# Language-based audio retrieval with textual embeddings of tag names

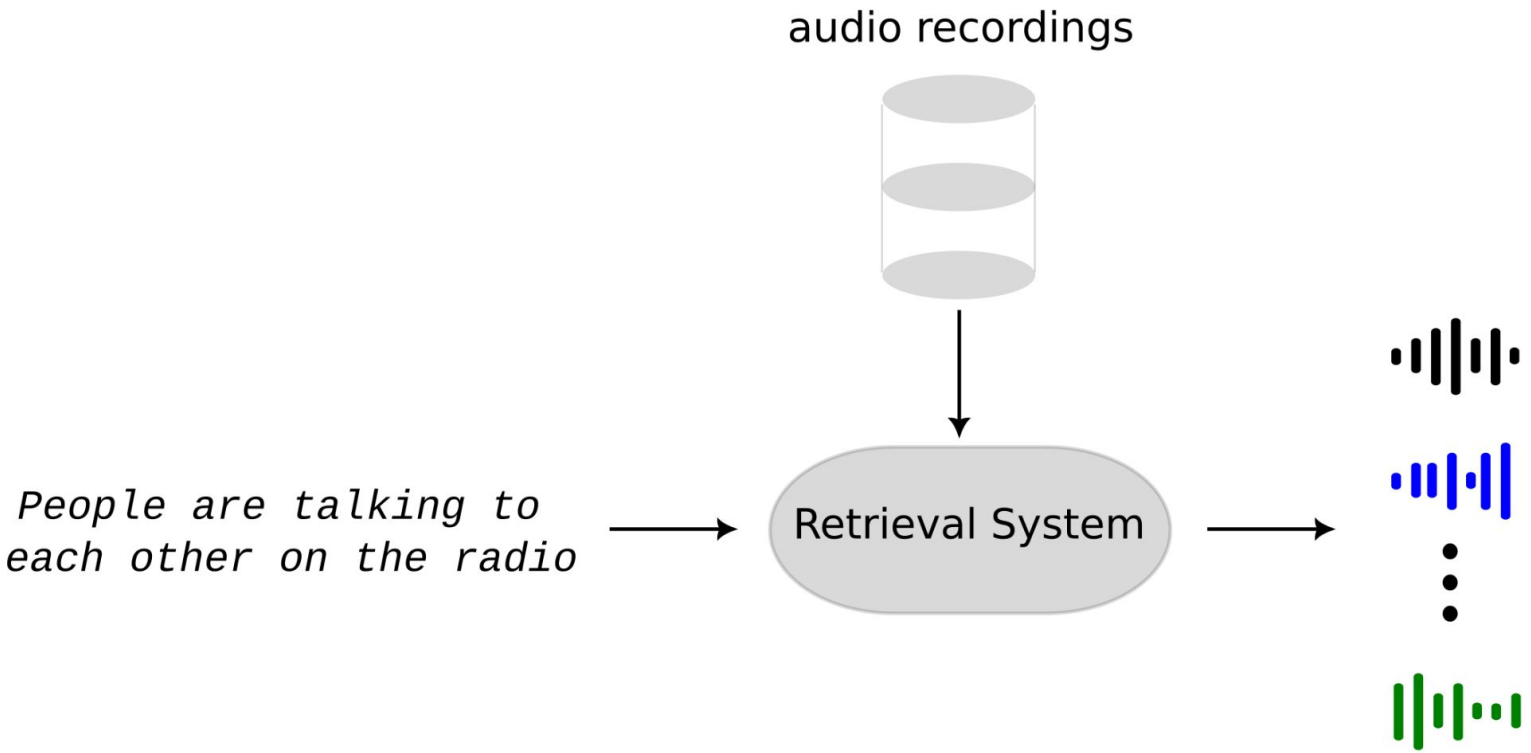
Thomas Pellegrini

thomas.pellegrini@irit.fr

Nancy, Nov. 4 2022



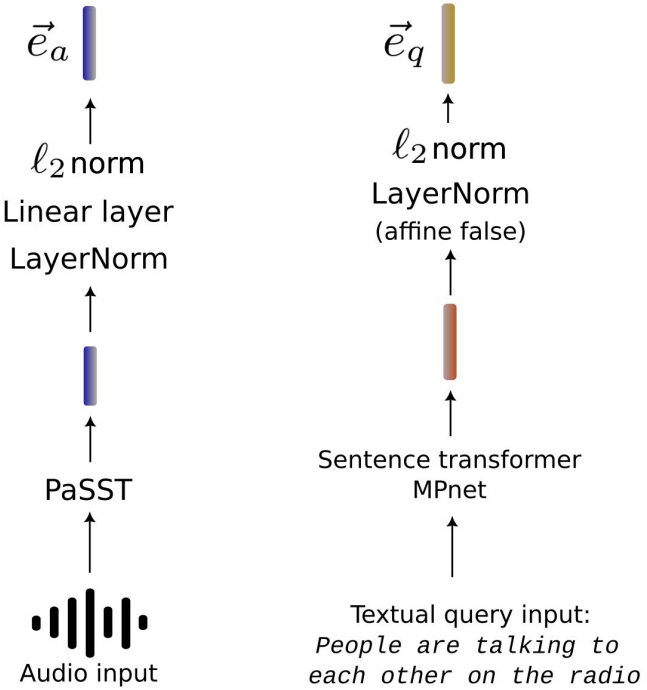

# What is Language-based Audio Retrieval?



# First system

Two frozen pretrained transformers, used once offline to precompute the embeddings

- Audio encoder: **AudioSet logits** obtained with PaSST [Khoutini *et al.*, 2022], **527-d**
- Text encoder: **sentence embeddings** with MPnet [sbert.net], **768-d**



# Adding the tag embeddings of the 527 AudioSet tags