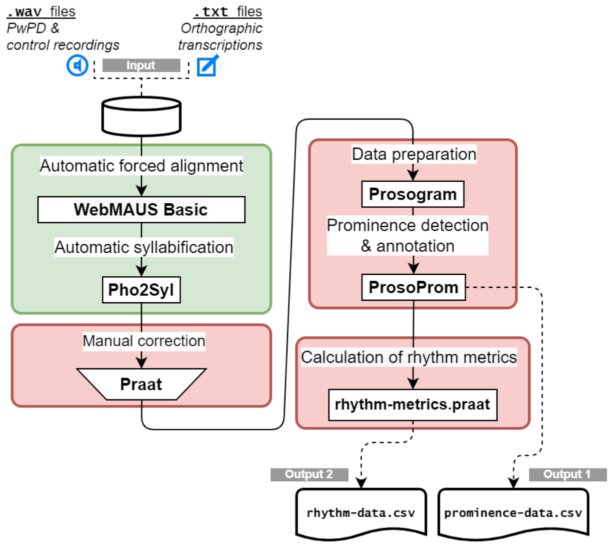

Figure 1: Data processing workflow.

2.2.1. Steps 1 and 2: speech annotation and segmentation

This step is carried out with two online services provided by the WebMAUS platform (in green in Figure 1): WebMAUS Basic [23] and Pho2Syll [24]. The audio recordings (.wav) and orthographic transcriptions (.txt) of target sentences are fed into WebMAUS Basic for semi-automatic speech annotation, and subsequently into Pho2Syll for automatic syllabification. The resulting annotated TextGrid files are then open in Praat (in red in Figure 1), verified and manually corrected if needed to address potential alignment errors.

2.2.2. Step 3: automatic prominence detection

Audio and annotation data are then fed to the Prosogram [25] and ProsoProm tools, part of the ProsoBox3 package [26]. These tools are a Praat plugin which allows to compute prominence as a gradual feature by analysing silent pauses along with the relative duration and pitch averages within an adjustable window of analysis: in this study, we set this to 2 syllables before and 1 syllable after. These relative measurements can also be carried out cumulatively to consider the relative weight of each of these acoustic cues and calculate a global score for each syllable. The duration threshold of silent pauses can be modified as well, in order to better adapt it to the study of pathological speech. We carried out our analysis with the default threshold of 250 ms.

ProsoProm's output is a TextGrid file with four extra tiers: one tiers marks syllables that are considered categorically prominent following the given threshold, and three tiers specify the cumulative contribution of relative duration, relative pitch and intra-syllable pitch movement to the gradual prominence of every syllable, whether they are classified as categorically prominent or not.

2.2.3. Step 4: computation of rhythm metrics

The output files from the previous step are then fed into a Praat script which computes the following rhythm metrics:

- %V: the percentage over which speech is vocalic for each speaker's recording.
- ΔC, ΔV: the standard deviation of the duration of consonant and vowel intervals.
- VarcoC, VarcoV: the variation coefficient for consonants and vowel intervals.
- rPVI-C, nPVI-V: raw and normalized pairwise variability index for the durations of consonantal and vocalic intervals.
- VarcoVC, rPVI-VC, nPVI-VC: variation coefficient and pairwise variability index for vocalic + consonantal intervals (i.e., measures reflecting the duration variability of pseudo-syllables).

The output datasets from steps 3 and 4 are saved in tabular format (.csv). Prominence-data.csv gives prominence data for every syllable, by speaker and by priming condition. Rhythm-data.csv gives average rhythm metrics by speaker and by priming condition. The ad-hoc Praat and R scripts we have developed are made available on GitHub at the following address: https://github.com/lcontrerasroa/EntrainPark. They can be freely used to replicate our protocol.

2.3. Results

So far, our analysis pipeline has been applied twice: once in a pilot case study [27], and once to the recordings of a larger-scale sample of participants: 5 PwPD and 7 controls who performed the task described in Section 2.1. Preliminary prominence results obtained through ProsoProm are shown in Fig. 2 which illustrates F0 movement absolute values at the four expected stressed syllables (1, 2, 3 and 4) of the read-aloud sentences.

The expected underlying prosodic structure of sentences can be seen in the productions of control subjects, who display broader pitch movements at positions 2 and 4 which correspond to prosodic boundaries. The same predicted rises are either less prominent or not observed at all in utterances produced by PwPD, particularly in

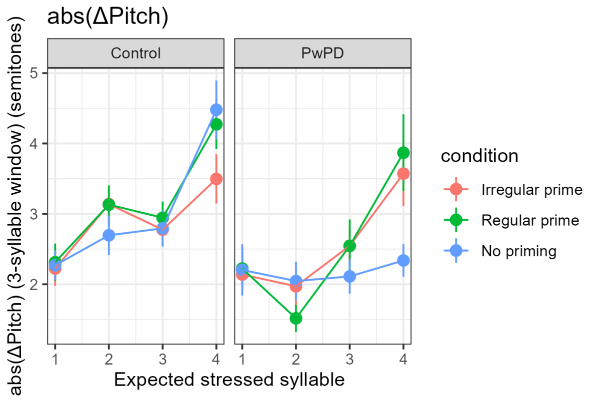

Figure 2: Absolute values of pitch movement (in semitones) at expected syllables, by group and priming condition.

absence of rhythmic priming. Two rhythm metric results (consonantal and vocalic pairwise variability indexes) are illustrated in Fig. 3.

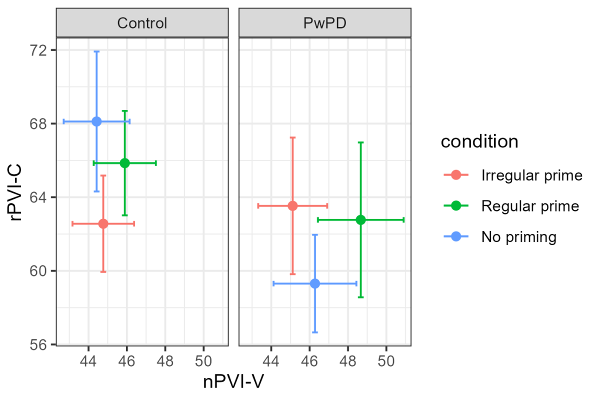

Figure 3: rPVI-C versus nPVI-V (means and standard error), by group and priming condition.

Consonantal metrics in particular (y-axis) react to priming condition, though in different directions: control subjects display higher pairwise variability scores without priming, whereas these values are lower for PwPD. Vocalic variability (x-axis) is considerably lower across priming conditions for control participants, whereas PwPD show more variation, especially between the irregular and regular priming conditions.

3. DISCUSSION

We set up a protocol to semi-automatically analyse the results of rhythm priming experiment. We ran it on preliminary data of an ongoing larger project about the effects of rhythm priming on PwPD, with promising results. The protocol can easily be replicated following the workflow instructions given in this article, and using the scripts available on the GitHub space associated to our work. The protocol and the script are by no means restricted to PwPD speech; they can be applied to any study investigating rhythm priming on any population. We believe that the main strength and originality of this protocol consists in trying to give a multimodal approach of speech rhythm, combining two levels of analysis: a first level is given by rhythm metrics, which reflect intersegmental/syllabic rhythm; a second analysis operates at the suprasegmental level, giving acoustic measures that are supposed to reflect prosodic prominence. Since both levels have been proven to be relevant for the perception of speech rhythm, such combined results can be used to create more robust statistical models to explore the interaction, and potential correlation between segmental/syllabic rhythmic and prosodic prominence. Furthermore, the combination of these two types of metrics and their contextualized analysis following the prosodic hierarchies of French allows us to bridge the gap and work at the interface of the phonetics and the phonology of speech rhythm.

The results of our preliminary study on PwPD speech are promising and will hopefully be confirmed and expanded on data that we are currently collecting. Thus far, rhythmic priming seems to have an effect on both intersegmental and prosodic cues, which display different behaviors for control subjects and PwPD. In the future, we may develop an online service with graphic user interface centralising all the steps of the analysis. This would certainly make it more accessible, and may be made available for other purposes. One possibility that we have in mind is to expand the protocol to study the effects of rhythm priming on L2/L3 acquisition, potentially also from a typological perspective, for example by comparing learners from L1s belonging to different rhythmic typologies. For now, we content ourselves with presenting our workflow combining a rhythmic analysis on two levels, and of making it available to the scientific community.

4. ACKNOWLEDGEMENT

This work was supported by French National Research Agency grant ANR-21-CE28-0011 (EntrainPark project).

5. REFERENCES

[1] J. Volín *et al.*, “Appeal and disrepute of the so-called global rhythm metrics,” *Acta Universitatis Carolinae Philologica*, no. 3, pp. 79–94, 2017.

[2] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, no. 3, pp. 265–292, 1999.

[3] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Papers in laboratory phonology*, vol. 7, no. 1982, pp. 515–546, 2002.

[4] V. Dellwo, P. Wagner, M. Solé, D. Recasens, and J. Romero, "Relations between language rhythm and speech rate," *Proc. of the 15th ICPhS*, pp. 471–474, 2003.

[5] M. Pettorino, M. Maffia, E. Pellegrino, M. Vitale, and A. De Meo, "Vtov: A perceptual cue for rhythm identification," in *Proceedings of the Prosody-Discourse Interface Conference*, 2013, pp. 101–106.

[6] L. He, "Development of speech rhythm in first language: The role of syllable intensity variability," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. EL463–EL467, 2018.

[7] L. Polyanskaya, M. G. Busà, and M. Ordin, "Capturing cross-linguistic differences in macro-rhythm: the case of italian and english," *Language and Speech*, vol. 63, no. 2, pp. 242–263, 2020.

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

[9] F. Nolan and H.-S. Jeon, "Speech rhythm: a metaphor?" *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130396, 2014.

[10] A. Tortel and D. Hirst, "Rhythm metrics and the production of english l1/l2," in *Speech Prosody 2010-Fifth International Conference*, 2010.

[11] J. M. Liss, L. White, S. L. Mattys, K. Lansford, A. J. Lotto, S. M. Spitzer, and J. N. Caviness, "Quantifying speech rhythm abnormalities in the dysarthrias," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 5, 2009.

[12] A. Lowit, A. Marchetti, S. Corson, and A. Kuschmann, "Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks," *Journal of Communication Disorders*, vol. 72, p. 26, 2018.

[13] A. Di Cristo, *La prosodie de la parole*. De Boeck Superieur, 2013.

[14] A. Michelas and M. d'Imperio, "Durational cues and prosodic phrasing in french: evidence for the intermediate phrase," in *Speech Prosody*, 2010, p. 4.

[15] P. M. Bertinetto and C. Bertini, "On modeling the rhythm of natural languages," in *Proceedings of the Fourth International Conference on Speech Prosody*, 2008, pp. 427–430.

[16] S. Skodda, H. Rinsche, and U. Schlegel, "Progression of dysprosody in parkinson's disease over time - a longitudinal study," *Movement disorders: official journal of the Movement Disorder Society*, vol. 24, no. 5, pp. 716–722, 2009.

[17] A. Basirat, C. Patin, and C. Moreau, "Relationship between Perception and Production of Intonation of French in Parkinson's Disease," in *Proceedings of the 9th International Conference on Speech Prosody 2018*, 2018, pp. 809–813.

[18] S. Frota, M. Cruz, R. Cardoso, I. Guimarães, J. J. Ferreira, S. Pinto, and M. Vigário, "(Dys)Prosody in Parkinson's Disease: Effects of Medication and Disease Duration on Intonation and Prosodic Phrasing," *Brain Sciences*, vol. 11, no. 8, p. 1100, Aug. 2021.

[19] A. Lowit, A. Marchetti, S. Corson, and A. Kuschmann, "Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks," *Journal of Communication Disorders*, vol. 72, pp. 26–39, 2018.

[20] L. Przybylski, N. Bedoin, S. Krifi-Papoz, V. Herbillon, D. Roch, L. Léculier, S. A. Kotz, and B. Tillmann, "Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders." *Neuropsychology*, vol. 27, no. 1, pp. 121–131, 2013.

[21] N. Cason and D. Schön, "Rhythmic priming enhances the phonological processing of speech," *Neuropsychologia*, vol. 50, no. 11, pp. 2652–2658, Sep. 2012.

[22] A. Fiveash, N. Bedoin, R. L. Gordon, and B. Tillmann, "Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders," *Neuropsychology*, Aug. 2021.

[23] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[24] U. D. Reichel, "Language-independent grapheme-phoneme conversion and word stress assignment as a web service," *Elektronische Sprachverarbeitung. Studientexte zur Sprachkommunikation*, vol. 71, pp. 42–49, 2014.

[25] P. Mertens, "The prosogram: Semi-automatic transcription of prosody based on a tonal perception model," in *Speech prosody 2004, international conference*, 2004.

[26] J.-P. Goldman and A. C. Simon-Hustinx, "Prosobox, a praat plugin for analysing prosody," in *Proceedings of the 10th International Conference on Speech Prosody*, 2020, pp. 1009–1013.

[27] L. Contreras Roa, P. Mairano, C. Moreau, and A. Basirat, "Impact de l'amorçage rythmique sur la production de la parole chez des personnes atteintes de la maladie de Parkinson : Étude pilote," in *Proc. XXXIVe Journées d'Études sur la Parole – JEP 2022*, 2022, pp. 578–587.