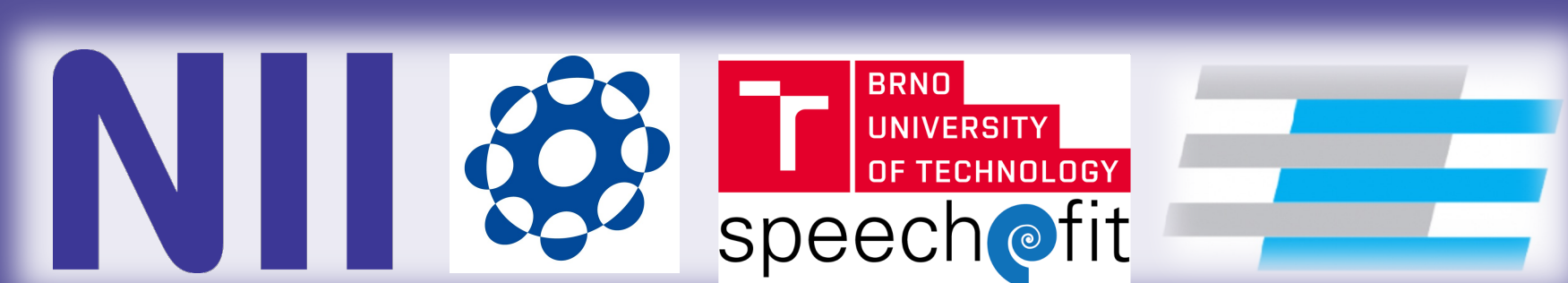


Spoof Diarization: “What Spoofed When” in Partially Spoofed Audio



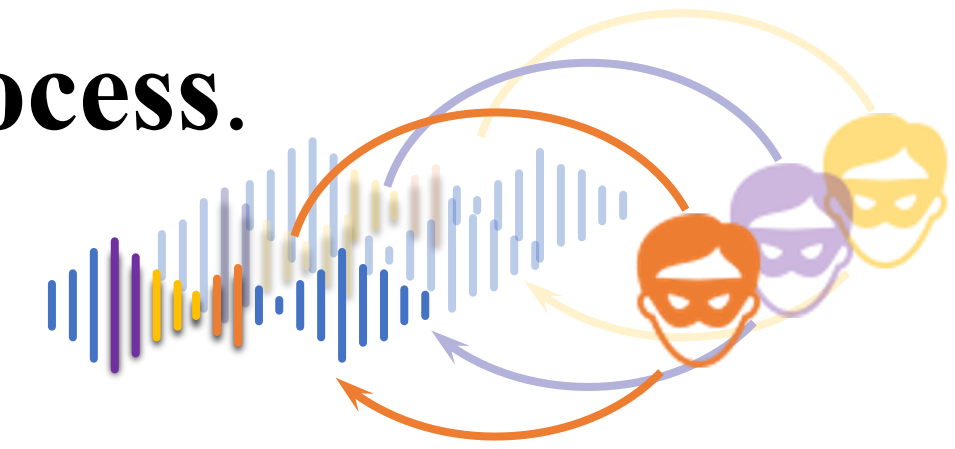
Lin Zhang^{1,2}, Xin Wang¹, Erica Cooper¹, Mireia Diez³, Federico Landini³, Nicholas Evans⁴, Junichi Yamagishi^{1,2}
¹NII, Japan ²SOKENDAI, Japan ³BUT, Czechia, ⁴EURECOM, France

New topic is waiting for you

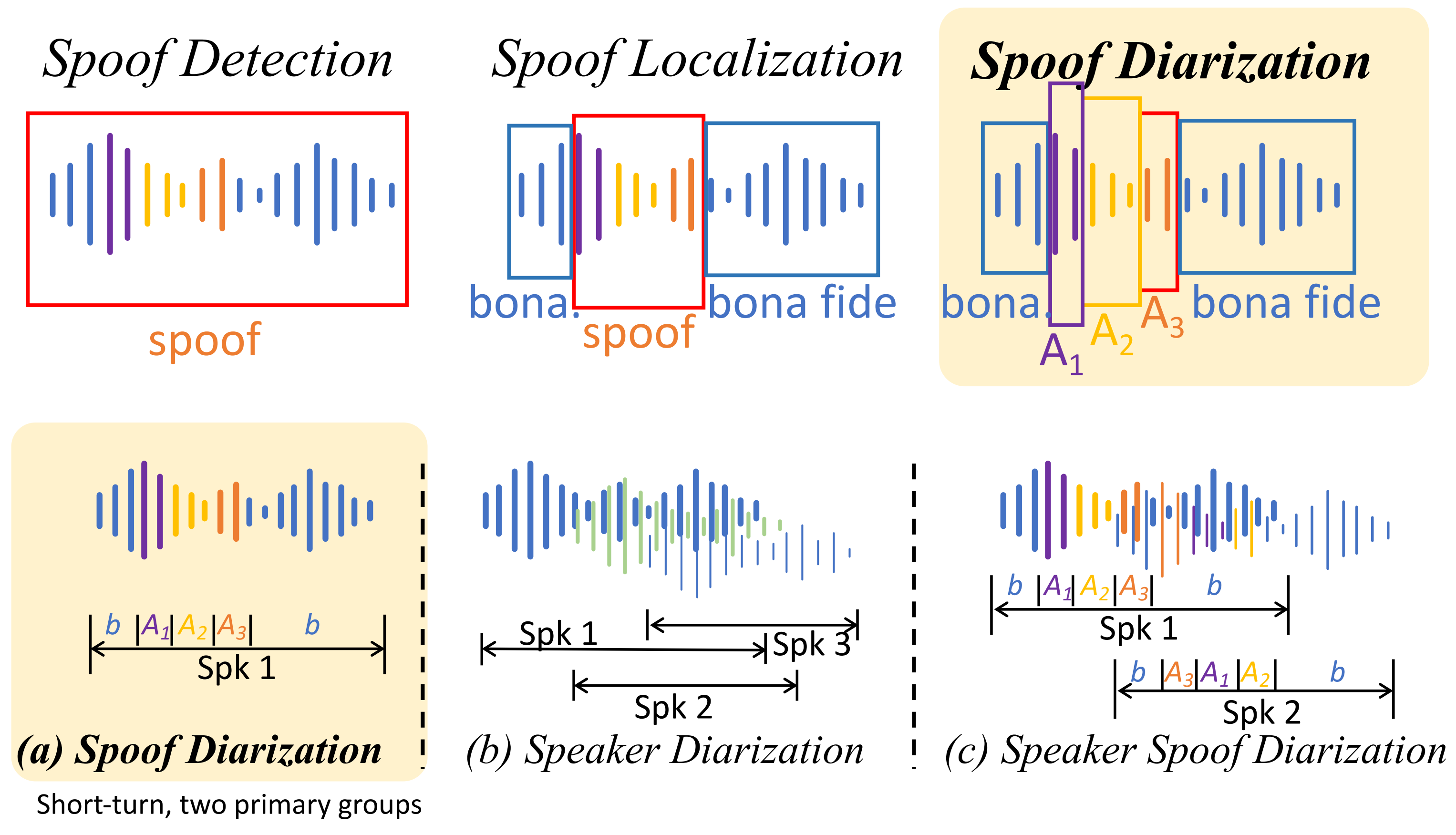
What is, and Why Spoof Diarization?

Definition & Motivation

- Audio can be altered by modifying part(s) of it,
- could through **multiple generative models** or through a **recursive generation process**.
- To obtain detailed information about spoofed region
 - cues for the spoofing algorithms,
 - tracing the origin or creator of the spoof.

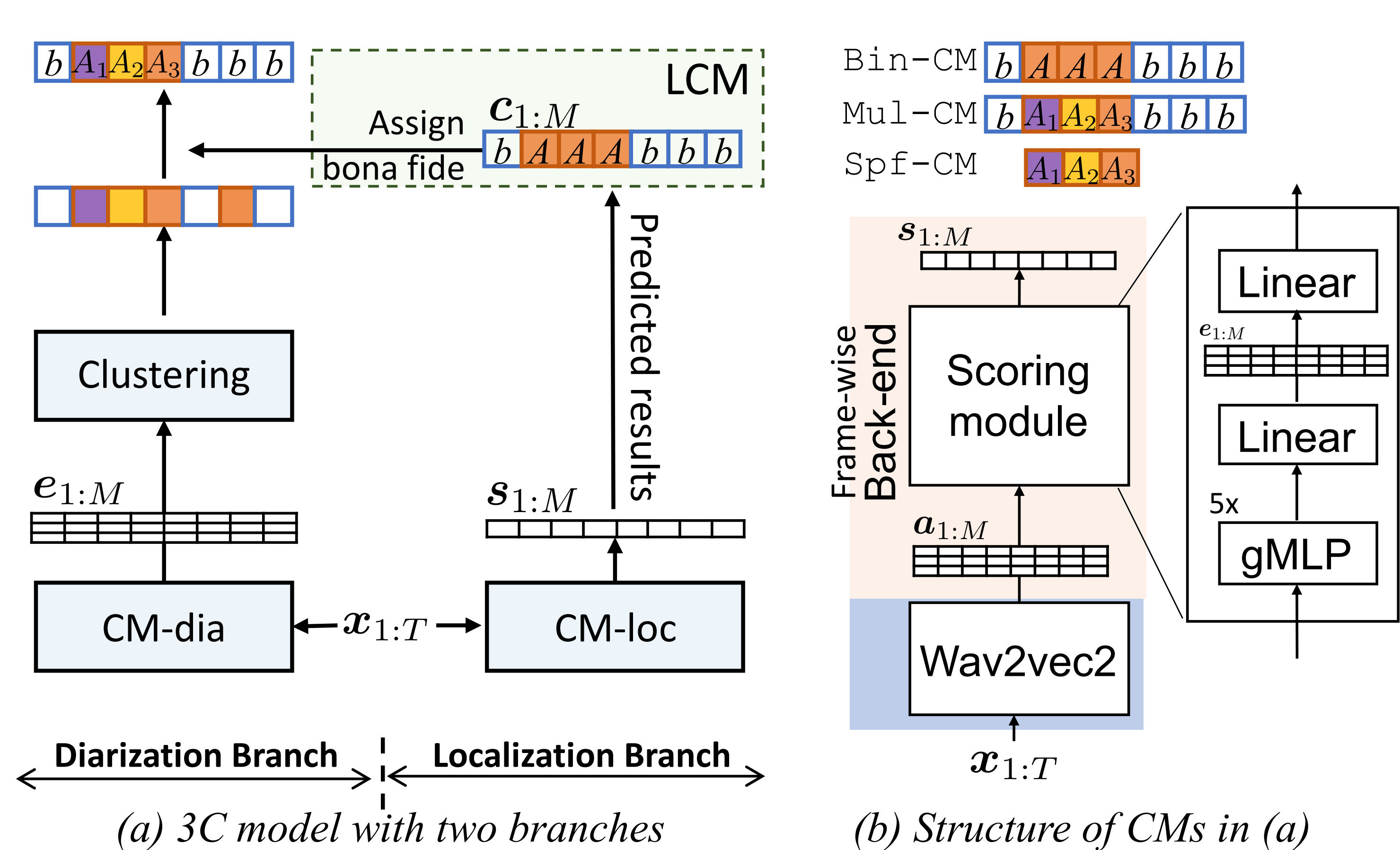


Spoof diarization can help with traceability!



Benchmark Model & Results

3C model: CM-Condition Clustering



Results on the PartialSpoof database.

Model		Evaluation set	
Dia.	Loc.	JI _{g.bona} (%)	JER _{g.spoof} (%)
	/	16.85	33.13
	/	19.66	28.05
	/	32.30	38.51
	Bin-CM	15.15↓	34.13
	Mul-CM	17.08↓	35.34
	Bin-CM	15.18↓	36.03↓
	Mul-CM	17.10↓	37.78↓

- w/o localization branch:
 - Bin-CM is specialized in **distinguishing bona fide** from spoof.
 - Specific treatment of the bona fide** class is needed during system training.
- w/ localization branch:
 - Although it negatively impacts diarizing spoofing methods on Mul-CM, **it brings improvements in most cases**.
 - Mul-CM < Spf-CM: Dia-Mul-CM **extracts better embeddings** for spoof diarization.

How to Measure Spoof Diarization?

- Diarization Error Rate (DER) vs. Jaccard Error Rate (JER)?

$$DER = \frac{FA + MD + CONF}{Total(speech)}$$

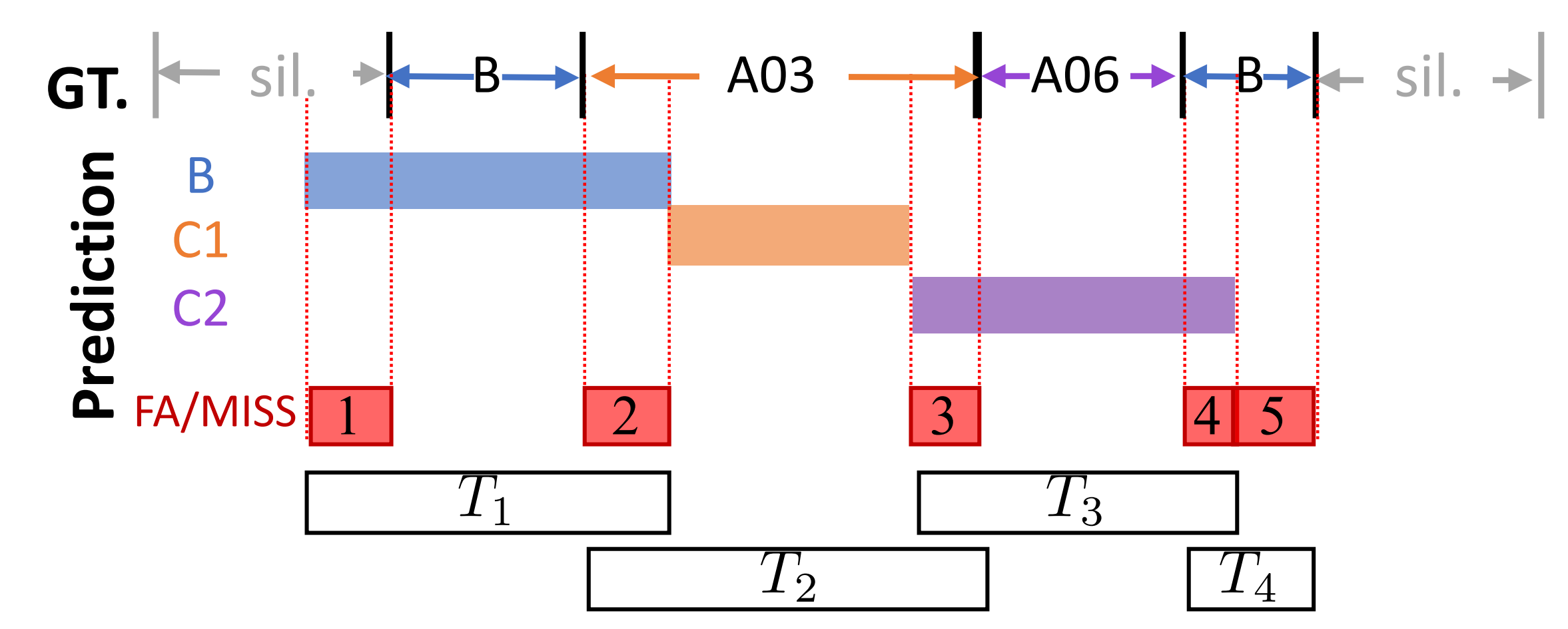
DER accounts for all errors with respect to the total duration of speech in recordings

Mistake on the **short turn** is also important for spoof diarization!

- JER
- For each class:
- JER gives **equal** weight to all classes, independently of their relative activity.

Spoof Jaccard Error Rate

- JI_{bona} : differentiating spoofed from bona fide segments
- JER_{spoof} discriminating different spoofing methods



$$JI_{global.bona} = \frac{1}{|D|} \sum_{j \in D} JI_{bona,j}$$

$$JI_{bona,j} = \frac{FA_{bona} + MISS_{bona}}{TOTAL_{bona}} = \frac{1+2+4+5}{T_1+T_4}$$

$$JER_{global.spoof} = \frac{1}{\sum_{j \in D} |A_j|} \sum_{j \in D} \sum_{A_i \in A_j} JER_{A_i,j}$$

$$JER_{spoof,j} = \frac{1}{2} \left(\frac{FA_{A03} + MISS_{A03}}{TOTAL_{A03}} + \frac{FA_{A06} + MISS_{A06}}{TOTAL_{A06}} \right)$$

$$= \frac{1}{2} \left(\frac{Null + 2+3}{T_2} + \frac{3+4+Null}{T_3} \right)$$

Spoof diarization, aiming to answer “what spoofed when,” is an **essential** task with **high complexity**. More methodologies are **waiting for your** exploration!



[2] Spoof Diarization: “What Spoofed When” in Partially Spoofed Audio. (This one)
 [1] Lin Zhang, Xin Wang, Erica Cooper, Nicholas Evans, Junichi Yamagishi, (2023) The PartialSpoof Database and Countermeasures for the Detection of Short Fake Speech Segments Embedded in an Utterance, in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 813-825, 2023, doi:10.1109/TASLP.2022.3233236.
 [0] Lin Zhang’s thesis: “Whether, When, What”: Detection, Localization, and Diarization of Partially Spoofed Audio ([2] is the latest version of the sec. 8)