

A corpus of spontaneous dialogues in L2 English by French and Japanese L1 speakers for automated assessment of fluency

Sylvain COULANGE^{1,2}, Takayuki KONISHI³, Tsuneo KATO², Mariko SUGAHARA², Solange ROSSATO¹, Monica MASPERI¹

¹ Grenoble Alpes Univ.; ² Doshisha Univ.; ³ Waseda Univ.

Context:

- CAPT tools rarely deal with spontaneous speech, and even more rarely with speech in real discussion situation.
- Lack of L2 spontaneous speech corpus.
- Lack of speech in peer dialogue situations.

Creation of a speech corpus:

- 2- or 3-student role play type argumentative discussion on a contentious topic.
- Different topics were used, such as security cameras, animal testing, the use of technology in classrooms, part-time jobs...
- Each candidate assumed a specific given role, either advocating for or against the subject.
- 2-5 minutes of preparation before the talk.
- Objective: negotiate, exchange viewpoints, and eventually work towards a compromise.

Processing Pipeline

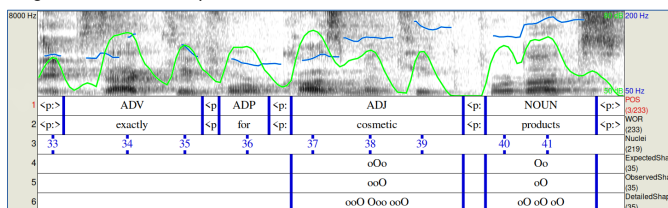
- We released an open-source automated processing pipeline specifically design for processing multi-speaker spontaneous L2 English speech. [3]
- The processing steps are as follows:

- Speech detection and neural speaker diarization (Pyannote)
- ASR & word-level alignment (WhisperX)
- Morphosyntactic analysis (SpaCy)

- **Localisation of pauses** with POS context and constituency analysis (Benepar)

- **Syllable nuclei detection** [4]
- **Syllabic parameter extraction** (intonation, intensity, duration ; speaker norm.)
- **Comparison of prosodic shape of words** with a reference dictionary

- Insight of a TextGrid output:



pipeline
on
GitHub

Data Available for Academic Research

CLES French-L1 corpus

(Public portion of the CLES corpus of spontaneous L2 English) [1,2]

- 128 speakers
- French as mother tongue: 93% (other: Albanese, Arabic, Chinese, Georgian, Indonesian, Latvian, Persian, Spanish, Ukrainian)
- 48% F, 52% M
- 62 groups (3-speaker: 4, 2-speaker: 58)
- Proficiency: B1~B2
- Speech duration: 10 hours (mean 9'35", min 5'12", max 14'30")

Data available for academic research: coordination-nationale@certification-cles.fr

Waseda-Doshisha Japanese-L1 corpus

- 30 speakers
- 100% Japanese-L1
- 60% F, 40% M
- 16 groups (16 pairs, including 2 with a Japanese-L1 English teacher)
- Proficiency: B1~C2
- Speech duration: 4 hours (mean 16:39, min 9:54, max 33:52)

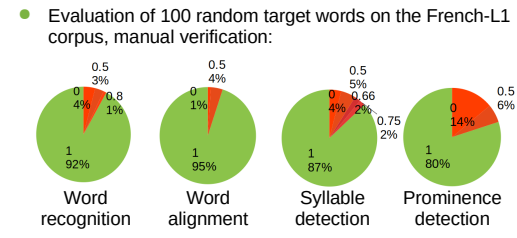
Data available for academic research: msugahar@mail.doshisha.ac.jp

Doshisha English-L1 corpus

- 14 speakers
- 100% English-L1
- 64% F, 36% M
- 7 groups
- Speech duration: 2 hours (mean 17:24, min 13:20, max 21:18)

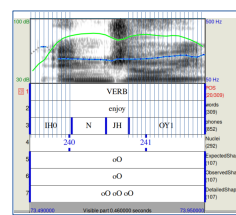
Data available for academic research: msugahar@mail.doshisha.ac.jp

Pipeline Evaluation

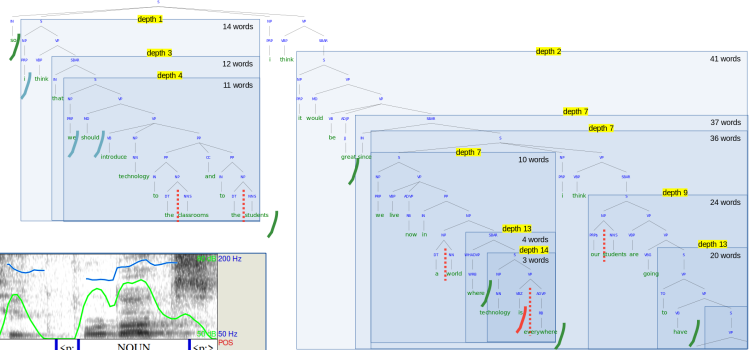


New Version of the Pipeline

- A new version of the pipeline is currently being developed. It now uses a **phoneme-level alignment step** to measure prominence at **vowel intervals**, as well as F0 dynamics.



Pause Position Analysis



- Pauses are categorized into **inter-clause**, **inter-phrase** and **intra-phrase** pauses, along with POS context, syntactic depth and nb. of words of adjacent constituents.

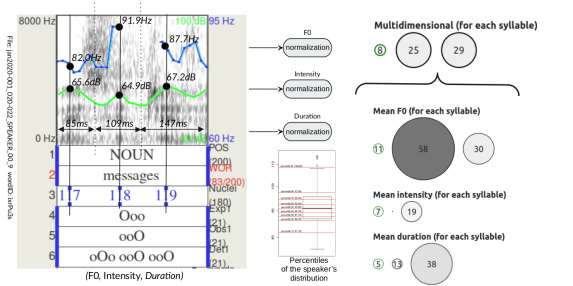
Visualisation Platform

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



Lexical Stress Analysis

- Lexical stress is estimated from prosodic prominence of syllables, based on measures of F0, intensity and duration.



References:

- [1] CLES official website: <https://www.certification-cles.fr/english/>
- [2] Coulange, S., Fries, M.-H., Masperi, M., Rossato, R. (submitted). A corpus of spontaneous L2 English speech for real-situation speaking assessment. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 20-25 May, Torino, Italy.
- [3] Coulange, S., Kato, T., Rossato, R., Masperi, M. (in press). Enhancing Language Learners' Comprehensibility through Automated Analysis of Pause Positions and Syllable Prominence. In Mairano, P. & Schwab, S (eds.) *Languages, Special Issue 'Speech Analysis and Tools in L2 Pronunciation Acquisition'*.
- [4] 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.