

# UWCL at MediaEval 2013: Similar Segments in Social Speech Task

Gina-Anne Levow  
Department of Linguistics  
University of Washington  
Box 352425 Seattle, WA USA  
levow@uw.edu

## ABSTRACT

This paper describes the participation of the University of Washington Computational Linguistics Laboratory (UWCL) in the Similar Segments in Social Speech task at MediaEval 2013. Participants in this task develop systems that, given a span of speech from a recorded conversation, aim to identify all and only highly similar regions in other recordings. As this was a new task for this year, the goal was to establish a baseline and a framework for future experimentation. The approach aimed to address two particular challenges posed by the task: the lack of prior segmentation of the conversations and the limited material provided by a single brief example segment. To this end, the system employed a query-by-example information retrieval framework using passage retrieval to identify segments dynamically and query expansion to support robust retrieval. Query expansion provided substantial gains when applied to both manual and automatic transcriptions; results using automatic transcripts were competitive with those using manual ones.

## 1. INTRODUCTION

Recent years have seen a dramatic growth in the use of social media as well as the sharing of multimedia materials in venues ranging from Facebook to YouTube. Users increasingly share personal content in these social media settings. However, flexible search and targeted access to this material remains challenging, relying largely on manually assigned metadata, such as titles and tags, to identify content, rather than directly indexing the content of these multimedia materials themselves. Furthermore, not only is extraction of content from multimedia streams more challenging than from text, but skimming or browsing in a media stream is slower and more difficult than in text.

The Similar Segments in Social Speech (SSSS) Task developed for MediaEval 2013 aims to overcome these limitations in information access. As described in the task overview paper [6], the task requires participating systems to identify similar spans of speech given an exemplar span. The resulting spans can be viewed as jump-in points for listeners searching or browsing through a multi-media stream.

In contrast to the significant prior work on spoken document retrieval [3] and topic detection and tracking [7], this task applies a more general and abstract notion of similarity, rather than focusing on retrieval of documents related to

particular topics or events. In addition much of that prior work emphasized retrieval from broadcast news sources. Retrieval from less formal audio sources has focused on voice-mail [2] and oral history interviews [5].

The remainder of the paper is organized as follows. Section 2 presents key challenges in the task and UWCL's approach to addressing them. Section 3 describes the experimental configuration, official runs, and results, along with discussion. Section 4 concludes with plans for future work.

## 2. CHALLENGES AND APPROACH

The SSSS task posed a wide range of interesting challenges. These issues included:

- Task modeling: would the task be best modelled as retrieval, clustering, ranking, or something else?
- Sources of similarity: How should similarity be assessed—through lexical or acoustic information or some combination?
- Segmentation: how should segments be identified—via fixed segmentation, agglomeration, or other means?
- Generalization: given a simple example segment, how can we overcome differences across speakers?
- Transcription: what is the effect of transcription type—manual or automatic—on task effectiveness?

For this challenging new task, UWCL's approach built on and extended existing methodologies. In particular the approach adopts an information retrieval perspective, using the text transcriptions of the spoken data. From this perspective, system design focused on the latter three issues identified above: segmentation, generalization, and transcription.

**Segmentation** Much prior work on spoken document retrieval has either provided a gold standard segmentation or assumed its existence. In contrast, the SSSS task does not provide a segmentation, and one could imagine different segmentations based on different notions of similarity. Thus, the strategy aimed to create segments and jump-in points sensitive to the similarity measure and to the exemplar segment. The UWCL system exploits passage retrieval [1] to extract overlapping windowed spans within recordings, with fixed length and step in words, that have high similarity with the example. Overlapping and adjacent retrieved passages are merged and receive the rank of the highest ranked contributing passage. Based on experimentation on the training corpus, retrieval returned the top 75 passages, which were 60 terms in length with a 30 term step.

**Generalization** Differences in lexical choice between materials being searched and the searcher's specification of their

Name	Trans.	Exp.	Exp. set	NSUR	F
uwclman	man	no	n.a.	0.57	0.58
uwclauto	auto	no	n.a.	0.57	0.58
uwclmanexp	man	yes	man	0.82	0.81
uwclautoexp	auto	yes	man	0.66	0.68
uwclauto2exp	auto	yes	auto	0.796	0.80

**Table 1: Contrastive official run settings and results**

information need are a well-known issue in information retrieval. The segments in the SSSS task, which average about 50 seconds in the training set and about 30 seconds in the test set, are not particularly short. However, it seems likely that variation between speakers and the broad notion of similarity will make lexical match highly challenging. To address this issue, the UWCL system investigates the use of query expansion [8]. In pseudo-relevance feedback query expansion, the original query is used in a preliminary search pass. The query is then augmented and, one hopes, improved by adding highly ranked terms from the top-ranked spans which are presumed to be relevant. The resulting query is used for final retrieval. In the UWCL system, the training set data is used to augment the small test set during expansion. The procedure used the top five passages retrieved to create a relevance model and selected the ten terms with highest likelihood under that model for expansion.

**Transcription** Both manual and automatic transcripts of the spoken data are employed.

### 3. EXPERIMENTATION

#### 3.1 Experimental Setup

The UWCL system employed the INDRI/LEMUR information retrieval engine (<http://www.lemurproject.org>) for indexing and retrieval with default settings [4]. The LEMUR system provides a sophisticated query language, has built-in support for passage retrieval, and supports pseudo-relevance feedback query expansion. We made use of two different transcriptions of the conversations: manual transcripts provided by the task organizers and automatic transcripts generously provided by the University of Edinburgh. Each conversation was converted to a single TREC-format text document for indexing. For query formulation, the system extracted all tokens in any time-aligned span which overlapped the exemplar segment. These terms were then linked through unweighted combination (the *#combine* operator). Manual transcriptions were aligned by turn; conversion of automatic transcriptions relied on alignments at the word level.

#### 3.2 Experiment runs and results

Five official runs on the test data were submitted and scored. As shown in Table 1, contrasting conditions explored the impact of transcription (manual/automatic), query expansion (yes/no), and expansion corpus (manual/automatic). The official results are also tabulated, for the primary metrics, Normalized Searcher Utility Ratio (NSUR) and F-measure, as described in the task overview [6].

#### 3.3 Discussion

We find that, although the baseline query formulation achieves modest effectiveness, query expansion using pseudo-relevance feedback based on a matched corpus yielded sub-

stantially increased effectiveness. With the mismatched expansion corpus, the divergence between manual and automatic transcription led to a smaller, but still noticeable, improvement. Finally, it is interesting to note that, with suitable query expansion, a configuration based on automatic transcription greatly outperformed one using manual transcripts without query expansion and was highly competitive with one using manual transcripts with query expansion.

## 4. CONCLUSIONS

UWCL’s approach to the MediaEval 2013 SSSS task employed a text-based information retrieval approach, using passage retrieval to create segments dynamically. Automatic query expansion yielded strong improvements for both manual and automatic transcripts. While these approaches showed promise, many avenues for improvement remain. In addition to tuning retrieval factors, such as passage length and retrieval models, I plan to explore the integration of acoustic, especially acoustic-prosodic, evidence into measures of segment similarity, in addition to the lexical evidence already in use. Such measures could be particularly helpful in recognizing segments with similarity based less on topical content than on emotional or attitudinal content.

## Acknowledgments

Many thanks to the task organizers and also to Steve Renals for providing the high-quality automatic transcriptions.

## 5. REFERENCES

- [1] J. P. Callan. Passage-level evidence in document retrieval. In *Proceedings of the 17th annual ACM SIGIR conference*, pages 302–310, 1994.
- [2] J. Hirschberg, M. Bacchiani, D. Hindel, P. Isenhour, A. Rosenberg, L. Stark, L. Stead, S. Whittaker, and G. Zamchick. Scanmail: Browsing and searching speech data by content. In *Proceedings of EUROSPEECH 2001*, 2001.
- [3] M. Larson and G. J. F. Jones. Spoken content retrieval: A survey of techniques and technologies. *Foundations and Trends in Information Retrieval*, 5(4–5):235–422, 2012.
- [4] D. Metzler, T. Strohman, and W. B. Croft. Indri at TREC 2006: Lessons learned from three terabyte tracks. In *Proceedings of TREC 2006*, 2006.
- [5] D. W. Oard, J. Wang, G. J. Jones, R. W. White, P. Pecina, D. Soergel, X. Huang, and I. Shafran. Overview of the CLEF-2006 cross-language speech retrieval track. In *CLEF-2006*, 2006.
- [6] N. G. Ward, S. D. Werner, D. G. Novick, E. E. Shriberg, C. Oertel, L.-P. Morency, and T. Kawahara. The similar segments in social speech task. In *Proceedings of MediaEval 2013*, Barcelona, Spain, October 18-19 2013.
- [7] C. Wayne. Multilingual topic detection and tracking: Successful research enabled by corpora and evaluation. In *Language Resources and Evaluation Conference (LREC) 2000*, pages 1487–1494, 2000.
- [8] J. Xu and W. Croft. Query expansion using local and global document analysis. In *Proceedings of the 19th Annual International ACM SIGIR Conference*, 1996.