The LIMSI SDR System for TREC-9

Jean-Luc Gauvain, Lori Lamel, Claude Barras, Gilles Adda, and Yannick de Kercardio

Spoken Language Processing Group (http://www.limsi.fr/tlp)
LIMSI-CNRS, B.P. 133, 91403 Orsay cedex, France
{gauvain,lamel,gadda,barras,kercadio}@limsi.fr

ABSTRACT

In this paper we describe the LIMSI Spoken Document Retrieval system used in the TREC-9 evaluation. This system combines an adapted version of the LIMSI 1999 Hub-4E transcription system for speech recognition with text-based IR methods. Compared with the LIMSI TREC-8 system, this year's system is able to index the audio data without knowledge of the story boundaries using a double windowing approach. The query expansion procedure of the information retrieval component has been revised and makes use of contemporaneous text sources.

Experimental results are reported in terms of mean average precision for both the TREC SDR'99 and SDR'00 queries using the same 557h data set. The mean average precision of this year's system is 0.5250 for SDR'99 and 0.3706 for SDR'00 for the focus unknown story boundary condition with a 20% word error rate.

1. INTRODUCTION

This paper describes the LIMSI broadcast news indexating and retrieval system developed for the TREC-9 Spoken Document Retrieval track. Compared with the LIMSI TREC-8 SDR system, both the speech transcription system and the information retrieval component have been improved. Concerning the speech recognizer, we have both sped up the decoder and slightly reduced the word error rate. The query expansion procedure of the information retrieval component has been revised and the capability to index non-segmented audio streams for the unknown story boundaries condition has been added.

During our developement work we investigated the impact of various system parameters on the IR results including: the transcriber speed, the epoch of the texts used for query expansion, the query expansion term weighting strategy, the query length, and the use of non-lexical information.

Most of the reported results here were obtained using the TREC-8 SDR'99 conditions, i.e. the TREC-8 data collection consisting of 557 hours of broadcast news from the period of February through June 1998. This data includes 21750 stories and has an associated set of 50 queries.

The remainder of this paper is as follows: In the next three sections we provide an overview of the broadcast news indexation and information retrieval components, followed by an investigation of the impact of decoding speed and the consequence of the word error rate on the information retrieval process. The subsequent two sections address query expansion and the use of non-lexical information. We then describe how we addressed the unknown story boundary condition and the terse query condition in this year's evaluation. Comparative results are given on the development queries from SDR'99 and this year's query set, and some conclusions are made.

2. TRANSCRIPTION SYSTEM OVERVIEW

The LIMSI broadcast news transcription system [5] consists of an audio partitioner [10] and a speech recognizer [11, 12]. The goal of audio partitioning is to divide the acoustic signal into homogeneous segments, labeling and structuring the acoustic content of the data. Partitioning consists of identifying and removing non-speech segments, and then clustering the speech segments and assigning bandwidth and gender labels to each segment. The result of the partitioning process is a set of speech segments with cluster, gender and telephone/wideband labels, which can be used to generate metadata annotations. The partitioning approach used in the LIMSI BN transcription system relies on an audio stream mixture model [10]. Each component audio source, representing a speaker in a particular background and channel condition, is modeled by a GMM. The segment boundaries and labels are jointly identified by an iterative maximum likelihood segmentation/clustering procedure using GMMs and agglomerative clustering.

For each speech segment, the word recognizer determines the sequence of words in the segment, associating start and end times and an optional confidence measure with each word. The speaker-independent large vocabulary, continuous speech recognizer makes use of n-gram statistics for language modeling and of continuous density HMMs with Gaussian mixtures for acoustic modeling. Word recognition is usually performed in three steps: 1) initial hypothesis generation, 2) word graph generation, 3) final hypothesis generation. The hypotheses are used in cluster-based acoustic model adaptation using the MLLR technique [16] prior to word graph generation, and all subsequent decoding passes. The final hypothesis is generated using a 4-gram language

model.

For all the experimental results given in this paper, the following training conditions were used. The acoustic models were trained on about 150 hours of American English broadcast news data. The phone models are position-dependent triphones, with about 11500 tied-states for the largest model set. The state-tying is obtained via a divisive, decision tree based clustering algorithm. Wideband and telephone band sets of gender-dependent acoustic models were built using MAP adaptation of SI seed models. Fixed language models were obtained by interpolation of n-gram backoff language models trained on 3 different data sets: 203 M words of BN transcripts; 343 M words of NAB newspaper texts and AP Wordstream texts; 1.6 M words corresponding to the transcriptions of the acoustic training data. The interpolation coefficients of these LMs were chosen so as to minimize the perplexity on the Hub4 Nov98 evaluation data. The 4-gram LM contains 7M bigrams, 14M trigrams and 11M fourgrams.

The recognition word list contains 65122 words. The word pronunciations are based on a 48 phone set (3 of them are used for silence, filler words, and breath noises). A pronunciation graph is associated with each word so as to allow for alternate pronunciations, including optional phones. Frequent inflected forms have been verified to provide more systematic pronunciations. As done in the past, compound words for about 300 frequent word sequences subject to reduced pronunciations were included in the lexicon as well as the representation of the most frequent acronyms as words.

3. INFORMATION RETRIEVAL

The automatically generated partition and word transcription can be used for indexation and information retrieval purposes. Techniques commonly applied to automatic text indexation were applied to the automatic transcriptions of the broadcast news radio and TV documents. These classical techniques are based on document term frequencies, where the terms are obtained after standard text processing, such as text normalization, tokenization, stopping, stemming and named-entity identification.

In order to be able to apply the same IR system to different text data types (automatic transcriptions, closed captions, additional texts from newspapers or newswires), all of the documents are preprocessed in a homogeneous manner. This preprocessing, or tokenization, is the same as the text source preparation for training the speech recognizer language models [7], and attempts to transform the texts to be closer to the observed American speaking style. The basic operations include translating numbers and sums into words, removing all the punctuation symbols, removing case distinctions and detecting acronyms and spelled names. However removing all punctuations implies that certain hyphenated words such as *anti-communist*, *non-profit* are rewritten as *anti communist* and *non profit*. While this offers advantages for speech recognition, it can lead to IR errors. To avoid IR problems due to this type of transformation, the output of the tokenizer (and recognizer) is checked for common prefixes, in order to rewrite a sequence of words such as *anti communist* as a single word. The prefixes that are handled include *anti*, *co*, *bi*, *counter*. A rewrite lexicon containing compound words formed with these prefixes and a limited number of named entities (such as *Los-Angeles*) is used to transform the texts. Similarly all numbers less than one hundred are treated as a single entity (such as *twenty-seven*).

In order to reduce the number of lexical items for a given word sense, each word is translated into its stem (as defined in [2, 21]) or, more generally, into a form that is chosen as being representative of its semantic family. The stemming lexicon (derived from the UMass 'porterized' lexicon) [2] contains about 32000 entries and was constructed using Porter's algorithm on the most frequent words in the collection, and then manually corrected.

Two approaches for IR were explored for SDR'99 and this year, the first based on the popular TF$*$IDF weigthing scheme and the second using a Markovian term weighting [14, 17, 19].

For the TF$*$IDF approach, the score of document d for a query is given by the Okapi-BM25 formula[22, 23]. It is the sum over all the terms t in the query of:

$$cw_{t,d} = \frac{(K+1) * tf_{t,d}}{K * (1 - b + b * L_d) + tf_{t,d}} * \log \frac{N}{N_t} * qtf_t$$

where $tf_{t,d}$ is the number of occurrences of term t in document d (i.e. term frequency in document), N_t is the number of documents containing term t at least once, N is the total number of documents in the collection, L_d is the length of document d divided by the average length of the documents in the collection, and qtf_t the number of occurrences of term t in the query.

For the second approach the score of a story is obtained by summing the query term weights $mw_{t,d}$ which are the unigram log probabilities of the terms given the story model once interpolated with a general English model:

$$mw_{t,d} = qtf_t * \log(\alpha \Pr(t|d) + (1-\alpha) \Pr(t)).$$

The text of the query may or may not include the index terms associated with relevant documents. One way to cope with this problem is to use query expansion based on terms present in retrieved documents on the same (Blind Relevance Feedback, BRF) or other (Parallel Blind Relevance Feedback, PBRF) data collections [24]. For SDR'99 we combined the two approches in our system. For PBRF we used 6 months of commercially available broadcast news transcripts from the period jun-dec 1997 [1]. This corpus contains 50000 stories and 49.5 M words. For a given query, the

terms found in the top B documents from the baseline search are ranked by their *offer weight* [23], and the top T terms are added to the query. Since only the T terms with best offer weights are kept, the terms are filtered using a stop list of 144 common words, in order to increase the likelihood that the resulting terms are relevant.

Table 1 gives the results for both *cw* and *mw* term weightings for the SDR'99 data set. Four experimental configurations are reported: baseline search (*base*), query expansion using BRF (*brf*), query expansion with parallel BRF (*pbrf*) and query expansion using both BRF and PBRF (*brf+pbrf*). For BRF and PBRF, the terms are added to the query with a weight of 1. For BRF+PBRF, the terms from each source are added with a weight of 0.5. The results clearly demonstrate the interest of using both BRF and PBRF expansion techniques, as consistent improvements are obtained over the baseline system for the two conditions (R1 and S1). BRF is found to be more effective for both the S1 condition (the recognizer transcripts) and the R1 condition (the manual transcripts).

data	*meth.*	*base*	*brf*	*pbrf*	*brf+pbrf*
R1K	*tf∗idf*	0.4711	0.5318	0.5147	0.5487
	unigram	0.4691	0.5354	0.5098	0.5430
S1K	*tf∗idf*	0.4327	0.5239	0.4919	0.5350
	unigram	0.4412	0.5302	0.4943	0.5398

Table 1: Comparison of IR results on the SDR'99 data set using both Okapi and Markovian term weightings (b=0.86, K=1.1, B=15, T=10, α=0.5). R1: reference transcript. S1: automatic speech transcription. K: known story boundary condition.

The two IR approaches are seen to yield comparable results [13]. Only small differences in information retrieval performance as given by the mean average precision were observed for automatic and manual transcriptions when the story boundaries are known.

4. DECODING SPEED

Processing time is an important factor in making a speech transcription system viable for automatic indexation of radio and television broadcasts. When only concerned by the word error rate, it is common to design systems that run in 100 times real-time or more. Although it is usually assumed that processing time is not a major issue since computer power has been increasing continuously, it is also known that the amount of data appearing on information channels is increasing very rapidly. Therefore processing time is an important factor in making a speech transcription system viable for audio data mining and other related applications. Constraints on the computational resources led us to reconsider some of the system design issues, particularly those concerning the acoustic models and the decoding strategy. We investigated the design of a system which performs well with computational resources in the range 1 to 10xRT on commonly available platforms. A new decoder was implemented with which broadcast data can be transcribed in few times real-time with only a slight increase in word error rate when compared to our best system.

A 4-gram single pass dynamic network decoder has been developed. It is a time-synchronous Viterbi decoder with dynamic expansion of LM state conditioned lexical trees [3, 18, 20] with acoustic and language model lookaheads. The decoder can handle position-dependent, cross-word triphones and lexicons with contextual pronunciations. It makes use of various pruning techniques to reduce the search space and computation time, including three HMM-state pruning beams and fast Gaussian likelihood computations. It can also generate word graphs and rescore them with different acoustic and language models. Faster than real-time decoding can be obtained using this decoder with a word error under 30%, running in less than 100 Mb of memory on widely available platforms such Pentium III or Alpha machines.

The decoder by itself does not solve by itself the problem of reducing the recognition time as proper models have to be used in order to optimize the recognizer accuracy at a given decoding speed. In general, better models have more parameters, and therefore require more computation. However, since the models are more accurate, it is often possible to use a tighter pruning level (thus reducing the computational load) without any loss in accuracy. Thus, limitations on the available computational resources affect the design of the acoustic and language models. For each operating point, the right balance between model complexity and pruning level must be found.

In order to assess the effect of the recognition time on the information retrieval results we transcribed the 557 hours of broadcast news data (the TREC SDR'99 data set – epoch Feb98 to Jun98) using two decoder configurations: a single pass 1.4xRT system and a three pass 10xRT system. The SDR'99 test data consists of 21750 stories and an associated set of 50 queries with on average 14 words. Although story boundaries are available, this information is not used by the speech recognizer. The information retrieval results are given in term of mean average precison (MAP), as is done for the TREC benchmarks. Word error rates are measured on a 10h test subset [6]. For comparison, results are also given for manually produced closed captions. In order for the same IR system to be applied to different text data types (automatic transcriptions, closed captions, additional texts from newspapers or newswires), all of the documents are preprocessed in a homogeneous manner. This preprocessing, or tokenization, is the same as the text source preparation for training the speech recognizer language models.

Table 2 gives the word error rates and IR results for the three sets of transcriptions with and without query expansion. Query expansion uses blind relevance feedback (BRF)

Transcriptions	*Werr*	*Base*	*BRF*
Closed-captions	-	0.4691	0.5430
10xRT	20.5%	0.4528	0.5385
1.4xRT	32.6%	0.4090	0.4938

Table 2: Impact of the word error rate on the mean average precision using the SDR'99 conditions using a 1-gram document model.

pbrf'99	*brf+pbrf'99*	*pbrf'00*
0.5017	0.5397	0.5956

Table 3: Comparison of query expansion schemes on the SDR'99 data with known story boundaries.

on both the audio document collection and some commercially available broadcast news transcripts predating the audio corpus (Jun-Dec 1997 vs Feb-Jun 1998). With query expansion comparable IR results are obtained using the closed captions and the 10xRT transcriptions, and a small degradation (4% absolute) is observed using the 1.4xRT transcriptions.

5. QUERY EXPANSION

In our SDR'99 system query expansion was done by adding terms present in retrieved documents on the same data collection and in an independent set of texts. For PBRF we made use of 6 months of commercially available broadcast news transcripts for covering the period of June through December 1997 [1] (50000 stories and 49.5 M words). However, the SDR'00 specifications (as well as the SDR'99 specifications) allow us to use texts (except for BN transcripts) covering exactly the same epoch of the audio data. Therefore this year we implemeted PBRF using 3 sources of contemporary newspaper data: the New York Times, the Los Angeles Times and the Washington Post. The parallel corpus conatined a total of 42 M words and 78 K documents between Jan98 and Jun98. Experiments with these texts on the SDR'99 show that PBRF using contemporary texts offers a significant performance gain compared with a PBRF using texts predating the audio data. In fact we found that we no longer needed to combine both BRF and PBRF, since PRBF with the new texts gave comparable benefits.

This year we also changed the term weighting used with query expansion, using a weight proportional to the *offer weight* as defined in [23, 15]. This approach allowed us to significantly increase the number of expansion terms, going from 10 terms with the previous approach to 25 terms with the term weighting. The sum of the weights for the expansion terms is set to the number of added terms, i.e., 25. Table 3 shows the combined improvment obtained with the new query expansion scheme on the SDR'99 data. These results were obtained using the Okapi term weighting with a parameter setting (b=0.7, K=1.2) and a slighlty different stemmer from that used for the results reporter earlier in this paper.

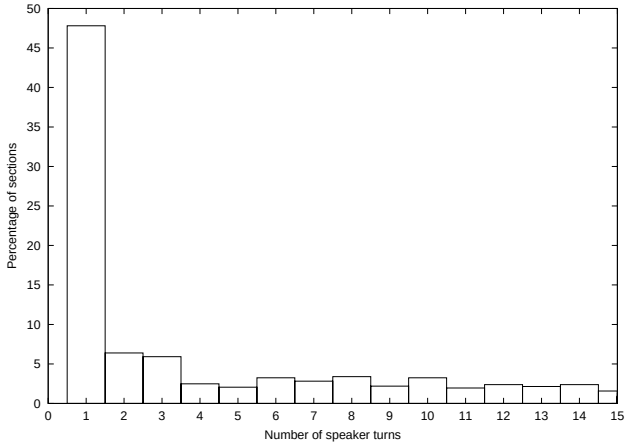

Figure 1: Histogram of the number of speaker turns per section in the 1997 Hub-4 data set.

6. NON-LEXICAL INFORMATION

The broadcast news transcription system also provides non-lexical information along with the word transcription. This information is available in the partition of the audio track, which identifies speaker turns. We investigated the use of automatically detected speaker changes for locating document boundaries. Statistics were made on the 1997 English Hub-4 training data set, which consists of about 100 hours of radio and television broadcast news with manual transcription and speaker identification. On this set, 2096 sections were manually marked as report sections and used as documents for the SDR'98 evaluation. Among them, 817 sections (about 40%) start without a manually annotated speaker change. This means that using only speaker change information for detecting document boundaries would result in 40% missed boundaries. This figure would likely increase with the use of automatically detected speaker changes. At the same time, 11,160 of the total of 12,439 speaker turns occur in the middle of a document, which gives almost a 90% false alarm rate. A more detailed analysis shows that about 50% of the sections involve a single speaker, but that the distribution of the number of speaker turns per section falls off very gradually from 2 to 20 speakers (cf. Figure 1). False alarms are not as harmful as missed detections, since it is possible to merge adjacent turns into a single document in subsequent processing. However these results show clearly that even perfect speaker turn boundaries cannot be used as the primary cue for locating document boundaries. They can be used to refine the placement of a document boundary located near a speaker change.

Besides speaker turns, changes in the background acoustic conditions can be detected by the audio partitioner and can be considered as indicators of story boundaries. We did not investigate this because the background conditions were not manually marked in the 1997 English Hub-4 corpus.

We investigated using simple statistics on the durations of

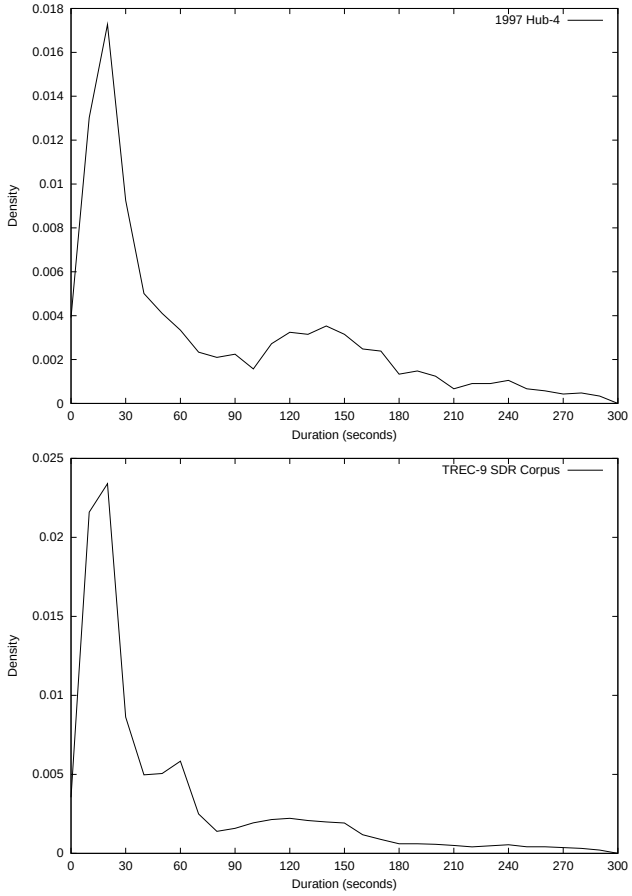

Figure 2: Distribution of document durations in the Hub4'97 and SDR'00 data sets.

the documents in the SDR'98 data set. A histogram of the 2096 sections is shown in Figure 2. One third of the sections are shorter than 30 seconds. The histogram has a sharp peak around 20 seconds, and a smaller, flat peak around 2 minutes, resulting in a bimodal distribution of document length. Very short documents are typical of headlines which are uttered by single speaker, whereas longer documents are more likely to contain data from multiple talkers. This distribution led us to consider using a multi-scale segmentation of the audio stream into documents. Similar statistics were measured on the SDR'99 data using the known document boundaries. The distribution, shown in lower part of Figure 2 is quite similar to that of the SDR'98 data, with an additional, small peak at 60 seconds.

7. UNKNOWN STORY BOUNDARY CONDITION

As proposed in [9], we first segmented the audio stream into overlapping documents of a fixed duration. As a result of optimization using the TREC-8 SDR queries, we chose a 30 second window duration with a 15 second overlap. Since there are many stories significantly shorter than 30s in broadcast shows (see Figure 2) we conjunctured that it may be of interest to use a double windowing system in order to better target short stories. The window size of the smaller window was selected to be 10 seconds. So for each query, we independently retrieved two sets of 2700 documents, one set for each window size. Then for each document set, document recombination is done by merging overlapping documents until no further merges are possible. The score of a combined document is set to maximum score of any one of the components. For each document derived from the 30s windows, we produce a time stamp located at the center point of the document. However, if any smaller documents are embedded in this document, we take the center of the best scoring document. This way we try to take avantage of both window sizes. The MAP using a single 30s window and the double windosing strategy are shown in Table 4.

Mode	*30s*	*30s + 10s*
baseline	0.3673	0.3791
PBRF	0.5001	0.5260

Table 4: Unkown story boundary condition development results on SDR'99 data.

8. TERSE QUERIES

A new component of this year's evaluation was the use of terse queries for indexation. Since terse forms of the 1999 queries were not available, we generated a set for use in system development. These were generated based on the instructions given to the assessors that developed the SDR'00 short and terse queries.[1] Different group members used these general instructions to independently generate terse versions of the SDR'99 queries. These were then compiled and a single form was selected. The resulting SDR'99 terse queries contain on average 3.3 words per query to be compared to 13.7 words for the regular "short" queries.

We carried out retrieval experiments with these terse queries using the system parameter values tuned for the short queries. The retrieval results are given on Table 5 for both the known and unknown story boundary conditions on the SDR'99 data. We can see that there is only about a 1% absolute reduction of the mean average precision when the short queries are replaced by the terse queries. Given this small degradation we did not try to modify our system to better optimize performance on the terse queries.

9. RESULTS

Retrieval results for the SDR'00 evaluation system are given in Tables 6 and 7 for both SDR'99 and SDR'00 queries. It is clear from these results that the system behavoir is quite different on the two query sets. First the SDR'00

[1] Although no specific written guidelines were available, John Garofolo kindly described the instructions given to the assessors.

Mode	*short queries*	terse queries
R1K	0.5975	0.5852
S1K	0.5956	0.5795
S1U	0.5260	0.5147

Table 5: Retrieval results with short and terse queries on the SDR'99 data. R1: reference transcript. S1: automatic speech transcription. K: known story boundary condition. U: unknown story boundary condition.

queries appear to be significantly more difficult, with a 25% relative reduction in the mean average precision compared to the SDR'99 queries. Second, we get significantly better results with the terse queries than with the short queries, while we observed a slight loss on our SDR'99 terse queries. The average length of the SDR'00 terse queries (3.0) is not significantly different from the average length of our SDR'00 terse queries (3.3), but there is a substantial difference in the number of new words compared to the short queries. The SDR'00 terse queries introduce 54 new words with 85 words in common the the SDR'00 short queries, whereas we had only 17 new words in our SDR'99 terse queries with 181 words in common. These numbers show that our SDR'99 terse queries were essentially shorter versions of the corresponding short query, whereas the SDR'00 terse queries appear to be a reformulation of the SDR'00 short queries.

Mode	*Queries'99*		*Queries'00*	
	short	*terse*	*short*	*terse*
R1K	0.5975	0.5852	0.4636	0.5132
S1K	0.5956	0.5795	0.4327	0.4812

Table 6: Retrieval results on SDR'99 and SDR'00 data with known story boundaries. R1: reference transcript. S1: automatic speech transcription. K: known story boundary condition.

Mode	*Queries'99*		*Queries'00*	
	short	*terse*	*short*	*terse*
R1U	0.5233	-	0.4027	0.4283
B1U	0.5034	-	0.3712	0.3922
S1U	0.5260	0.5147	0.3706	0.3982

Table 7: Retrieval results on SDR'99 and SDR'00 data with unknown story boundaries. R1: reference transcript. B1: baseline automatic speech transcription. S1: automatic speech transcription. U: unknown story boundary condition.

10. CONCLUSION

In this paper we have described the LIMSI TREC-9 spoken document retrieval system. This system is based on the 1999 LIMSI system, with a few substantial modifications. First, the decoder of the speech recognizer has been replaced by a new, faster decoder able to transcribes broadcast data in several (6 to 10) times real-time with only a slight increase in word error rate when compared to our best system and with a word error of about 30% for essentially real-time decoding. Second, the query expansion procedure of the information retrieval component has been revised and makes use of contemporaneous text sources. Thirdly, a double windowing approach has been developed to localize stories for the unknown boundary condition.

The experimental results show that only a moderate IR performance degradation is obtained in spoken document retrieval with a close to real-time system, and that generally speaking, the transcription quality of our system is not a limiting factor given todays IR techniques.

ACKNOWLEDGEMENTS

This work has been partially financed by the European Commission under the IST-1999-10354 ALERT project and the French Ministry of Defense. We also thank Patrick Paroubek for providing the terse versions of the SDR'99 query set used for system development.

REFERENCES

[1] http://www.thomson.com/psmedia/bnews.html

[2] ftp://ciir-ftp.cs.umass.edu/pub/stemming/

[3] X. Aubert, "One Pass Cross Word Decoding for Large Vocabularies Based on a Lexical Tree Search Organization," *Proc. ESCA Eurospeech'99*, **4**, pp. 1559-1562, Budapest, Hungary, September 1999.

[4] J. Davenport, L. Nguyen, S. Matsoukas, R. Schwartz, J. Makhoul, "The 1998 BBN Byblos 10x Real Time System," *Proc. DARPA Broadcast News Workshop*, Feb.-Mar. 1999.

[5] J.L. Gauvain, L. Lamel, and G. Adda, "Transcribing broadcast news for audio and video indexing," Communications of the ACM, 43(2), February 2000.

[6] J.S. Garofolo et al., "1999 Trec-8 Spoken Document Retrieval Track Overview and Results," *Proc. 8th Text Retrieval Conference TREC-8*, Gaithersburg, MD, November 1999.

[7] J.L. Gauvain, L. Lamel, M. Adda-Decker, "The LIMSI Nov93 WSJ System," *Proc. ARPA Spoken Language Technology Workshop*, March, 1994.

[8] J.L. Gauvain, Y. de Kercadio, L.F. Lamel, G. Adda "The LIMSI SDR system for TREC-8," *Proc. of the 8th Text Retrieval Conference TREC-8*, pp. 405-412, Gaithersburg, MD, November 1999.

[9] D. Abberley, S. Renals, Dan Ellis and T. Robinson, "The THISL SDR System at TREC-8", *Proc. of the 8th Text Retrieval Conference TREC-8*, Gaithersburg, MD, November 1999.

[10] J.L. Gauvain, L. Lamel, G. Adda, "Partitioning and Transcription of Broadcast News Data," *ICSLP'98*, **5**, pp. 1335-1338, December 1998.

[11] J.L. Gauvain, L. Lamel, G. Adda and M. Jardino, "The LIMSI 1998 Hub-4E Transcription System", *Proc.* DARPA *Broadcast News Workshop*, pp. 99-104, Herndon, VA, February 1999.

[12] J.L. Gauvain, L. Lamel, G. Adda, "Recent Advances in Transcribing Television and Radio Broadcasts," *Proc. Eurospeech'99*, **2**, pp. 655-658, Budapest, Hungary, September 1999.

[13] J.L. Gauvain, L. Lamel, Y. de Kercadio, and G. Adda, "Transcription and Indexation of Broadcast Data," *Proc. IEEE ICASSP'00*, **III**, pp. 1663-1666, Istanbul, Turkey, June 2000.

[14] D. Hiemstra, K. Wessel, "Twenty-One at TREC-7: Ad-hoc and Cross-language track," *Proc. of the 8th Text Retrieval Conference TREC-7*, Gaithersburg, MD, 1998.

[15] S.E. Johnson, P. Jourlin, K. Spärck Jones, P.C. Woodland, "Spoken Document Retrieval for TREC-8 at Cambridge University", *Proc. of the 8th Text Retrieval Conference TREC-8*, Gaithersburg, MD, November 1999.

[16] C.J. Leggetter, P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech & Language*, **9**(2), pp. 171-185, 1995.

[17] D. Miller, T. Leek, R. Schwartz, "Using Hidden Markov Models for Information Retrieval", *Proc. of the 8th Text Retrieval Conference TREC-7*, Gaithersburg, MD, 1998.

[18] H. Ney, R. Haeb-Umbach, B.H. Tran and M. Oerder, "Improvements in Beam Search for 10000-Word Continuous Speech Recognition," *Proc. IEEE ICASSP-92*, **I**, pp. 9-12, San Francisco, CA, March 1992.

[19] K. Ng, "A Maximum Likelihood Ratio Information Retrieval Model," *Proc. of the 8th Text Retrieval Conference TREC-8*, pp. 413-435, Gaithersburg, MD, November 1999.

[20] J.J. Odell, V. Valtchev, P.C. Woodland and S.J. Young, "A One Pass Decoder Design for Large Vocabulary Recognition," *Proc. ARPA Human Language Technology Workshop*, pp. 405-410, Princeton, NJ, March 1994.

[21] M. F. Porter, "An algorithm for suffix stripping", *Program*, **14**, pp. 130–137, 1980.

[22] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, M. Gatford, "Okapi at TREC-3", *NIST Special Publication 500-226: Overview of the Third Text REtrieval Conference (TREC-3)*, November 1994.

[23] K. Spärk Jones, S. Walker, S. E. Robertson, "A probabilistic model of information retrieval: development and status", *a Technical Report of the Computer Laboratory, University of Cambridge, U.K.*, 1998.

[24] S. Walker, R. de Vere, "Improving subject retrieval in online catalogues: 2. Relevance feedback and query expansion", *British Library Research Paper 72*, British Library, London, U.K., 1990.

[25] P.C. Woodland, J.J. Odell, T. Hain, G.L. Moore, T.R. Niesler, A. Tuerk, E.W.D. Whittaker, "Improvements in Accuracy and Speed in the HTK Broadcast News Transcription System," *Eurospeech'99*, pp. 1043-1046, September 1999.