Comprehensibility diagnosis of spontaneous L2 English:

Automated analysis of pausing and lexical stress patterns

Sylvain Coulange[1,2a,2b] , Tsuneo Kato [1], Solange Rossato [2a], Monica Masperi [2b]

1 Doshisha Univ., Spoken Language Processing Laboratory (SLPL), 610-0394 Kyoto, Japan
2a Univ. Grenoble Alpes, CNRS, Institute of Engineering, Grenoble Computer Science Laboratory (LIG), Grenoble, France
2b Univ. Grenoble Alpes, Laboratory of Linguistics and Didactics of Foreign and Mother Tongues (LIDILEM), Grenoble, France

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

科研費 KAKENHI This research is supported by JSPS Kakenhi n°23H00648

financé par IDEX Université Grenoble Alpes

Outlines

- Automated assessment of L2 English and comprehensibility
- Pauses and Lexical Stress Processing Pipeline (PLSPP)
- Premilinary studies
- Next steps and Discussion

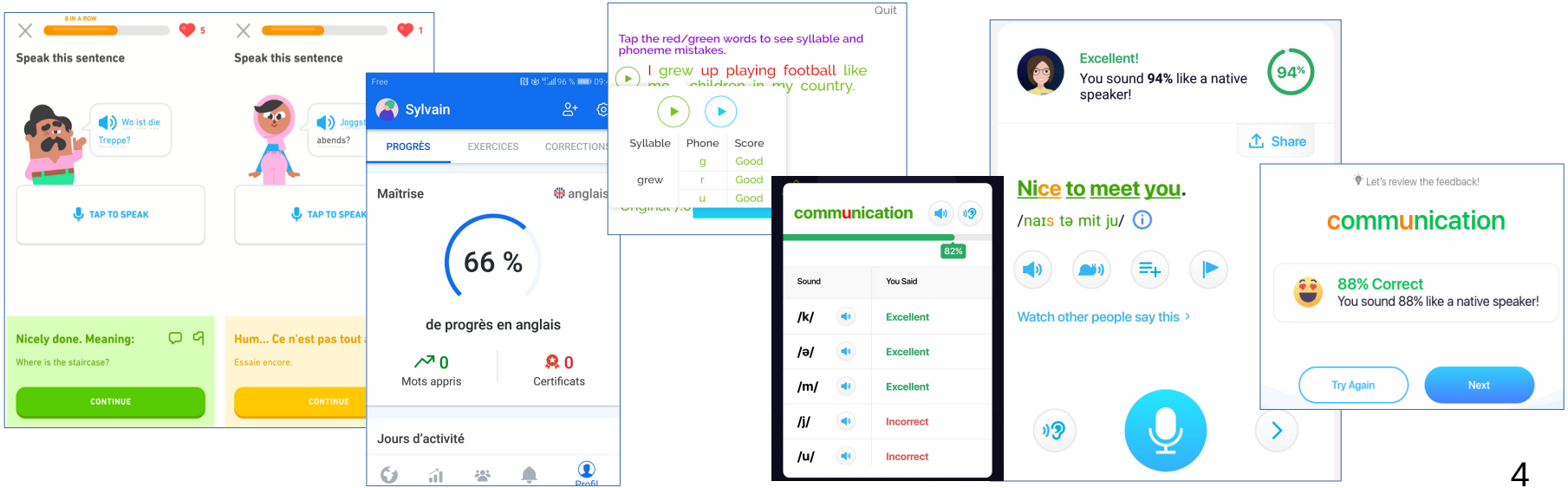
LIDILEM Université Grenoble Alpes
SLPL Spoken Language Processing Laboratory
UGA Université Grenoble Alpes
同志社大学 Doshisha University

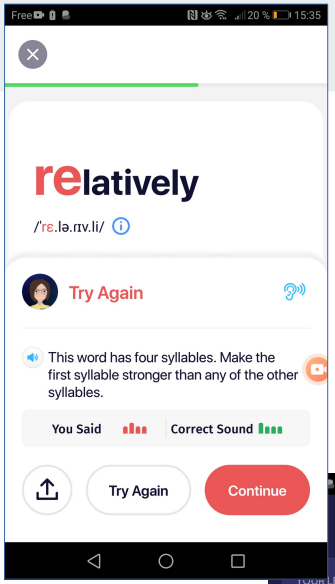
Automated assessment of L2 and Comprehensibility

Apps offering automated pronunciation feedback

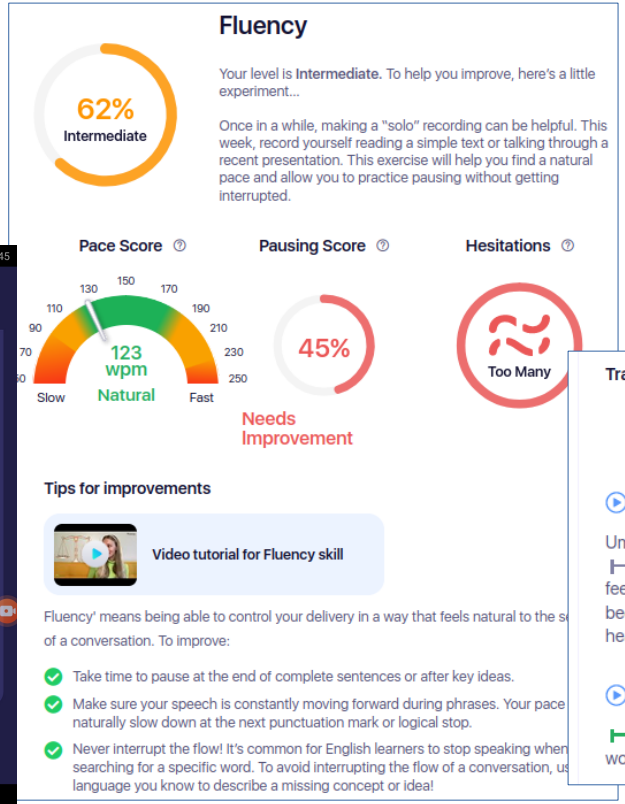
For a wider review:
Coulange (2023). Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up. Proc. EPIP7

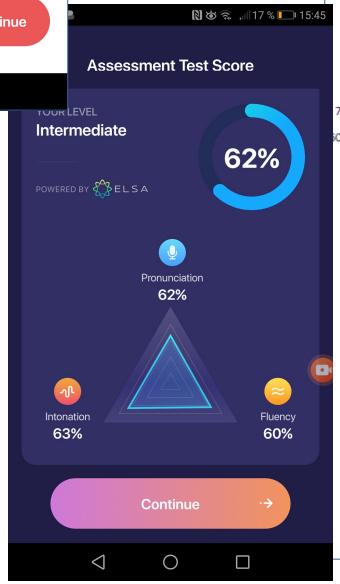
duolingo, MEM RISE, (Babbel), (Busuu), (Rosetta Stone), ELSA, IELTS Speaking Practice, LingoChamp, ...

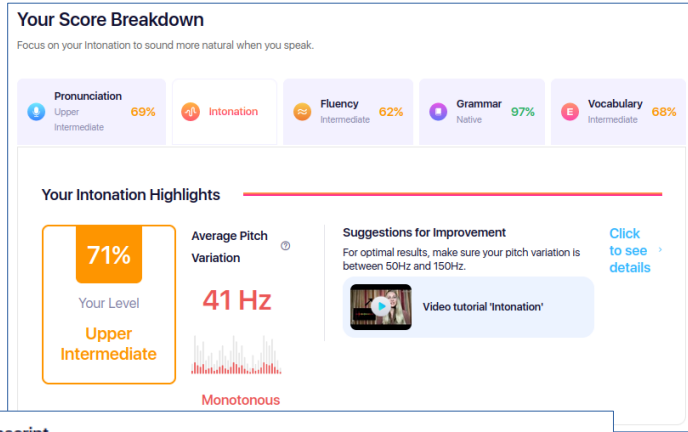


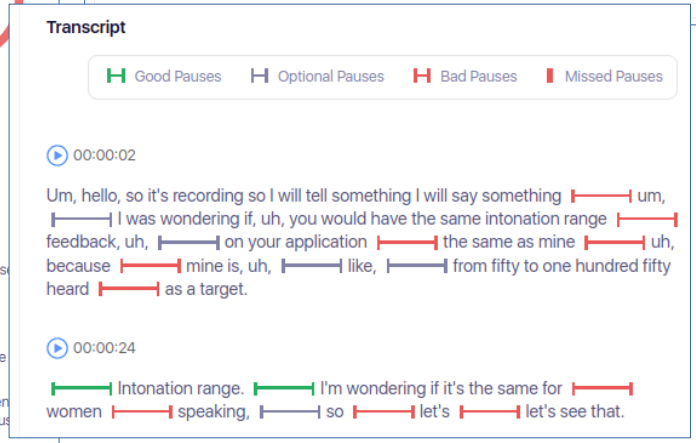


ELSA
Speech Analyzer





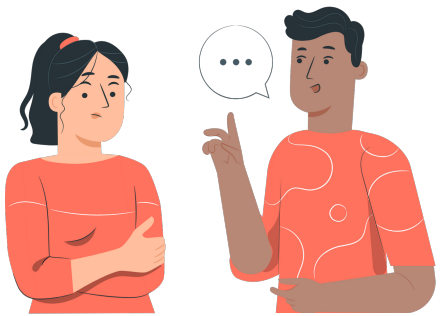




From nativelikeness to intelligibility

Native speaker as a target

Be (easily) understood



“Intelligibility”

“Comprehensibility”

Isaacs, T., Trofimovich, P., and Foote, J. A. (2018) Developing a user-oriented L2 comprehensibility scale for english-medium universities. Language Testing 35(2), 193–216.
Jenkins, J., Baker, W., & Dewey, M. (Eds.). (2017) The Routledge Handbook of English as a Lingua Franca (1st ed.). Routledge.
Frost, D., O’Donnell, J. (2018) Evaluating the essentials, the place of prosody in oral production. In J. Volín (ed.). Pronunciation of EFL.
Council of Europe (2020) Common European framework of reference for languages. Strasbourg, France.
Walker, R., Low, E., & Setter, J. (2021) English pronunciation for a global world. Oxford: Oxford University Press

Isaacs et al (2018)
Second Language English Comprehensibility Global and Analytic Scales, Version 1.0

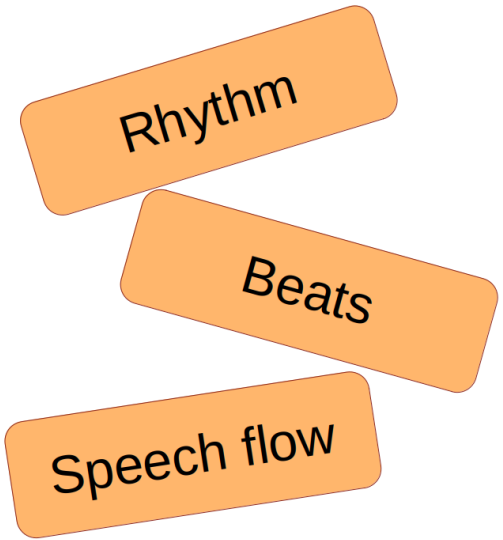
COMPREHENSIBILITY LEVEL	OVERALL DESCRIPTION OF COMPREHENSIBILITY
5	**Speech is effortless to understand** *Errors, are rare and do not interfere with the message* ***Sounding nativelike or producing hesitation- or error-free speech is not necessary to achieve a level 5 (highest level)*
4	**Speech requires little effort to understand** *Errors minimally interfere with the message*
3	**Speech requires some effort to understand** *Errors somewhat interfere with the message*
2	**Speech is effortful to understand** *Errors are detrimental to the message*
1	**Speech is painstakingly effortful to understand or indecipherable** *Errors are debilitating to the message* ***Not enough comprehensible language is generated for coherent communication, confining the speaker to level 1*
UR	**Unable to Rate the speech** *No assessable speech sample is produced (e.g., unresponsive to the task, no articulation of English-like sounds)*

COMP	PRONUNCIATION	FLUENCY	VOCABULARY	GRAMMAR
5	- pronunciation is **effortless to understand** - errors do not interfere with the message - pitch variation may make the speech sound lively or engaging - sounding nativelike is *not* expected	- fluent speech, which is optimally paced, is **effortless to understand** - hesitation markers are used at appropriate junctures or strategically to sustain listener attention	- precise lexical choice relevant to the task is **effortless to understand** - errors do not interfere with the message - nuanced idiomatic expressions may be present, depending on the task	grammatical use conveys precise meaning or nuance, resulting in speech that is **effortless to understand** - errors do not interfere with the message - complex sentences may be used, depending on the task
4	- pronunciation requires **little effort to understand** - errors minimally interfere with the message - speech may be characterized by too many or too few variations in pitch, sounding disjointed or monotone	- mostly fluent speech, which may be slightly too fast or slow, requires **little effort to understand** - hesitation markers are generally used at appropriate junctures	- sufficient lexical choice mostly relevant to the task requires **little effort to understand** - errors minimally interfere with the message - unusual or less familiar lexical expressions may be used	- grammatical use mostly conveys precise meaning, resulting in speech that requires **little effort to understand** - errors minimally interfere with the message - a mix of simple and complex sentences are used
3	- pronunciation requires **some effort to understand** - errors somewhat interfere with the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence)	- somewhat fluent speech, which is too fast or slow, requires **some effort to understand** - hesitation markers are occasionally used at inappropriate junctures	- simple lexical choice requires **some effort to understand** - errors somewhat interfere with the message - occasional gaps in vocabulary make the speech somewhat labored, although meaning is still roughly conveyed	- grammatical use conveys general meaning, resulting in speech that requires **some effort to understand** - errors somewhat interfere with the message - simpler sentences are used instead of more complex ones
2	- pronunciation is **effortful to understand** - errors are detrimental to the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence) - production difficulties may obscure the meaning of a few words	- speech, which is markedly dysfluent or too fast, is **effortful to understand** - hesitation markers are frequently used at inappropriate junctures - compensatory strategies are used to offset gaps in fluency (e.g., ideas are described in a roundabout way, self-correction)	- limited lexical choice and frequent lexical errors are **effortful to understand** - errors are detrimental to the message - frequent gaps in vocabulary may make the speech labored or unelaborated - lexical chunks may be used to compensate for limited vocabulary	- grammatical use may obscure meaning, resulting in speech that is **effortful to understand** - errors are detrimental to the message - only basic sentence structures are used
1	- pronunciation is **painstakingly effortful to understand** - errors are debilitating to the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence) - production difficulties may make words sound slurred or indistinct	- speech, which is extremely dysfluent or much too fast, is **painstakingly effortful to understand** - hesitation markers are very frequently used at inappropriate junctures, leading to halting or "broken" speech - no compensatory strategies are used to offset gaps in fluency	- extremely simplistic or limited lexical choice and very frequent lexical errors make the speech **painstakingly effortful to understand** - errors are debilitating to the message - frequent gaps in vocabulary make the speech unelaborated or indecipherable - no lexical chunks are used to compensate for limited vocabulary	- grammatical use obscures meaning, making the speech **painstakingly effortful to understand** - errors are debilitating to the message - only very basic or fragmented sentences are used
UR	Unable to Rate. Speaker does not produce an assessable sample of speech (e.g., unresponsive to the task, no articulation of English-like sounds)			

1 = low comprehensibility; 5 = high comprehensibility

Isaacs, T., Trofimovich, P., and Foote, J. A. (2018) Developing a user-oriented L2 comprehensibility scale for english-medium universities. Language Testing 35(2), 193–216.

From nativelikeness to intelligibility

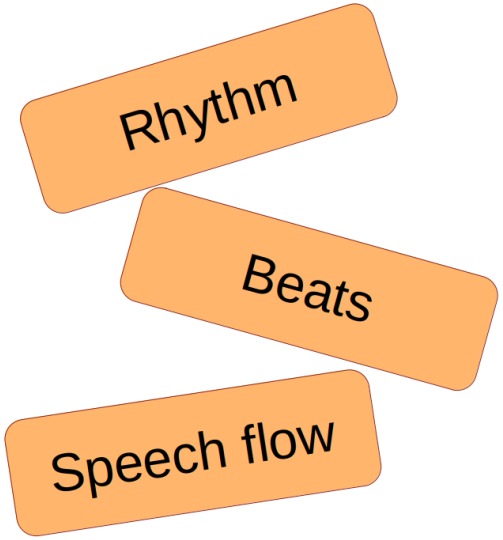


Parameters related to L2 English comprehensibility:

- Hesitation markers position (pauses, false starts, repetitions…)
- Lexical stress (presence, position, quality)
- Speech rate (not too fast, not too slow)
- Pitch variation (make the speech sound lively and engaging)
- Phonemes quality (depending on their functional load)

Isaacs, T., Trofimovich, P., and Foote, J. A. (2018) Developing a user-oriented L2 comprehensibility scale for english-medium universities. Language Testing 35(2), 193–216.
Jenkins, J., Baker, W., & Dewey, M. (Eds.). (2017) The Routledge Handbook of English as a Lingua Franca (1st ed.). Routledge.
Frost, D., O'Donnell, J. (2018) Evaluating the essentials, the place of prosody in oral production. In J. Volín (ed.). Pronunciation of EFL.
Council of Europe (2020) Common European framework of reference for languages. Strasbourg, France.
Walker, R., Low, E., & Setter, J. (2021) English pronunciation for a global world. Oxford: Oxford University Press

From nativelikeness to intelligibility



Parameters related to L2 English comprehensibility:

- Hesitation markers position (pauses, false starts, repetitions…)
- Lexical stress (presence, position, quality)

Université Grenoble Alpes (France) - 3rd year

Doshisha University (Japan)

Semi-automatic diagnosis of spontaneous English as a foreign language: the role of rhythm in speaker comprehensibility

Enlgish L1 vs. L2: Pause patterns

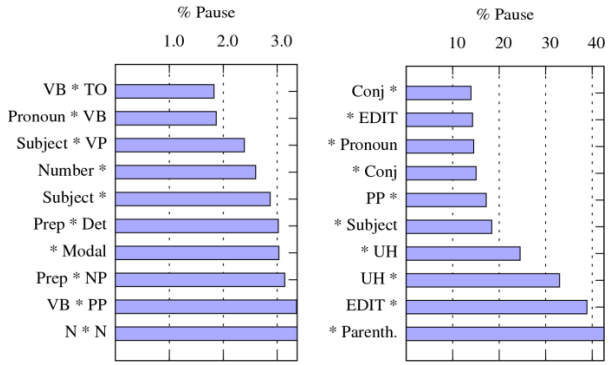
- Pauses allows to **regulate** speech flow and **structure** the discourse (Dodane & Hirsch 2018)
- Pause = silent or filled interruption of speech (hums, false starts, repetitions)
- Various duration thresholds from 100ms (Trouvain 2004) to 400ms (Tavakoli 2011)

- More pauses in L2 (Fauth & Trouvain 2018)
- More pauses at lower proficiency (Fauth & Trouvain 2018)

- Most pauses arise at **expected positions** (Candea 2000)
- Pauses at inadequate junctures → **Low** comprehensibility (Isaacs et al. 2017)
- Pauses at strategic junctures → **High** comprehensibility (Isaacs et al. 2017)

Candea, M. (2000). Contribution à l'étude des pauses silencieuses et des phénomènes dits «d'hésitation» en français oral spontané. Ph. D. thesis, Paris 3.
Dodane, C. & Hirsch, F. (2018). L'organisation spatiale et temporelle de la pause en parole et en discours. Langages, 211, 5-12.
Fauth, C., & Trouvain, J. (2018). Détails phonétiques dans la réalisation des pauses en Français : étude de parole lue en langue maternelle vs en langue étrangère. Langages, N°211(3), 81–95.
Isaacs, T., Trofimovich, P., Foote, J. (2018). Developing a user-oriented second language comprehensibility scale for english-medium universities. Language Testing 35(2), 193–216.
Tavakoli, P. (2010). Pausing patterns: differences between L2 learners and native speakers. ELT Journal 65(1), 71–79.
Trouvain, J. (2004). Tempo Variation in Speech Production: Implications for Speech Synthesis. Ph. D. thesis, Saarland University.

Enlgish L1 vs. L2: Pause patterns



Rarest (left) and most frequent (right) pause contexts in L1 spontaneaous English (Tauberer 2008:407)

L1: (Switchboard corpus, Tauberer 2008)

- Pauses **unexpected** between S*VP, Prep*Det, Prep*NP, S*
- Pauses **expected** next to hesitations, conjunctions, before subjects

Successful public speakers: (Cao & Chen 2019)

- Most pauses between **main and subordinate clauses**

Pause position in successful public speakers' speech (Cao & Chen 2019:2050)

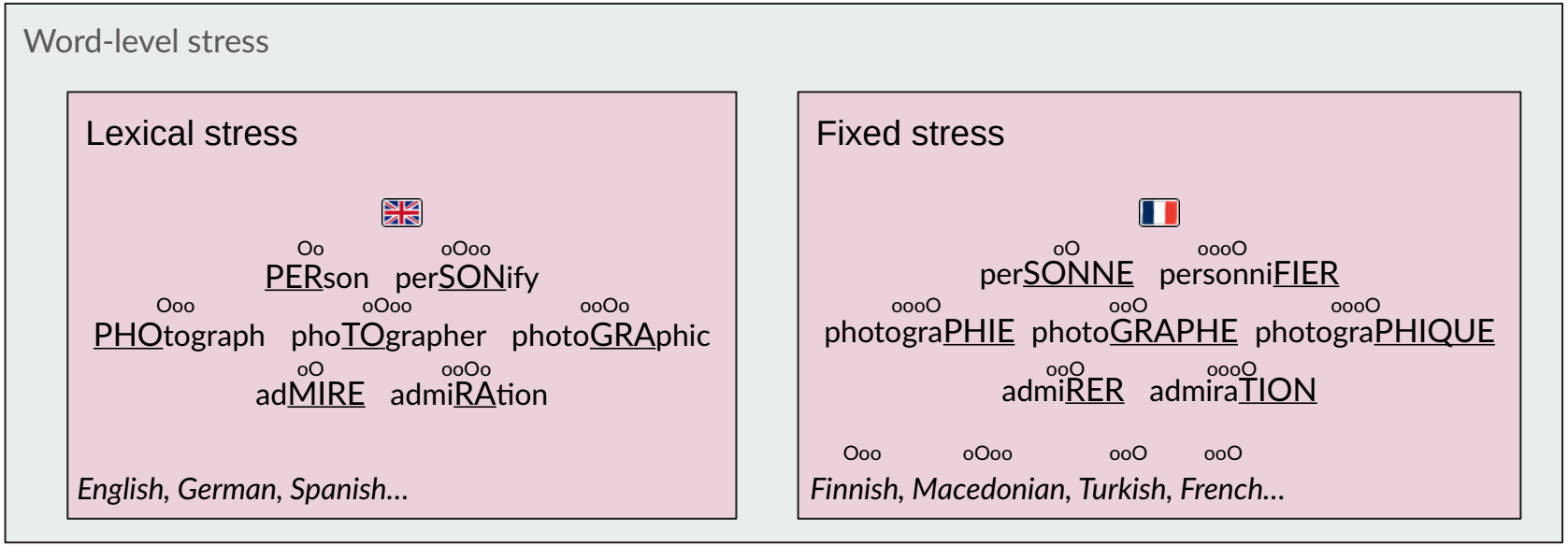
	Speakers	Example	Structure
1	*British*	be it through the Commonwealth Games /{518 ms}/ **which** begin in a few months' time on Australia's Gold Coast	main clause // attributive clause
2	*American*	we must never forget /{501ms}/ **that** those heroes /{314 ms}/ **who** fought against evil /{568 ms}/ also fought for /{450 ms}/ the nations /{433 ms}/ **that** they loved	main clause // object clause // attributive clause // attributive clause
3	*Nigerian*	**when** you are speaking /{202 ms}/ I put my ear down to understand what you say	adverbial clause // main clause
4	*Japanese*	**If** it is not available in your area /{398 ms}/ you can also use ham instead	conditional clause // main clause
5	*American*	we must never forget // that those heroes // who fought against evil // **also fought for** // **the nations** // **that they loved**	independent tone group indicting emphatic function

Disclaimer:

- *Pause often used for **emphasize** (Cao & Chen 2019)*
- *Pause use is highly linked to **phonostyles***
- *Pause often used for **dialogue** management*
- *One pause might have **several purposes/causes***
- ***Acoustic** pauses ≠ **Perceived** pauses (Dodane & Hirsch 2018)*

Cao, Y., Chen, H. (2019). World englishes and prosody: Evidence from the successful public speakers. APSIPA ASC, 2048–2052.
Dodane, C. & Hirsch, F. (2018). L'organisation spatiale et temporelle de la pause en parole et en discours. Langages, 211, 5-12.
Tauberer, J. (2008). Predicting intrasentential pauses: is syntactic structure useful? In Speech Prosody 2008, pp. 405–408.

Enlgish L1 vs. L2: Lexical stress



Cutler, A., & Jesse, A. (2021) “Word stress in speech perception.” John Wiley & Sons, Ltd.
Tortel, A. (2021) “Le rythme en anglais oral : considérations théoriques et illustrations sur corpus.” Recherche et pratiques pédagogiques en langues - Cahiers de l’APLIUT.

English L1 vs. L2: Lexical stress

Lexical stress

Oo
PERson perSONify

Ooo oOoo ooOo
PHOtograph phoTOgrapher photoGRAphic

oO ooOo
adMIRE admiRAtion

English, German, Spanish…

Roles of lexical stress

- Word segmentation

it was **COLD**, *and the* **LI***ttle* **FISH TWIS***ted and* **TUM***bled in the* **WA***ter; the* **BIRDS** *were* **ALL QUI***et, and the* **PROUD LI***ons* **ROA***red.*

- Grammatical disambiguation

PER*son* vs. per**SO***nify*

Plain words → stressed
Functional words → unstressed

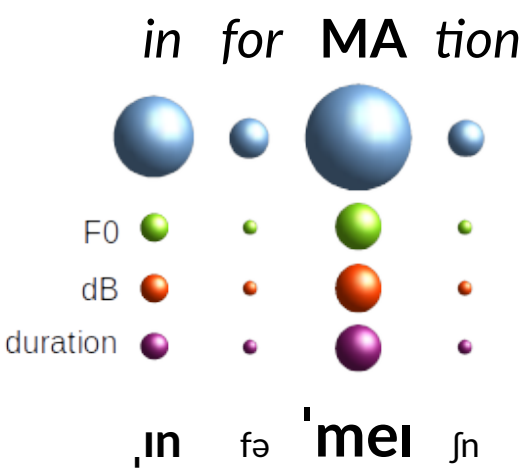
- (Word disambiguation)

DE*sert* vs. de**SSERT**

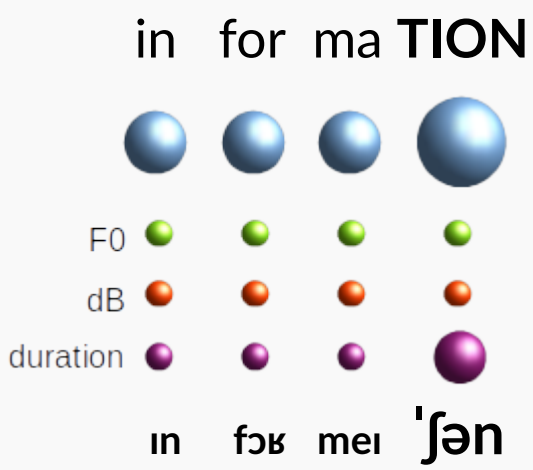
Cutler, A., & Jesse, A. (2021) "Word stress in speech perception." John Wiley & Sons, Ltd.
Cutler, A. (2015) "Lexical stress in english pronunciation." In The handbook of english pronunciation. (pp. 106–124). Hoboken, NJ:John Wiley & Sons, Inc.

English L1 vs. L2: Lexical stress

Expected pronunciation:



Observed pronunciation:



Case of French speakers of English

- Stress shift to the final syllable
- No vowel reduction
- Stress deafness

Tortel, A., & Hirst, D. (2010) "Rhythm metrics and the production of English L1/L2." Speech Prosody 2010, Paper 959.
Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997) "A Destressing Deafness in French?". Journal of Memory and Language, 36: 3, 406-421.
Tortel, A. (2021) "Le rythme en anglais oral : considérations théoriques et illustrations sur corpus." Recherche et pratiques pédagogiques en langues - Cahiers de l'APLIUT.

Pauses and Lexical Stress Processing Pipeline (PLSPP)

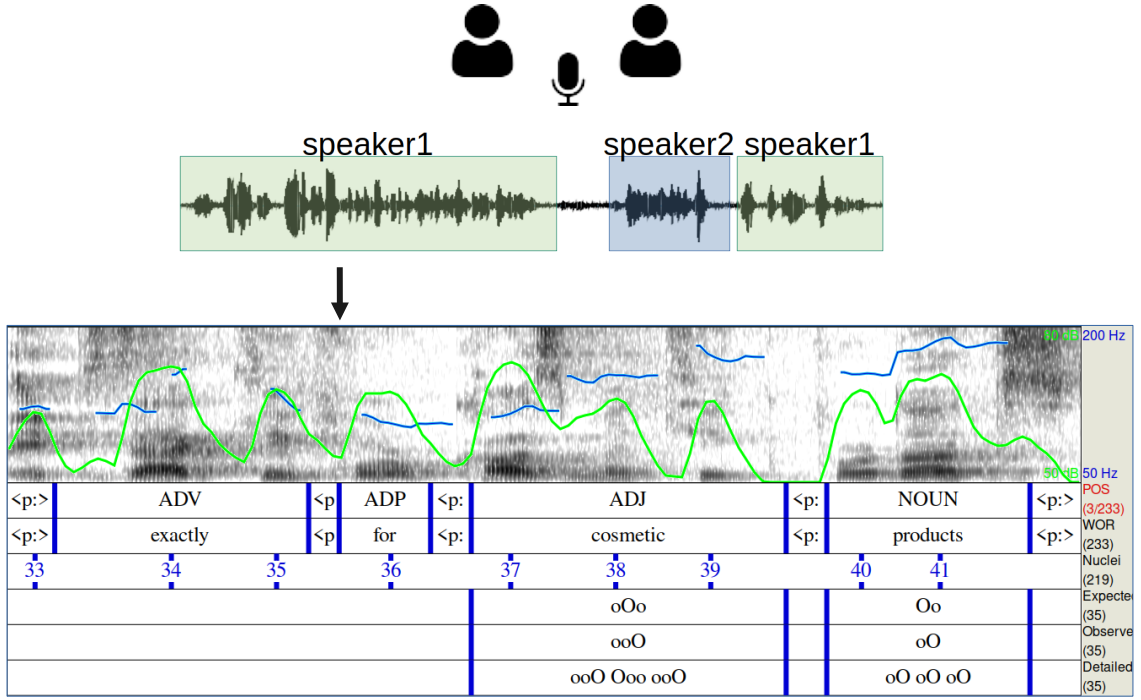
Pauses and Lexical Stress Processing Pipeline (PLSPP)

Pipeline*

- Speech detection and neural speaker diarization (Pyannote)
- ASR & Forced Alignment (WhisperX)
- Morphosyntactic analysis (SpaCy)

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

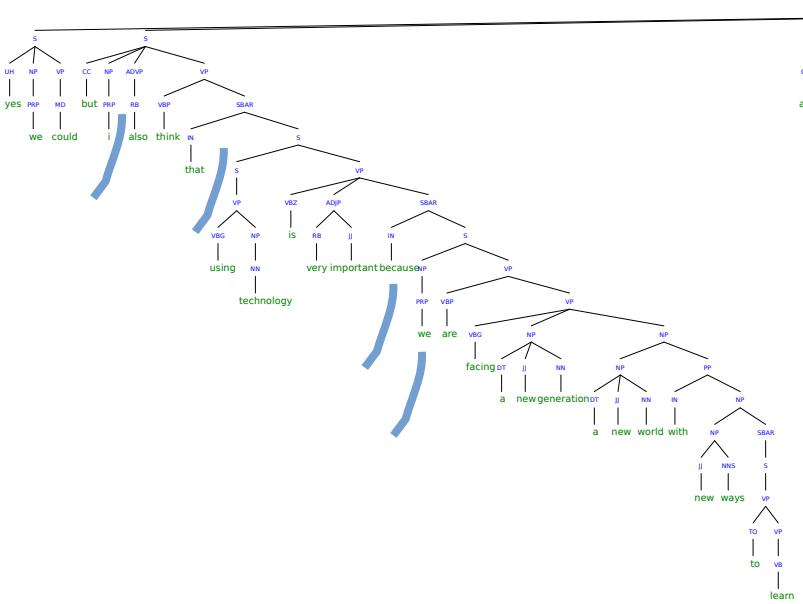
- **Syllable** nuclei detection (De Jong et al., 2021)
- Syllabic **parameter extraction** (intonation, intensity, duration ; speaker normalization)
- **Comparison** of prosodic shape of nouns, verbs, adjectives with a reference dictionary

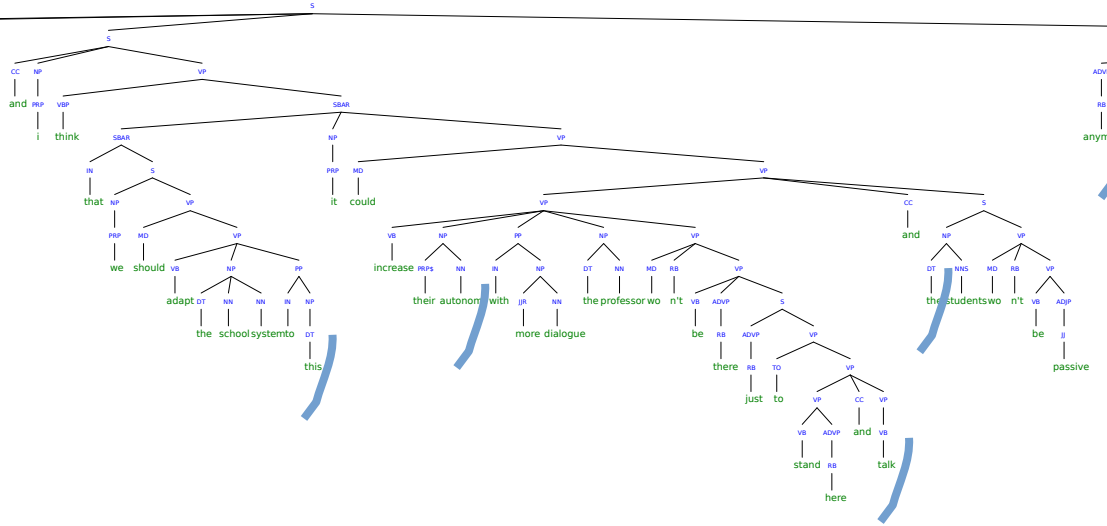


*The full pipeline is available here: https://gricad-gitlab.univ-grenoble-alpes.fr/lidilem/plspp

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.

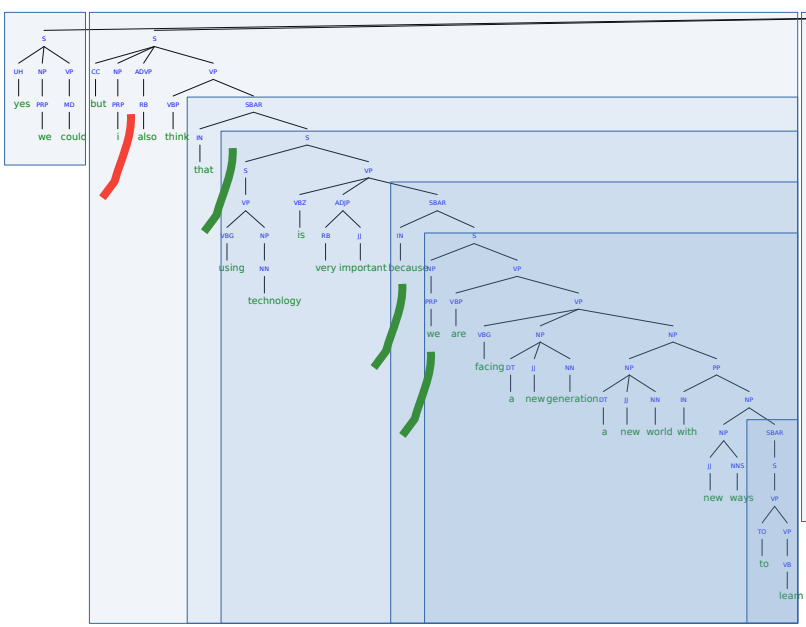
Pauses processing

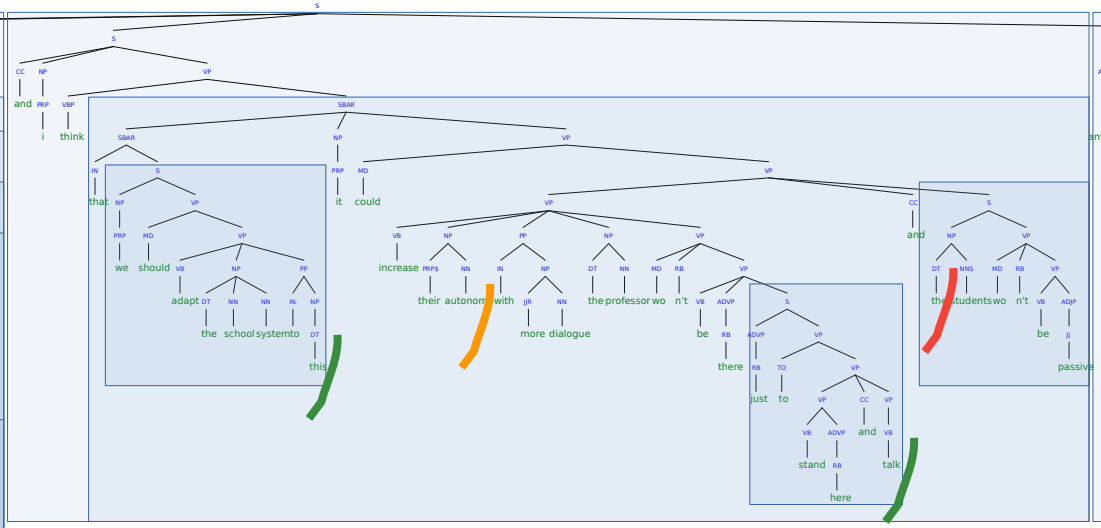




file: dec2022-003_039-040_SPEAKER_01_5
Speaker total speech duration: 6’33’’

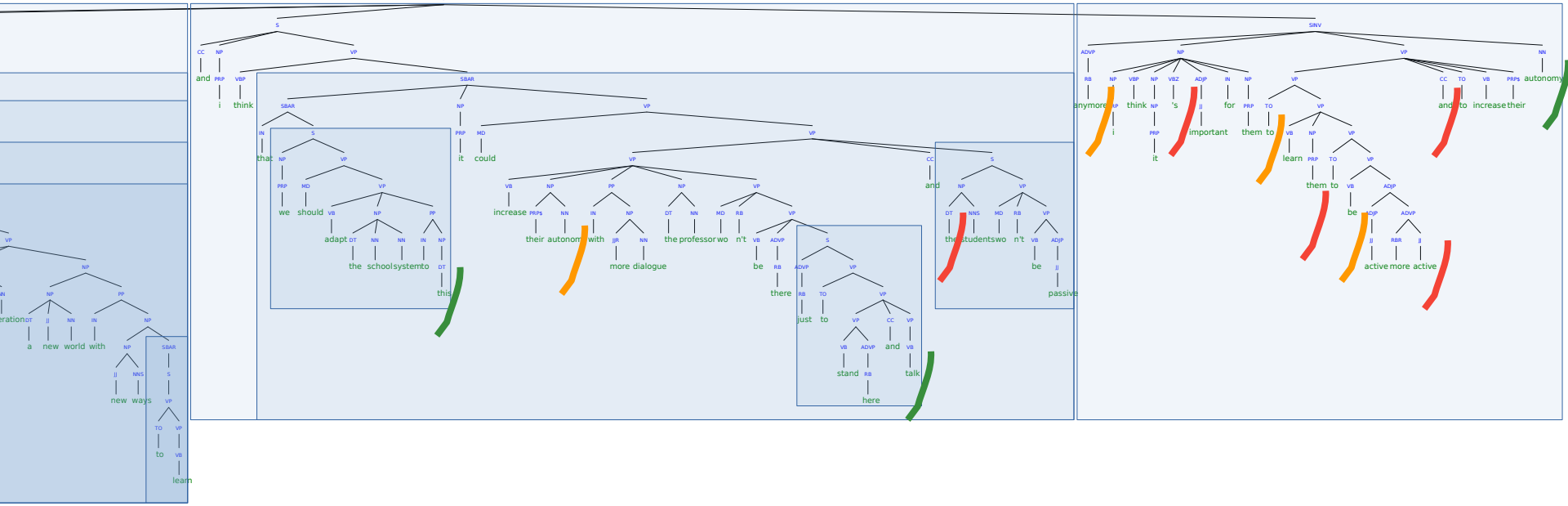
Pauses processing





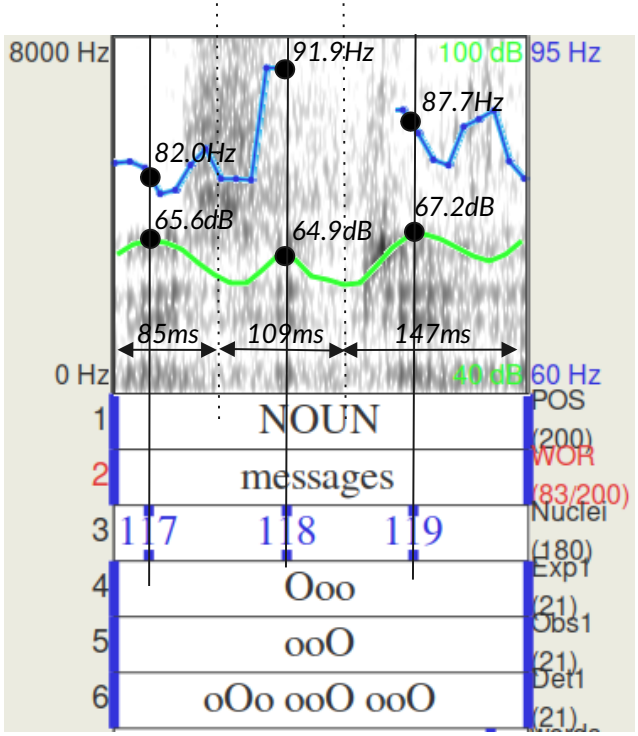
file: dec2022-003_039-040_SPEAKER_01_5
Speaker total speech duration: 6’33’’

Pauses processing

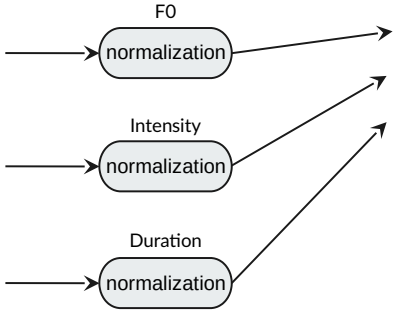


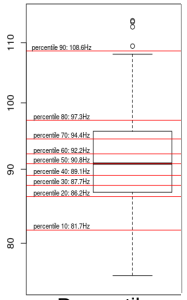
file: dec2022-003_039-040_SPEAKER_01_5
Speaker total speech duration: 6'33''

Stress processing



(F0, Intensity, *Duration*)





Percentiles of the speaker's distribution

	117	118	119
F0:	(11),	(58),	(30)
dB:	(7),	(4),	(19)
dur:	(5),	(13),	(38)

Observed pattern

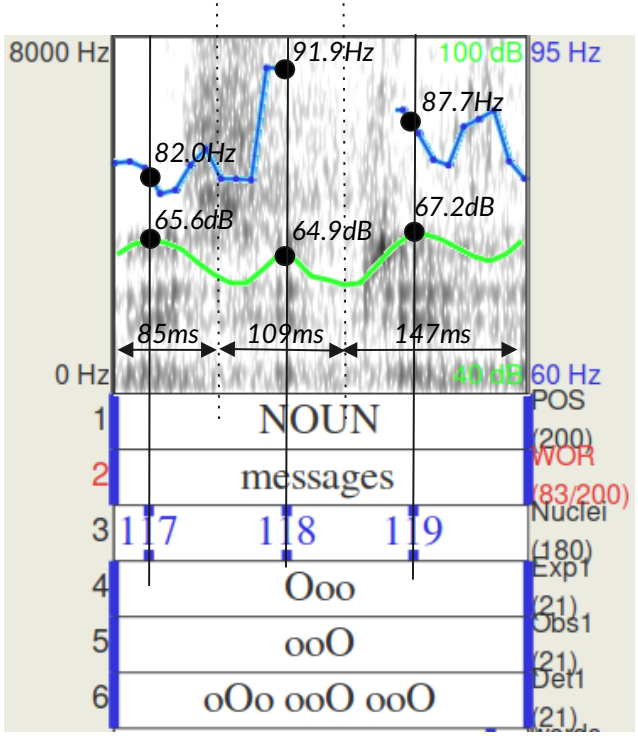
(8), (25), (29)

o o **O**

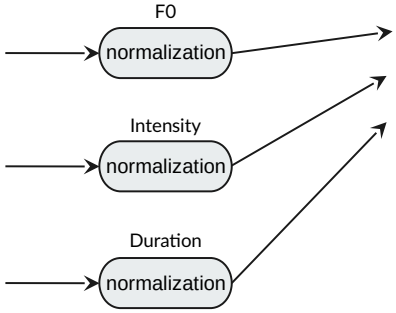
Reference pattern

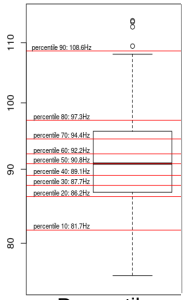
O o o

Stress processing



(F0, Intensity, *Duration*)

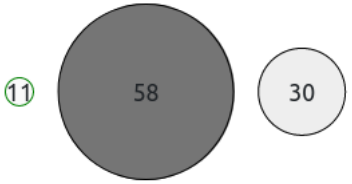




Percentiles of the speaker's distribution

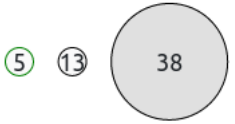
	117	118	119
F0:	(11),	(58),	(30)
dB:	(7),	(4),	(19)
dur:	(5),	(13),	(38)

Mean F0 (for each syllable)



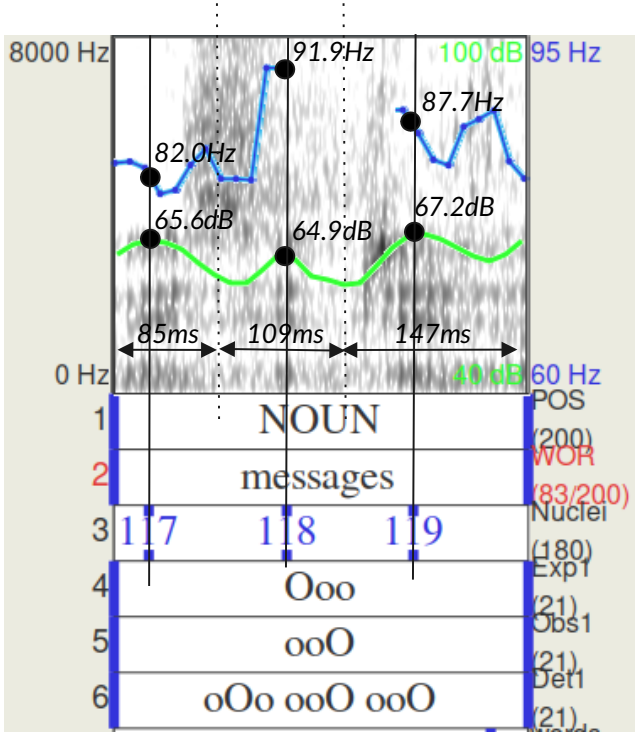
Mean intensity (for each syllable)

Mean duration (for each syllable)

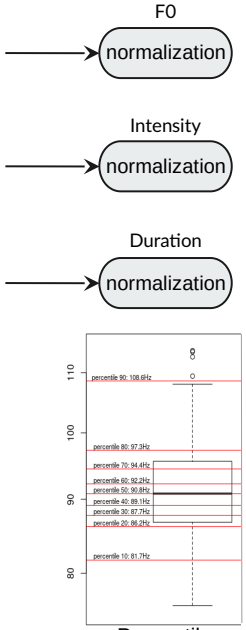


File: jan2020-001_020-022_SPEAKER_00_9 wordID: ixrPu3s

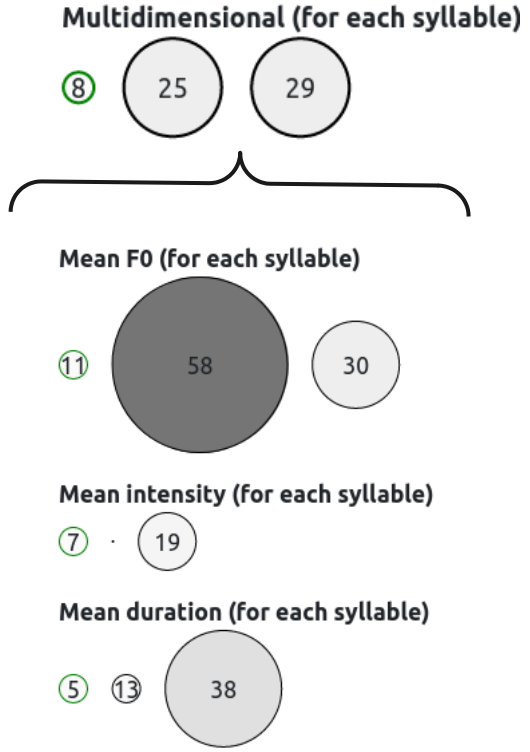
Stress processing



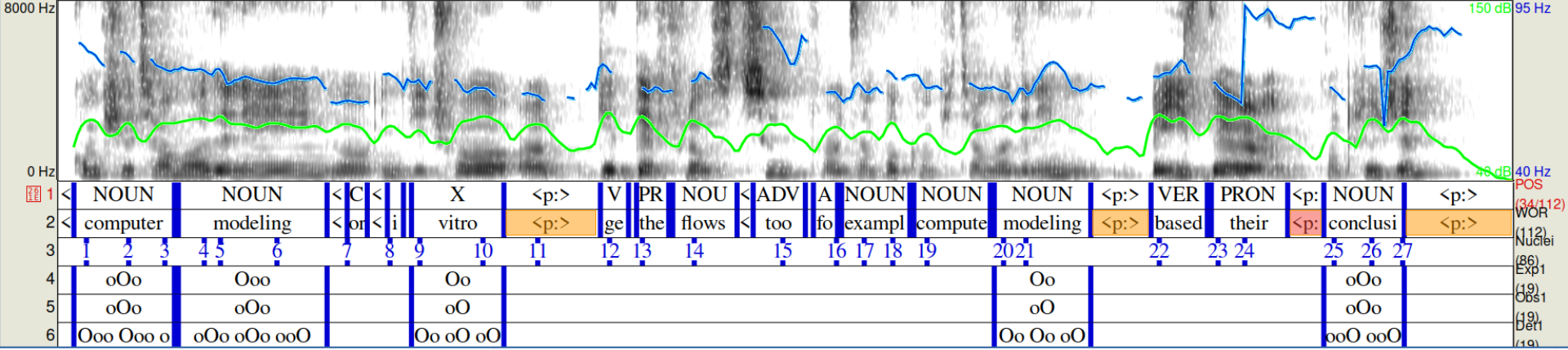
(F0, Intensity, *Duration*)



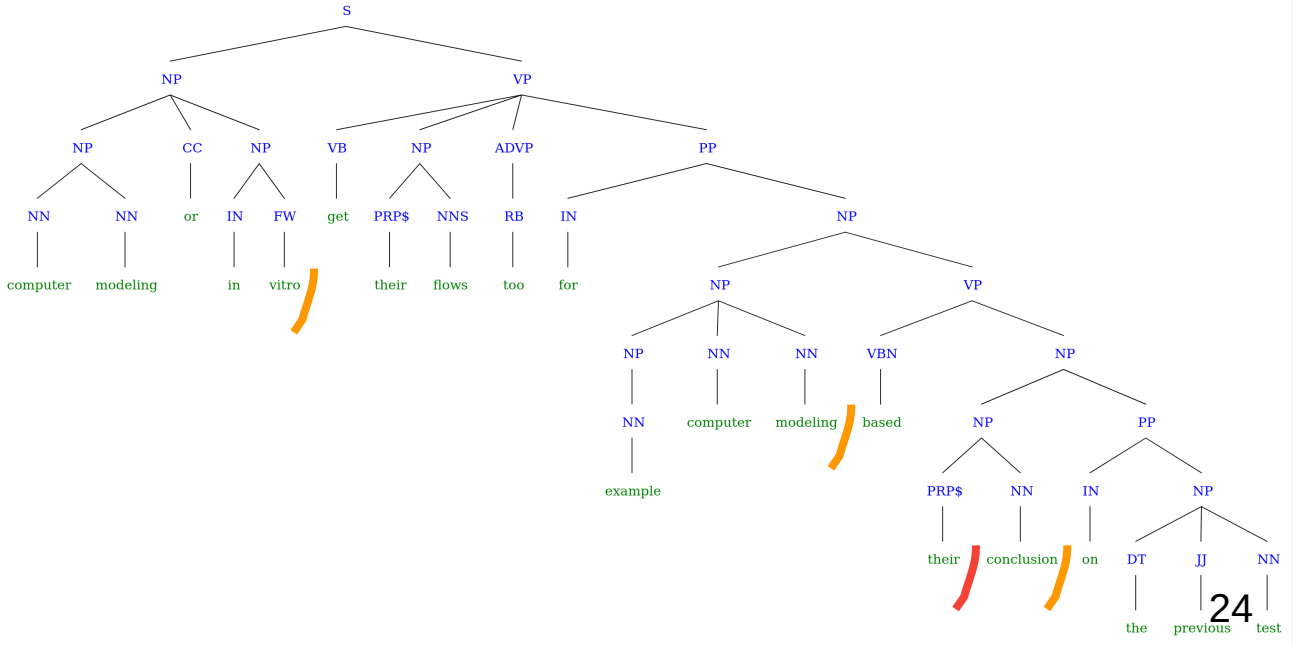
Percentiles of the speaker's distribution

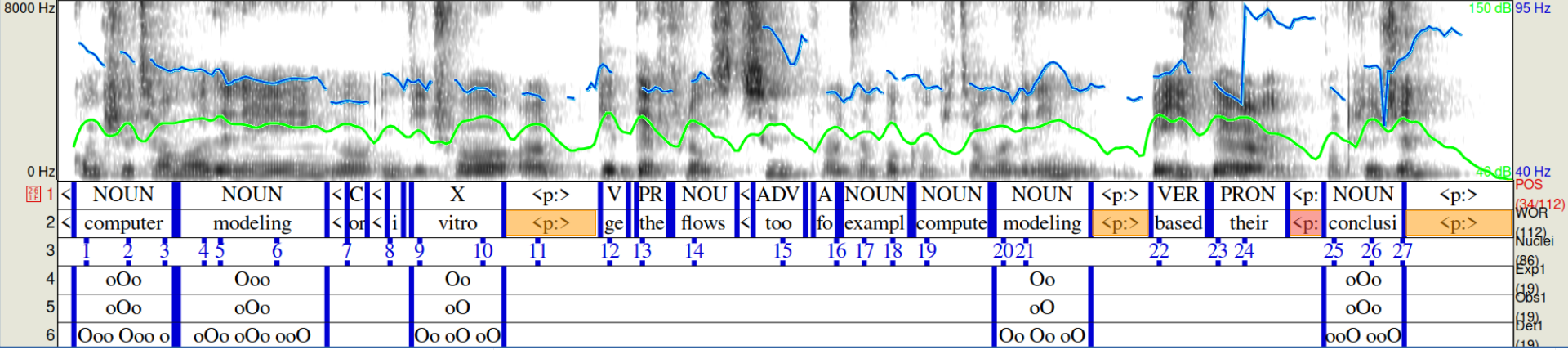


File: jan2020-001_020-022_SPEAKER_00_9 wordID: ixrPu3s

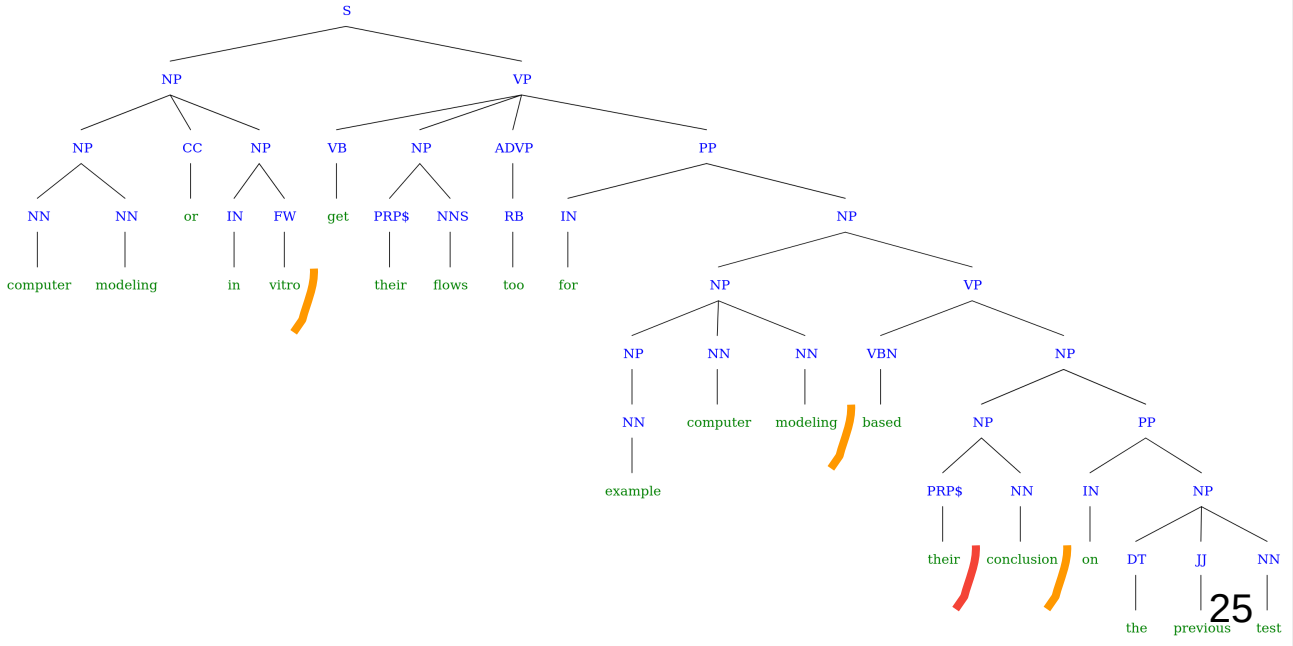


File: jan2023-301_019-086_SPEAKER_01_7





File: jan2023-301_019-086_SPEAKER_01_7



Preliminary studies

- Spont. speech by French univ. students (Coulange, Kato, Masperi, Rossato)
- Recitations by Japanese primary school students (Kimura[Doshisha])
- Read speech by Korean and Japanese univ. students (Sugahara[Doshisha])
- Read speech by French univ. students (Frost[Grenoble])
- Spont. speech by Japanese univ. students (Konishi[Waseda])

Current PhD experiment

Corpus:

- ✔ L2 English spontaneous speech from 176 French learners recorded during CLES certification speaking session.
- ✔ Situation: 2 or 3 candidates discussing a polemical topic (role play) during 10min.

- ➢ 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

Hypothesis:

- **Pauses:**
  - More random pauses with B1
  - More structurant pauses with B2
- **Stress:**
  - Stress position accuracy B2>B1
  - Stress shift to last syllable
  - Stress mainly by duration change
    - F0 and intensity used mainly by high proficiency speakers

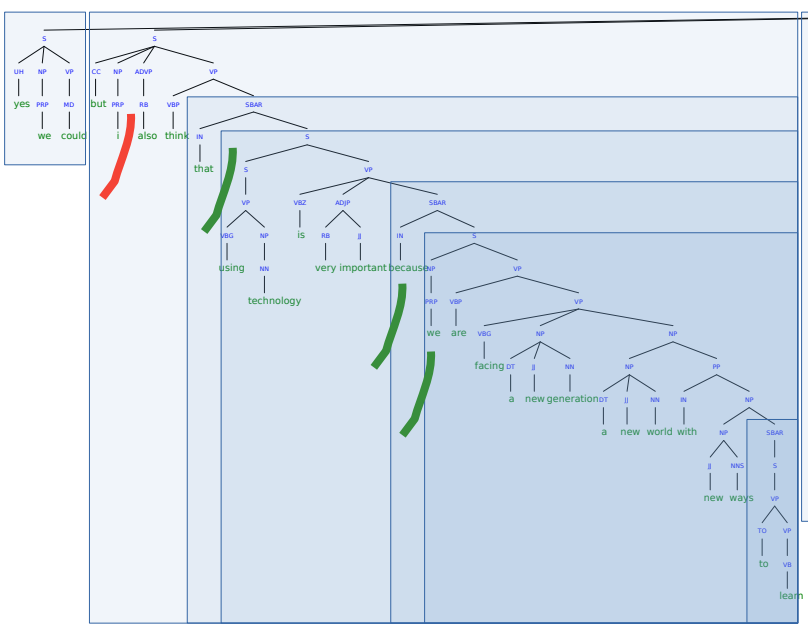
CLES official website: https://www.certification-cles.fr/english/
Raw data is available for research: coordination-nationale@certification-cles.fr
See Coulange, S., Fries, M.-H., Masperi, M., Rossato, R. (submitted). A corpus of spontaneous L2 English speech for real-situation speaking assessment. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 20-25 May, Torino, Italy.

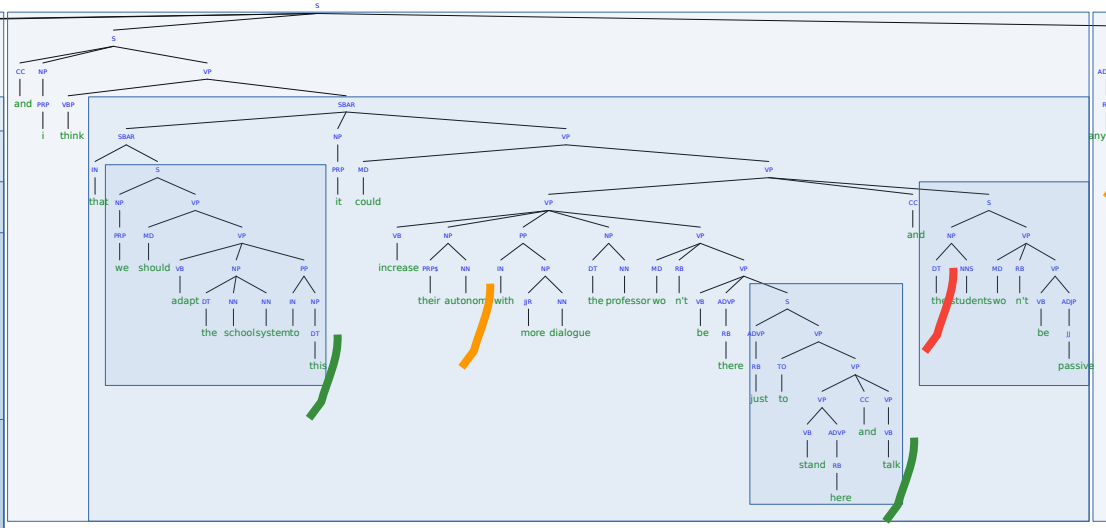
LIDILEM Université Grenoble Alpes | SLPL Spoken Language Processing Laboratory | UGA Université Grenoble Alpes | 同志社大学 Doshisha University

176 B1/B2 French native speakers, **6 350 target words, 21 831 pauses**

spk	lab	pos	lenSyllxpo	expectedShapes	expectedShape	observedShape	globalDeciles	F0shape	dBshape	durshape	sylIF0	F0Deciles	sylldB
dec2022-205_111-055_SPEAKER_01	building	VERB	2	Oo	Oo	Oo	[65, 47.7]	oO	Oo	oO	[164.58407143093706, 166.3141791271218]	[49, 55]	[51.94780216931707, 45.10209764459848]
dec2022-205_111-055_SPEAKER_01	process	NOUN	2	Oo	Oo	Oo	[41.3, 34.7]	Oo	Oo	oO	[158.90480130818477, 145.63261964387854]	[28, 5]	[46.86708137706048, 43.63292251684199]
dec2022-205_111-055_SPEAKER_01	learning	VERB	2	Oo	Oo	oO	[28.3, 47.7]	oO	oO	oO	[159.30461799896116, 160.63498969193438]	[30, 35]	[47.6733118974855O5, 48.96136466667488]
dec2022-205_111-055_SPEAKER_01	technologies	NOUN	4	oOoo	oOoo	Oooo	[37.3, 25, 18, 19]	oOoo	ooOo	Oooo	[160.39467991666788, 161.7943211660824, 153.50491615562396, 134.5222324481085]	[33, 37, 19, 1]	[41.50049965943468, 47.15541760100778, 47.625
dec2022-203_070-157_SPEAKER_02	okay	INTJ	2	oO	oO	Oo	[27.7, 27.3]	oO	Oo	oO	[123.716657042490O5, 127.11303305731948]	[1, 3]	[41.59605312309949, 36.42920139418712]
dec2022-002_038-016_SPEAKER_01	possible	ADJ	3	Ooo	Ooo	ooO	[70.3, 54, 73.3]	ooO	Ooo	ooOo	[145.85607534783227, 149.86689022197467, 150.9323445061917]	[85, 87, 91]	[70.5651611297956, 62.53410099882898, 69.689
dec2022-002_038-016_SPEAKER_01	major	ADJ	2	Oo	Oo	oO	[53.3, 67]	oO	Oo	oO	[135.55139946444592, 136.32959616052827]	[60, 62]	[68.76335981009956, 67.10563600922826]
dec2022-002_038-016_SPEAKER_01	problem	NOUN	2	Oo	Oo	oO	[58.7, 67]	Oo	oO	oO	[139.30615950696262, 128.339410212376]	[75, 37]	[66.8799452573924, 67.59833926326664]
dec2022-002_038-016_SPEAKER_01	technologies	NOUN	4	oOoo	oOoo	oOoo	[27.7, 28.3, 27.3, 27.3]	oOoo	Oooo	oooO	[124.1831595354649, 128.29012582299262, 122.66517082128361, 114.00234256688725]	[18, 36, 13, 1]	[63.9608494230991O4, 61.107467116788887, 63.0
jan2023-401_152-116_SPEAKER_00	thinking	VERB	2	Oo	Oo	Oo	[44.7, 37.7]	Oo	Oo	oO	[226.32342974284O4, 187.64787355642056]	[62, 14]	[66.87374249768531, 63.35600686464649]
jan2023-401_152-116_SPEAKER_00	shouldn't	AUX+PART	2	Oo	Oo	oO	[43.7, 53.7]	Oo	oO	oO	[227.01904156866225, 223.9988124395869]	[63, 58]	[67.71400716033487, 68.54225155615447]
jan2023-401_152-116_SPEAKER_00	subjects	NOUN	2	Oo/oO	Oo	oO	[17.3, 33.7]	Oo	Oo	oO	[201.86306680771358, 181.30901219209392]	[25, 13]	[60.24457093328994, 58.728443325093124]
jan2023-401_152-116_SPEAKER_00	level	NOUN	2	Oo	Oo	oO	[40.7, 45.7]	oO	oO	oO	[213.35631898373876, 213.647852366661O8]	[46, 47]	[71.02170532956757, 71.638767367728155]
jan2023-401_152-116_SPEAKER_00	discomfort	NOUN	3	oOo	oOo	oOo	[14.3, 64.7, 51.7]	oOo	oOo	ooO	[197.8377581196835, 221.7600421333367, 210.01484073713692]	[18, 55, 39]	[61.88381651214592, 67.89710446878011, 64.42
jan2023-401_152-116_SPEAKER_00	prevent	VERB	2	oO	oO	oO	[55.3, 81]	oO	oO	oO	[235.38622493209223, 238.822967463371335]	[70, 77]	[71.71229493988947, 71.77453942234565]
jan2023-401_152-116_SPEAKER_00	testing	VERB	2	Oo	Oo	Oo	[58, 47]	Oo	Oo	oO	[244.60630303893134, 211.30902826830314]	[79, 42]	[69.62557445419516, 66.109906637259344]
jan2023-401_152-116_SPEAKER_00	humans	NOUN	2	Oo	Oo	oO	[72.7, 81]	oO	Oo	oO	[286.9904148807544, 283.733258335888761]	[95, 95]	[75.31753133351741, 70.27701274448296]
jan2023-401_152-116_SPEAKER_00	humans	NOUN	2	Oo	Oo	Oo	[58, 48.3]	Oo	Oo	oO	[253.87162980124415, 209.4704071887013]	[86, 38]	[72.11279795223446, 66.91735871042349]
jan2023-401_152-116_SPEAKER_00	humans	NOUN	2	Oo	Oo	oO	[52, 52.7]	Oo	Oo	oO	[231.4656034332194, 211.81640554077833]	[67, 43]	[68.98237170689472, 67.3687859395835]
jan2023-401_152-116_SPEAKER_00	very	ADV	2	Oo	Oo	oO	[58, 83.7]	oO	oO	oO	[256.02968776053626, 275.28195747500394]	[87, 93]	[73.31656456249104, 73.54565499093779]
jan2023-401_152-116_SPEAKER_00	simple	ADJ	2	Oo	Oo	oO	[71, 79.7]	Oo	oO	oO	[270.4394536487946, 247.5986814936122]	[91, 81]	[70.70305774215379, 72.0054758925618]
jan2023-401_152-116_SPEAKER_00	member	NOUN	2	Oo	Oo	Oo	[55.3, 46]	Oo	Oo	oO	[239.081612634621222, 201.15559526076024]	[78, 24]	[69.1799784961076, 67.78274126077542]
jan2023-401_152-116_SPEAKER_00	medicine	NOUN	3	Ooo	Ooo	ooO	[68.3, 65.3, 68.7]	oOo	oOo	ooO	[302.110947898042, 322.24241055289775, 306.31265106415367]	[96, 98, 97]	[73.56787767185958, 74.27410946995222, 72.485
jan2023-401_152-116_SPEAKER_00	tested	VERB	2	Oo	Oo	Oo	[59.7, 43.7]	Oo	Oo	oO	[237.43401901757832, 205.0988374351786]	[74, 33]	[68.92683224905485, 67.56983437057023]
jan2023-401_152-116_SPEAKER_00	doesn't	AUX+PART	2	Oo	Oo	oO	[49.7, 82]	oO	oO	oO	[212.75193172309997, 236.51819023353454]	[45, 72]	[64.20804160009475, 71.74739893964298]
jan2023-401_152-116_SPEAKER_00	scientists	NOUN	2	Ooo/Oo	Oo	oO	[51.7, 62.3]	oO	Oo	oO	[236.02498820280482, 259.5758429815999]	[71, 89]	[64.97782978558745, 60.492400212704624]
jan2023-401_152-116_SPEAKER_00	alzheimer's	NOUN+PART	3	Ooo	Ooo	oOo	[30.3, 39.3, 21.7]	oOo	Ooo	oOo	[208.24914162596494, 214.04976347341497, 198.284607583205631]	[36, 49, 20]	[62.6966749647739, 62.60810262117271, 62.094
jan2023-401_152-116_SPEAKER_00	doesn't	AUX+PART	2	Oo	Oo	oO	[56.3, 70.7]	oO	Oo	oO	[252.922477112926226, 267.8225829488548]	[84, 90]	[68.07329184147216, 67.04841834212498]
jan2023-401_152-116_SPEAKER_00	cases	NOUN	2	Oo	Oo	oO	[9.7, 11]	OO	oO	Oo	[195.50575989510463, 193.1324378920989]	[16, 16]	[61.07065018323635, 63.74257106916825]
jan2023-201_141-155_SPEAKER_00	methods	NOUN	2	Oo	Oo	oO	[49.7, 67.7]	oO	oO	oO	[190.77830943505793, 192.81136294312833]	[38, 44]	[52.4664001800547O6, 53.53115271697205]
dec2022-206_082-097_SPEAKER_01	final	ADJ	2	Oo	Oo	oO	[62, 59.7]	Oo	oO	oO	[133.74845462266765, 114.93322089045768]	[79, 26]	[62.97939843552732, 60.03190079558409]
dec2022-206_082-097_SPEAKER_01	because	SCONJ	2	oO/oo	oo	oO	[33.3, 47.3]	Oo	oO	oO	[125.10019431942614, 117.80222186761557]	[53, 32]	[52.83851934108151, 53.35584530435036]
dec2022-206_082-097_SPEAKER_01	internet	NOUN	3	Ooo	Ooo	ooO	[56.7, 70.7, 26.7]	oOo	Ooo	oOo	[125.53323889509299, 141.90687281351978, 109.89092038214284]	[54, 92, 14]	[52.34538219389984, 56.627361354321025, 51.0
dec2022-206_082-097_SPEAKER_01	difference	NOUN	2	Ooo/Oo	Oo	oO	[53.3, 81]	oO	oO	oO	[138.8312532132902, 148.22613308574103]	[84, 94]	[54.68585262256626, 57.63074011898019]
dec2022-206_082-097_SPEAKER_01	different	ADJ	2	Ooo/Oo	Oo	oO	[15, 63.3]	oO	oO	oO	[107.49499377891266, 124.527859782459]	[7, 51]	[45.97001257230532, 53.49484644888938]
jan2023-202_140-137-113_SPEAKER_0	facing	VERB	2	Oo	Oo	oO	[35, 47]	oO	oO	Oo	[139.38107415276, 153.7235707791737]	[13, 37]	[45.74395187046916, 50.85383722413109]
jan2020-000_028-031_SPEAKER_01	about	ADP	2	oO	oO	Oo	[62, 44.7]	Oo	oO	oO	[120.51122029501467, 119.16298050648578]	[24, 13]	[54.16485650612178, 47.86968071256431]
jan2020-000_028-031_SPEAKER_01	insist	VERB	2	oO	oO	Oo	[71, 23]	Oo	Oo	Oo	[130.98623458623004, 126.75839062614749]	[78, 58]	[50.048400302497285, 35.047540246980691]
jan2020-000_028-031_SPEAKER_01	because	SCONJ	2	oO/oo	oo	oO	[49, 58.3]	Oo	oO	oO	[162.34241629719682, 124.69960359111944]	[97, 41]	[49.186920712858826, 50.406450673919481]
jan2020-000_028-031_SPEAKER_01	people	NOUN	2	Oo	Oo	oO	[29, 49.3]	Oo	oO	oO	[123.18563949537102, 122.85725367090205]	[33, 31]	[48.44224368768189, 50.471925424154136]
jan2020-000_028-031_SPEAKER_01	working	VERB	2	Oo	Oo	oO	[13.3, 42.3]	oO	oO	oO	[119.19173058095213, 126.25722766008792]	[14, 52]	[44.51063670842647, 46.13339647165502]
jan2020-000_028-031_SPEAKER_01	smoking	NOUN	2	Oo	Oo	oO	[33, 34.3]	oO	Oo	Oo	[117.05744710151721, 132.71556203853203]	[6, 79]	[47.6865615740066, 43.72081574678038]
dec2022-204_121-079_SPEAKER_00	teacher	NOUN	2	Oo	Oo	oO	[61, 85]	oO	oO	oO	[137.1688210191352, 156.4533918271153]	[65, 93]	[62.42878091960817, 63.632369360705795]
dec2022-204_121-079_SPEAKER_00	provide	VERB	2	oO	oO	oO	[32.7, 57.3]	oO	oO	oO	[126.21170477487756, 126.90693658176173]	[27, 29]	[60.532886837761895, 62.920288507166646]
dec2022-204_121-079_SPEAKER_01	student	NOUN	2	Oo	Oo	oO	[60.7, 66]	oO	Oo	oO	[122.91761324931839, 124.691962652415]	[83, 85]	[54.03715448952302, 57.391774942074671]
dec2022-204_121-079_SPEAKER_01	themselves	PRON	2	oO	oO	oO	[53, 82.7]	oO	Oo	oO	[111.935887637079O9, 133.32731167243986]	[50, 92]	[61.249467495037905, 60.547372094696286]
dec2022-204_121-079_SPEAKER_01	progress	NOUN	2	Oo/oO	Oo	oO	[56.3, 76.3]	Oo	oO	oO	[112.92601357780111, 112.47676917286252]	[55, 51]	[62.242958622338858, 64.1212876351483]
dec2022-204_121-079_SPEAKER_01	rhythm	NOUN	2	Oo	Oo	oO	[55.3, 58]	oO	Oo	oO	[119.28430489413957, 138.75571834390655]	[78, 94]	[61.75995800591919, 60.32981171526333]
dec2022-204_121-079_SPEAKER_01	classes	NOUN	2	Oo	Oo	oO	[74.7, 49.3]	Oo	Oo	Oo	[119.47296048649666, 114.36540733683364]	[79, 59]	[61.14108898356506, 56.95720781181572]
dec2022-204_121-079_SPEAKER_01	progressed	VERB	2	oO	oO	oO	[53, 75.3]	Oo	oO	oO	[115.88393537325244, 112.83115093147421]	[66, 54]	[61.742195785392, 64.028190975235555]
dec2022-204_121-079_SPEAKER_01	themselves	PRON	2	oO	oO	oO	[43, 73.3]	oO	oO	oO	[109.25184454099674, 126.39493920494806]	[38, 86]	[60.6758743608343, 60.20195419327548]
dec2022-205_057-094_SPEAKER_01	because	SCONJ	2	oO/oo	oo	oO	[34, 70.7]	Oo	oO	oO	[137.3939483987555, 129.83942813747552]	[79, 57]	[58.962704315218005, 70.59421372742151]
dec2022-205_057-094_SPEAKER_01	single	ADJ	2	Oo	Oo	oO	[47.3, 83.3]	oO	oO	oO	[126.33009213655106, 129.38107898076223]	[41, 56]	[59.113767323235294, 71.05167186252965]
dec2022-205_057-094_SPEAKER_01	technology	NOUN	4	oOoo	oOoo	oooO	[27, 45.7, 38.3, 53.7]	oooO	oooO	ooOo	[114.58121670199516, 127.28078181543432, 115.71195813275646, 138.41372335806776]	[10, 47, 11, 82]	[61.86211879239944, 63.68770235203321, 63.89
dec2022-205_057-094_SPEAKER_01	single	ADJ	2	Oo	Oo	oO	[80, 36.3]	Oo	Oo	Oo	[136.67534906947103, 127.27863776192741]	[76, 46]	[67.76862052232468, 61.48747118485634]
dec2022-205_057-094_SPEAKER_01	using	VERB	2	Oo	Oo	oO	[52.3, 72.3]	oO	oO	oO	[134.17936777926118, 134.7702097085192]	[69, 73]	[68.51272377541504, 71.69970[illegible]903]
dec2022-205_057-094_SPEAKER_01	paper	NOUN	2	Oo	Oo	oO	[64.3, 71]	oO	oO	oO	[148.92604924465337, 142.39063640502772]	[91, 88]	[65.61828657746719, 72.49604[illegible]697]
dec2022-205_057-094_SPEAKER_01	because	SCONJ	2	oO/oo	oo	oO	[8.3, 34.7]	oO	oO	oO	[110.26311746713587, 112.47954881161151]	[4, 8]	[61.723362434325835, 63.42290907331175]
dec2022-205_057-094_SPEAKER_01	computers	NOUN	3	oOo	oOo	ooO	[32, 25.3, 78]	ooO	ooO	ooO	[120.85354276580024, 117.33182359344595, 132.0618641942411]	[27, 16, 66]	[65.81891805457329, 62.25691191922651, 67.94

Where do students pause?





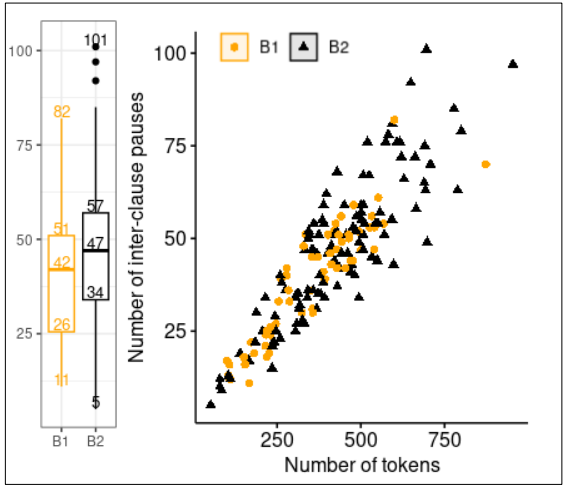
Focus on:

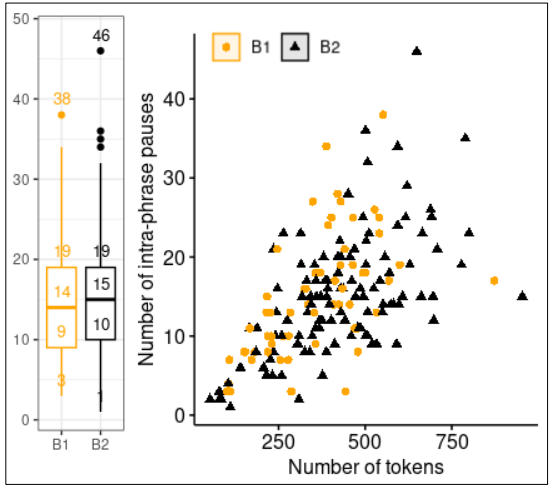
- Inter-clause pauses
- Intra-phrase pauses

Where do students pause?

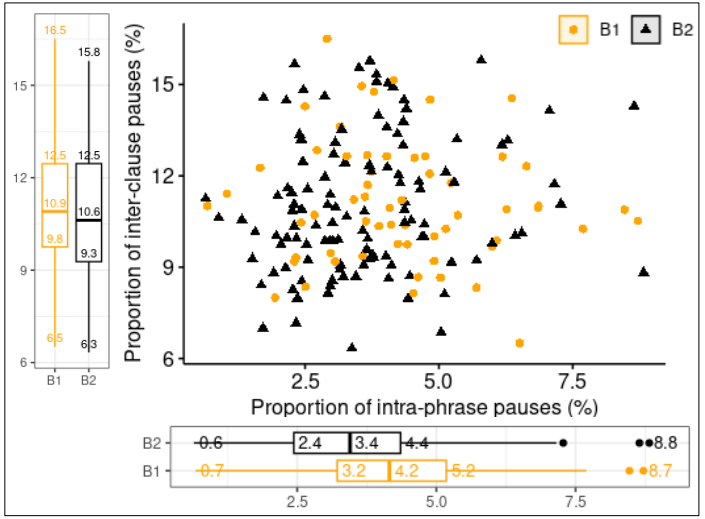
Results from the structural analysis:

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





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

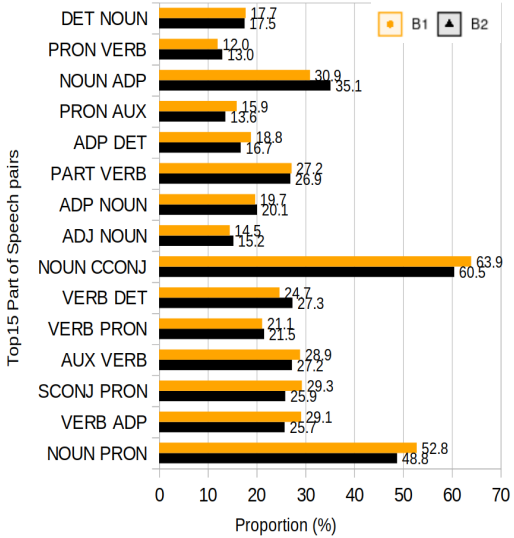


Proportion of inter-clause and intra-phrase pauses per speaker *(nb pauses / nb tokens)*

Where do students pause?

Results from the lexical analysis:

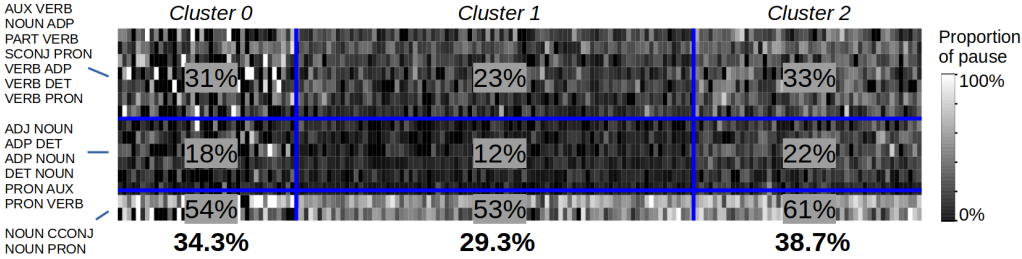
- Pauses in same proportions for B1 and B2 for top15 most frequent POS contexts;
- B2 speakers make generally less pauses in these contexts.



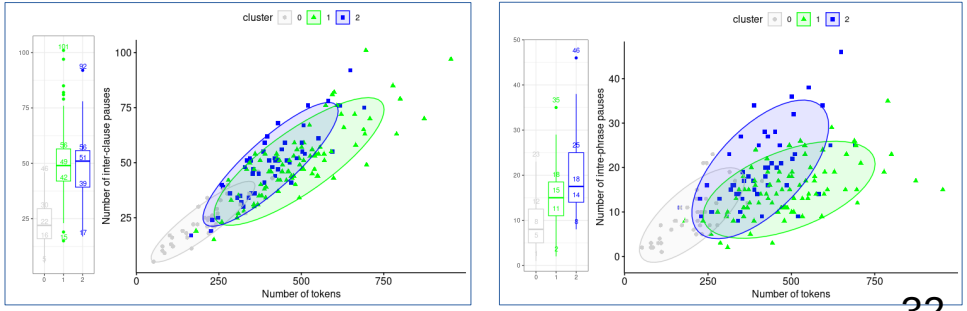
Proportion of pause for the 15 most frequent part-of-speech contexts

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).



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



Absolute number of inter-clause (left) and intra-phrase (right) pauses per speaker

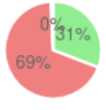
Where do students pause?

Discussion:

- **Limited contrast** between B1 and B2 speakers;
- Instead, **large inter-speaker diversity** in pausing pattern, **especially within phrases**;
- Need for investigating **intra-speaker variability**;
- Need for investigating the relationship between pause position and **comprehensibility.**

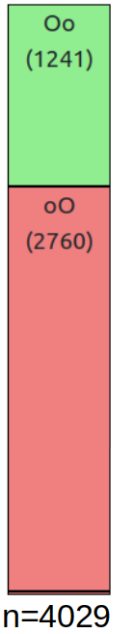
Stress position

Expected pattern →



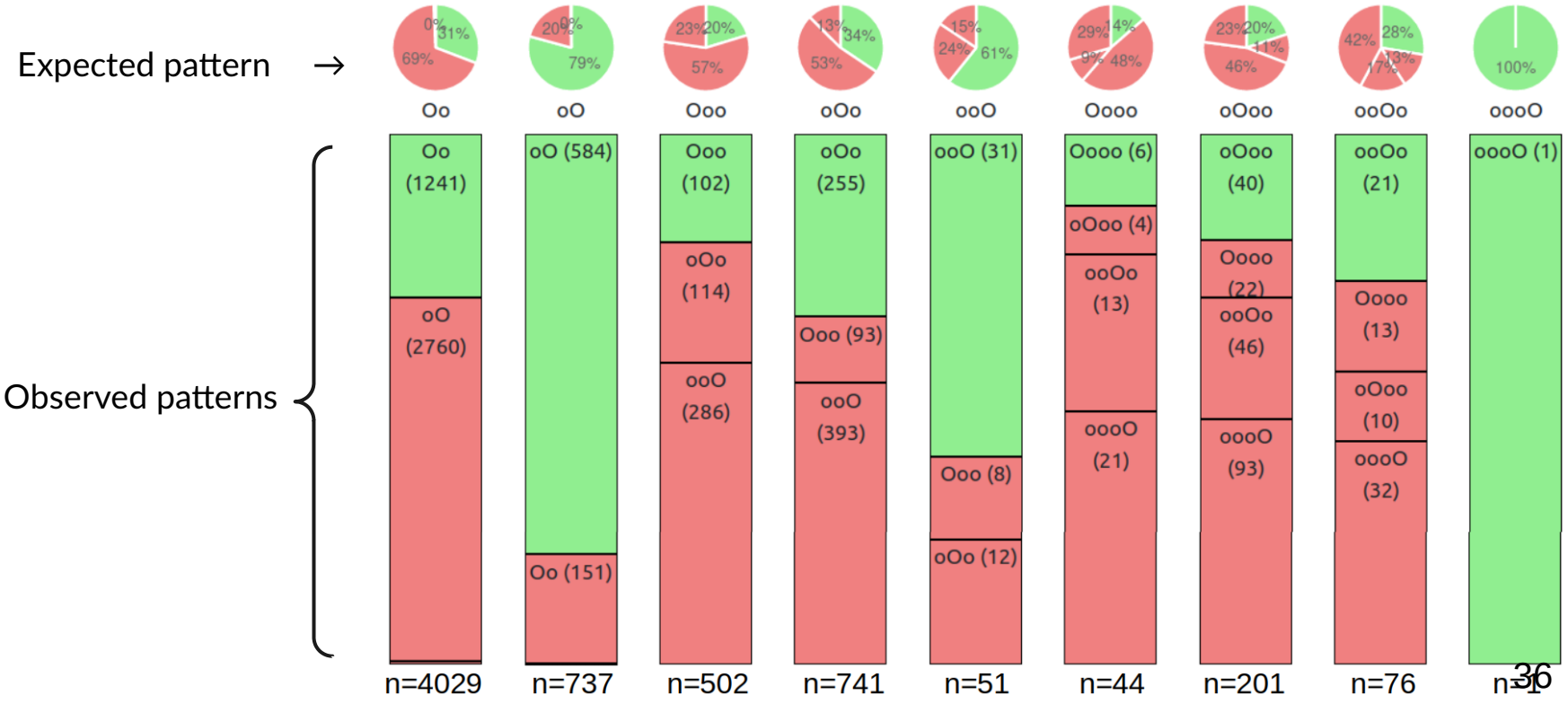
Oo

Observed patterns

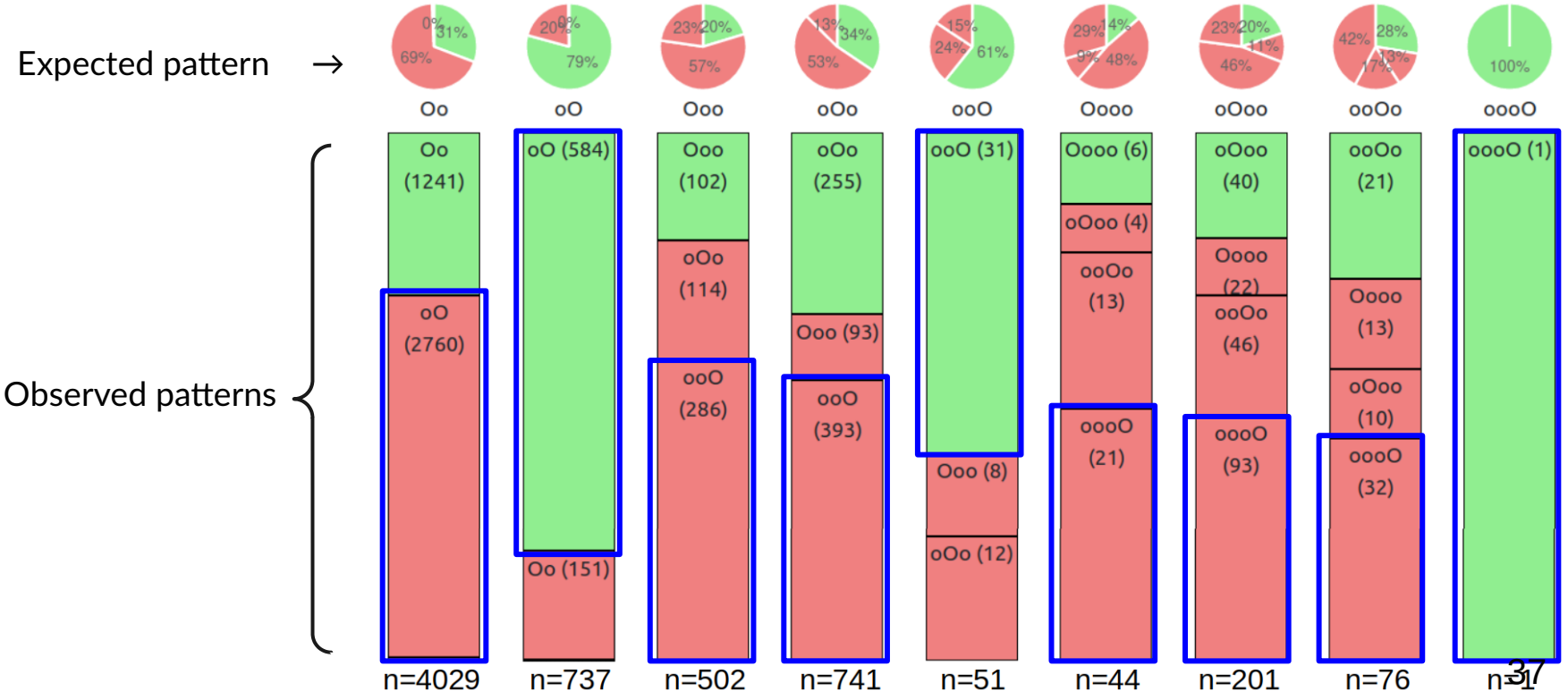


n=4029

Stress position

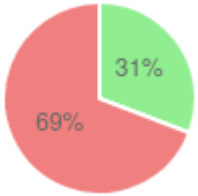


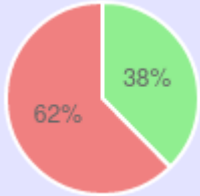
Stress position



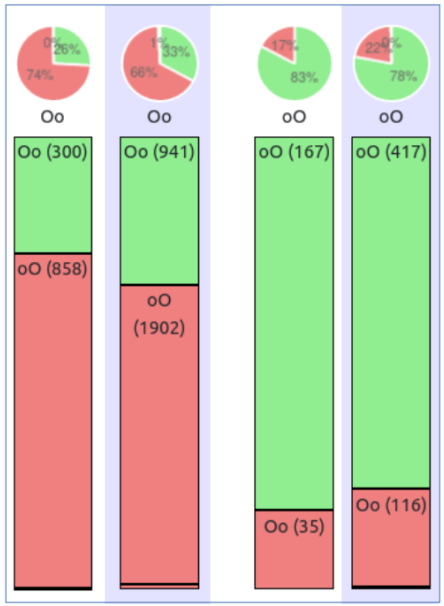
Stress position

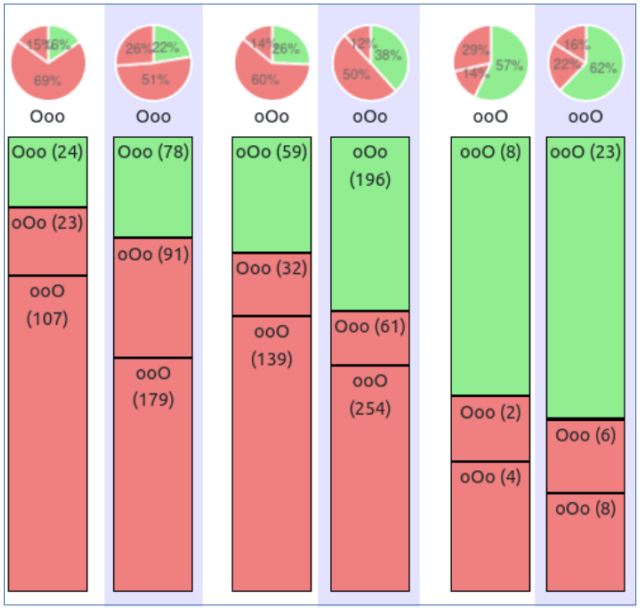
B1 speakers
spk=59
words=1873

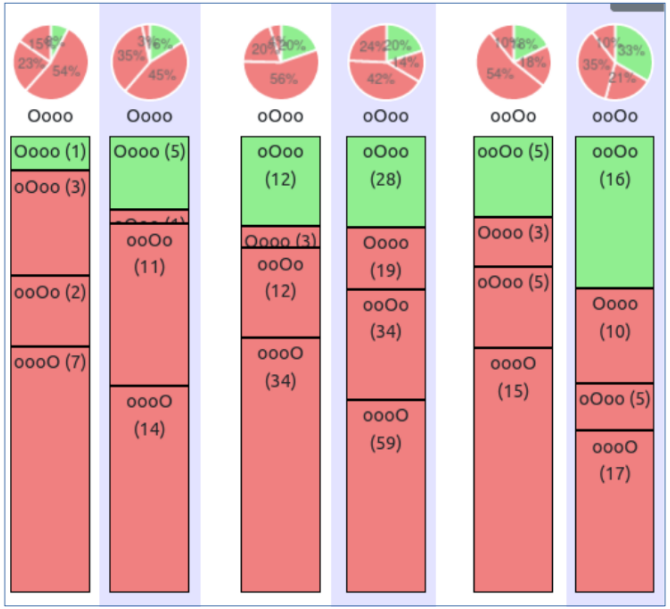




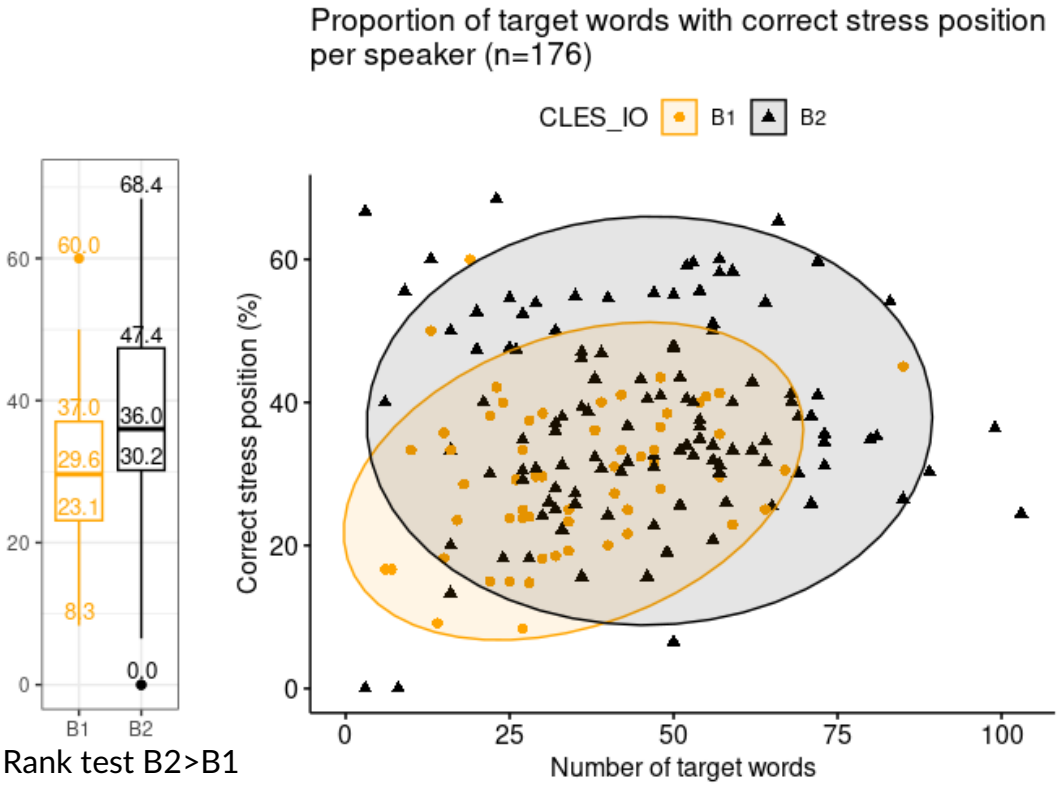
B2 speakers
spk=117
words=4551







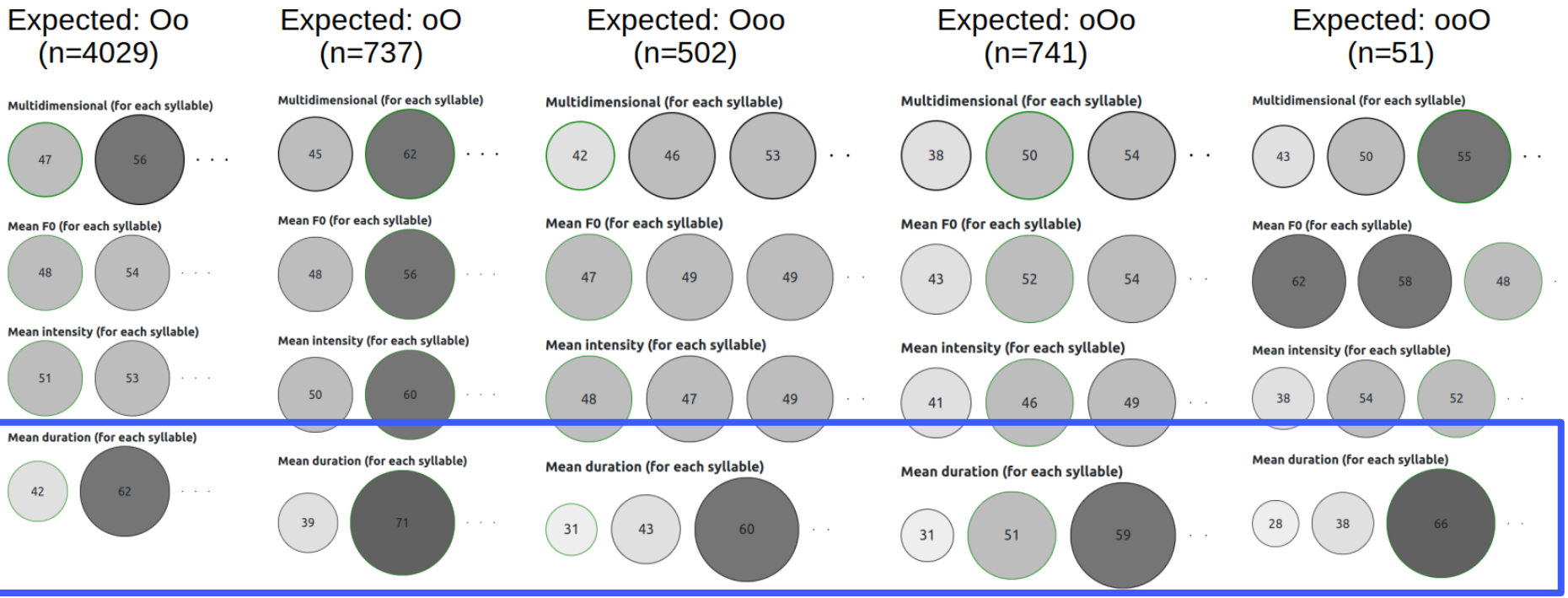
Stress position accuracy per speaker



Rank test B2>B1
p<0.0001

- *Mean stress position accuracy: 35.4 %*
- *Stress accuracy per speaker: 0 % ～ 68.4 %*
- *Stress accuracy per CEFR level: B1 = 29.6 % B2 = 36 % (+ 6.4, p<.0001)*

Stress quality: dimension



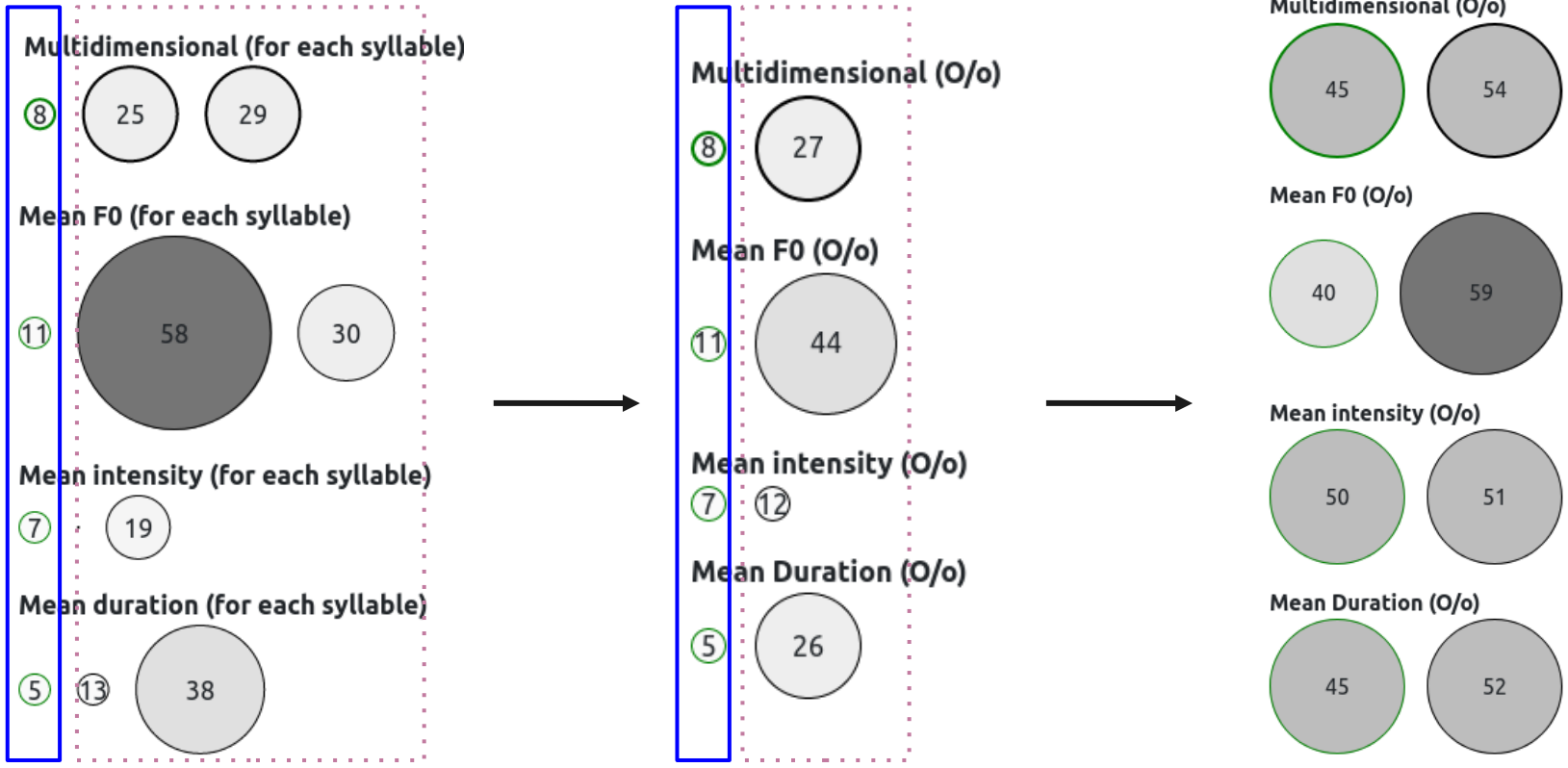
全ての話者（176 人）

Stress quality: dimension

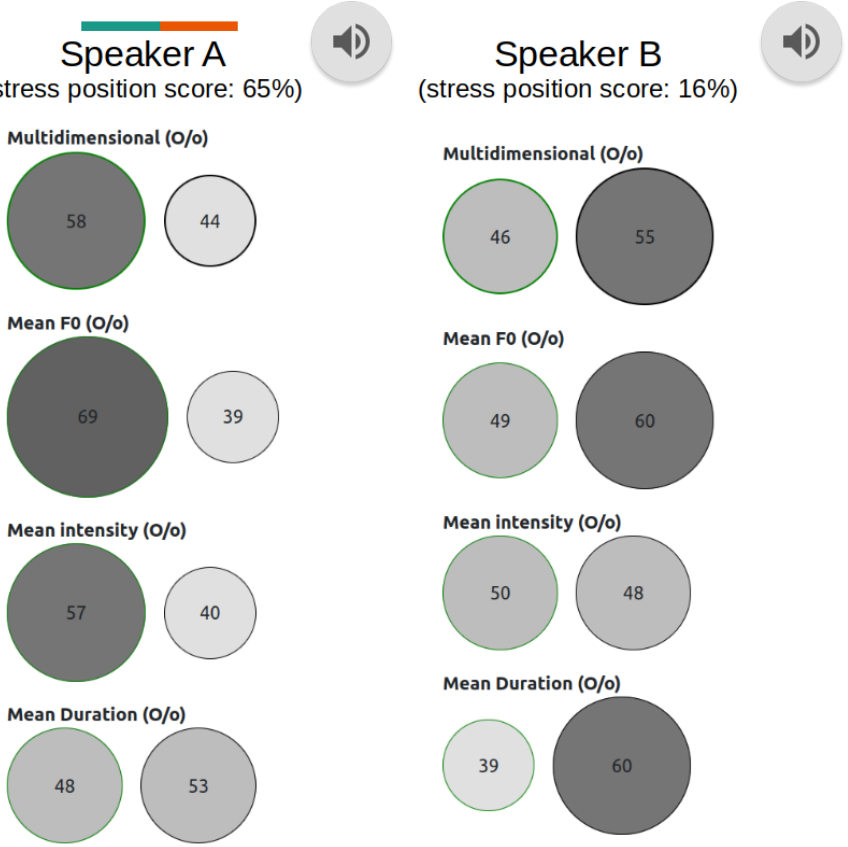
Speaker jan2020-001_020-022_SPEAKER_00

- 42 target words
- Stress position accuracy: 19%
- Mean prosodic contrast: -9 points

Expected Ooo



Stress quality: contrast



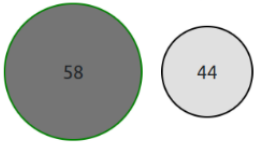
SpeakerA

SpeakerB

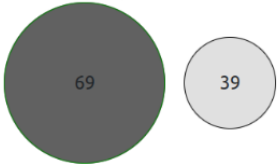
Stress quality: contrast

Speaker A
(stress position score: 65%)

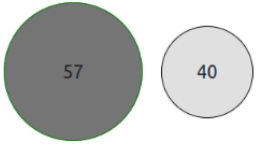
Multidimensional (O/o)



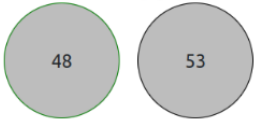
Mean F0 (O/o)



Mean intensity (O/o)

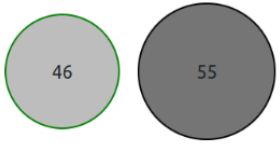


Mean Duration (O/o)

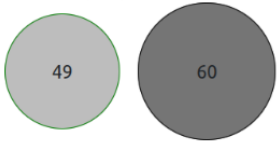


Speaker B
(stress position score: 16%)

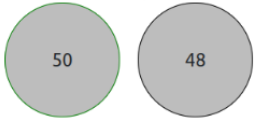
Multidimensional (O/o)



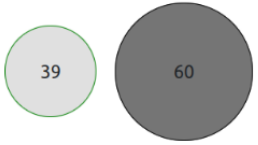
Mean F0 (O/o)

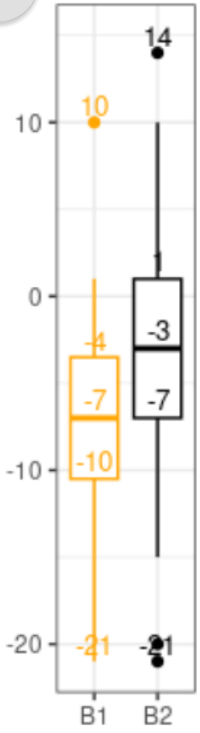


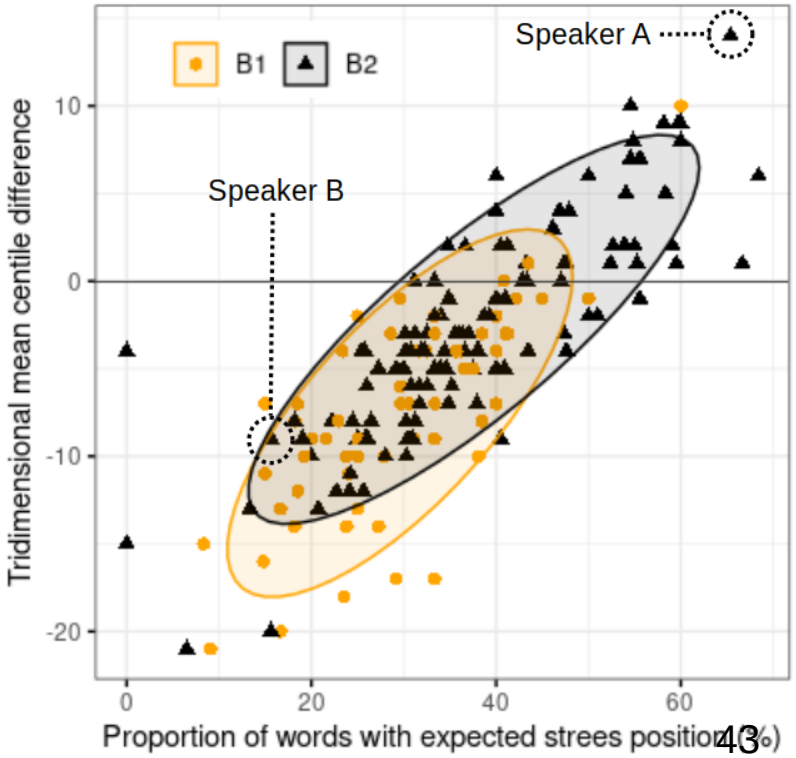
Mean intensity (O/o)



Mean Duration (O/o)

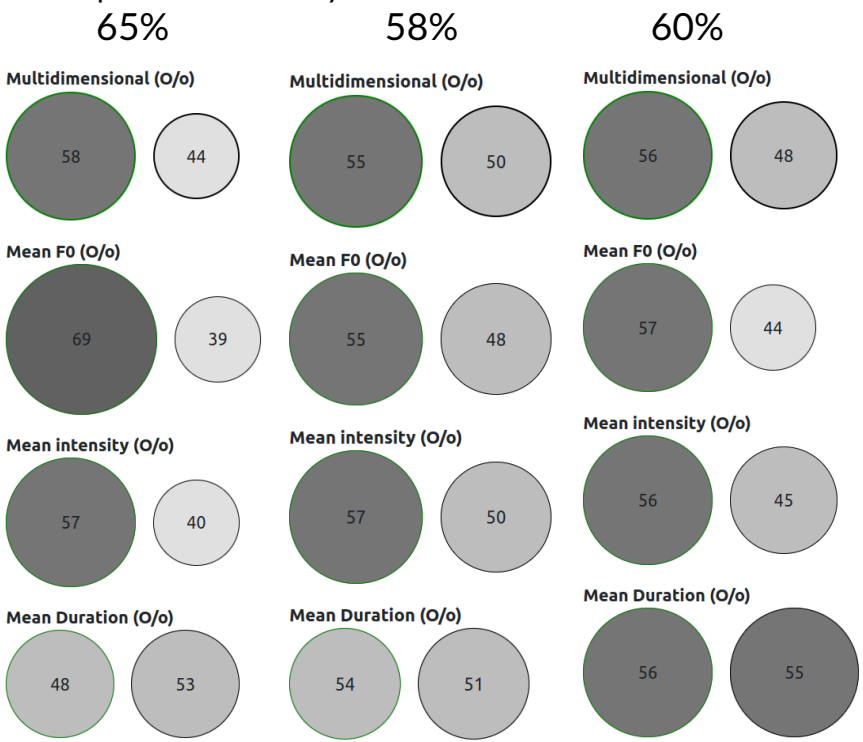


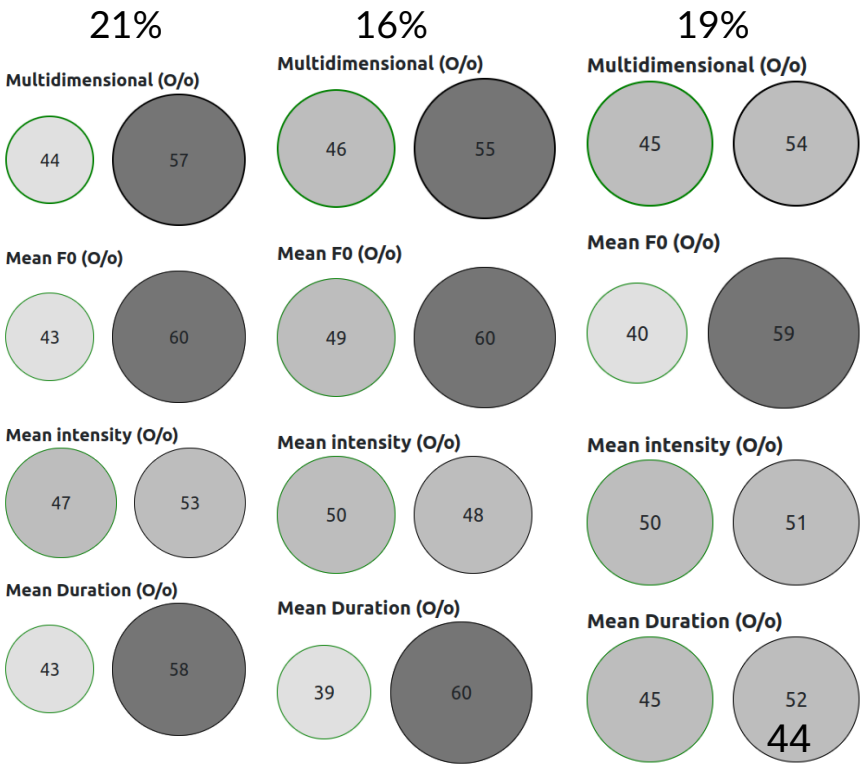




Stress quality: contrast

Stress position accuracy:





Discussion

Wrap up

- Creation of the Pauses and Lexical Stress Processing Pipeline
- Analysis of B1 and B2 speaking level French-L1 university students
  11 hours of speech 6350 target words 21 831 pauses

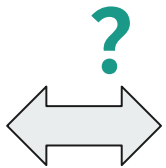
➢ Pause position:
- Great variation of number of intra-phrasal pauses, less with inter-clausal pauses
- B2 speakers make less intra-phrasal pauses than B1 speakers
- Difference between B1 and B2 is small
- High intra-speaker variability

➢ Lexical stress position:
- *Mean stress position accuracy: 35.4 %*
- *Stress accuracy per speaker: 0 % ~ 68.4 %*
- *Stress accuracy per CEFR level:*
  B1 = 29.6 % B2 = 36 % (+ 6.4, p<.0001)
- *Frequent stress shift to the last syllable*

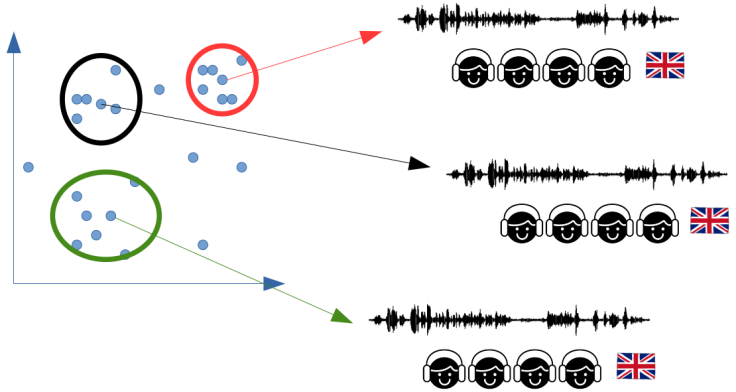
➢ Lexical stress quality:
- *Low accuracy speakers: lengthening of the last syllable*
  tendency to make it higher
  No change in intensity
- *High accuracy speakers: the expected syllable is higher in F0 and intensity*
  No change in duration

Next step

- High number of intra-phrase pauses
- Low number of inter-clause pauses
- Low lexical stress position accuracy
- Low stress contrast



Comprehensibility



Incremental judgment (yuck response)

Click when you make an effort
to understand what the speaker says

I'm struggling

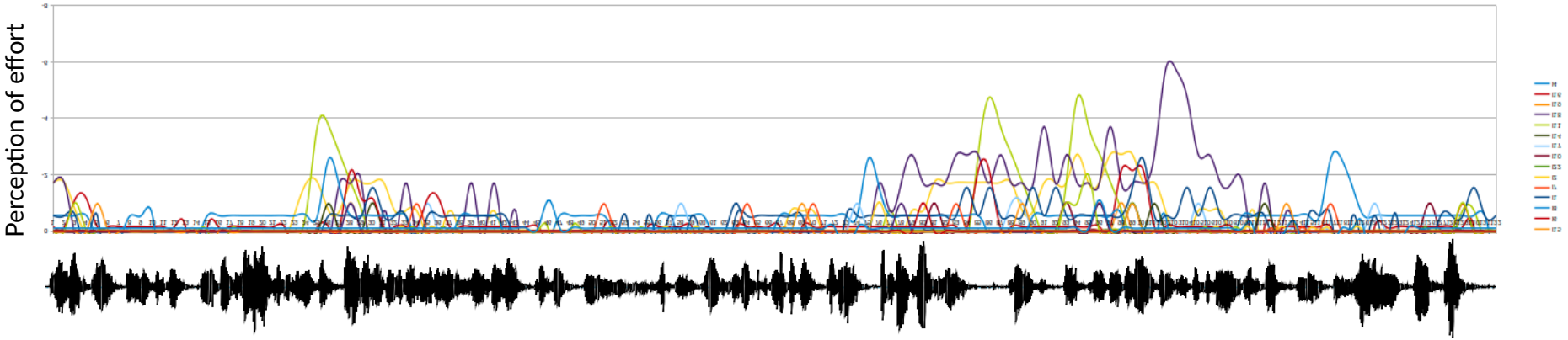
de Kok, I.A. (2013). Listening Heads. Ph. D. thesis, University of Twente.
Nagle, C., Trofimovich, P., Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. Studies in Second Language Acquisition 41(4), 647–672.

Next step

- High number of intra-phrase pauses
- Low number of inter-clause pauses
- Low lexical stress position accuracy
- Low stress contrast

?

Comprehensibility

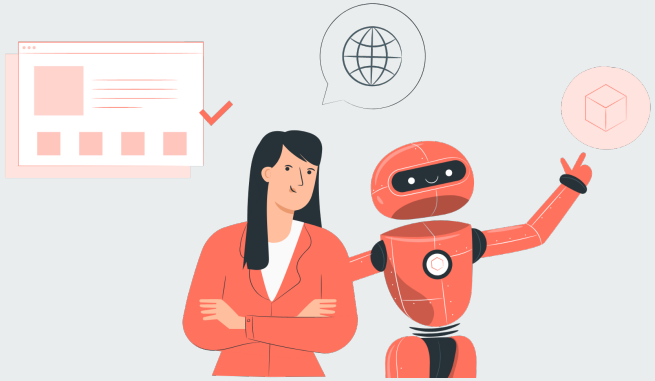


de Kok, I.A. (2013). Listening Heads. Ph. D. thesis, University of Twente.
Nagle, C., Trofimovich, P., Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. Studies in Second Language Acquisition 41(4), 647–672.

Ongoing studies using PLSPP

Lexical stress realization in recited speech by Japanese-L1 elementary school students **T. Kimura (Doshisha Univ.)** 8 speakers Recited speech (text) 10 English native raters	Stress awareness vs. stress production: Comparison of primary stress assignment to English words between Japanese and Korean university students **M. Sugahara (Doshisha Univ.)** 54 speakers Read speech (carrier phrases)	Prosody, intelligibility and communication: pronunciation assessment before and after a training session **D. Frost (U. Grenoble Alpes)** 280 speakers (corpus PIC) Read speech (text)	A corpus of spontaneous L2 English speech by Japanese university students **T. Konishi (Waseda Univ.)** Similar conditions with the CLES corpus

Thank you!



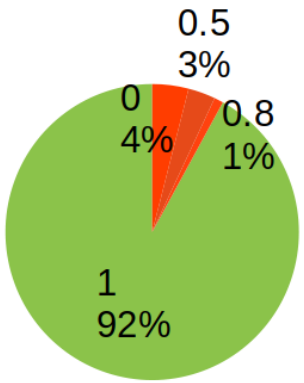
Link to the pipeline: https://gricad-gitlab.univ-grenoble-alpes.fr/lidilem/plspp
To get the public part of the corpus: coordination-nationale@certification-cles.fr

Sylvain COULANGE
sylvain.coulange@univ-grenoble-alpes.fr

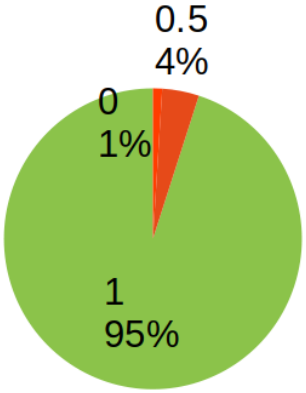
First step of PLSPP evaluation

28 random files
100 target words, manual verification

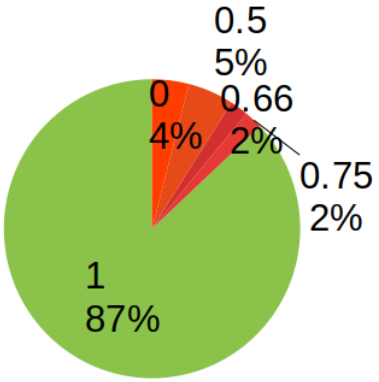
Word-recognition



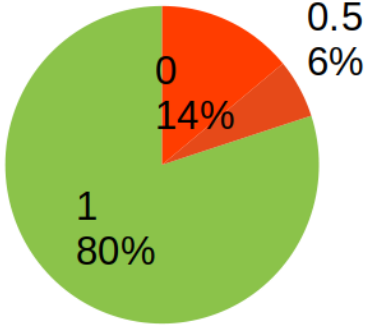
Word-alignment



Syllable-detection



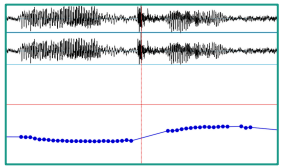
Prosodic-shape



Currently ongoing: manual transcription of random files by Master students

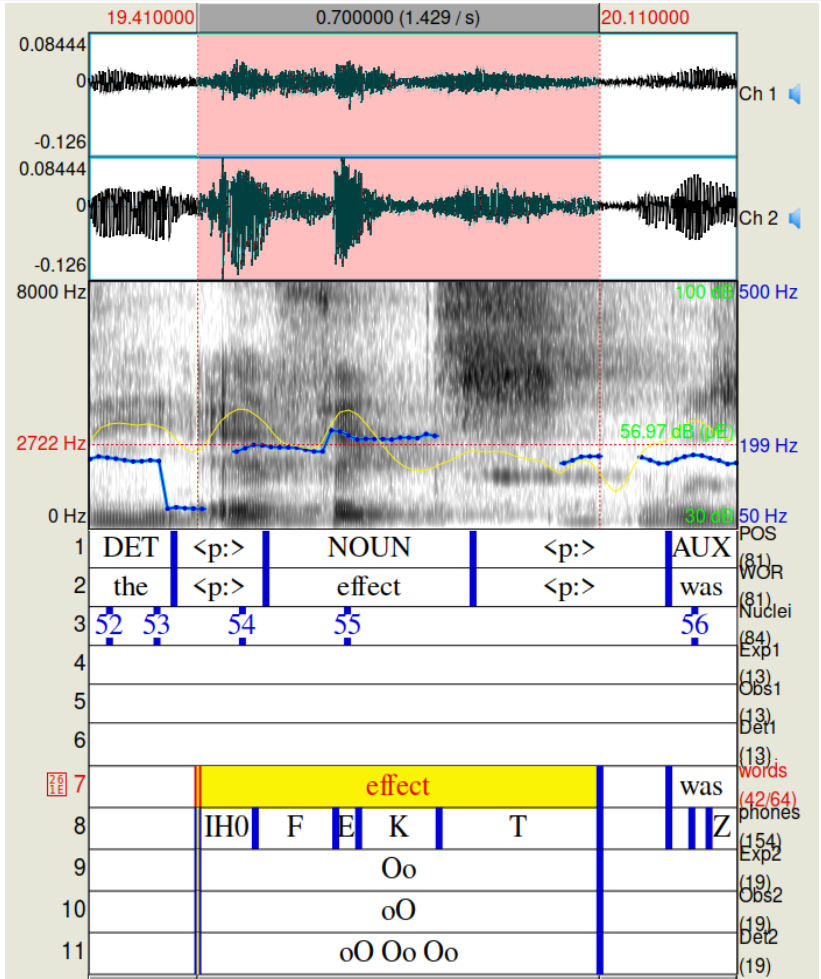
Implementation of MFA

- Better word-level alignment
- Acoustic parameter extraction on the vowel interval instead of syllable nuclei
- Consider F0 variation within the vowel
- F0 interpolation for devoiced vowels
- No more influence from final consonant lengthening



Linear interpolation if no F0 value at syllable point

Whisper
+
Montreal Forced Aligner
+
Parameter extraction on vowel intervals



改善したパイプラインにおける韻律特徴量抽出

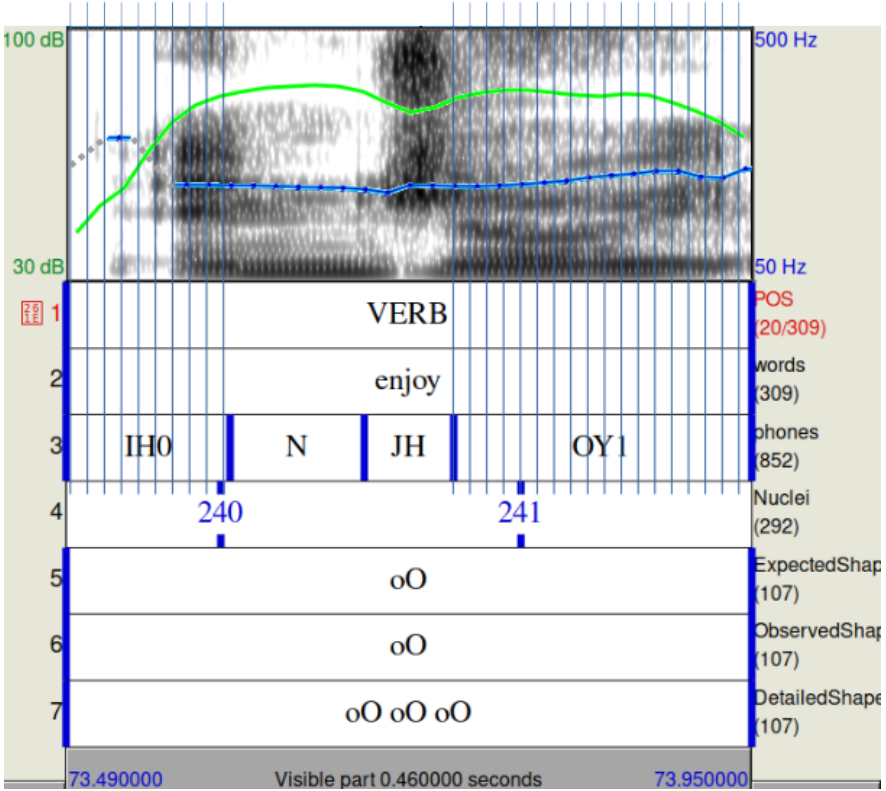
time_step = 10ms
(customizable)

F0
- mean(F0s)
- (Min, max, sd)

pitch linear interpolation

Intensity
- max(dBs)

Duration
- Length of vowel interval



改善したパイプラインにおけるワードアライメント

Better consideration of initial phoneme

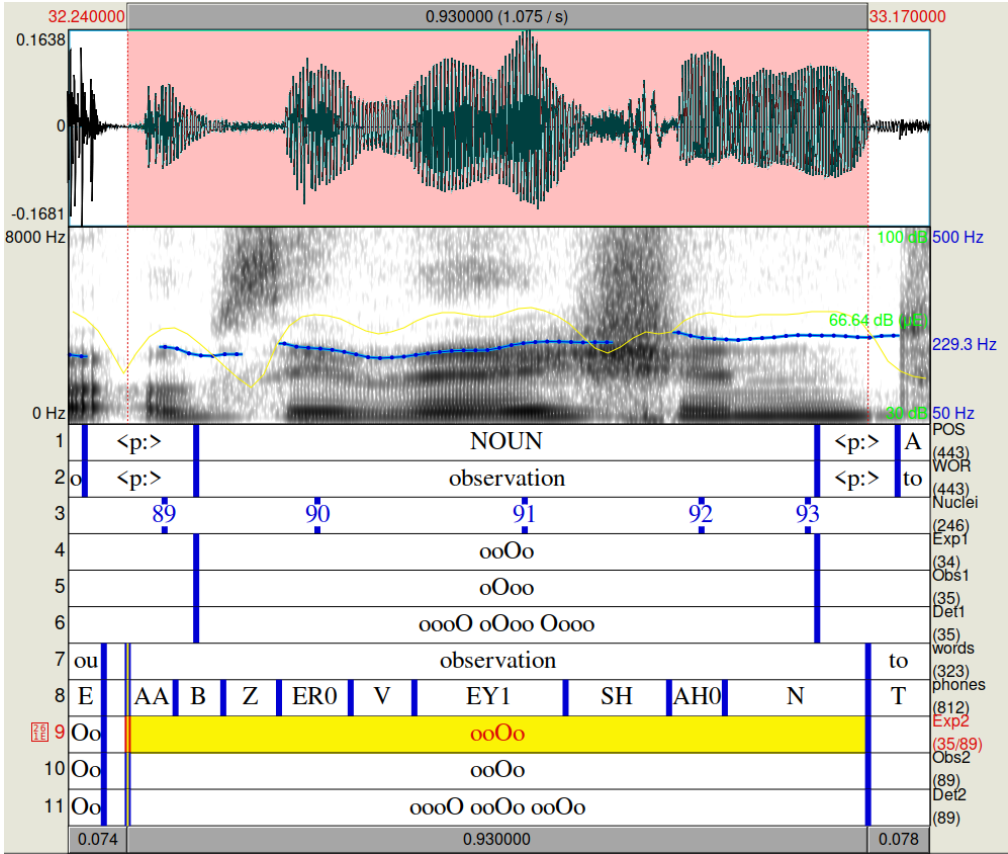
Final consonant included until it ends

Good alignment vowel-to-syllable nucleus in general

General alignment needs to be improved
(especially with spontaneous speech)

Stress analysis even if inadequate nb of syllable nuclei

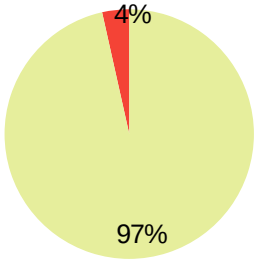
Extra nuclei might be present (hesitation, epenthesis, bad recognition…)
→possibility to filter in order to keep only words with adequate nb of syllable nuclei



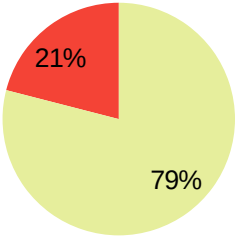
Word alignment precision

Number of target words with totally wrong alignment,
among the first 200 plain target words in the visualization interface:

plspp : 7 target words

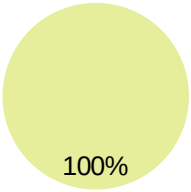


plspp_mfa : 42 target words

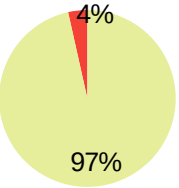


Corpus PIC (Frost, D.)
(280 speakers Read speech
~1min20s/spk)

Plspp: 0 words



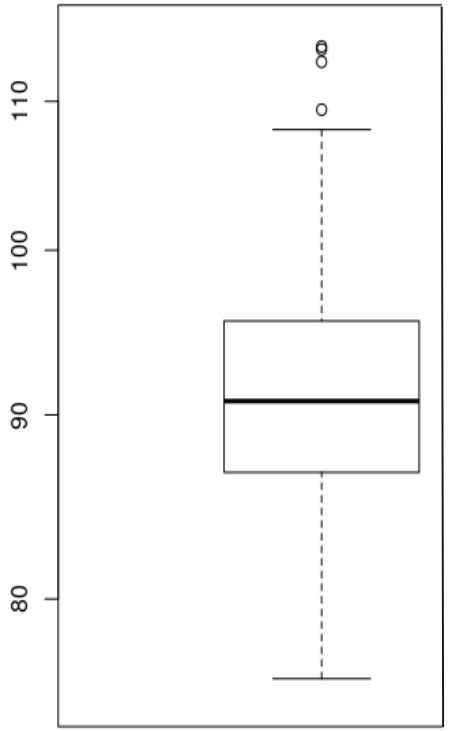
plspp_mfa: 7 words



話者正規化

正規化

Speaker: jan2020-001_020-022_SPEAKER_00

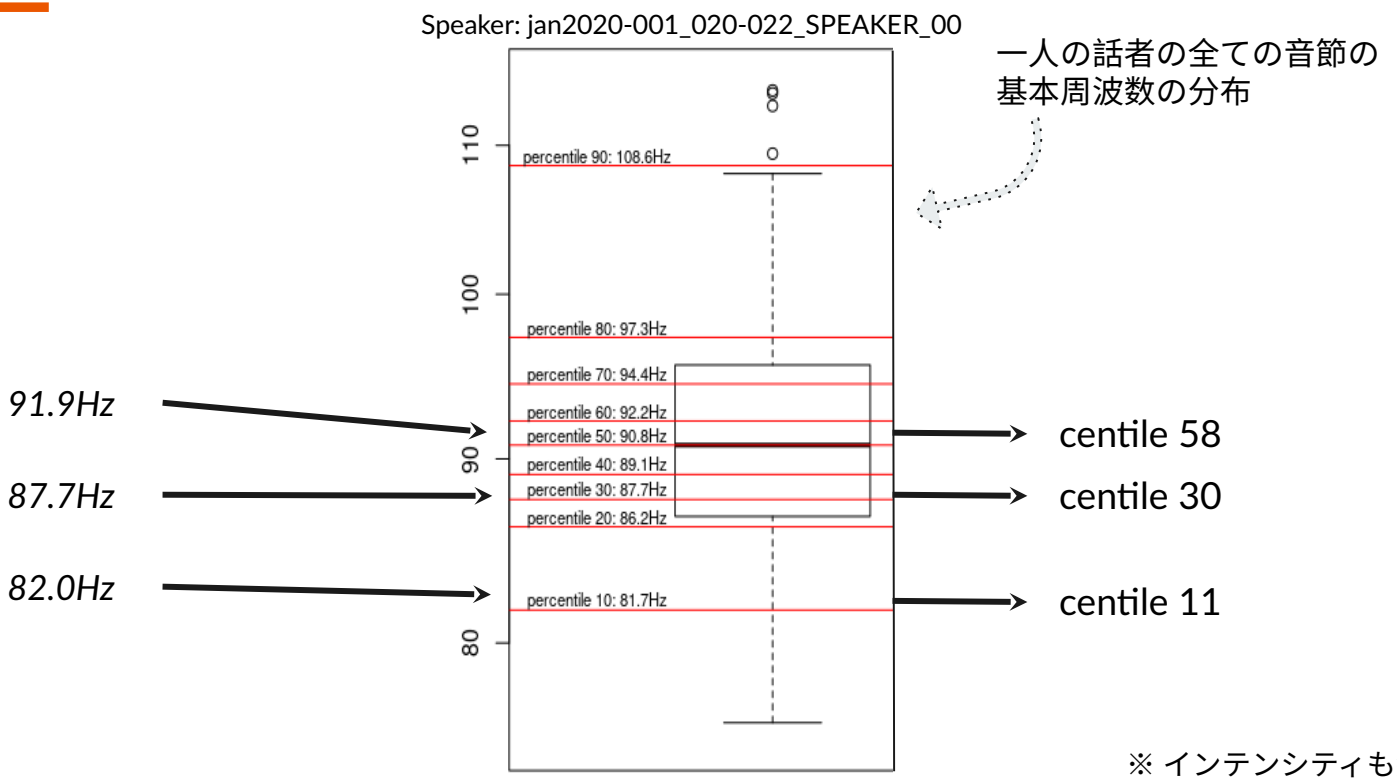


一人の話者の全ての音節の
基本周波数の分布

※ インテンシティも継続時間も
同様に正規化を行う

話者正規化

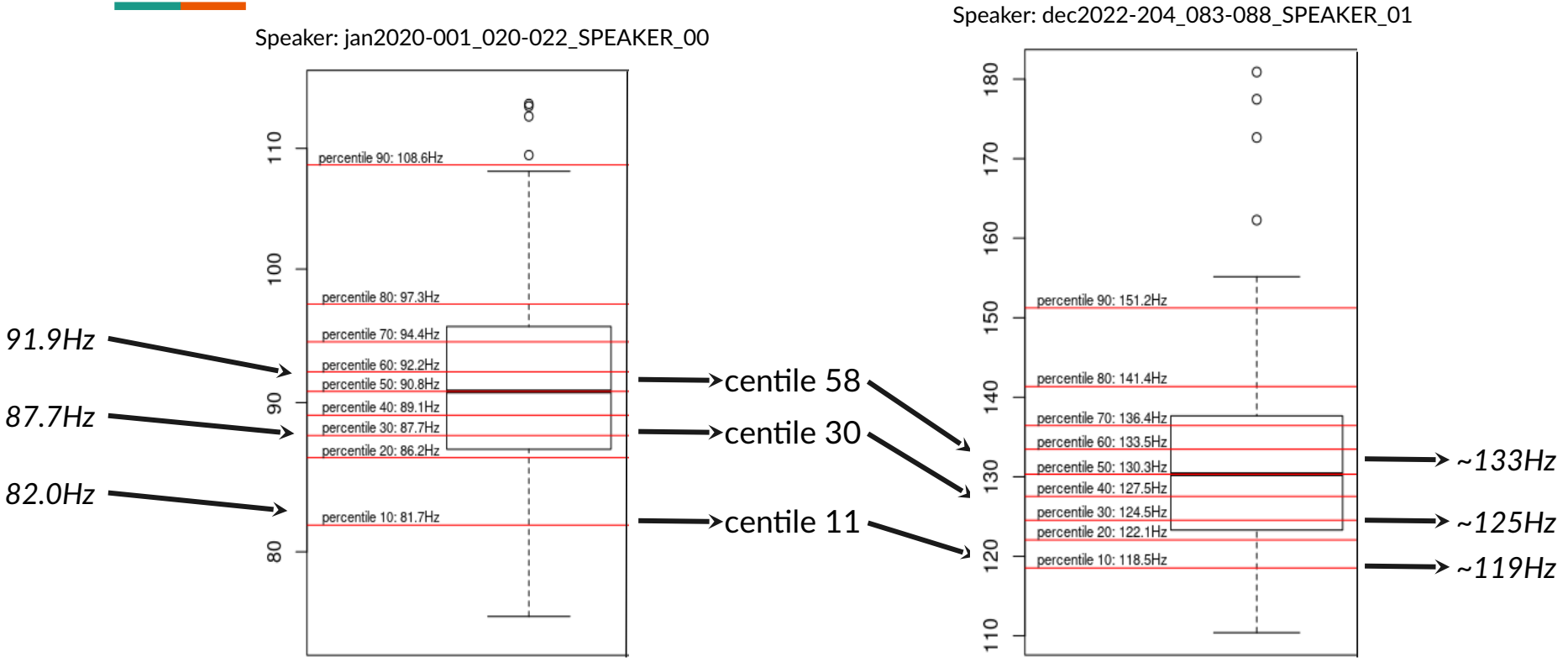
正規化



※ インテンシティも継続時間も
同様に正規化を行う

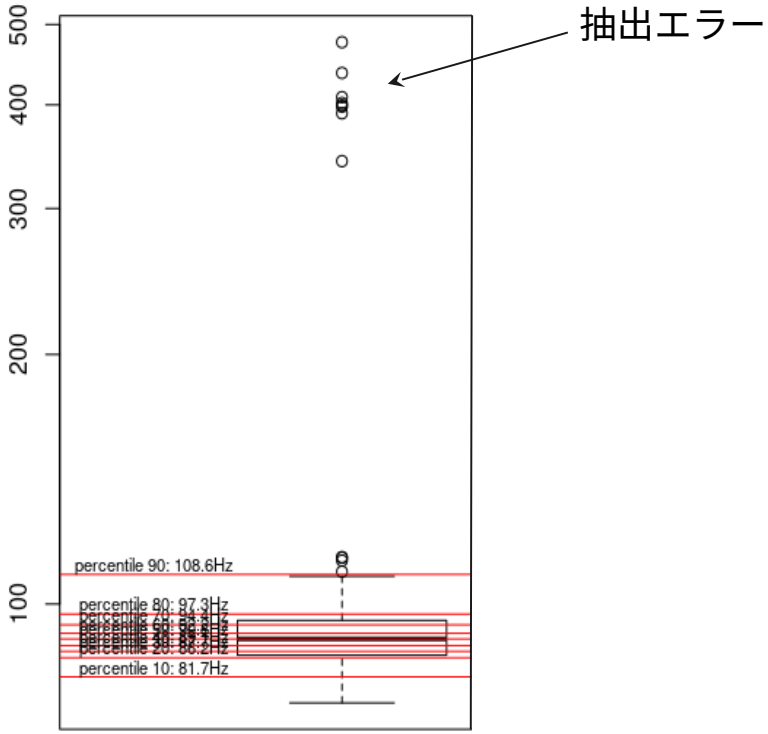
話者正規化

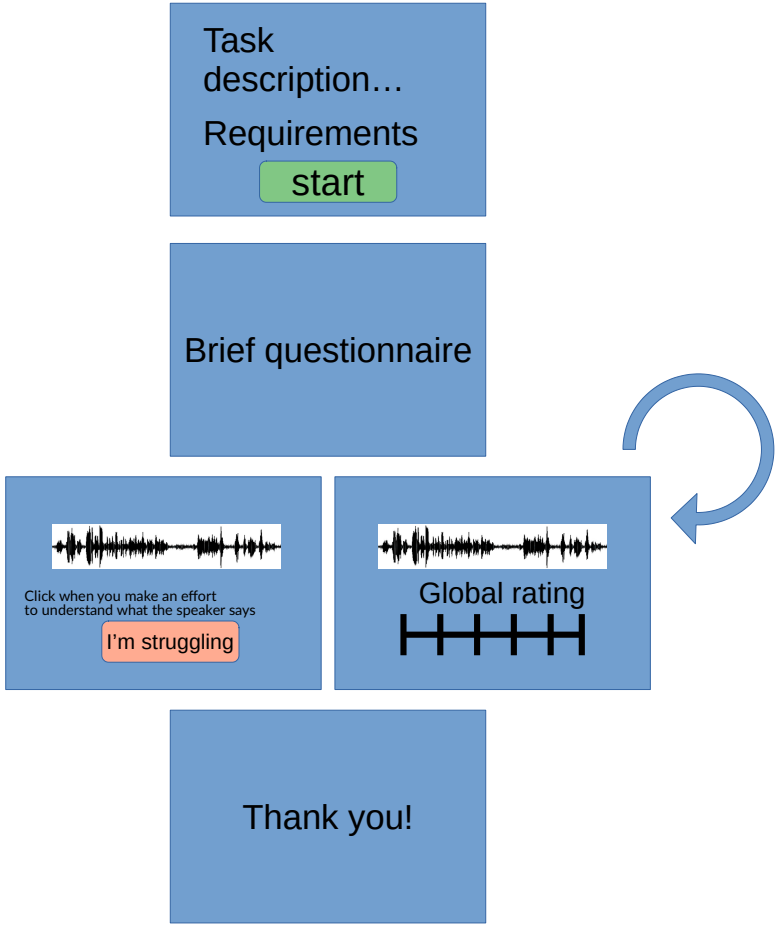
正規化

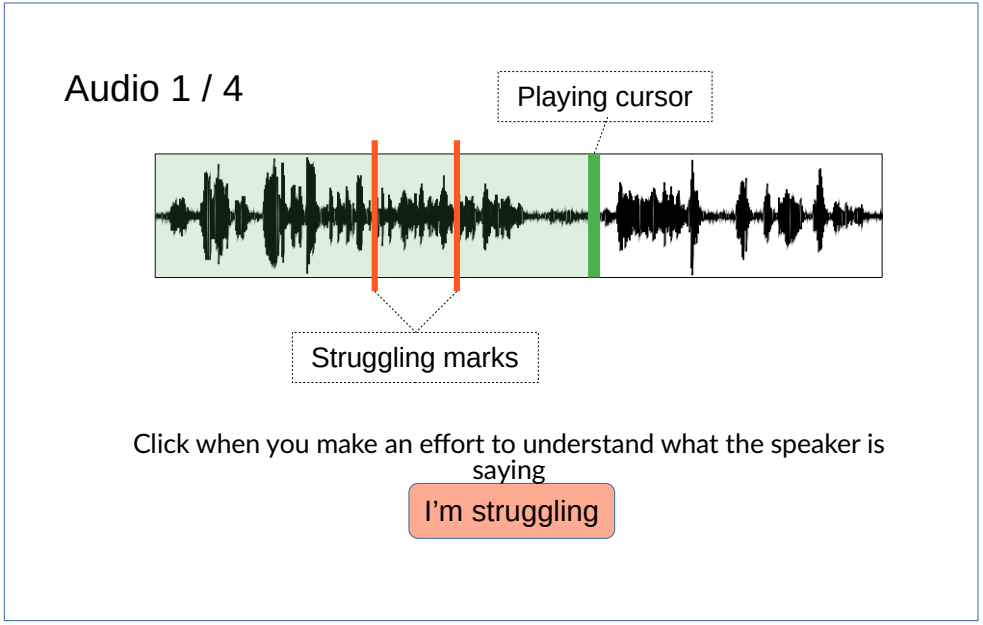

Speaker: jan2020-001_020-022_SPEAKER_00
percentile 90: 108.6Hz
percentile 80: 97.3Hz
percentile 70: 94.4Hz
percentile 60: 92.2Hz
percentile 50: 90.8Hz
percentile 40: 89.1Hz
percentile 30: 87.7Hz
percentile 20: 86.2Hz
percentile 10: 81.7Hz
110
100
90
80
91.9Hz
87.7Hz
82.0Hz
centile 58
centile 30
centile 11
Speaker: dec2022-204_083-088_SPEAKER_01
percentile 90: 151.2Hz
percentile 80: 141.4Hz
percentile 70: 136.4Hz
percentile 60: 133.5Hz
percentile 50: 130.3Hz
percentile 40: 127.5Hz
percentile 30: 124.5Hz
percentile 20: 122.1Hz
percentile 10: 118.5Hz
180
170
160
150
140
130
120
110
~133Hz
~125Hz
~119Hz

話者正規化

正規化


抽出エラー
500
400
300
200
100
percentile 90: 108.6Hz
percentile 80: 97.3Hz
percentile 10: 81.7Hz


Task description…
Requirements
start
Brief questionnaire
Click when you make an effort to understand what the speaker says
I'm struggling
Global rating
Thank you!


Audio 1 / 4
Playing cursor
Struggling marks
Click when you make an effort to understand what the speaker is saying
I'm struggling

LIDILEM Université Grenoble Alpes

SLPL Spoken Language Processing Laboratory

UGA Université Grenoble Alpes

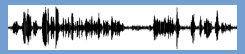
同志社大学 Doshisha University

Task description…

Requirements

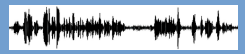
start

Brief questionnaire



Click when you make an effort to understand what the speaker says

I'm struggling

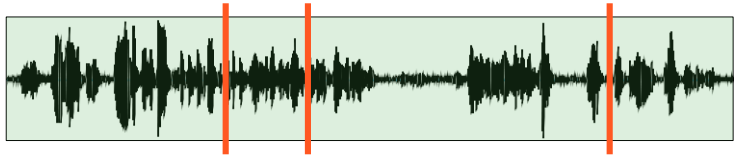


Global rating

Thank you!

Contact info...

Audio 1 / 4



Tell us about your global impressions:

Overall pronunciation accuracy

Very poor pronunciation — Very good pronunciation

Overall fluency

Very poor fluency — Very fluent

The speaker was...

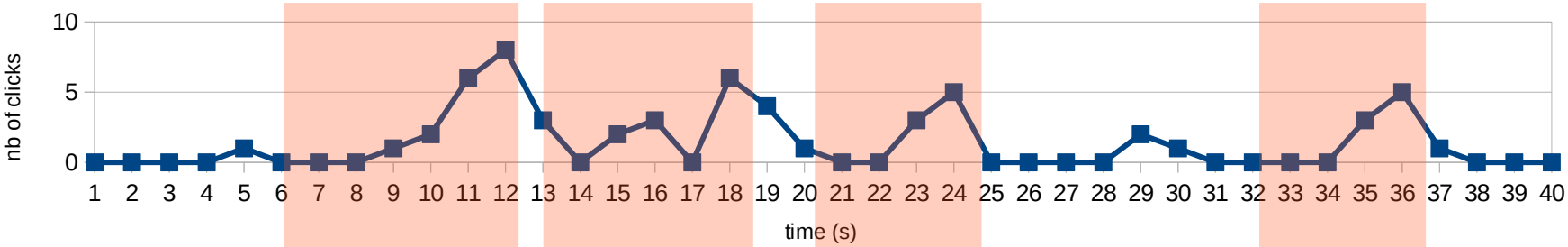
Very hard to understand — Very easy to understand

Optional: what did disturb you most in the student

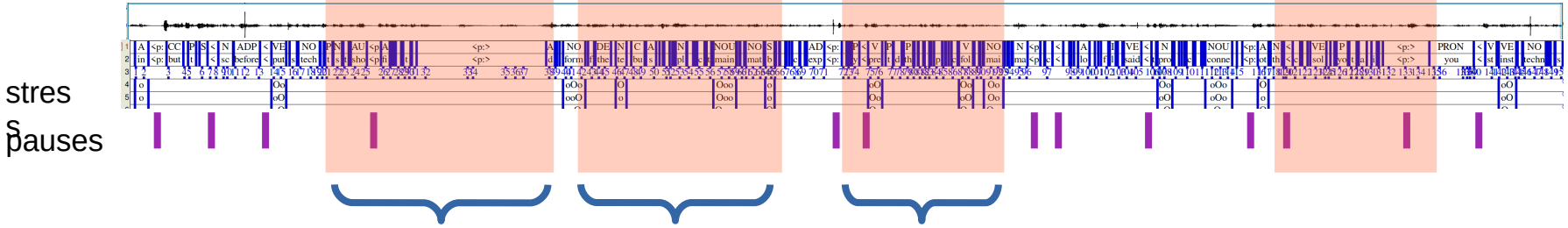
next

LIDILEM Université Grenoble Alpes | SLPL Spoken Language Processing Laboratory | UGA Université Grenoble Alpes | 同志社大学 Doshisha University

Expected output (all raters together):



Pipeline output:



- Nb words with inapopriate stress pattern
- Prosodic contrast stressed vs. unstressed syllables
- Nb of pauses, nb of intra-phrase pauses

...Then compare normal vs red zones