

Mixed-Source Multi-Document Speech-to-Text Summarization

Ricardo Ribeiro
INESC ID Lisboa/ISCTE/IST
Spoken Language Systems Lab
Rua Alves Redol, 9
1000-029 Lisboa, Portugal
rdmr@l2f.inesc-id.pt

David Martins de Matos
INESC ID Lisboa/IST
Spoken Language Systems Lab
Rua Alves Redol, 9
1000-029 Lisboa, Portugal
david@l2f.inesc-id.pt

Abstract

Speech-to-text summarization systems usually take as input the output of an automatic speech recognition (ASR) system that is affected by issues like speech recognition errors, disfluencies, or difficulties in the accurate identification of sentence boundaries. We propose the inclusion of related, solid background information to cope with the difficulties of summarizing spoken language and the use of multi-document summarization techniques in single document speech-to-text summarization. In this work, we explore the possibilities offered by phonetic information to select the background information and conduct a perceptual evaluation to better assess the relevance of the inclusion of that information. Results show that summaries generated using this approach are considerably better than those produced by an up-to-date latent semantic analysis (LSA) summarization method and suggest that humans prefer summaries restricted to the information conveyed in the input source.

1 Introduction

News have been the subject of summarization for a long time, demonstrating the importance of both the subject and the process. Systems like NewsInEssence (Radev et al., 2005), Newsblaster (McKeown et al., 2002), or even Google

News substantiate this relevance that is also supported by the spoken language scenario, where most speech summarization systems concentrate on broadcast news (McKeown et al., 2005). Nevertheless, although the pioneering efforts on summarization go back to the work of Luhn (1958) and Edmundson (1969), it is only after the renaissance of summarization as a research area of great activity—following up on the Dagstuhl Seminar (Endres-Niggemeyer et al., 1995)—that the first multi-document news summarization system, SUMMONS (McKeown and Radev, 1995), makes its breakthrough (Radev et al., 2005; Spärck Jones, 2007). In what concerns speech summarization, the state of affairs is more problematic: news summarization systems appeared later and still focus only on single document summarization (McKeown et al., 2005). In fact, while text summarization has attained some degree of success (Hovy, 2003; McKeown et al., 2005; Spärck Jones, 2007) due to the considerable body of work, speech summarization still requires further research, both in speech and text analysis, in order to overcome the specific challenges of the task (McKeown et al., 2005; Furui, 2007). Issues like speech recognition errors, disfluencies, and difficulties in accurately identifying sentence boundaries must be taken into account when summarizing spoken language. However, if on the one hand, recognition errors seem not to have a considerable impact on the summarization task (Murray et al., 2006; Murray et al., 2005), on the other hand, spoken language summarization systems often explore ways of minimizing that impact (Zechner and Waibel, 2000; Hori et al., 2003; Kikuchi et al., 2003).

We argue that by including related solid background information from a different source less prone to this kind of errors (e.g., a textual source)

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

in the summarization process, we are able to reduce the influence of recognition errors on the resulting summary. To support this argument, we developed a new approach to speech-to-text summarization that combines information from multiple information sources to produce a summary driven by the spoken language document to be summarized. The idea mimics the natural human behavior, in which information acquired from different sources is used to build a better understanding of a given topic (Wan et al., 2007). Furthermore, we build on the conjecture that this background information is often used by humans to overcome perception difficulties. In that sense, one of our goals is also to understand what is expected in a summary: a comprehensive, shorter, text that addresses the same subject of the input source to be summarized (possibly introducing new information); or a text restricted to the information conveyed in the input source.

This work explores the use of phonetic domain information to overcome speech recognition errors and disfluencies. Instead of using the traditional output of the ASR module, we use the phonetic transliteration of the output and compare it to the phonetic transliteration of solid background information. This enables the use of text, related to the input source, free from the common speech recognition issues, in further processing.

We use broadcast news as a case study and news stories from online newspapers provide the background information. Media monitoring systems, used to transcribe and disseminate news, provide an adequate framework to test the proposed method.

This document is organized as follows: section 2 briefly introduces the related work; section 3 presents a characterization of the speech-to-text summarization problem and how we propose to address it; section 4 explicits our use of phonetic domain information, given the previously defined context; the next section describes the case study, including the experimental set up and results; conclusions close the document.

2 Related Work

McKeown et al. (2005) depict spoken language summarization as a much harder task than text summarization. In fact, the previously enumerated problems that make speech summarization such a difficult task constrain the applicability of text

summarization techniques to speech summarization (although in the presence of planned speech, as it partly happens in the broadcast news domain, that portability is more feasible (Christensen et al., 2003)). On the other hand, speech offers possibilities like the use of prosody and speaker identification to ascertain relevant content.

Furui (2007) identifies three main approaches to speech summarization: sentence extraction-based methods, sentence compaction-based methods, and combinations of both.

Sentence extractive methods comprehend, essentially, methods like LSA (Gong and Liu, 2001), Maximal Marginal Relevance (Carbonell and Goldstein, 1998), and feature-based methods (Edmundson, 1969). Feature-based methods combine several types of features: current work uses lexical, acoustic/prosodic, structural, and discourse features to summarize documents from domains like broadcast news or meetings (Maskey and Hirschberg, 2005; Murray et al., 2006; Ribeiro and de Matos, 2007). Even so, spoken language summarization is still quite distant from text summarization in what concerns the use of discourse features, and shallow approaches is what can be found in state-of-the-art work such as the one presented by Maskey and Hirschberg (2005) or Murray et al. (2006). Sentence compaction methods are based on word removal from the transcription, with recognition confidence scores playing a major role (Hori et al., 2003). A combination of these two types of methods was developed by Kikuchi et al. (2003), where summarization is performed in two steps: first, sentence extraction is done through feature combination; second, compaction is done by scoring the words in each sentence and then a dynamic programming technique is applied to select the words that will remain in the sentence to be included in the summary.

3 Problem Characterization

Summarization can be seen as a reductive transformation ϕ that, given an input source I , produces a summary S :

$$S = \phi(I),$$

where $len(S) < len(I)$ and $inf(S)$ is as close as possible of $inf(I)$; $len()$ is the length of the given input and $inf()$ is the information conveyed by its argument.

The problem is that in order to compute S , we are not using I , but \tilde{I} , a noisy representation of I .

Thus, we are computing \tilde{S} , which is a summary affected by the noise present in \tilde{I} :

$$\tilde{S} = \phi(\tilde{I}).$$

This means that

$$\begin{aligned} \inf(\tilde{S}) &\subset \inf(S) \subset \inf(I), \text{ whereas} \\ \text{len}(\tilde{S}) &\approx \text{len}(S) < \text{len}(I). \end{aligned}$$

Our argument is that using a similar reductive transformation ψ , where solid background information B is also given as input, it is possible to compute a summary \hat{S} :

$$\begin{aligned} \hat{S} &= \psi(\tilde{I}, B), \text{ such that} \\ \inf(\hat{S}) &\subset (\inf(\tilde{S}) \cap \inf(S)) \subset \inf(I), \text{ with} \\ \text{len}(\hat{S}) &\approx \text{len}(\tilde{S}) \approx \text{len}(S) < \text{len}(I). \end{aligned}$$

As seen in section 2, the most common method to perform these transformations is by selecting sentences (or extracts) from the corresponding input sources.

Thus, let the input source representation \tilde{I} be composed by a sequence of extracts e_i ,

$$\tilde{I} = e_1, e_2, \dots, e_n$$

and the background information be defined as a sequence of sentences

$$B = s_1, s_2, \dots, s_m.$$

The proposed method consists of selecting sentences s_i form the background information B such that

$$\text{sim}(s_i, e_j) < \varepsilon \wedge 0 \leq i \leq m \wedge 0 \leq j \leq n,$$

with $\text{sim}()$ being a similarity function and ε an adequate threshold. The difficulty lies in defining the function and the threshold.

4 Working in the phonetic domain

The approach we introduce minimizes the effects of recognition errors through the selection, from previously determined background knowledge, of sentence-like units close to the ones of the news story transcription. In order to select sentence-like units, while diminishing recognition problems, we compute the similarity between them at the phonetic level. The estimation of the threshold is based on the distance, measured in the phonetic

Feature	Values
Type	vowel, consonant
Vowel length	short, long, diphthong, schwa
Vowel height	high, mid, low
Vowel frontness	front mid back
Lip rounding	yes, no
Consonant type	stop, fricative, affricative, nasal, liquid
Place of articulation	labial, alveolar, palatal, labio-dental, dental, velar
Consonant voicing	yes, no

Table 1: Phone features.

domain, between the output of the ASR and its hand-corrected version.

The selection of sentences from the background information is based on the alignment cost of the phonetic transcriptions of sentences from the input source and sentence from the background information. Sentences from the background information with alignment costs below the estimated threshold are selected to be used in summary generation.

4.1 Similarity Between Segments

There are several ways to compute phonetic similarity. Kessler (2005) states that phonetic distance can be seen as, among other things, differences between acoustic properties of the speech stream, differences in the articulatory positions during production, or as the perceptual distance between isolated sounds. Choosing a way to calculate phonetic distance is a complex process.

The phone similarity function used in this process is based on a model of phone production, where the phone features correspond to the articulatory positions during production: the greater the matching between phone features, the smaller the distance between phones. The phone features used are described in table 1.

The computation of the similarity between sentence-like units is based on the alignment of the phonetic transcriptions of the given segments. The generation of the possible alignments and the selection of the best alignment is done through the use of Weighted Finite-State Transducers (WFSTs) (Mohri, 1997; Paulo and Oliveira, 2002).

4.2 Threshold Estimation Process

To estimate the threshold to be used in the sentence selection process, we use the algorithm presented in figure 1. The procedure consists of comparing automatic transcriptions and their hand-corrected versions: the output is the average difference between the submitted inputs.

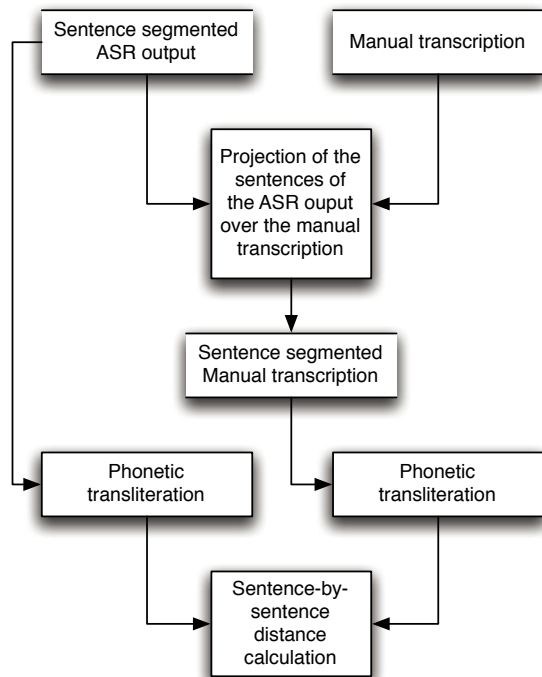


Figure 1: Threshold estimation process.

The idea is that the phonetic distance between the automatic transcription and its hand-corrected version would be similar to the phonetic distance between the automatic transcription and the background information. Even though this heuristic may appear naive, we believe it is adequate as a rough approach, considering the target material (broadcast news).

5 A Case Study Using Broadcast News

5.1 Media Monitoring System

SSNT (Amaral et al., 2007) is a system for selective dissemination of multimedia contents, working primarily with Portuguese broadcast news services. The system is based on an ASR module, that generates the transcriptions used by the topic segmentation, topic indexing, and title&summarization modules. User profiles enable the system to deliver e-mails containing relevant news stories. These messages contain the name

of the news service, a generated title, a summary, a link to the corresponding video segment, and a classification according to a thesaurus used by the broadcasting company.

Preceding the speech recognition module, an audio preprocessing module, based on Multi-layer Perceptrons, classifies the audio in accordance to several criteria: speech/non-speech, speaker segmentation and clustering, gender, and background conditions.

The ASR module, based on a hybrid speech recognition system that combines Hidden Markov Models with Multi-layer Perceptrons, with an average word error rate of 24% (Amaral et al., 2007), greatly influences the performance of the subsequent modules.

The topic segmentation and topic indexing modules were developed by Amaral and Trancoso (2004). Topic segmentation is based on clustering and groups transcribed segments into stories. The algorithm relies on a heuristic derived from the structure of the news services: each story starts with a segment spoken by the anchor. This module achieved an *F-measure* of 68% (Amaral et al., 2007). The main problem identified by the authors was boundary deletion: a problem which impacts the summarization task. Topic indexing is based on a hierarchically organized thematic thesaurus provided by the broadcasting company. The hierarchy has 22 thematic areas on the first level, for which the module achieved a correctness of 91.4% (Amaral et al., 2006; Amaral et al., 2007).

Batista et al. (2007) inserted a module for recovering punctuation marks, based on maximum entropy models, after the ASR module. The punctuation marks addressed were the “full stop” and “comma”, which provide the sentence units necessary for use in the title&summarization module. This module achieved an *F-measure* of 56% and *SER* (Slot Error Rate, the measure commonly used to evaluate this kind of task) of 0.74.

Currently, the title&summarization module produces a summary composed by the first n sentences, as detected by the previous module, of each news story and a title (the first sentence).

5.2 Corpora

Two corpora were used in this experiment: a broadcast news corpus, the subject of our summarization efforts; and a written newspaper corpus, used to select the background information.

Corpus	Stories	SUs	Tokens	Duration
train	184	2661	57063	5h
test	26	627	7360	1h

Table 2: Broadcast news corpus composition.

The broadcast news corpus is composed by 6 Portuguese news programs, and exists in two versions: an automatically processed one, and a hand-corrected one. Its composition (number of stories, number of sentence-like units (SUs), number of tokens, and duration) is detailed in table 2. To estimate the threshold used for the selection of the background information, 5 news programs were used. The last one was used for evaluation.

The written newspaper corpus consists of the online version a Portuguese newspaper, downloaded daily from the Internet. In this experiment, three editions of the newspaper were used, corresponding to the day and the two previous days of the news program to be summarized. The corpus is composed by 135 articles, 1418 sentence-like units, and 43102 tokens.

5.3 The Summarization Process

The summarization process we implemented is characterized by the use of LSA to compute the relevance of the extracts (sentence-like units) of the given input source.

LSA is based on the singular vector decomposition (SVD) of the term-sentence frequency $m \times n$ matrix, M . U is an $m \times n$ matrix of left singular vectors; Σ is the $n \times n$ diagonal matrix of singular values; and, V is the $n \times n$ matrix of right singular vectors (only possible if $m \geq n$):

$$M = U\Sigma V^T$$

The idea behind the method is that the decomposition captures the underlying topics of the document by means of co-occurrence of terms (the latent semantic analysis), and identifies the best representative sentence-like units of each topic. Summary creation can be done by picking the best representatives of the most relevant topics according to a defined strategy.

For this summarization process, we implemented a module following the original ideas of Gong and Liu (2001) and the ones of Murray, Renals, and Carletta (2005) for solving dimensionality problems, and using, for matrix operations, the

GNU Scientific Library¹.

5.4 Experimental Results

Our main objective was to understand if it is possible to select relevant information from background information that could improve the quality of speech-to-text summaries. To assess the validity of this hypothesis, five different processes of generating a summary were considered. To better analyze the influence of the background information, all automatic summarization methods are based on the up-to-date LSA method previously described: one taking as input only the news story to be summarized (*Simple*) and used as baseline; other taking as input only the selected background information (*Background only*); and, the last one, using both the news story and the background information (*Background + News*). The other two processes were human: extractive (using only the news story) and abstractive (understanding the news story and condensing it by means of paraphrase). Since the abstractive summaries had already been created, summary size was determined by their size (which means creating summaries using a compression rate of around 10% of the original size).

As mentioned before, the whole summarization process begins with the selection of the background information. Using the threshold estimated as described in section 4.2 and the method described in section 4.1 to compute similarity between sentence-like units, no background information was selected for 11 of the 26 news stories of the test corpus. For the remaining 15 news stories, summaries were generated using the three automatic summarization strategies described before.

In what concerns the evaluation process, although ROUGE (Lin, 2004) is the most common evaluation metric for the automatic evaluation of summarization, since our approach might introduce in the summary information that it is not present in the original input source, we found that a human evaluation was more adequate to assess the relevance of that additional information. A perceptual evaluation is also adequate to assess the perceive quality of the summaries and a better indicator of the what is expected to be in a summary.

We asked an heterogeneous group of sixteen people to evaluate the summaries created for the 15 news stories for which background information

¹<http://www.gnu.org/software/gsl/>

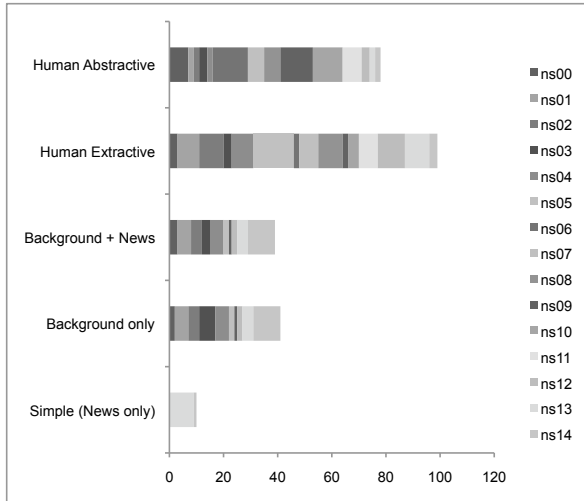


Figure 2: Overall results for each summary creation method (*nsnn* identifies a news story).

was selected. Each evaluator was given, for each story, the news story itself (without background information) and five summaries, corresponding to the five different methods presented before. The evaluation procedure consisted in identifying the best summary and in the classification of each summary (1–5, 5 is better) according to its content and readability (which covers issues like grammaticality, existence of redundant information, or entity references (Nenkova, 2006)).

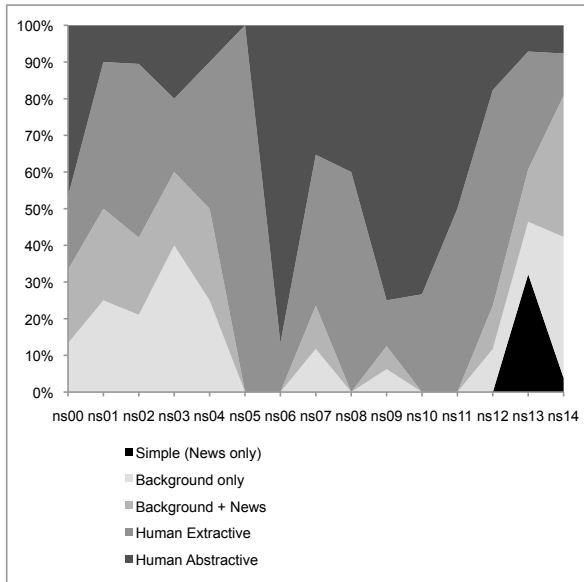


Figure 3: Relative results for each news story (*nsnn* identifies a news story; stack order is inverse of legend order).

Surprisingly enough (see figures 2 and 3), in general, the extractive human summaries were pre-

ferred over the abstractive ones. Moreover, the summaries generated automatically using background information (exclusively or not) were also selected as best summary (over the human created ones) a non-negligible number of times. The poorest performance was attained, as expected, by the simple LSA summarizer, only preferred on two news stories for which all summaries were very similar. The results of the two approaches using background information were very close, a result that can be explained by the fact the summaries generated by these two approaches were equal for 11 of the 15 news stories (in the remaining 4, the average distribution was 31.25% from the news story versus 68.75% from the background information).

Figure 4 further discriminates the results in terms of content and readability.

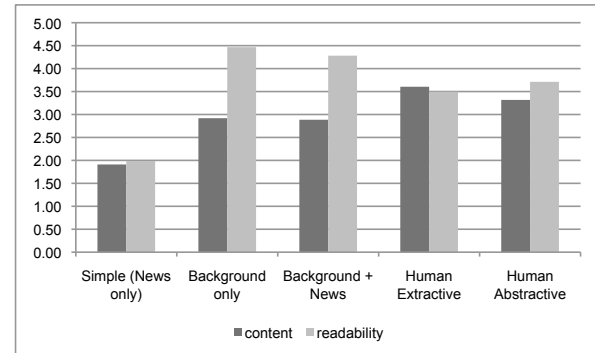


Figure 4: Average of the content and readability scores for each summary creation method.

Regarding content, the results suggest that the choice of the best summary is highly correlated with its content, as the average content scores mimic the overall ones of figure 2. In what concerns readability, the summaries generated using background information achieved the best results. The reasons underlying these results are that the newspaper writing is naturally better planned than speech and that speech transcriptions are affected by the several problems described before (and the original motivation for the work), hence the idea of using them as background information. However, what is odd is that the result obtained by the human abstractive summary creation method is worse than the ones obtained by automatic generation using background information, which could suffer from coherence and cohesion problems. One possible explanation is that the human abstractive summaries tend to mix both informa-

tive and indicative styles of summary.

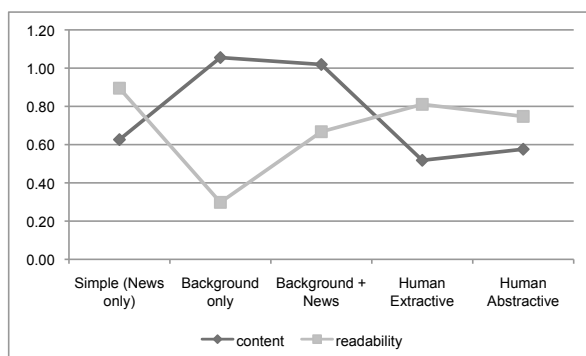


Figure 5: Standard deviation of the content and readability scores.

Figure 5 presents the standard deviation for content and readability scores: concerning content, automatically generated summaries using background information achieved the highest standard deviation scores (see also figure 6 for a sample story). That is in part supported by some commentaries made by the human evaluators on whether a summary should contain information that is not present in the input source. This aspect and the obtained results, suggest that this issue should be further analyzed, possibly using an extrinsic evaluation setup. On the other hand, readability standard deviation scores show that there is a considerable agreement in what concerns this criterion.

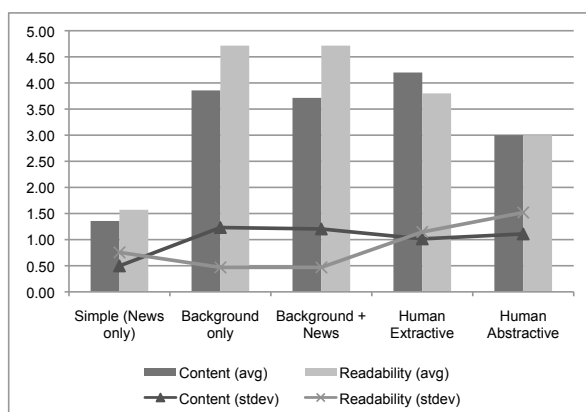


Figure 6: Average and standard deviation of the content and readability scores for one news story.

6 Conclusions

We present a new approach to speech summarization that goes in the direction of the integration of text and speech analysis, as suggested by McKeown et al. (2005). The main idea is the inclusion

of related, solid background information to cope with the difficulties of summarizing spoken language and the use of multi-document summarization techniques in single document speech-to-text summarization. In this work, we explore the possibilities offered by phonetic information to select the background information and conducted a perceptual evaluation to assess the relevance of the inclusion of that information.

The results obtained show that the human evaluators preferred human extractive summaries over human abstractive summaries. Moreover, simple LSA summaries attained the poorest results both in terms of content and readability, while human extractive summaries achieved the best performance in what concerns content, and a considerably better performance than simple LSA in what concerns readability. This suggests that it is still relevant to pursue new methods for relevance estimation. On the other hand, automatically generated summaries using background information were significantly better than simple LSA. This indicates that background information is a viable way to increase the quality of automatic summarization systems.

References

- Amaral, R. and I. Trancoso. 2004. Improving the Topic Indexation and Segmentation Modules of a Media Watch System. In *Proceedings of INTERSPEECH 2004 - ICSLP*, pages 1609–1612. ISCA.
- Amaral, R., H. Meinedo, D. Caseiro, I. Trancoso, and J. P. Neto. 2006. Automatic vs. Manual Topic Segmentation and Indexation in Broadcast News. In *Proc. of the IV Jornadas en Tecnologia del Habla*.
- Amaral, R., H. Meinedo, D. Caseiro, I. Trancoso, and J. P. Neto. 2007. A Prototype System for Selective Dissemination of Broadcast News in European Portuguese. *EURASIP Journal on Advances in Signal Processing*, 2007.
- Batista, F., D. Caseiro, N. J. Mamede, and I. Trancoso. 2007. Recovering Punctuation Marks for Automatic Speech Recognition. In *Proceedings of INTERSPEECH 2007*, pages 2153–2156. ISCA.
- Carbonell, J. and J. Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *SIGIR 1998: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–336. ACM.
- Christensen, H., Y. Gotoh, B. Kolluru, and S. Renals. 2003. Are Extractive Text Summarisation Techniques Portable To Broadcast News? In *Proceedings*

- of the *IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 489–494. IEEE.
- Edmundson, H. P. 1969. New methods in automatic abstracting. *Journal of the Association for Computing Machinery*, 16(2):264–285.
- Endres-Niggemeyer, B., J. R. Hobbs, and K. Spärck Jones, editors. 1995. *Summarizing Text for Intelligent Communication—Dagstuhl-Seminar-Report 79*. IBFI.
- Furui, S. 2007. Recent Advances in Automatic Speech Summarization. In *Proceedings of the 8th Conference on Recherche d’Information Assistée par Ordinateur (RIAO)*. Centre des Hautes Études Internationales d’Informatique Documentaire.
- Gong, Y. and X. Liu. 2001. Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. In *SIGIR 2001: Proceedings of the 24st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 19–25. ACM.
- Hori, T., C. Hori, and Y. Minami. 2003. Speech Summarization using Weighted Finite-State Transducers. In *Proceedings of the 8th EUROSPEECH - INTERSPEECH 2003*, pages 2817–2820. ISCA.
- Hovy, E., 2003. *The Oxford Handbook of Computational Linguistics*, chapter Text Summarization, pages 583–598. Oxford University Press.
- Kessler, B. 2005. Phonetic comparison algorithms. *Transactions of the Philological Society*, 103(2):243–260.
- Kikuchi, T., S. Furui, and C. Hori. 2003. Two-stage Automatic Speech Summarization by Sentence Extraction and Compaction. In *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR-2003)*, pages 207–210. ISCA.
- Lin, C. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81. ACL.
- Luhn, H. P. 1958. The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2):159–165.
- Maskey, S. and J. Hirschberg. 2005. Comparing Lexical, Acoustic/Prosodic, Structural and Discourse Features for Speech Summarization. In *Proceedings of the 9th EUROSPEECH - INTERSPEECH 2005*, pages 621–624. ISCA.
- McKeown, K. R. and D. Radev. 1995. Generating Summaries of Multiple News Articles. In *SIGIR 1995: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 74–82. ACM.
- McKeown, K. R., R. Barzilay, D. Evans, V. Hatzivassiloglou, J. L. Klavans, A. Nenkova, C. Sable, B. Schiffman, and S. Sigelman. 2002. Tracking and Summarizing News on a Daily Basis with Columbia’s Newsblaster. In *Proc. of the 2nd International Conference on Human Language Technology Research*, pages 280–285. Morgan Kaufmann.
- McKeown, K. R., J. Hirschberg, M. Galley, and S. Maskey. 2005. From Text to Speech Summarization. In *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, volume V, pages 997–1000. IEEE.
- Mohri, M. 1997. Finite-State Transducers in Language and Speech Processing. *Computational Linguistics*, 23(2):269–311.
- Murray, G., S. Renals, and J. Carletta. 2005. Extractive Summarization of Meeting Records. In *Proceedings of the 9th EUROSPEECH - INTERSPEECH 2005*, pages 593–596. ISCA.
- Murray, G., S. Renals, J. Carletta, and J. Moore. 2006. Incorporating Speaker and Discourse Features into Speech Summarization. In *Proceedings of the HLT/NAACL*, pages 367–374. ACL.
- Nenkova, A. 2006. Summarization Evaluation for Text and Speech: Issues and Approaches. In *Proceedings of INTERSPEECH 2006 - ICSLP*, pages 1527–1530. ISCA.
- Paulo, S. and L. C. Oliveira. 2002. Multilevel Annotation Of Speech Signals Using Weighted Finite State Transducers. In *Proc. of the 2002 IEEE Workshop on Speech Synthesis*, pages 111–114. IEEE.
- Radev, D., J. Otterbacher, A. Winkel, and S. Blair-Goldensohn. 2005. NewsInEssence: Summarizing Online News Topics. *Communications of the ACM*, 48(10):95–98.
- Ribeiro, R. and D. M. de Matos. 2007. Extractive Summarization of Broadcast News: Comparing Strategies for European Portuguese. In *Text, Speech and Dialogue – 10th International Conference. Proceedings*, volume 4629 of *Lecture Notes in Computer Science (Subseries LNAI)*, pages 115–122. Springer.
- Spärck Jones, K. 2007. Automatic summarising: The state of the art. *Information Processing and Management*, 43:1449–1481.
- Wan, X., J. Yang, and J. Xiao. 2007. CollabSum: Exploiting Multiple Document Clustering for Collaborative Single Document Summarizations. In *SIGIR 2007: Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 143–150. ACM.
- Zechner, K. and A. Waibel. 2000. Minimizing Word Error Rate in Textual Summaries of Spoken Language. In *Proceedings of the 1st conference of the North American chapter of the ACL*, pages 186–193. Morgan Kaufmann.