Thematic Segmentation of Meetings through Document/Speech Alignment

Dalila Mekhaldi Denis Lalanne Rolf Ingold

DIVA/DIUF
Chemin du musée, 3
CH-1700 Fribourg, Switzerland
+41 26 429 66 78 (65 96)

{Dalila.Mekhaldi, Denis.Lalanne, Rolf.Ingold} @unifr.ch

ABSTRACT

This article proposes a multimodal approach for segmenting meeting recordings. This bi-modal method takes advantages of the alignment of speech transcript with documents, in the context of meetings or lectures, where documents are discussed. The method first displays the alignment results as a set of nodes in a 2D space, where the two axes represent respectively the documents content and the speech transcript. The most connected regions in this graph are detected using a clustering method. The final clusters are then projected on the speech axis. Finally, the obtained sequence of segments is considered as the thematic structure of the speech transcript. In this article, we present our bi-modal method and compare it with two other mono-modal thematic segmentation methods.

Categories and Subject Descriptors

H.3.1 **[Content Analysis and Indexing]**: indexing methods; H.3.3 **[Information Search and Retrieval]**: Clustering; Search process; I.7.5 **[Document Capture]**: Document analysis; I.5.3 **[Clustering]**: Similarity measures

General Terms

Algorithms, Measurement, Performance, Experimentation.

Keywords

Multimedia information retrieval, multimodal thematic alignment, thematic segmentation, clustering techniques, document analysis, meeting dialogs structuring.

1. INTRODUCTION

The thematic alignment between printable documents and other media data (audio, video) appears as an important and necessary step for a full understanding of meeting dialogs. Since multimedia data are time dependent, and not documents, it is necessary to bridge a temporal link between them (see Figure 1). The speech transcript contains timestamps for each speech utterance and each speaker turn. Thus, matching documents content with speech transcript can enrich documents with temporal indexes. Further it can synchronize documents with other medias, sharing the same meeting time. Our previous researches focused on the thematic alignment of meeting documents with the transcription of the speech [9] [13], which bridges temporal links between documents and speech transcript. Further, it facilitates the synchronization between all the meetings data, and improves their indexing and retrieval. In particular, it can help answering questions such as: *"When was a specific document or document part discussed?"* and *"What was said about it?"*, *"What was the document being discussed at time T?"* or *"What are all the documents related to the document discussed at time T?"*, etc. Further, documents being highly thematic, we believe that segmenting meetings according to documents' parts will lead to a robust topic segmentation of meetings, especially if the meeting dialogs are centered around documents.

The current paper demonstrates the close link that exists between the document/speech thematic alignment and the speech thematic segmentation. The results of our segmentation method revealed that the more modalities are used for segmenting, the more robust the segmentation is. Indeed, the evaluation section shows that our bi-modal segmentation method performed better in comparison to other standard mono-modal segmentation methods.

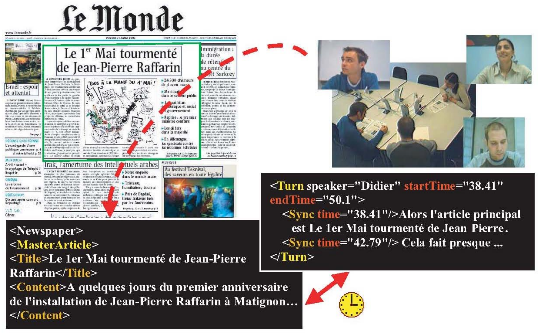
The paper is organized as follow, in section 2, our meetings data set is presented, then in section 3 a brief description of the thematic alignment process is given, which is the basis of our thematic segmentation method. In section 4, a short state-of-the-art of the existing thematic segmentation methods is presented. Finally in section 5, we present our bi-modal segmentation method, with the obtained evaluation results, compared to other mono-modal methods.

2. EXPERIMENTAL DATA SET

Our meeting room is equipped with 8 camera/microphone pairs (one pair for each participant in the meeting), a video projector, a camera for the projection screen capture and several cameras to capture documents on the table.

MM'04, October 10–16, 2004, New York, New York, USA.

At the time of writing, about 30 meetings have been recorded. The research presented in this article is focusing on press review meetings. About 22 French press review meetings have been recorded for that purpose. In each meeting, between 3 to 6 speakers discuss, during 15 minutes, about various French daily newspaper cover pages, which contain many small heterogeneous articles. Other document types (e.g. agenda, slides, etc.) will be considered in the future. The information contained in the documents, in PDF form, and archived before the meeting, is first extracted and then automatically converted into a canonical multi-layered structure containing text, layout and logical structure mainly [5]. On the other hand, the speech is currently manually transcribed. In the near future, we plan to test our alignment techniques with automatic transcriptions, in which the WER (word error rate) is important.


Le Monde
Le 1er Mai tourmenté de Jean-Pierre Raffarin
Immigration : la durée de rétention au centre du projet Sarkozy
Israël : espoir et attentat
TOUS À LA MANIF DU 1er MAI !
Irak, l'amertume des intellectuels arabes
Au festival Teknival, des raveurs en toute légalité

```
<Newspaper>
<MasterArticle>
<Title>Le 1er Mai tourmenté de Jean-Pierre
Raffarin</Title>
<Content>A quelques jours du premier anniversaire
de l'installation de Jean-Pierre Raffarin à Matignon…
</Content>
```

```
<Turn speaker="Didier" startTime="38.41"
endTime="50.1">
   <Sync time="38.41"/>Alors l'article principal
     est Le 1er Mai tourmenté de Jean Pierre.
   <Sync time="42.79"/> Cela fait presque ...
</Turn>
```

Figure 1. Thematic linking between documents and audio/video meeting data.

3. THEMATIC ALIGNMENT

The thematic alignment of printable documents with meeting dialogs is a complex process, since it highly depends on the quality of the various segmentations of both document and speech transcript into a set of units. Our alignment process can be described as follow [9]: first the document is segmented into syntactic and logical structure, and the speech transcript into utterances and turns. The alignment links can then be of three kinds, either a citation, a reference in the speech to a document part, or a thematic similarity between documents and speech transcript units. Our work is focusing on the thematic alignment, in which various similarity methods are used in order to find similarity links between the documents and the speech transcript units. After proper stop-words removal and stemming, both documents and speech transcript units, represented as vectors of weighted terms (e.g. U1, U2), are compared using similarity metrics. These metrics count the co-occurrences of terms, in respect to their respective weights, and compute a similarity distance. They can be described as follow:

- Cosine = $|U1 \cap U2| / \sqrt{(|U1|.|U2|)}$
- Jaccard = $|U1 \cap U2| / |U1 \cup U2|$
- Dice = $2 * |U1 \cap U2| / (|U1| + |U2|)$

The relevant pairs of document/speech units were then chosen according to two strategies: the *best-one* and the *multiple* alignments strategy.

In the *best-one* strategy, the best speech unit (resp. document) was computed for each document unit (resp. speech), i.e. the target unit that returns the maximum similarity value. This alignment strategy is thus oriented. For this simple alignment method, we compared the three similarity metrics presented above [9]. The *Jaccard* method gave back the most promising results, on a recall and precision basis, when the units to compare have a similar size (utterances/sentences: recall 076, precision 0.84 and sentences/utterances: recall 0.64, precision: 0.80). On the other hand, the *Cosine* method performed better for the remaining pairs (sentences with turns, utterances with logical blocks, etc.). The best results were obtained with Cosine when matching speaker turns with document logical blocks (recall: 0.84, precision: 0.85). Finally, *Dice* was always below the two other metrics. For this reason, we used a combination of *Cosine* and *Jaccard* in order to compute a robust similarity distance. More details on this evaluation are given in [9].

In the *multiple* alignment strategy, all the best matches, whose similarity value overcomes a defined threshold, are retained. This last strategy thus generates symmetrical alignment results, i.e. the results obtained from the document to the speech transcript are the same than the ones obtained in the other direction. We will see later that this symmetrical property is important for our bi-modal segmentation method.

4. THEMATIC SEGMENTATION

The thematic segmentation, i.e. the decomposition of a given text into topics or homogeneous segments, has been the subject of many research works. Salton and al. [16] used a text relationship map that establishes similarity links between the text excerpts (sentences or paragraphs), which are represented as nodes in the map. In order to define textual thematic segments, all the triangles are located in the full relationship map. Many triangles can be merged when the similarity between their corresponding vectors centroids exceeds a defined threshold. This map provides information about homogeneity of the text, so that if there are many links between adjacent paragraphs, this proves the homogeneity treatment of topics.

Hearst's *TextTiling* method [6] divides the text into tokens; i.e. individual lexical units. The adjacent pairs of blocks, i.e. sequences of tokens, are compared using a similarity method. The topics boundaries are then defined according to the change in the sequence of similarity scores.

Ferret has used similar method based on the boundaries detection and on similarity measure between adjacent units [3]. Ferret's method is enriched with a lexical co-occurrence network built from a large corpus. This work reinforces the descriptors (vectors representing the units) by linking words that have semantic relationships.

5. METHODOLOGY

5.1 Thematic alignment vs. segmentation

The thematic segmentation of spoken dialogs is still a difficult task that has not yet been completely solved [3][6][16]. In the thematic alignment process described in section 3 and [9][13], the thematic segmentation of the resources being aligned could have improved our results. However, we tested various state-of-the-art methods, and we did not obtain satisfactory results. At this point,

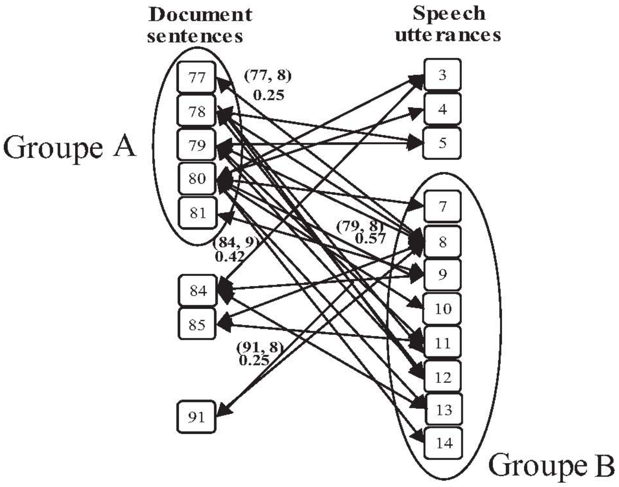


Figure 2. A bi-graph representing multiple alignments.

we made the hypothesis that the document/speech alignment and the thematic segmentation of meeting dialogs are highly related, and thus the good alignment results obtained could help getting a drastically better thematic segmentation of both documents and meeting dialogs. This is why we have tried to deduce some ideas from the preliminary results of this thematic alignment, or more particularly from the *multiple* alignment. The *multiple* alignment as it was defined in section 3 considers all the best alignments, i.e. all the similarity links that overcome a prefixed threshold, between document's units and speech transcript's unit.

Looking at Figure 2, which is a visualization of the generated alignment results, each unit is represented by a node, documents units on the left of the bi-graph, and speech transcript units on the right. The similarities between these units are represented by edges between their nodes, where the edge weight value represents the alignability value (e.g. sentence 79 with utterance 8 has a similarity value of 0.57).

The generated graph is a bi-graph, since each node belongs to one of the two modalities. When analyzing this bi-graph, it appears that some regions are denser than others (e.g. group A and group B on Figure 2). Nodes on each side of the bi-graph are respecting their appearance order, spatial order or adjacency for the document units (e.g. sentences), and temporal order for the speech units (e.g. utterances). For this reason, the denser regions can be explained by the fact that a group of successive units from the document is thematically linked to a group of successive units from the speech transcript (e.g. group A with group B). Our hypothesis is that groups A and B may share the same theme and may represent respectively the thematic segments of the document and the speech transcript. Since an elementary unit (e.g. a sentence or an utterance) very rarely belongs to two different thematic regions, we decided first to use the alignment results between sentences and utterances, which are respectively the smallest units for documents and speech transcript.

Using the bi-graph, presented in Figure 2, the first step in our bi-modal thematic segmentation is to extract the densest regions in the graph, which can be obtained by isolating them from the entire graph. The second step consists in the extraction, by projection, of the corresponding thematic segments for the speech transcript. Finally the obtained results are evaluated. Figure 3 illustrates the overall structure of our method for extracting the thematic structure of the meeting dialogs.

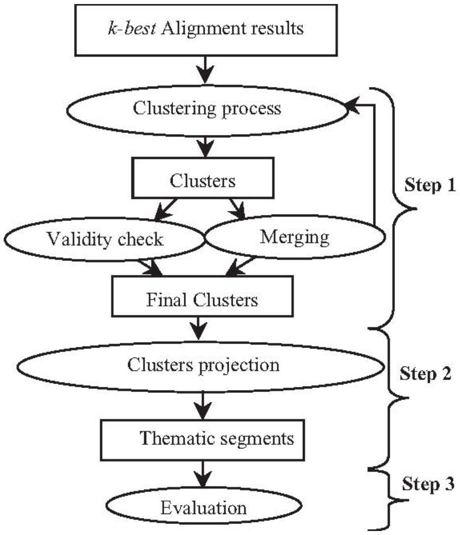


Figure 3. The thematic segmentation process.

5.2 Densest regions extraction

Our first attempt to solve this thematic regions extraction problem was the use of the intersection graphs [4], which were generated by the projection of the bi-graph on each side. An intersection graph for the document and another for the speech transcript, were thus generated. These intersection graphs group each two units from one source, which are related to a same intermediate unit from the second source. At first, we thought that edges weight could be fortified by detecting the intersection graphs until a fixed nth level, which means: from a source vertex v1, we have to go through n intermediate vertexes from the other source before reaching the target vertex v2. Unfortunately, we noticed that this method was not efficient. The main reason is that in the generated intersection graph, all the nodes were related. Another reason for abandoning this representation is that it does not contain all the information available in the original alignment, i.e. the information offered by the thematic links to the other part of the bi-graph, which can be useful in the segmentation process.

Therefore, we looked at a more suitable solution in the clustering field. Our two resources are represented on two axes: X and Y, where the identifiers of their respective units are represented by the X values and Y values, and the alignment links between these units is represented by nodes in this 2D graph (see Figure 4), so that the grouped regions in Figure 2 are represented by clusters of nodes (e.g. cluster AxB). The weight value (i.e. the alignability) of each link is represented by the node sizes; a big node has a larger weight than a small node.

As we can see in Figure 4, this 2D illustration is a complete representation of the alignment data, since all the alignment information is plotted: spatial attributes or adjacency are represented by the X values, temporal attributes by the Y values, and the alignability values are represented by the nodes' size.

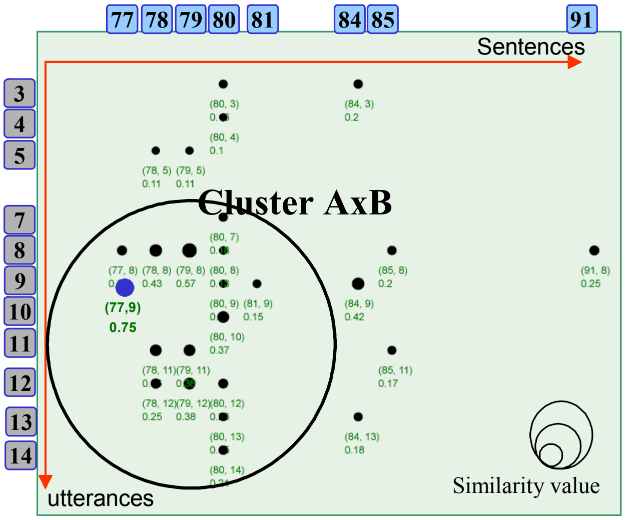


Figure 4. 2D representation of the alignment results.

Thus, much information can be deduced from this representation such as: in which temporal order the meeting was played, if there are overlapping of topics, etc. This information helps us not only to detect the thematic segments of the speech transcript, but also to detect the temporal links between the segments of the document. Now that the problem of thematic segmentation has been transformed into a clustering problem, as it is known in this field [7][18], many relative attributes of the object being clustered have to be considered, such as the spatial attributes of the document and the temporal attributes of the speech units. Even if many clustering methods are available [7][15], we have chosen the most standard one in order to bootstrap our method, which is the original K-Means method [12] and its improved version [11].

5.2.1 K-means algorithm

Our objective is to identify the denser clusters, and the most separated ones from the others. For this reason, the original K-Means method is the first method that had been used. However, this method has many drawbacks [11]:

1. The K value, number of clusters, must be fixed at the beginning, which is not easy since the number of clusters changes from a graph to another, which is the major disadvantage of this method.
2. The internal criteria are not considered:
    - The compactness of clusters, taking into account the distance between the centroids and the nodes of the clusters,
    - The density of the clusters, taking into consideration not only the distance to the centroid but the nodes weights and their number too.
3. The external criteria are not considered:
    - The distance between the clusters centroids.
    - The average cluster density in case of overlapping.

For all these reasons, the application of this version of the K-Means algorithm is not sufficient. In [11] , an improved version is presented, which considers the mentioned internal and external criterion, but it does not consider neither the node weights nor the clusters density. In this K-Means version, many thresholds that must be initialized by the user, are considered:

1. Defining the K centroids randomly is always followed by a merging method (linking process), which checks the non-closeness of generated centroids; otherwise it merges the clusters with closest centroids. This merging task is based on the distance value between the centroids, using a defined threshold.
2. The number of nodes in a cluster is significant, so that the minimum value needs to be defined by the user, by observing a sample of clustering. In our work we have fixed this threshold to 2 nodes per cluster.

Having these two rules, it is more practical to define a large K value, since it decreases if there are close or non-significant clusters.

3. A third parameter is considered: the clustering validity measure. The convergence of this clustering method to the best result is measured by the variance formula XB [17]. This measure checks the change of the centroids positions.

With these additional parameters, the clusters internal criteria (i.e. compactness) and external criteria (i.e. well separated each one from the other), as well the validity indexes are well considered.

Using the improved K-Means method, our clustering process takes as input, a vector of nodes representing the bi-graph. It associates to each node 3 information, the identifier in the document (s_i) and the identifier in the speech transcript (u_j) and the weight w (similarity value). And as output, it generates a set of clusters. The clustering process is repeated until the clusters centroids reach a stable state. Currently, the clustering of the (s_i,u_j,w) vectors is only based on the Euclidean distance between the spatio-temporal components (s_i,u_j). The weight w will be considered in the near future and hopefully shall improve the clustering.

Since the improved K-Means algorithm does not consider the nodes weights, we have implemented an enhanced method that filters the weakest clusters in regards to their density. The density of each cluster is based on its nodes' weight, nodes' number, and nodes' Euclidean distance from the final clusters' centroids. This way, a cluster with a given number of nodes in a large surface, is less significant than a cluster with the same number of nodes but in a smaller surface. In the same way, a cluster with heavy nodes is more significant than a cluster having the same number of nodes but with lighter weights. Filtering the weak densities is based on a dynamically defined threshold, according to the clusters densities. Since the filtering is applied on the final clusters, the nodes corresponding to weak density clusters are not reassigned to any cluster.

The final filtered clusters may represent the various meeting topics, where each topic links a speech theme to a similar document theme. Figure 5 displays the results of this clustering process for a given meeting. The circle, around each cluster centroid, represents the cluster density, where the radius increases relatively with its density value.

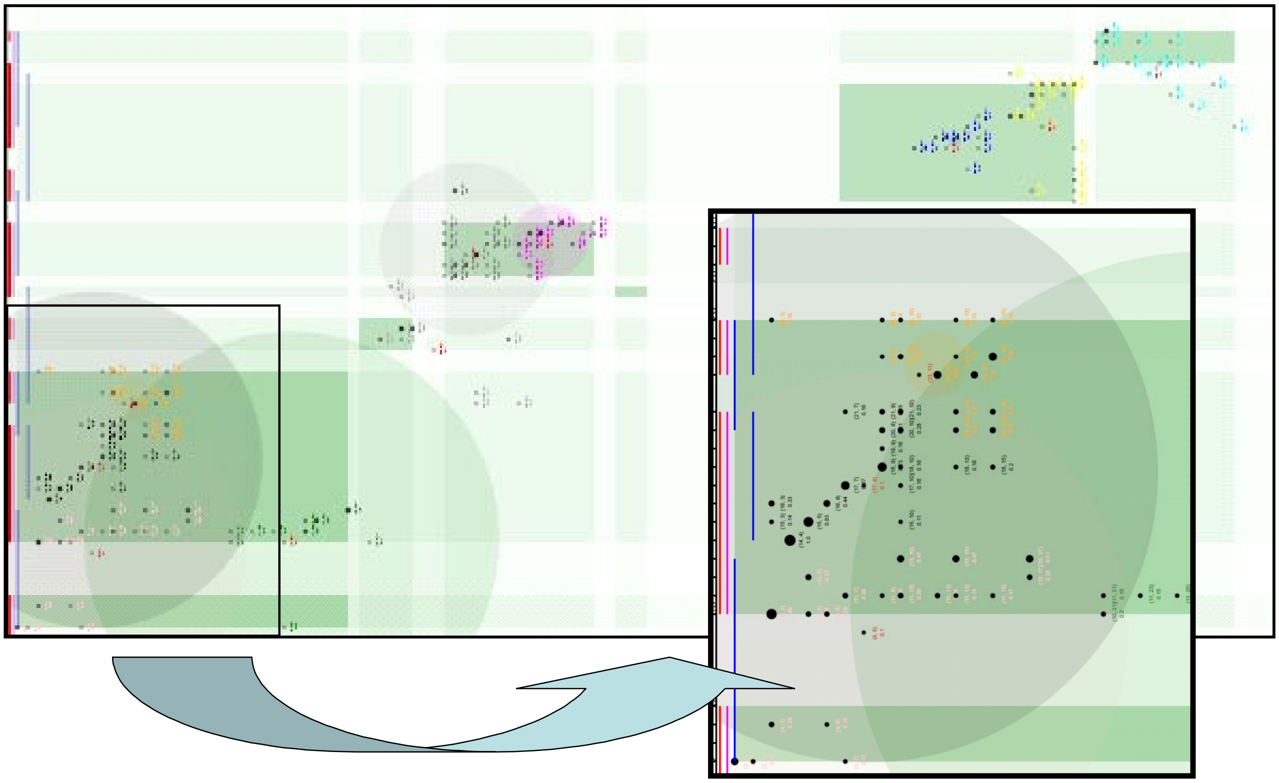


Figure 5. An example of final clustering for a stereotyped meeting. Speech utterances are plotted on the vertical axis and document sentences on the horizontal axis. The ground-truth thematic segments are displayed as vertical and horizontal bars (resp. documents and meeting dialogs).

5.3 Thematic segments extraction

Knowing that the final clusters, representing the denser regions in the 2D graph, are composed by many adjacent nodes, then decomposing a given cluster A by projecting it on the speech axis (Figure 6), generates a group S_A of adjacent Y values, where: $S_A=(y_j, y_{j+1}, y_{j+2}, .., y_{j+l})$ is a speech thematic segment, delimited by the utterances (y_j) and (y_{j+l}). It should be mentioned that the documents thematic segments can also be obtained, by projection of the clusters on the document axis. In the next section, the obtained segmentation evaluation results are presented.

5.4 Evaluation and results

5.4.1 Test data

As input for our experiments, we have used the data of 10 French-speaking meeting recordings, with a total of 1280 utterances and 1133 sentences. These meetings have been classified into two groups (stereotyped vs. non-stereotyped meetings), according to the number of oral interactions, that we characterized using the average number of utterances per speaker turn:

1. Stereotyped meetings, where each speaker presents one or more document articles, with rare interruptions by other speakers. 5 meetings from this category were tested (Set1), with an average ratio of 2.7 speech utterances per speaker turn. In general, we consider that this ratio should be superior to 2 for this category.
2. Non-stereotyped meetings, i.e. speech with numerous debates (Set2). In this category, the speakers debate the news, by commenting the various articles of the document. This increases drastically the number of speaker turns, the 5 meetings tested from this category have an average ratio of 1.3 speech utterances per turn. In this category, this ratio should be inferior to 2.

Table 1. Bi-modal thematic segmentation, comparing to other mono-modal methods.

	Pk	
	Set 1	**Set 2**
Bi-modal	.24 (.40)	.42
TextTiling	.51	.68
Speaker-Turns	.40	.62

5.4.2 Evaluation measures

In order to evaluate our method, the Pk (Beeferman) metric [1] has been used in respect to a prepared manual ground-truth. For a perfect segmentation, the metric value is 0. The Pk metric measures the probability that a randomly chosen pair of units, at a distance of k units apart, is inconsistently classified in respect to the ground truth. For this experiment, the parameter k has been fixed to 4, which corresponds to the minimum size of a relevant thematic segment. This Pk metric is more adequate than the

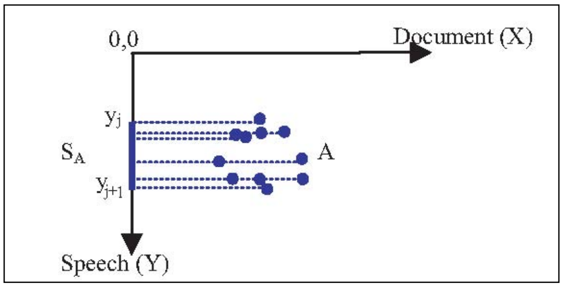


Figure 6. Clusters projection on the speech axis.

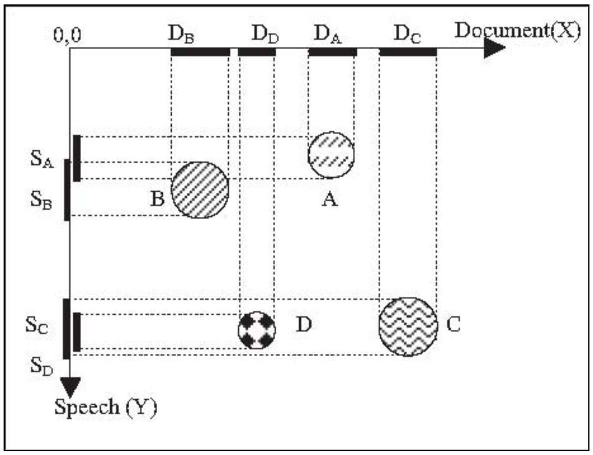


Figure 7. The segments overlapping problem.

recall/precision metric which has many disadvantages [8][14], especially because it measures just the correctness of boundaries detection, without considering the distribution within the generated segments. Table 1 presents the evaluation results obtained for two meetings categories (Set1 and Set2, respectively stereotyped and non-stereotyped meetings) using the Pk metric. In order to avoid the effect of the random choice of initial centroids of clusters, for each meeting we have computed the average value for the Pk measure on 10 tests.

Our method has been compared with two mono-modal segmentation methods: the *TextTiling* [6] method described in section 4, and a *Speaker-Turns* method, in which each speaker turn whose size overcomes a fixed threshold, is considered as a speech transcript thematic segment.

As we can observe in table 1, the segmentation results using the three methods depend on the kind of meeting tested. With our bi-modal method, the Pk evaluation is generally satisfactory for the two meetings categories, especially in comparison to the *TextTiling* method. For stereotyped meetings our method was first unsatisfactory (Pk=.40) in comparison to the *Speaker-turns* segmentation. We explain in the section 5.4.4 how the filtering of the isolated nodes improved our segmentation results to Pk=.24, which is an encouraging result. Our method is also more efficient for non-stereotyped meetings, mainly because often in this meeting category, a topic can be composed of various small turns. Thus, our bi-modal segmentation method is more accurate in detecting the exact number of thematic segments, which is not the case for neither the *TextTiling* method nor the *Speaker-Turns* method, which generate many extra segments. This can be explained by the fact that using the document modality limits the number of possible themes, which constrains the segmentation, and thus helps in computing the exact number of the speech thematic segments.

Finally, even if the results are already usable for stereotyped meetings (Pk=.24), they are still preliminary for non-stereotyped meetings. However, the clustering process can be drastically improved, taking into consideration the nodes' weight, and we are thus confident that the Pk value will be improved. An aspect that makes us particularly optimistic about this bi-modal method is that the alignment results matches perfectly well with the manually segmented meetings, when visualized on our SVG representation (Figure 5). Further, we have seen in a recent user evaluation that visualizing links between documents and meeting dialogs improves browsing performances on meeting archives [10]. Finally, in this evaluation, we used non-logically structured documents (hierarchy of title, section, article, etc.). Thus, both the speech modality and the documents can benefit from the alignments in order to gain a thematic structure. In the future, we plan to use logically structured documents, which should drastically improve meeting dialogs segmentation.

5.4.3 Discussion

Despite the fact that the node weights (the alignment similarity) are not yet considered in our clustering process, the comparison between our bi-modal method and the two mono-modal methods, tends to prove that using documents improves considerably meetings thematic segmentation. Moreover, this bi-modal segmentation method represents an important advantage, which is the detection of all the potential thematic links between non-adjacent speech transcript segments. This happens when these non-adjacent segments are linked to a same document thematic segment.

This important aspect can be seen only through a 2D representation of the thematic alignment. However, a segments overlapping problem occurred during the clusters projection step. This problem has a negative impact on the segmentation results. In the next paragraph, a brief presentation of this problem is given, along with the preliminary filtering method we implemented to by-pass the problem.

5.4.4 Segments overlapping

During the segment extraction step, while projecting the final clusters, we have faced a problem with overlapping clusters. This happens when the projection of some clusters on the speech axis generates overlapped segments (see Figure 7). The overlapping within the speech segments (e.g. S_C with S_D) can be explained either by the thematic similarity between their corresponding themes in the document (theme D_C with D_D), or by the fact that one of these document themes may be referenced while the other was discussed.

An example is given in the following utterances (in the first utterance, the journalist talks about the after-Saddam Iraq, if it will be liberated or occupied. In the second utterance another speaker refers to another article describing the point of view of Blair about this issue)

- « Et.. il y a un article dans lequel le journaliste commence réfléchir à l'après-Saddam. Euh.. voir qu'est-ce qui va se passer, si l'Irak sera occupé ou libéré »
- « Justement j'ai un petit article sur ce point-là, donc selon Tony Blair, euh.. l'après-Saddam, c'est-à-dire l'Irak de l'après-Saddam va être géré par des irakiens. »

Given two overlapped segments (S1, S2), two kinds of overlapping exist:

- A segment includes another (S_C contains S_D)
- The two segments partially overlap (S_A/S_B).

Our first attempt to resolve this overlapping problem, and thus to improve the clustering results, was based on Gaussian probabilistic. But this approach did not improve the results. We have recently implemented and evaluated another approach that consists in the elimination of the isolated nodes, i.e. very distant from the other nodes from the same cluster, taking into consideration both the distances on the document and on the speech transcript axes. The weight of the isolated node has also been considered, in order to avoid filtering important document/speech links. The gap between these isolated nodes and the other nodes of the same cluster can generally be explained by the existence of some transitional utterances between the spoken articles. Futher, we have observed that these isolated nodes are generally the source of the overlapping problem, since they extend the thematic segments more than what they should. Finally, this new filtering method has drastically improved our bi-modal segmentation method results; for the stereotyped meetings (Set1), the Pk value has increased from 0.40 to 0.24. However, for non-stereotyped meetings (Set2), this filtering method did not improve the results significantly.

6. CONCLUSION

In this paper, we have presented a new bi-modal method for thematically segmenting meeting recordings. The method is based on the thematic alignment of documents with meeting dialogs. The evaluation of our bi-modal method in comparison with two other mono-modal methods, has shown very promising results and tends to prove that a good speech/document thematic alignment can lead to a correct meeting thematic segmentation. However, it should be mentioned that this method was more profitable for speech transcript segmentation than for document segmentation, which is partial, since not all the documents articles are discussed in our meetings.

In the future, the node weights (i.e. the similarity values) have to be considered in the clustering process, in order to improve the segmentation. Concerning the alignment results used in the segmentation process, other pairs, such as the alignment of sentences with speech turns, or turns with document logical blocks, will be considered.

7. ACKNOWLEDGEMENT

We would like to thank the University of Applied Sciences of Fribourg for helping to set up the meeting capture environment, and Fati Arabchahi for the work she did on meetings annotation and transcription.

8. REFERENCES

[1] Beeferman D., Berger A., Lafferty J., Statistical Models for Text Segmentation, Machine Learning, Vol. 34, No1/3, 1999, pp. 177-210.

[2] Ding C., He X., Zha H., Gu M. and Simon H., A Min-max Cut Algorithm for Graph Partitioning and Data Clustering, IEEE International Conference on Data Mining, California 2001, pp.107-114.

[3] Ferret O., Grau B. and Masson N., Thematic Segmentation of Texts: Two Methods for Two Kinds of Text, 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Canada 1998, pp.392-396.

[4] Golumbic M.C., Algorithmic Graph Theory and Perfect Graphs Second Edition, Edition Hardcover, Publisher Academic Press 1997, ISBN: 0-444-51530-5,

[5] Hadjar K., Rigamonti M., Lalanne D. and Ingold R., Xed: a new tool for eXtracting hidden structures from Electronic Documents, Palo Alto, California January 2004, pp.212-224.

[6] Hearst M., Multi-Paragraph Segmentation of Expository Text, In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico 1994, pp.9-16.

[7] Jain A., Dubes R., Algorithms for Clustering Data, Edition Hardcover, Publisher Prentice Hall College Div, ISBN 013022278X, 1988.

[8] Kehagias A., Pavlina F. and Petridis V., Linear Text Segmentation using a Dynamic Programming Algorithm, Proceedings of 10th Conference of the European Chapter of the Association for Computational Linguistics 2003, pp.171-178.

[9] 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, USA, 2004, pp.82-91.

[10] Lalanne, D., Ingold, R., von Rotz, D., Behera, A., Mekhaldi, D., Popescu-Belis, A. (in press) - "Using static documents as structured and thematic interfaces to multimedia meeting archives". In Bourlard H. & Bengio S., eds. (2004), Multimodal Interaction and Related Machine Learning Algorithms, LNCS, Springer-Verlag, Berlin, 8 p.

[11] Looney C., Interactive clustering and merging with a new fuzzy expected value, Pattern Recognition, Vol. 35, No11,August 2002, pp.2413-2423.

[12] McQueen, J., Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1967, pp.281-297.

[13] Mekhaldi D., Lalanne D. and Ingold R., Thematic Alignment Of Recorded Speech With Documents, Proceedings of the ACM symposium on Document engineering, France 2003, pp.52-54.

[14] Pevzner L. and Marti Hearst, A Critique and Improvement of an Evaluation Metric for Text Segmentation, Computational Linguistics, Vol.28, No1, 2002, pp.19-36.

[15] Perner P., Data Mining on Multimedia Data, Edition & publisher Springer Verlag 2002, ISBN: 3-540-00317-7.

[16] Salton G., Singhal A., Buckley C. and Mitra M., Automatic Text Decomposition Using Text Segments and Text Themes. In Proceedings of the Hypertext '96 Conference, USA, pp.53-65.

[17] Xie X.L., Beni G., A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and machine Intelligence, Vol. 13, No4, August 1991, pp.841-847.

[18] Zhao Y. and Karypis G., Criterion Functions for Document Clustering, University of Minnesota, Technical Report, February 2002.