

# Unity Is Strength: Coupling Media For Thematic Segmentation

Dalila Mekhaldi, Denis Lalanne and Rolf Ingold

Université de Fribourg , Chemin de musée 3, CH-1700 Fribourg  
{dalila.mekhaldi, denis.lalanne, rolf.ingold}@unifr.ch

**Abstract.** This paper presents the evaluation methods and the preliminary results of a combined thematic segmentation of (a) meeting documents and (b) meeting speech transcript. Our approach is based on a clustering method applied on a 2D representation of the thematic alignment, and then the projection of the extracted clusters on each axis, corresponding to meeting documents and the speech transcript. Finally, our bi-modal thematic segmentation method is evaluated, in regards to a mono-modal segmentation method (*TextTiling*).

## 1 Introduction

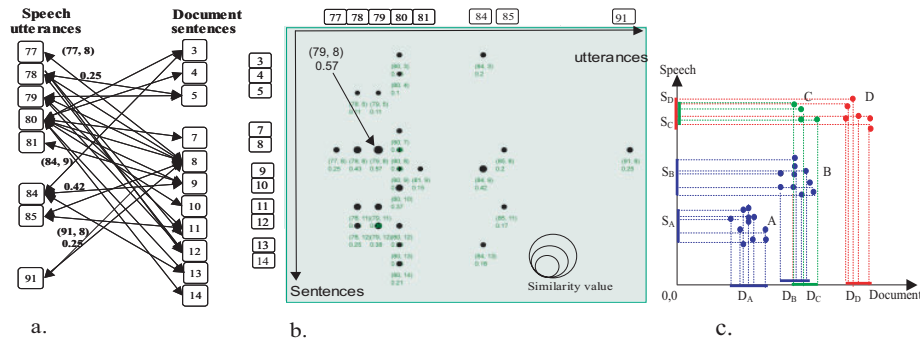
The document/speech thematic alignment and the thematic segmentation are closely related [4] [6]. The thematic alignment is building thematic links between documents units and speech units, which are semantically close. While thematic segmentation builds thematic links between units of a unique modality (document or speech). Thematic segmentation is thus an intra-modal segmentation, while thematic alignment is an inter-modal segmentation. Since the preliminary evaluation we have performed on state-of-the-art, thematic segmentation methods did not show good results, our assumption is that an inter-modal segmentation will be more efficient and will benefit from the various modalities information. In this article, we present briefly our bi-modal thematic segmentation method and its projection to each modality. A preliminary evaluation shows that our bi-modal segmentation is more efficient than a mono-modal segmentation.

## 2 Thematic alignment vs Thematic segmentation

Our document/speech alignment takes as input both the speech transcript of a meeting and the documents related to the meeting, and it generates a set of aligned pairs (document units, speech units)[4][6]. In our work, we are focusing on press reviews, where many speakers discuss a daily newspaper cover page. The information contained in the documents, in PDF form, is first extracted and then automatically converted into a multi-layered structure (layout and logical structure mainly) [2]. On the other hand, the speech is currently manually transcribed.

All the similar units in the target file (document or speech) for a given unit from the source file are selected (figure 1.a). This thematic alignment, which is a symmetrical relationship between document and speech units, can be represented by a 2-dimensional graph, where each dimension represents a distinct modality (figure 1.b). Each node in this representation is a relationship between the document and speech units (e.g. utterance 79 with sentence 8 has a similarity value of 0.57), and the node size represents their similarity value.

Using this 2D representation, a clustering process based on an improved *K-means* method [5] has been applied in order to bring to light the denser regions, that we believe that may represent the various topics of the meeting. This clustering method was enriched by a filtering step of the weak densities clusters, by considering the clusters size, the nodes weights and distances (e.g. Euclidean distance) from the clusters centroids. Once the denser regions are computed, they are projected on each axis in order to highlight the mono-modal thematic segments. In figure 1.c the cluster A corresponds to a document segment  $D_A$  and a speech segment  $S_A$ .



**Fig. 1.** a. Bi-graph representing the *k-best* alignment b. 2D representation of the bi-graph c. Clusters projection.

## 2.1 Experimental results

In order to evaluate our method, many metrics have been used, in respect to a prepared manual ground truth: the *entropy*, the *purity* and the  $P_k$  (Beeferman) metric [1]. The *entropy* measures the disorder of segments with an ideal value of 0. On the other hand, the *purity* measures the fraction of generated segments that don't contain incorrectly placed objects. The other metric used in this evaluation, the  $P_k$  metric measures the probability that a randomly chosen pairs of units at a distance of  $k$  units apart are inconsistently classified, i.e. if these two sentences are correctly labeled as being related or being unrelated. This metric is more adequate than a simple *recall* and *precision* that measures just

the boundaries detection. For this experiment, the  $k$  parameter has been fixed to 4 units, which corresponds to the minimum size of a relevant thematic segment.

Metrics	Document								Speech							
	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$	$D_8$	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$
Entropy	.14	.14	.14	.14	.38	.17	.23	.36	.25	.34	.33	.31	.20	.16	.13	.15
Purity	.82	.74	.82	.82	.60	.78	.67	.64	.78	.69	.68	.67	.79	.81	.85	.85
$P_4$	.41	.31	.38	.32	.54	.25	.43	.40	.36	.35	.39	.46	.42	.33	.43	.42

**Table 1.** Documents/Speech thematic segmentation evaluation.

Methods	Document								Speech							
	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$	$D_8$	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$
Bi-modal	.41	.31	.38	.32	.54	.25	.43	.40	.36	.35	.39	.46	.42	.33	.43	.42
Mono-modal	.74	.54	.69	.58	.59	.54	.52	.61	.54	.32	.45	.47	.77	.43	.65	.79

**Table 2.**  $P_k$  evaluation of a bi-modal segmentation, comparing to a mono-modal.

Table 1 shows the evaluation of the thematic segmentation of 8 meetings documents and speech transcripts. The generated *entropy* and *purity* values depend on the type of the meeting. Thus we distinguish two types:

1. If the speakers do not follow the linearized documents reading order, then the temporal indexes of the document segments are not adjacent. This reduces the number of overlapped segments, and as a result, it gives the satisfactory values for the *entropy* and *purity* (e.g. documents  $D_1$ ,  $D_2$ ,  $D_3$  and  $D_4$ ).
2. If the meeting is non stereotyped, i.e. speech with numerous debates, then there is less overlapped segments (e.g. speech transcripts  $S_6$ ,  $S_7$  and  $S_8$ ). This is due to the fact that the speech segments are well separated each one from the other. As results, their *entropy* and *purity* values are better, comparing to stereotyped meetings.

The  $P_k$  evaluation is generally satisfactory, especially in comparison to the *TextTiling*[3] method (see Table 2). Our bi-modal clustering method is more accurate in detecting the exact number of thematic segments, which is not the case for the *TextTiling* method that generates many extra segments.

## 2.2 Remarks

During the segments extraction process, overlapping problems often occurred. This kind of problems happens when a unit is assigned to many segments, and

it mainly appears in stereotyped meetings. The relationship between the overlapped segments can be one of two types: either one of them contains the other (e.g. in the figure 1.c,  $S_D$  contains  $S_c$ ), or they are partially overlapped (e.g.  $D_B$  with  $D_C$ ). Our contribution in resolving this problem is under work, and is based on the use of Gaussian probabilistic function. First, an overlapping coefficient is computed. Depending on this coefficient value, the corresponding segments are merged, or considered as two distinctive segments, using the Gaussian probabilistic. Other works are planned in order to improve this bi-modal thematic segmentation, such as the integration of the nodes weights in the clustering method, while computing the clusters centroids then while assigning the nodes to the clusters.

### 3 Conclusion and Future work

The current paper presents the results of the evaluation of a bi-modal thematic segmentation method, based on a preliminary thematic alignment of meetings documents with speech transcripts. The comparison of this method with a mono-modal method, i.e. *Text Tiling* method, shows promising results, despite the overlapping problem that affects the segmentation, and should be resolved. The segmentation quality can be improved by considering the nodes weights earlier in the clustering process. Other prospects are foreseen, such as the combination with other alignments, for instance the speech turns with the documents logical units, or references to documents in meeting dialogs, citations, etc. In a long term, we plan to integrate all the various types of alignments in a single framework.

This preliminary evaluation makes us believe that coupling modalities, in this meeting documents and speech transcripts, should considerably improve each involved modality segmentations.

### References

1. Beeferman D., Berger A., Lafferty J. (1999), Statistical Models for Text Segmentation, Machine Learning, Vol. 34(1/3), pp. 177-210.
2. Hadjar K., Rigamonti M., Lalanne D. and Ingold R., Xed: a new tool for eXtracting hidden structures from Electronic Documents. DIAL 2004, Palo Alto, California, January 2004.
3. Hearst M., Multi-Paragraph Segmentation of Expository Text, In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics 1994, Las Cruces, New Mexico.
4. Lalanne D., Mekhaldi D. and Ingold R. Talking about documents: revealing a missing link to multimedia meeting archives. Document Recognition and Retrieval XI, IS-T/SPIE's International Symposium on Electronic Imaging 2004, USA
5. Looney C., Interactive clustering and merging with a new fuzzy expected value, Pattern Recognition, August 2002.
6. Mekhaldi D., Lalanne D. and Ingold R., Thematic Alignment Of Recorded Speech With Documents, Proceedings of the 2003 ACM symposium on Document engineering, France