



**HAL**  
open science

# An automated pipeline for preprocessing spontaneous L2 English prosody

Sylvain Coulange, Tsuneo Kato, Solange Rossato, Monica Masperi

► **To cite this version:**

Sylvain Coulange, Tsuneo Kato, Solange Rossato, Monica Masperi. An automated pipeline for preprocessing spontaneous L2 English prosody. The 13th International Seminar on Speech Production (ISSP 2024), May 2024, Autrans, France. . hal-04595972

**HAL Id: hal-04595972**

**<https://hal.science/hal-04595972>**

Submitted on 31 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# An automated pipeline for preprocessing spontaneous L2 English prosody

Sylvain COULANGE<sup>1,2,3</sup>, Tsuneo KATO<sup>3</sup>, Solange ROSSATO<sup>2</sup>, Monica MASPERI<sup>1</sup>

1. Univ. Grenoble Alpes, Laboratory of Linguistics and Didactics of Foreign and Mother Tongues (LIDILEM) 38000 Grenoble, France

2. Univ. Grenoble Alpes, CNRS, Institute of Engineering, Grenoble Computer Science Laboratory (LIG) 38000 Grenoble, France

3. Doshisha Univ., Spoken Language Processing Laboratory (SLPL) 610-0394 Kyoto, Japan

sylvain.coulangue@univ-grenoble-alpes.fr, tsukato@mail.doshisha.ac.jp, solange.rossato@univ-grenoble-alpes.fr, monica.masperi@univ-grenoble-alpes.fr

## Context

- Prosody (e.g. fluency and rhythm) is crucial for intelligibility in L2 [1].
- Little work is done on prosody in classrooms.
- Teachers and learners both need tools helping measuring prosody.
- Existing Computer Assisted Pronunciation Training systems rarely deal with spontaneous speech and conversation situations.

## Proposition

- We developed a pipeline to help measuring speech segmentation and rhythm in spontaneous conversations involving learners of English.
- It combines most recent tools for speech recognition and syntactic analysis with lexical stress analysis and pause position analysis.

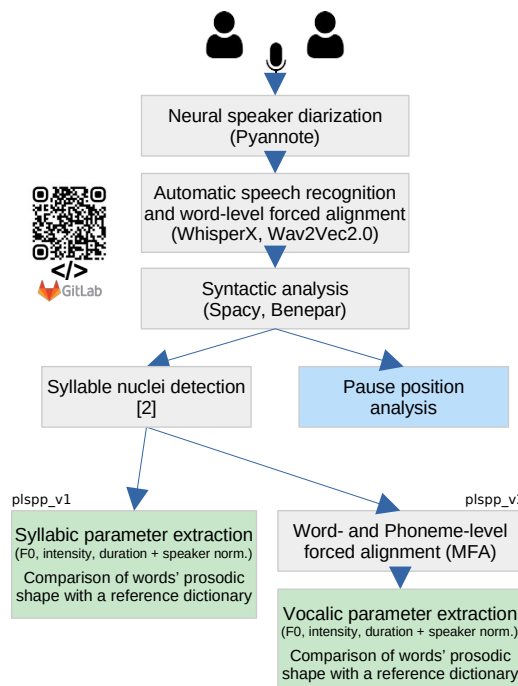
## Input

- Multispeaker spontaneous L2 English audio recordings.

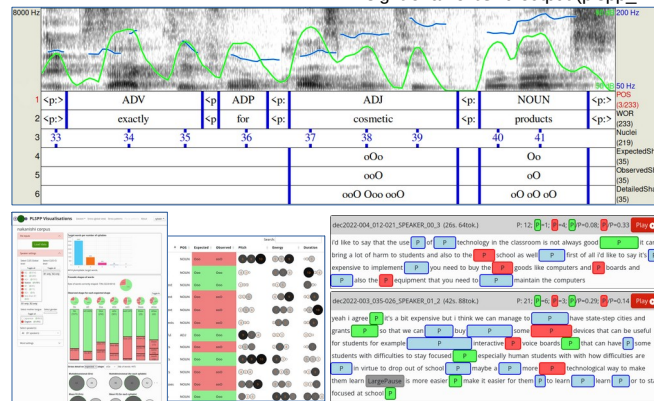
## Outputs

- TextGrid and csv files showing syllabic prominence, expected stress patterns and pause syntactic context.
- Visualisation of stress and pauses through an interactive web interface.

## The Pauses & Lexical Stress Processing Pipeline (PLSPP)



Insight of a TextGrid output (plspp\_v1)



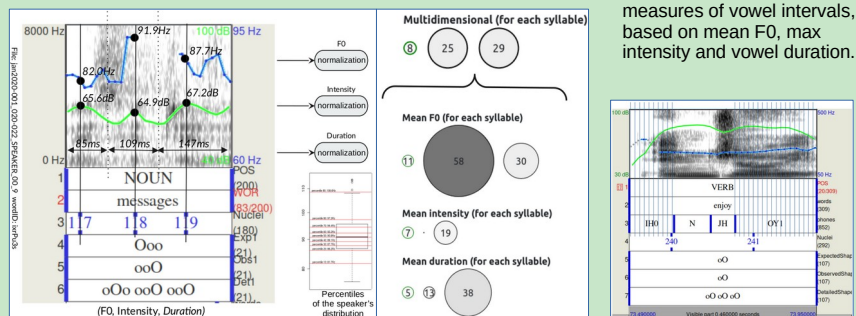
A server-based visualisation tool allows to easily view the processing outputs.

## Pipeline Evaluation & Limitations

- Precision of ASR and syntactic analysis varies greatly with speaking tasks and speakers' proficiency level. Because of MFA sensitivity to disfluencies, plspp\_v2 is less robust than v1 and works well only with read speech so far.
- With spontaneous conversations from [3] (plspp\_v1):
  - 41% of words have adequate number of syllable nuclei detected and thus analysed.
  - Manual evaluation of random 100 target words showed that 17% were miss-recognized or miss-aligned.
- With read-aloud monologues from [4] (plspp\_v2 with forced-alignment of reference texts):
  - 100% of words are analysed since it does not depend on syllable nuclei detection.
  - Manual evaluation of 400 words among 8 Japanese-L1 and 8 native speakers showed that 8.8% of words were miss-aligned (Japanese-L1: 9%, native speakers: 8.5%).

## Lexical Stress Analysis

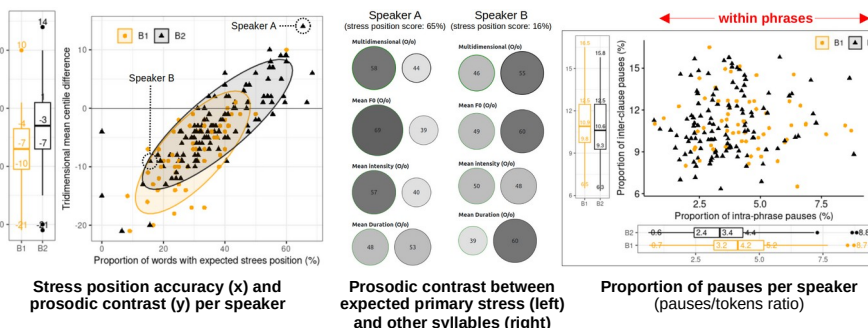
- plspp\_v1: Lexical stress is estimated from prosodic measures on syllable nuclei, based on F0, intensity and syllable duration.
- plspp\_v2: Lexical stress is estimated from prosodic measures of vowel intervals, based on mean F0, max intensity and vowel duration.



- In both versions, each prosodic measure is converted in speaker percentile, so that 50 means median prominence level for any speaker, 0 is minimum and 100 is maximum prominence level.

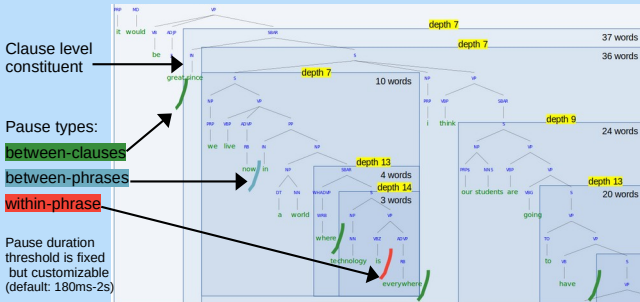
## Stress accuracy and pause positions in B1/B2 French-L1 spontaneous conversations [3]

- Corpus: 176 French-L1 university students (11 hours), 6350 polysyllabic target words, 21942 pauses.
- Main observations:
  - Mean stress position accuracy varies greatly among speakers (0-68.4%, mean: 35.4%).
  - B2 speakers perform better than B1 in terms of stress position accuracy (36% vs. 29.6%, rank test  $p < .001$ ) and prosodic contrast between expected primary stress syllable and mean of other syllables ( $p < .001$ ).
  - Strong impact of last syllable lengthening and pitch rise, the better the speaker mean stress position accuracy, the higher pitch and intensity of expected stressed syllable.
  - B2 speakers tend to make more but shorter pauses than B1 (median 34.3 pauses/min/speaker vs. 30.7,  $p < .01$ ; 592ms vs. 615ms,  $p < .01$ ), but proportionally less pauses within phrases than B1 speakers (4.2 pauses for 100 tokens vs. 3.4,  $p < .01$ ).



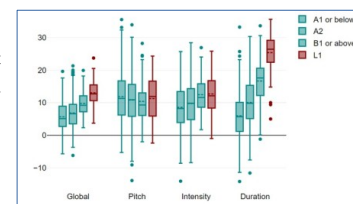
## Pause Position Analysis

- Pauses are categorized depending on their syntactic position in relation to clause and phrase level constituents, along with POS context, syntactic depth and nb. of words of adjacent constituents.



## Prosodic contrast between content and function words in Japanese and English-L1 monologues [4]

- Corpus: 42 Japanese-L1 speakers (A1-B2), 9 English-L1 professional narrators, 34 hours read-aloud speech.
- PLSPP\_v2, extended to monosyllabic words analysis, based on forced alignment of reference text.
- Main observations:
  - Contrast between most prominent syllable from content words and that of function words is higher for Native speakers than Japanese-L1 speakers ( $p < .001$ ).
  - The higher the proficiency level, the bigger the contrast.
  - Most of this contrast is due to vowel duration parameter.



## References:

- Levis, J. M. (2018). Cambridge applied linguistics: Intelligibility, oral communication, and the teaching of pronunciation.
- De Jong, N. H., Pacilly, J., Heeren, W. (2021) "Praat scripts to measure speed fluency and breakdown fluency in speech automatically," Assessment in Education: Principles, Policy & Practice, 28, 456-476.
- Coulange S, Kato T, Rossato S, Masperi M. (2024). Enhancing Language Learners' Comprehensibility through Automated Analysis of Pause Positions and Syllable Prominence. Languages 9(3):78
- Coulange, S., Nakanishi, M. (subm.). Measuring speech rhythm through automated analysis of syllabic prominences: Prosodic features of language learners' fluency (Speech Prosody WS), July 1, Leiden.