Audio Summarization with Audio Features and Probability Distribution Divergence

Carlos-Emiliano González-Gallardo[1], Romain Deveaud[1], Eric SanJuan[1] and Juan-Manuel Torres-Moreno[1,2]

[1]*LIA - Avignon Université, Avignon, France*

[2]*Département de GIGL, Polytechnique Montréal, Montréal, Canada*

{carlos-emiliano.gonzalez-gallardo, eric.sanjuan, juan-manuel.torres}@univ-avignon.fr

romain.deveaud@gmail.com

Introduction

- Audio summarization of massive online multimedia resources.
- Facilitate the understanding.
- Extractive audio summarization approaches: { - textual methods, - audio features, - hybrid
- Our proposition:
  - Represent the information within the text in terms of its audio features.
  - Hybrid during training phase; text independent during summary creation.

Probability Distribution Divergence for Audio Summarization

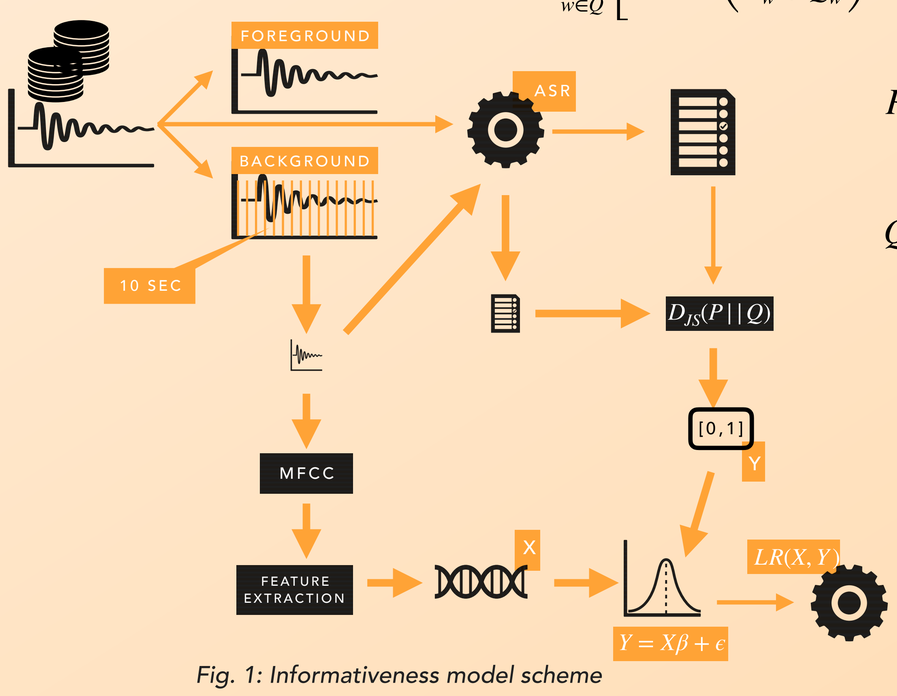
- Training phase:
  - Audio features (275 MFCC + 2) & textual information.
  - Mapping between audio features X and an informativeness value Y.

$$Y = D_{JS}(P||Q)$$

- Audio summary creation:
  - Audio features (275 MFCC + 2) ~~& textual information~~.
  - A score S_{Q_i} is computed to rank the pertinence of each segment $Q_1 \cdots Q_k$
  - Segments with higher S_{Q_i} scores are chosen until θ is reached.

Training phase (Informativeness model)

$$D_{JS}(P||Q) = \frac{1}{2}\sum_{w \in Q}\left[P_w \log_2\left(\frac{2P_w}{P_w+Q_w}\right) + Q_w \log_2\left(\frac{2Q_w}{P_w+Q_w}\right)\right]$$



$$P_w = \frac{C_w^P + \delta}{|P| + \delta \times \beta}$$

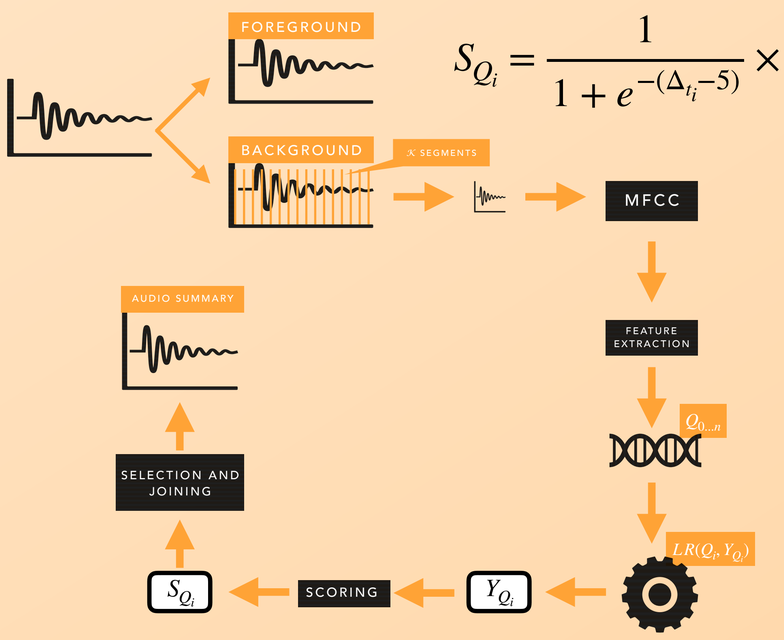
$$Q_w = \frac{C_w^Q + \delta}{|Q| + \delta \times \beta}$$

$$\beta = \frac{|V|}{2}$$

$$\delta = 0.0005$$

Fig. 1: Informativeness model scheme

Audio summary creation



$$S_{Q_i} = \frac{1}{1+e^{-(\Delta_{t_i}-5)}} \times \frac{|Q_i|}{|P|} \times e^{-\frac{t_{Q_i}}{\Delta_{t_i}}} \times e^{1-LR_{Q_i}}$$

$$\Delta_{t_i} = t_{Q_i+1} - t_{Q_i}$$

$$k = \frac{P_{length}}{60} \times 20$$

Fig. 2: Summary creation scheme

Experiments & Results

Training of Informativeness model

- 5,989 audio broadcasts (≈310 hours) in French, English and Arabic.
- Automatically obtained transcripts are treated with a stemming process.
- A linear least squares regression model is trained to map the audio features X of ≈111,600 training samples into an informativeness score Y.

Results

Score	Explanation
5	Full informative
4	Mostly informative
3	Half informateive
2	Quite informative
1	Not informative

Table 1: Evaluation scale

Sample	Length	Segments	Full Score	Average Score
1	3m19s	8	4.20	2.90
2	5m21s	13	3.50	2.78
3	2m47s	5	3.80	3.76
4	1m42s	5	3.60	2.95
5	8m47s	22	4.67	3.68
6	9m45s	30	4.00	2.49
7	5m23s	8	3.20	3.75
8	6m24s	20	3.75	2.84
9	7m35s	18	3.75	3.19
10	2m01s	4	2.75	2.63

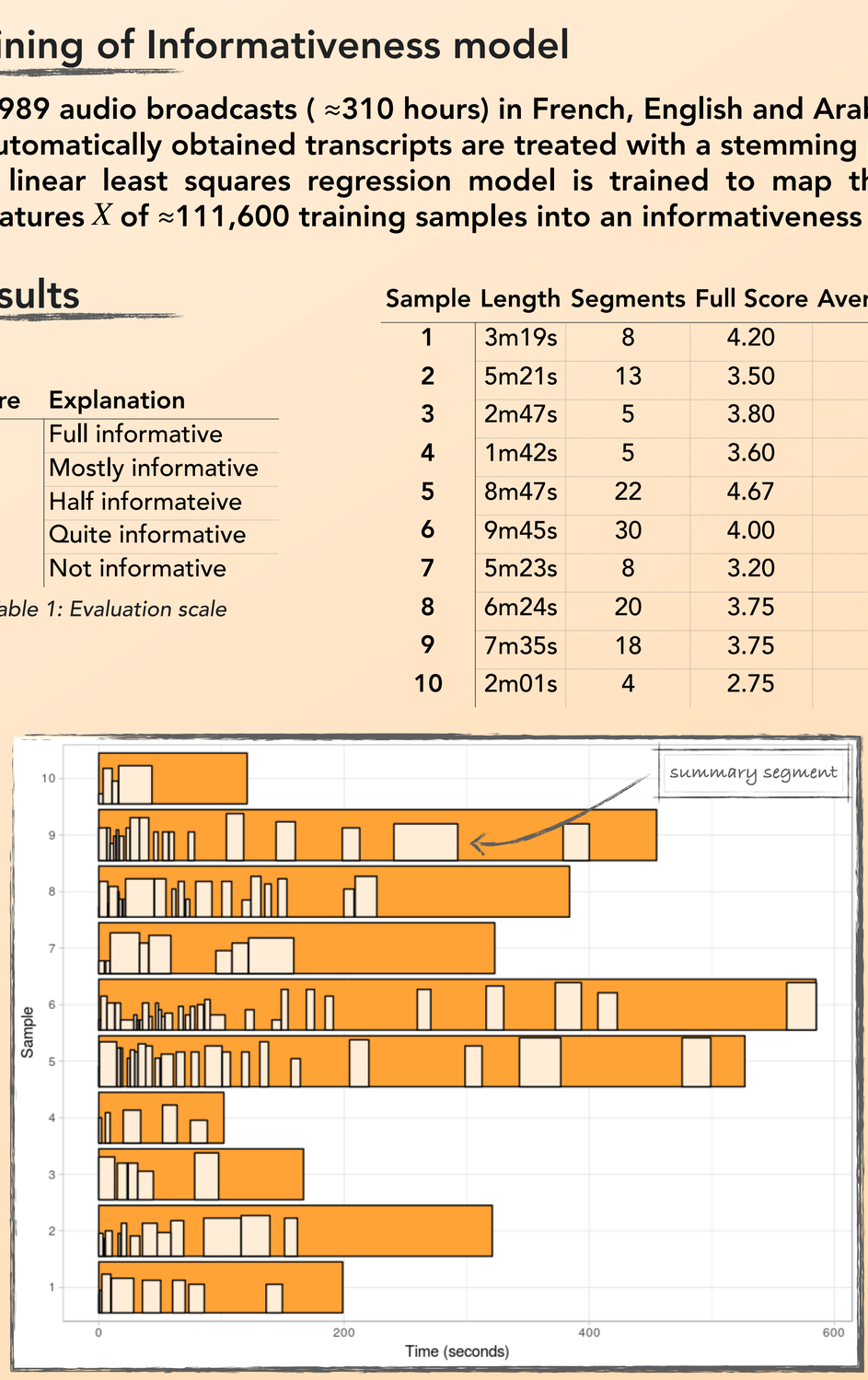


Fig. 3: Graphical representation of audio summarization performance

Conclusions & Perspectives

- Mapping informativeness from Mel-frequency Cepstral Coefficients (MFCC) features to their corresponding Jensen-Shannon (JS) divergence help to select those segments which are more relevant to the audio summary.
- Original approach; hybrid during training phase but text independent while creating summaries.
- It manages to generate at least half informative extractive summaries.
- Not a clear correlation between the quality of a summary and the quality of its parts.
- Future work will consider bigger evaluation datasets as well as French and Arabic summarization.

References

- Louis, A., Nenkova, A.: **Automatically evaluating content selection in summarization without human models**. In proceedings of EMNLP'09, (2009).
- Torres-Moreno, *et al.*: **Summary evaluation without references**. Polibits 42, 13-19 (2010).
- Rafii, Z., Pardo, B.: **Music/voice separation using the similarity matrix**. In proceedings of ISMIR'12, (2012).
- Jouvet, D., *et al.*: **Adaptation of speech recognition vocabularies for improved transcription of youtube videos.** ISGA 1(1), 1-9 (2018).

Acknowledgments

We would like to acknowledge the support of **CHIST-ERA** for funding this work through the **A**ccess **M**ultilingual **I**nformation opinion**S** (AMIS) project (France - Europe).

CICLING'19 - April 2019