
Measuring speech rhythm through automated analysis of syllabic prominences

Noriko Nakanishi
Kobe Gakuin University

nakanisi@gc.kobegakuin.ac.jp



Sylvain Coulangue
Université Grenoble Alpes

sylvain.coulangue@univ-grenoble-alpes.fr



Outline

1. Introduction

- Fluency and rhythm
- Rhythmic patterns in English and Japanese
- Hypotheses

2. Data analysis

- Corpus
- Speech processing

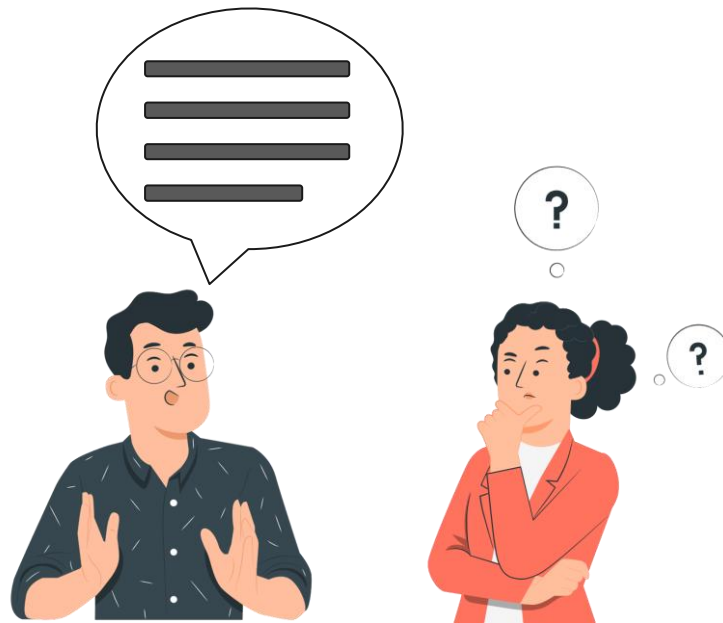
3. Results & Discussion

- Syllabic contrast within polysyllabic words
- Lexical contrast between content and function words

4. Conclusion

5. Limitations and future work

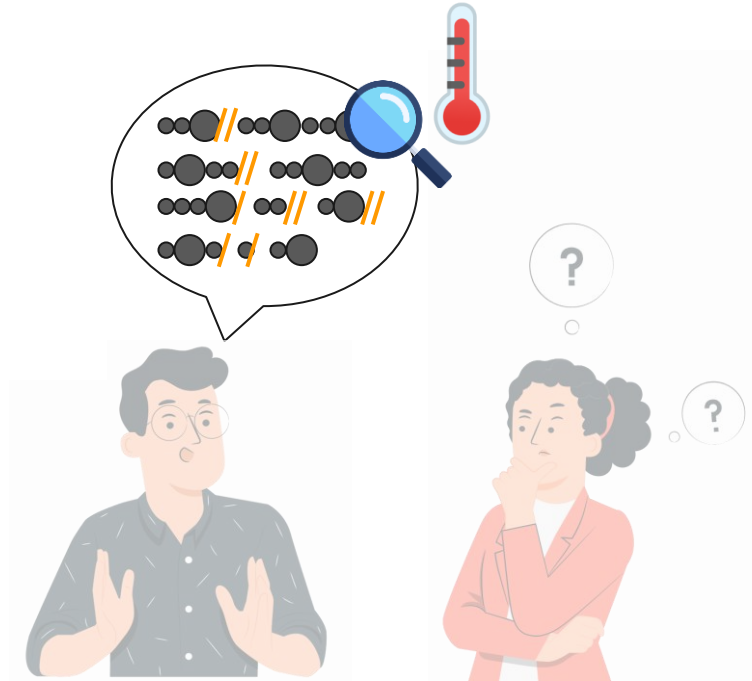
Introduction - Fluency & rhythm



Introduction - Fluency & rhythm



Introduction - Fluency & rhythm



Introduction - Rhythmic patterns

ENGLISH

rainbow



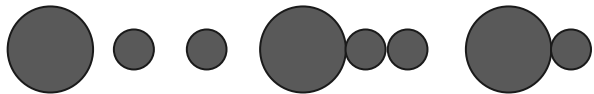
- **Stress** accent

Words can be distinguished **based on stress**.

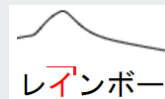
- **Syllable** as a rhythmic unit

Stressed syllables are usually **higher, louder, longer**.

Look at that beautiful rainbow!



JAPANESE

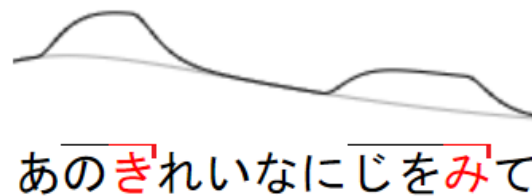


- **Pitch** accent

Words can be distinguished **based on pitch**.

- **Mora** as a rhythmic unit

Each mora has **similar duration** with **different pitch height**.



Introduction - Hypotheses

Compared to L1 English speakers (NE), Japanese English learners (JE) demonstrate...

- 1) lower prosodic **contrast between syllables** within polysyllabic words

-> Syllabic Contrast Scores: NE > JE



- 2) lower prosodic **contrast between content and function words**

-> Lexical Contrast Scores: NE > JE



Would it be possible to measure the syllabic prominences automatically?

Data analysis - Corpus

34-hour read-aloud corpus

- 877 JE samples recorded by 42 Japanese university students
- 91 NE samples recorded by 7 professional English narrators

21 topics

- Japanese students wrote their presentation manuscripts and read them aloud.
- Professional narrators read aloud model presentation passages.

Participants' English proficiency levels

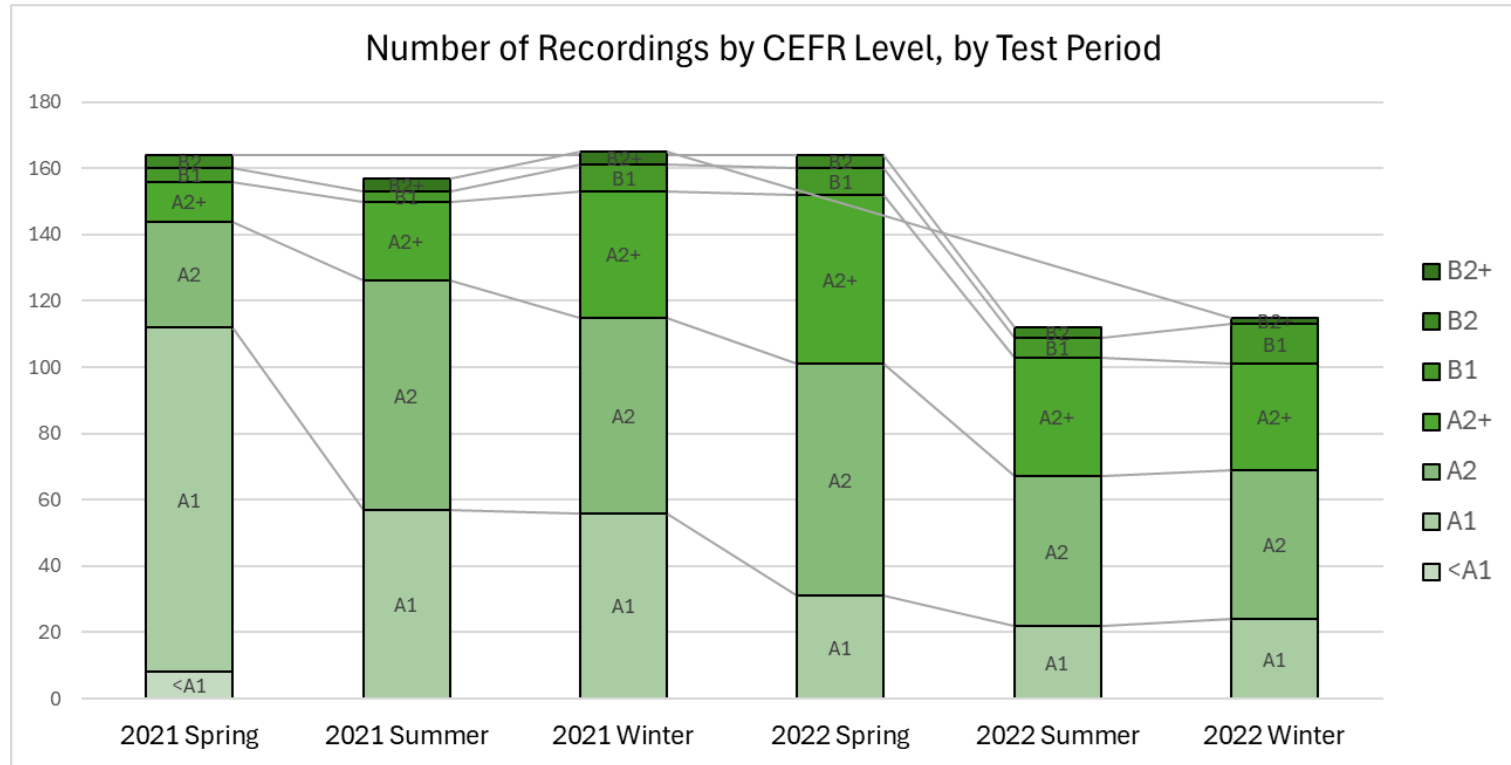
- Japanese students: ranging from CEFR A1 or below - B2
- Professional narrators: 5 AmE, 2 BrE; 3 males, 4 females

Data analysis - Corpus

2021 Spring VST	2022 Spring VST
U1 College Life	U1 Academic Research
U2 Understanding Copyright	U2 Social Issues in Japan
U3 Cyber Bullying	U3 Personal Safety
U4 Tourism	U4 Gender
2021 Summer VST	2022 Summer VST
U5 Foreign Encounters	U5 Religion
U6 Entertainment	
U7 International Affairs	U7 Career
U8 Technology	U8 Japanese Culture
2021 Winter VST	2022 Winter VST
U9 My Future	U9 Law and Peace
U10 Personal Finance	U10 Ethnicity
U11 Health	U11 Science and Scientists

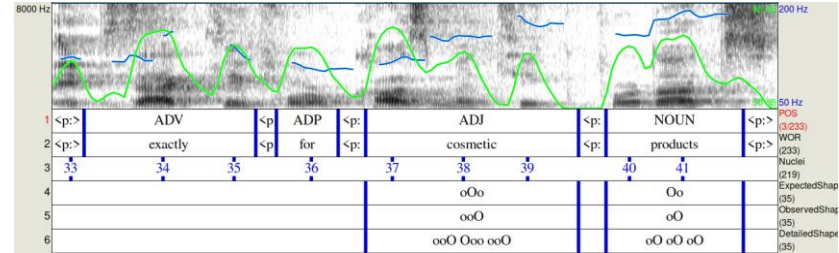
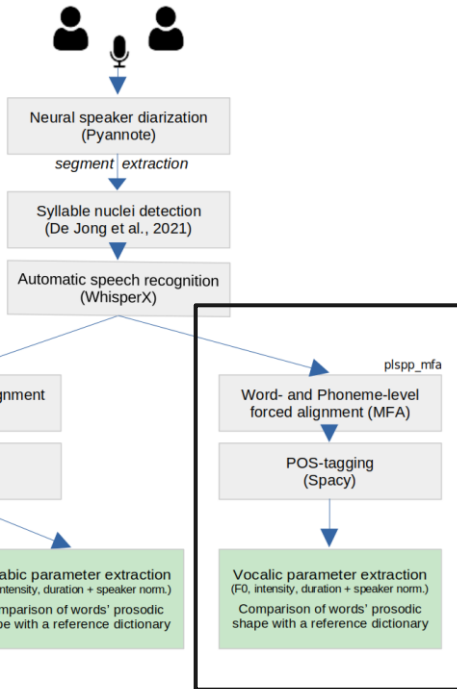


Data analysis - Corpus

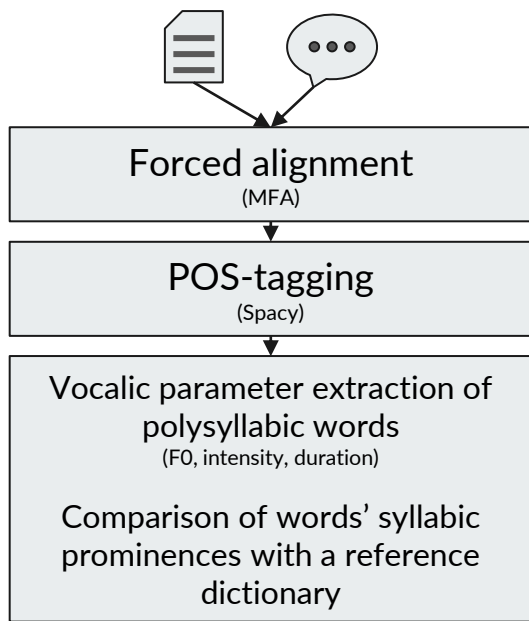


Data analysis - Speech processing

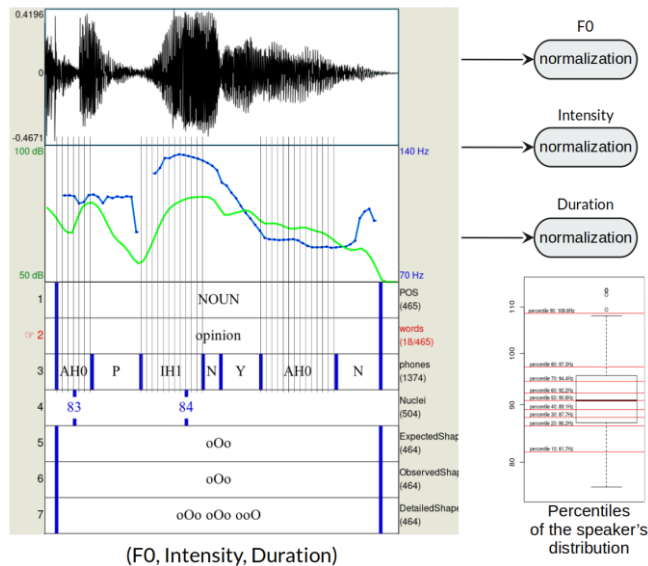
PLSPP
The Pauses and
Lexical Stress
Processing Pipeline



Data analysis - Speech processing

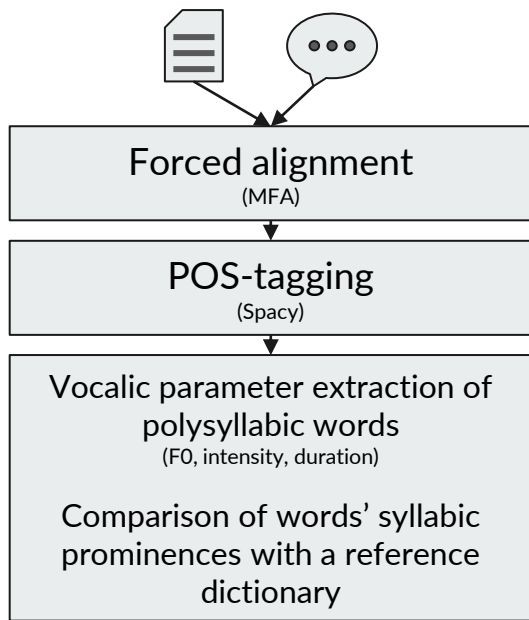


The Pauses and Lexical Stress Processing Pipeline (PLSPPP) v2



Speaker	Word	POS	Expected	Observed	Pitch	Energy	Duration
B121009_21u01	opinion	NOUN	oOo	oOo	6 9	4 5	2 6 8

Data analysis - Speech processing



The Pauses and Lexical Stress Processing Pipeline (PLSPPP) v2



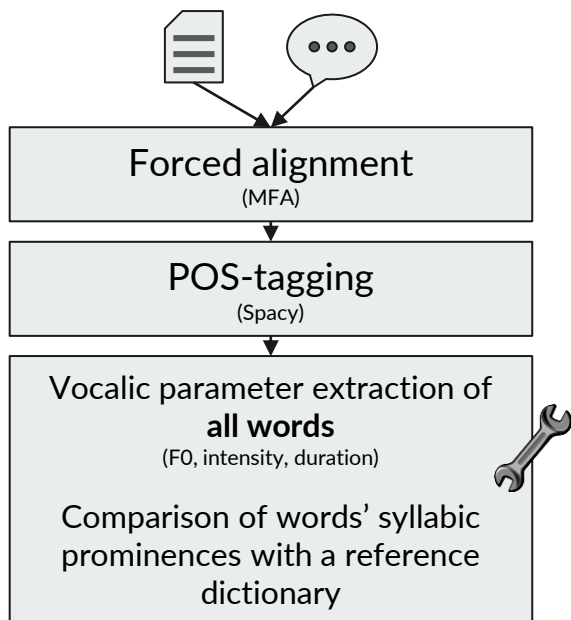
Stress Score (word-level)

$\text{mean}(\text{pitch, intensity, duration})_{\text{most prominent syllable}} - \text{mean}(\text{other syllables})$

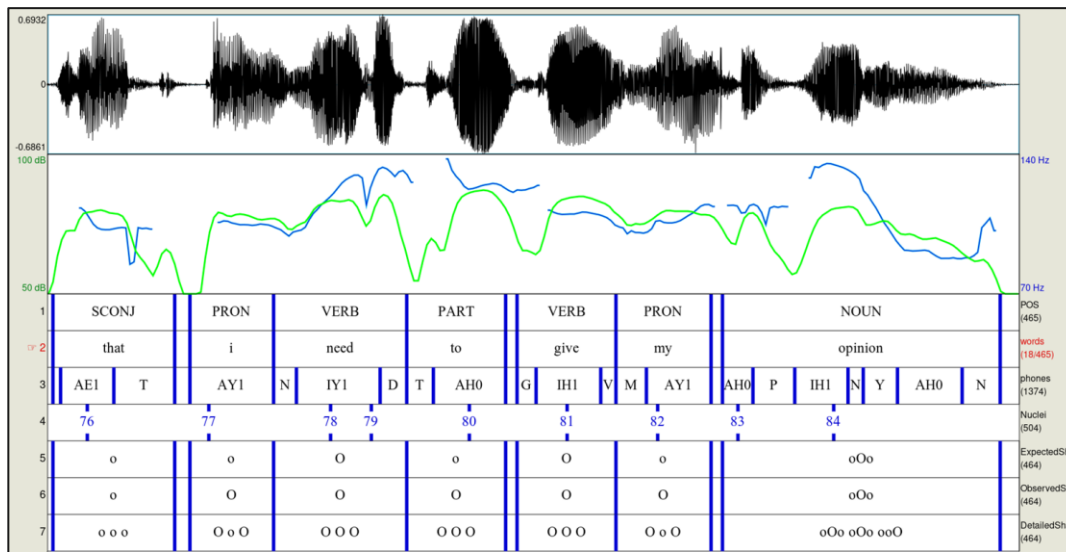
Syllabic Contrast (text-level)

$\text{mean}(\text{word stress scores})$

Data analysis - Speech processing



The Pauses and Lexical Stress Processing Pipeline (PLSPPP) v2+



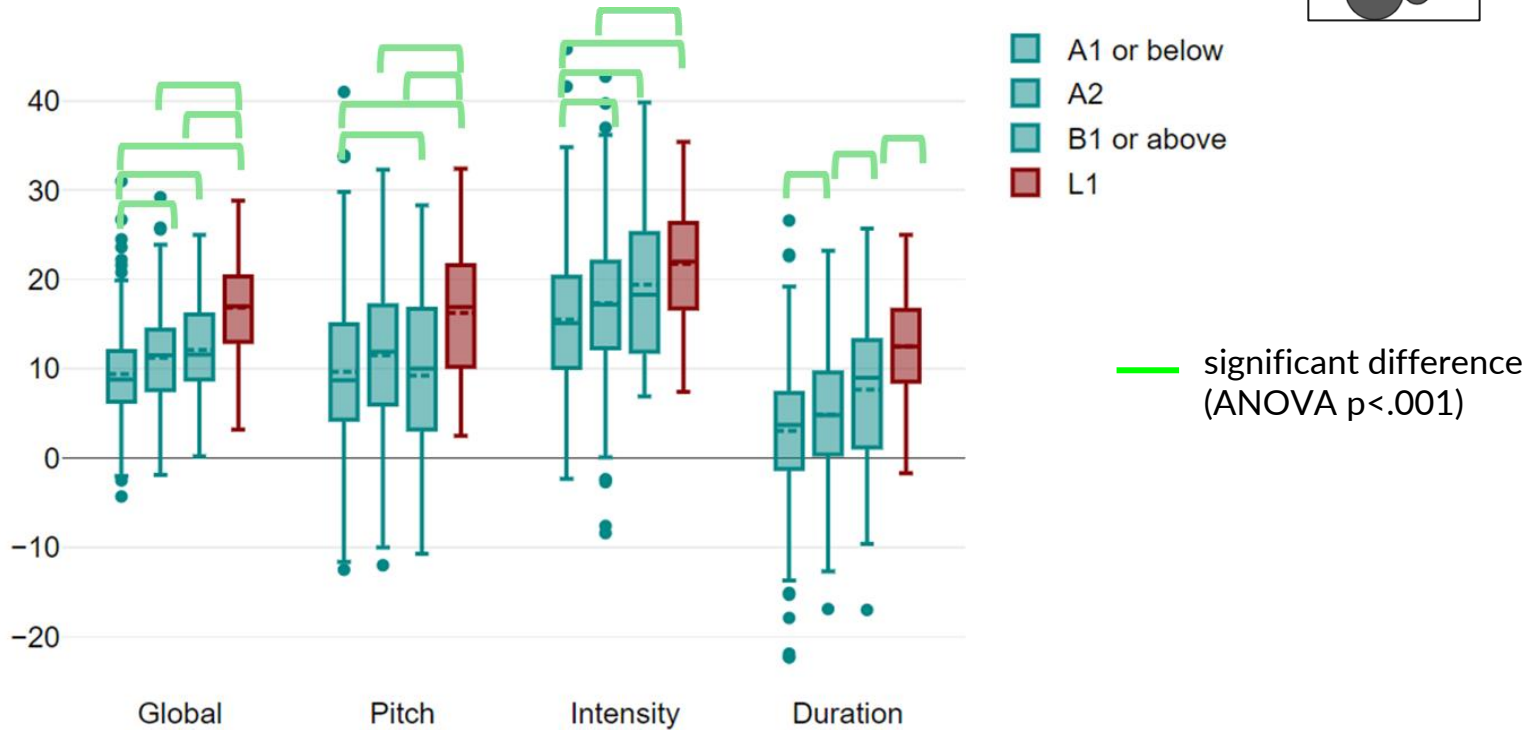
Lexical Contrast (text-level) = mean(most prominent syll. of content words)

- mean(most prominent syll. of

lexical words)

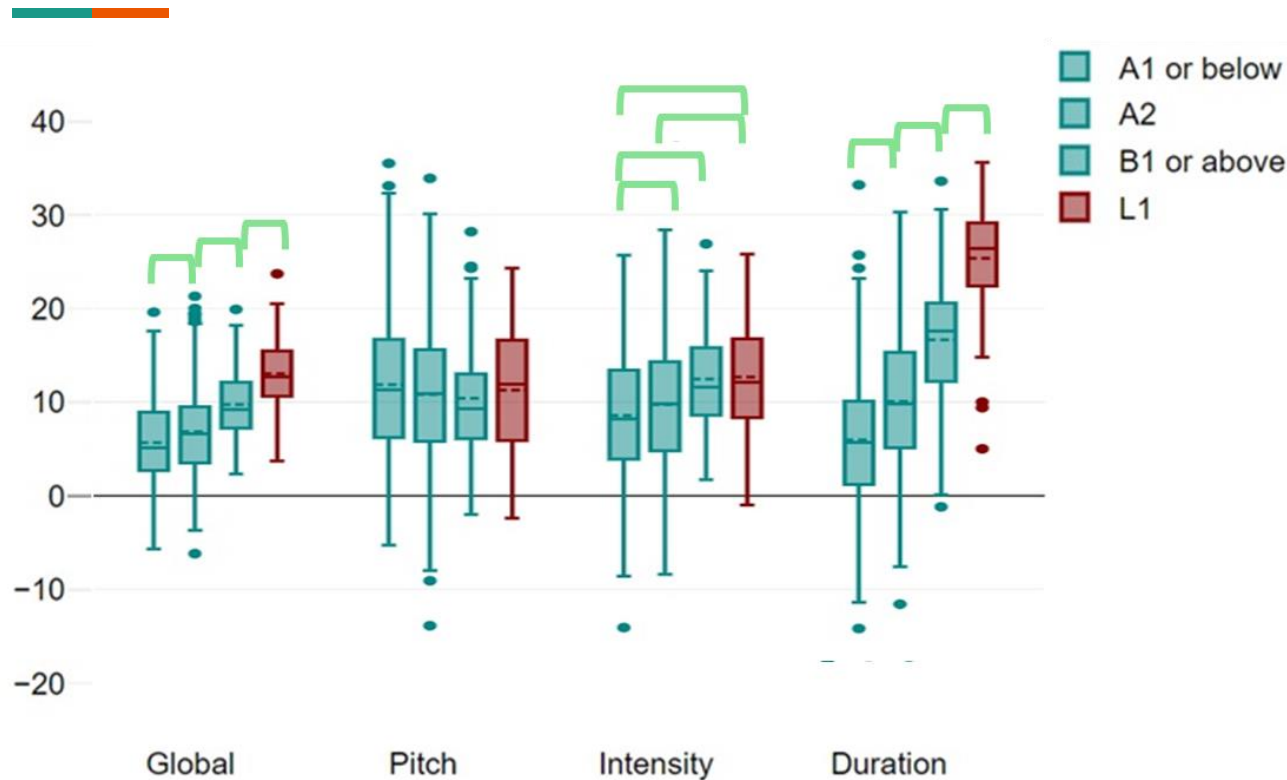
Results - Syllabic contrast

Degree of acoustical prominence contrast between stressed and unstressed syllables within polysyllabic words



Results - Lexical contrast

Degree of acoustical prominence contrast between content and function words



— significant difference (ANOVA $p < .001$)

Conclusion

Compared to L1 English speakers (NE), Japanese English learners (JE) demonstrate...

- 1) lower prosodic **contrast between syllables** within polysyllabic words

-> Syllabic Contrast Scores: NE > JE -> **YES**



- 1) lower prosodic **contrast between content and function words**

-> Lexical Contrast Scores: NE > JE

-> **Basically YES, but NOT with pitch contrast**

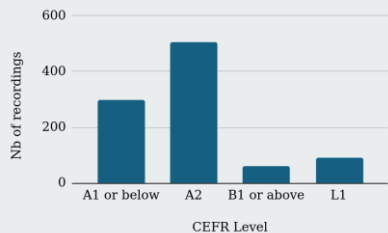


Would it be possible to **measure the syllabic prominences automatically?**

-> **YES**

Limitations and future work

Unbalanced groups

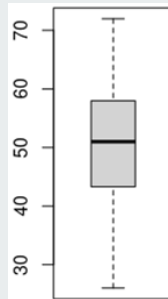


- Resampling to 55 recordings per group.

→ consider A2.1, A2.2...

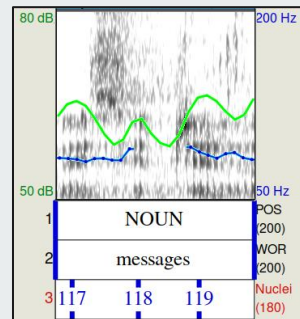
Monosyllabic words stress estimation method

- Monosyll. words are considered stressed if their mean acoustic prominence level is ≥ 50



→ compare to immediate context or utterance median value.

Stress estimation



- devoicing,
- natural lengthening
- vowel intrinsic duration

Context-dependent stress patterns

Look *at that* beautiful rainbow!



→ conduct stress analysis at the sentence level

→ investigate how learners emphasize essential information within their utterances



Thank you for your attention!

Noriko Nakanishi
Kobe Gakuin University

nakanisi@gc.kobegakuin.ac.jp

Sylvain Coulange
Université Grenoble Alpes

sylvain.coulange@univ-grenoble-alpes.fr

Illustrations by [Storyset](#)



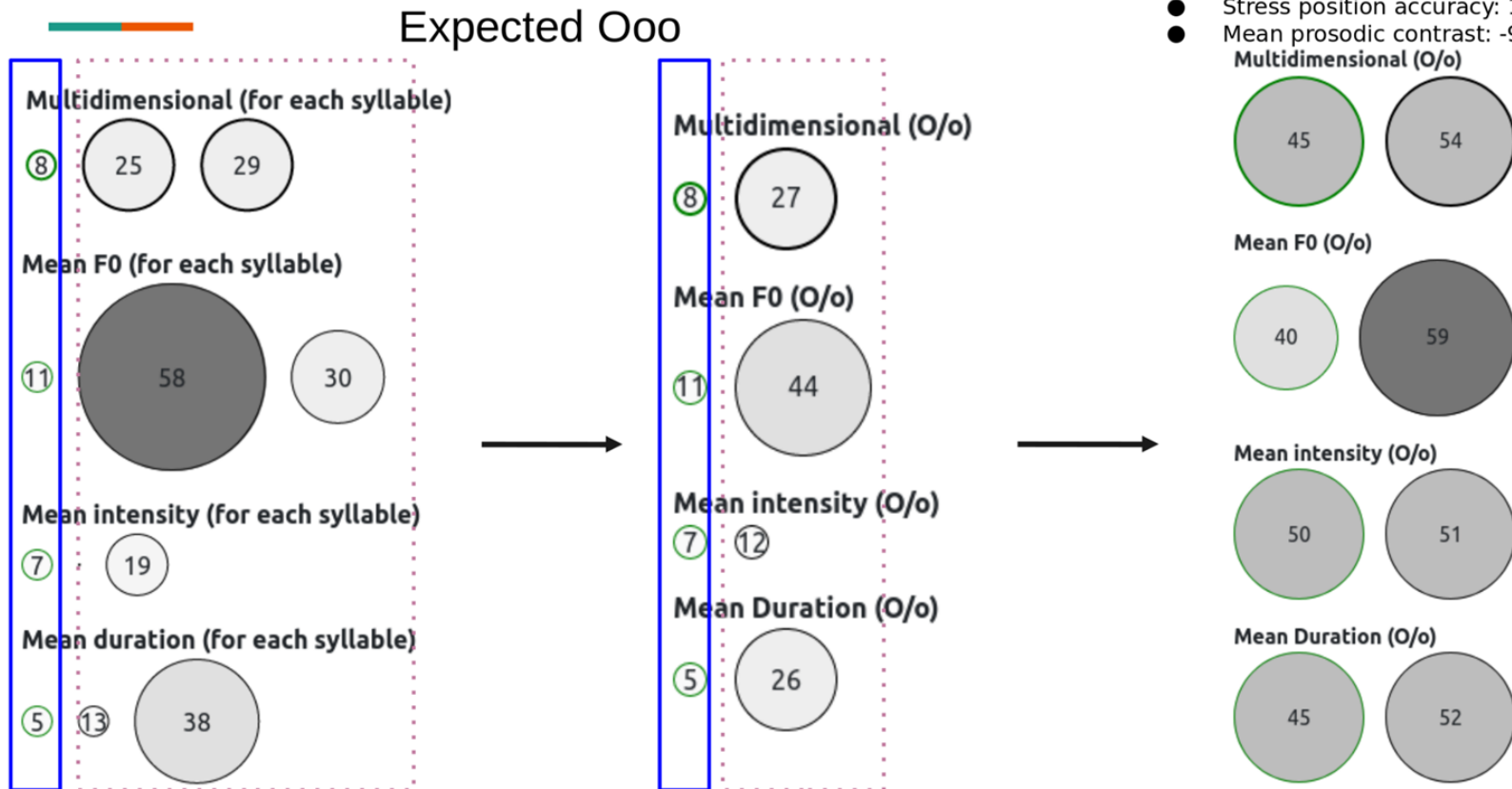
This research is supported by
JSPS Kakenhi Grant Number
24K04175

References

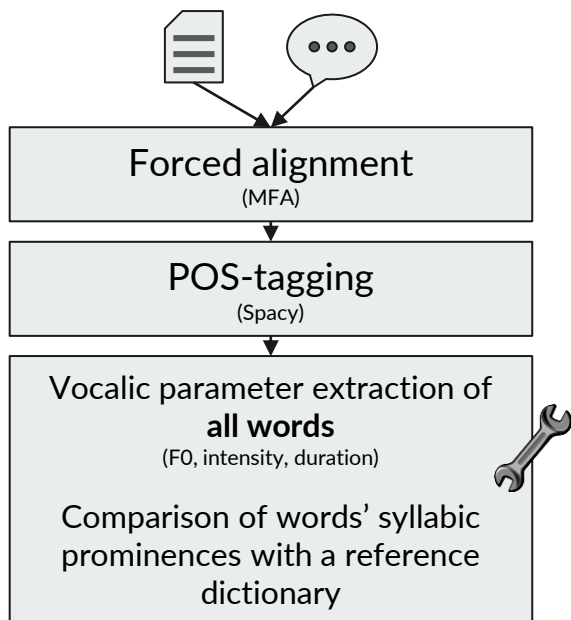
- Coulange S, Kato T, Rossato S, Masperi M. Enhancing Language Learners' Comprehensibility through Automated Analysis of Pause Positions and Syllable Prominence. *Languages*. 2024; 9(3):78
- Cutler, A. (2015) "Lexical stress in english pronunciation." In *The Handbook of English Pronunciation*. (pp. 106–124). Hoboken, NJ:John Wiley & Sons, Inc.
- Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America*, 105(3), 1877-1888.
- Hirano, H. et al. (2013). OJAD: A free online accent and intonation dictionary for teachers and learners of Japanese. *INTERSPEECH 2013*, 1875-1876.
- Nakanishi et.al. (2023). *Global Perspectives Listening & Speaking*. Seibido.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258-278.
- Pearson. (2024). *Versant by Pearson English Speaking and Listening Test: Test Description and Validation Summary*.
- Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency* (1st ed.). Routledge.
- Tortel, A. (2021) "Le rythme en anglais oral : considérations théoriques et illustrations sur corpus." *Recherche et pratiques pédagogiques en langues - Cahiers de l'APLIUT*.
- Walker, R., Low, E., & Setter, J. (2021) *English Pronunciation for a Global World*. Oxford: Oxford University Press.

Speaker jan2020-001_020-022_SPEAKER_00

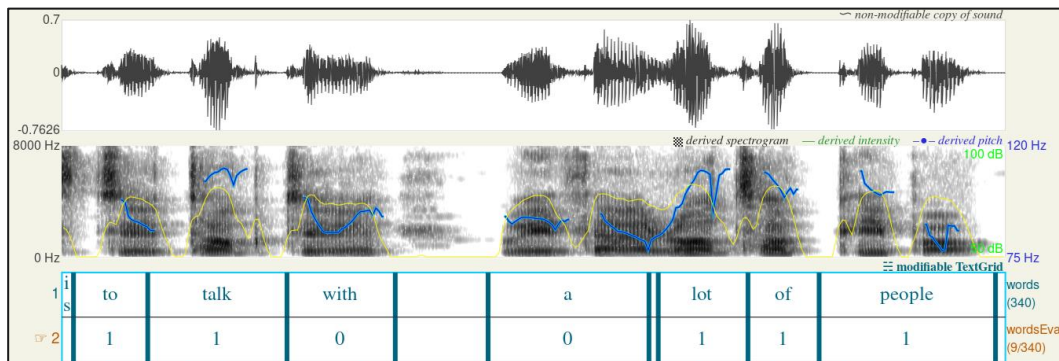
- 42 target words
- Stress position accuracy: 19%
- Mean prosodic contrast: -9 points



Data analysis - Speech processing



The Pauses and Lexical Stress Processing Pipeline (PLSPPP) v2+



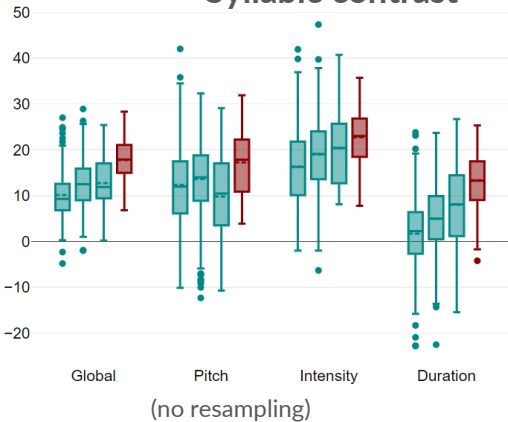
- All the files (966) manually checked for word-alignment precision

81% of words correctly aligned
 79% for JP and 92% for EN
 per recording: median at 90% for JP and 96% for EN)

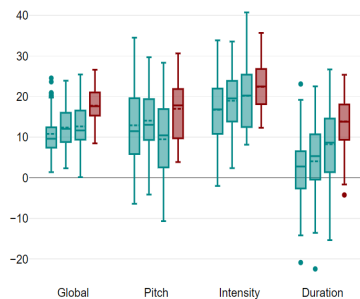


Keep only correctly aligned words (1) within recordings of at least 80% of correct alignment

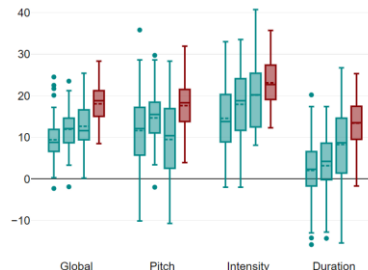
Syllabic contrast



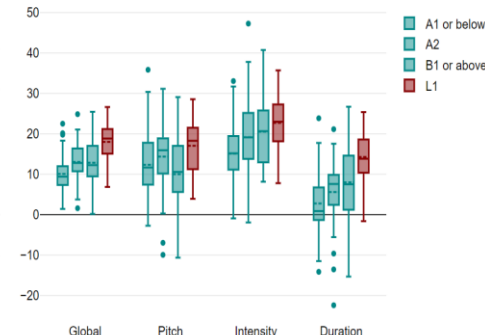
Legend:
 A1 or below
 A2
 B1 or above
 L1



(resampling iteration 1)

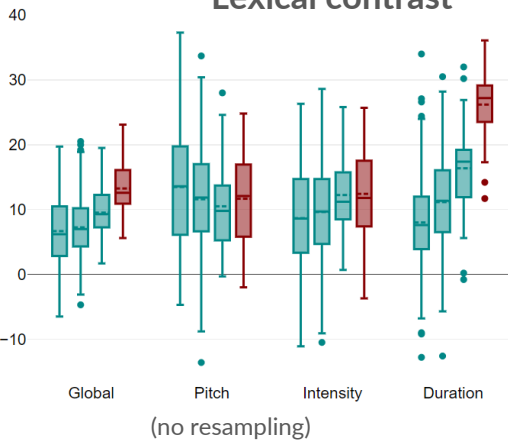


(resampling iteration 2)

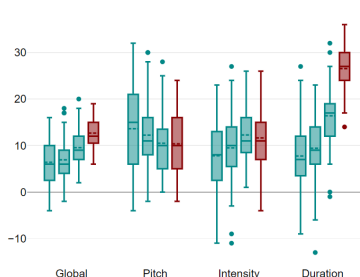


(resampling iteration 3)

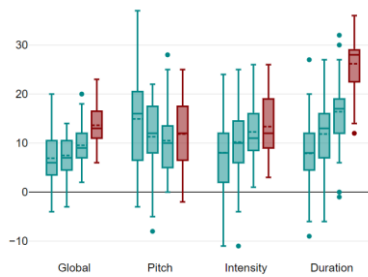
Lexical contrast



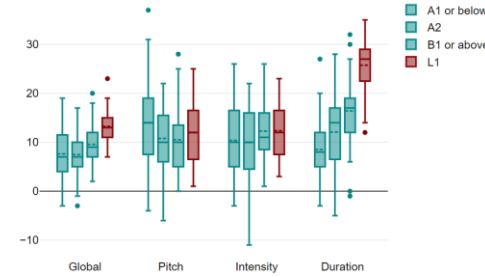
Legend:
 A1 or below
 A2
 B1 or above
 L1



(resampling iteration 1)

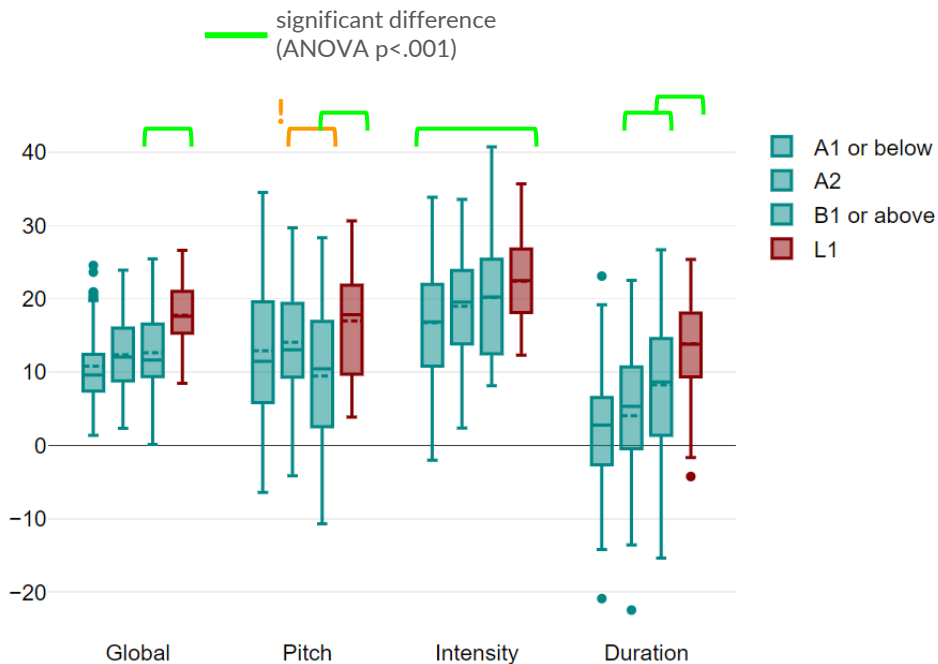


(resampling iteration 2)



(resampling iteration 3)

Results - Syllabic contrast

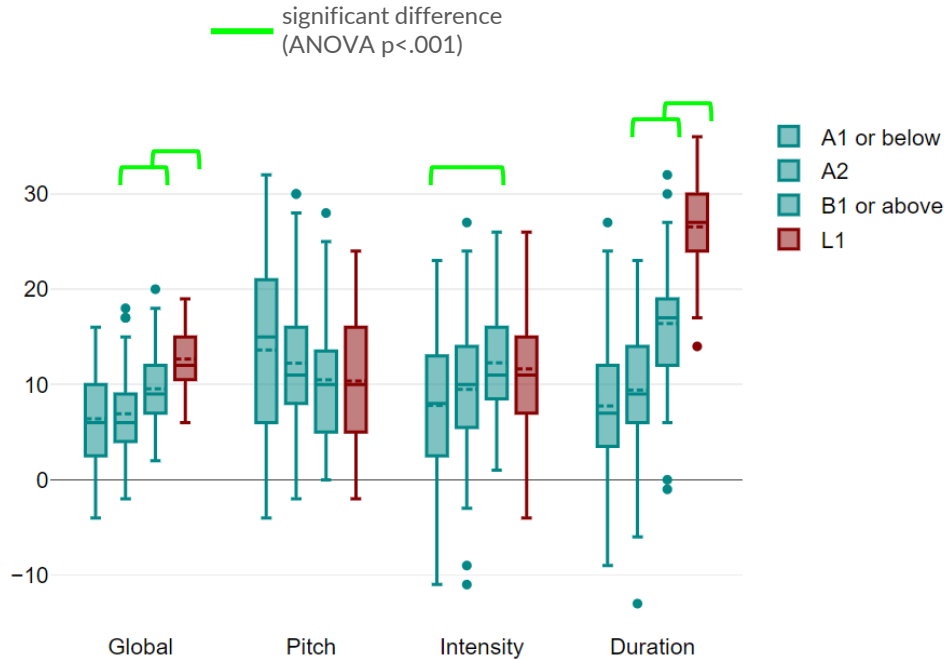


Random resampling to 55 recordings per group

Degree of acoustical prominence contrast between stressed and unstressed syllables within polysyllabic words

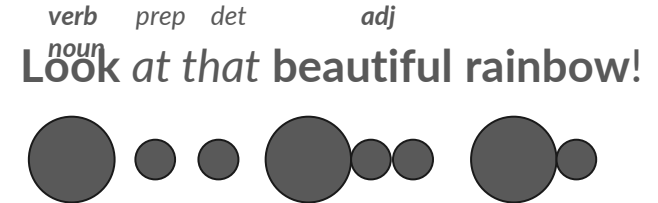
Speaker	Word	POS	Expected	Observed	Pitch	Energy	Duration
BaUSf01_21u06	academic	ADJ	ooOo	Oooo	●●●●●	●●●●●●●	●●●●●
BaUSf01_21u06	action	NOUN	Oo	Oo	●●	●●	●●
BaUSf01_21u06	ancient	ADJ	Oo	Oo	●●	●●	●●
BaUSf01_21u06	animals	NOUN	Ooo	Ooo	●●●	●●●	●●●
BaUSf01_21u06	animals	NOUN	Ooo	Ooo	●●●	●●●	●●●
BaUSf01_21u06	animals	NOUN	Ooo	ooO	●●●	●●●	●●●
BaUSf01_21u06	animals	NOUN	Ooo	Ooo	●●●	●●●	●●●
BaUSf01_21u06	another	DET	oOo	ooO	●●●	●●●	●●●
BaUSf01_21u06	appear	VERB	oO	oO	●●	●●	●●
BaUSf01_21u06	around	ADP	oO	oO	●●	●●	●●
BaUSf01_21u06	attacked	VERB	oO	oO	●●	●●	●●
BaUSf01_21u06	attitudes	NOUN	Ooo	Ooo	●●●	●●●	●●●
BaUSf01_21u06	camels	NOUN	Oo	oO	●●	●●	●●
BaUSf01_21u06	christmas	PROPN	Oo	Oo	●●	●●	●●
BaUSf01_21u06	circus	NOUN	Oo	Oo	●●	●●	●●
BaUSf01_21u06	circus	NOUN	Oo	Oo	●●	●●	●●
BaUSf01_21u06	circus	NOUN	Oo	Oo	●●	●●	●●

Results - Lexical contrast



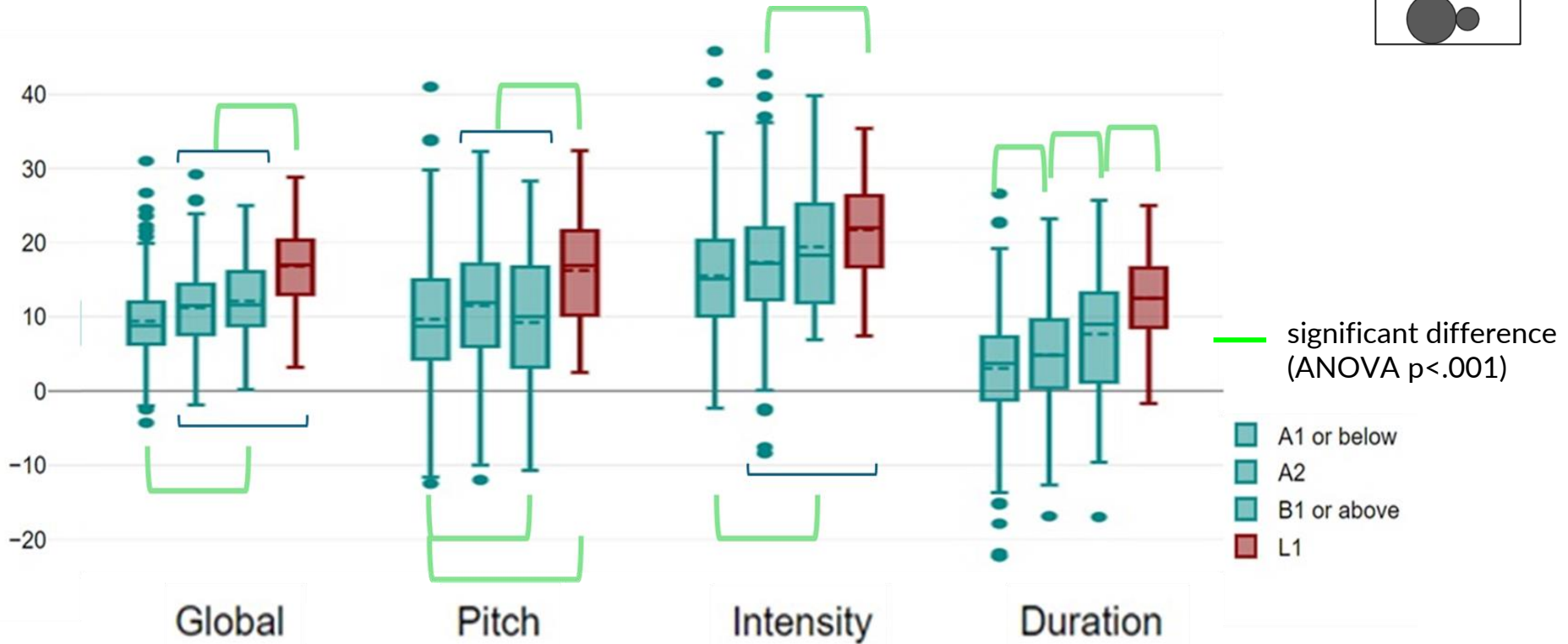
Random resampling to 55 recordings per group

Degree of acoustical prominence contrast between content and function words



Results - Syllabic contrast

Degree of acoustical prominence contrast between stressed and unstressed syllables within polysyllabic words



Results - Lexical contrast

Degree of acoustical prominence contrast between content and function words

Look at that beautiful rainbow!

