

Pause position analysis in spontaneous speech for L2 English fluency assessment

Sylvain Coulange, Tsuneo Kato

▶ To cite this version:

Sylvain Coulange, Tsuneo Kato. Pause position analysis in spontaneous speech for L2 English fluency assessment. 2023 Autumn Meeting of the Acoustic Society of Japan, Sep 2023, Nagoya, Japon, Japan. 150 2023 Autumn Meeting of the Acoustic Society of Japan, 2023. hal-04257261

HAL Id: hal-04257261

https://hal.science/hal-04257261

Submitted on 25 Oct 2023

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.

Copyright





Pause position analysis in spontaneous speech for L2 English fluency assessment

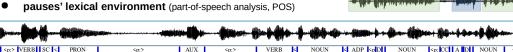
Sylvain COULANGE^{1,2}, Tsuneo KATO²

¹LIDILEM/LIG, Univ. Grenoble Alpes; ²SLPL, Doshisha Univ.

Context:

- Fluency assessment often deals with frequency and length of pauses,
- but a pause itself is not necessarily a disfluency. Pauses may help listeners to process the speech if they are strategically used, or they can disturb them if appearing at unexpected positions. [1]
- We present a pipeline for pause position analysis in spontaneous L2 English.
- Pauses are defined here as silent or filled speech interruptions of duration ranging from 180ms to 2s.
- This study analyses:
 - pauses between clauses (considered as structurant)
 - pauses within phrases (considered as disfluent)

Processing pipeline: Speech detection and neural speaker diarization (Pvannote) - ASR & Word-level Forced Alignment (WhisperX) Morphosyntactic analysis (SpaCy) Constituency analysis (Benepar) speaker2 speaker1



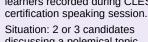
Corpus:

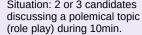


L2 English spontaneous speech from 176 French learners recorded during CLES









- Total 11 hours of continuous speech (per speaker: mean 3'44", min 32", max 6'51)
- Speaking B1 level: 34%, B2 level: 66%
- Speech duration: B1≈B2, Nb tokens: B1<B2, Nb pauses: B1<B2, Silence proportion: B1≈B2

Feature extraction:

- Pauses as inter-word <p:> tag of duration 180ms-2s.
- Total: 72,594 <p:> tags, including 21,942 pauses.
- For each <p:> tag:
 - get starting and ending biggest constituent label, number of words and syntactic depth;
 - get previous & next word's part-of-speech tag.

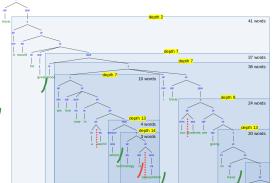
References:

- [1] Isaacs, T., Trofimovich, P. and Foote, J. (2018). Developing a user-oriented second language comprehensibility scale for english-medium universities. Language Testing 35(2), 193-216.
- [2] Nagle, C., Trofimovich, P. and Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. Studies in Second Language Acquisition 41(4), 647-672.

14 word: 12 words

Dynamic rating of comprehensibility [2]

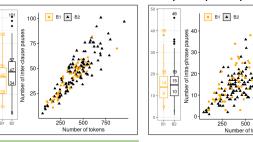




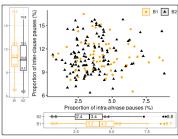
Structural analysis:

- **Great variation** of number of intra-phrasal pauses, less with inter-clausal pauses;
- No correlation between intra-phrasal and inter-clausal pause proportions:
- B2 speakers make less intra-phrasal pauses than B1 speakers;
- ...but difference between B1 and B2 is small.

Absolute number of inter-clause and intra-phrase pauses per speaker



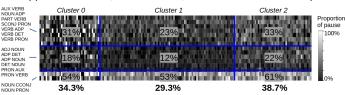
Proportion of inter-clause and intra-phrase pauses per speaker (nb pauses / nb tokens)



Lexical analysis:

- Pauses in same proportions for B1 and B2 for top15 most frequent POS contexts;
- B2 speakers make generally less pauses in these contexts.

Clustering output of pausing patterns in top 15 POS contexts (speakers in column, contexts in rows, with mean value of each block)



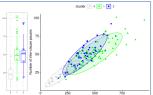
Proportion of pause for the 15 most frequent part-of-speech contexts

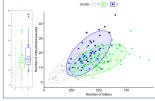
• B1 ▲ B2 NOUN ADP ADP DET ADP NOUN NOUN CCONT VERR PROM SCONT PROM VERB ADP 20 30

Grouping speakers according to pausing patterns:

- Ultimate segmentation is 3x3 groups;
- B1 and B2 speakers are mixed together in each group;
- Clusters 1 and 2 differenciate by overall frequency of pauses, Cluster 0 contains speakers with extreme values (too few occurrences) and intra-phrase pause frequency.

Absolute number of inter-clause and intra-phrase pauses per speaker





Discussion:

- Limited contrast between B1 and B2 speakers;
- Instead, large inter-speaker diversity in pausing pattern, especially within phrases;
- Need for inverstigating intra-speaker variability;
- Do more intra-phrasal pauses, and less inter-clausal pauses mean worse comprehensibility?
- To investigate it, we plan to confront these results with dynamic ratings of comprehensibility [2]