AMI Consortium

http://www.amiproject.org/

Funded under the EU Sixth Framework Programme

Multimodal interfaces action line of the IST Programme

Integrated Projects

AMI (IST-506811) and AMIDA (IST–033812)

State of the Art Report

Technologies for Automatic Summarization

November 6, 2007

AMI Consortium State of the Art Report

Technologies for Automatic Summarization

November 6, 2007

Abstract

In this document we provide a taxonomy of approaches to automatic summarization and a historical overview of both text and speech summarization. For speech summarization, we focus primarily on four popular domains of research: broadcast news, meetings, lectures and voicemail. The purpose of this document is primarily to place current speech summarization in the proper historical context, and to increase links between text summarization and speech summarization researchers.

1 Types of Summaries

One possible division of automatic summaries is between *extracts* and *abstracts*, where the former consists of units removed from the source text and concatenated together in a new, shorter document, and the latter consists of novel sentences representing the source document from a more high-level perspective. Rather than being a hard division, however, abstracts and extracts exist on a single continuum, and extracts can potentially be made more abstract-like through further interpretation or transformation of the data. Simple extracts can also be more than simply cutting and pasting; the extracted units can be compressed, made less disfluent, ordered to maximize coherence, and merged to reduce redundancy, to give a few examples.

Another possible division of summaries is between *indicative* and *informative* summaries. An *informative* summary is meant to convey the most important information of the source text, thus acting as a substitute for the original text. On the other hand, an *indicative* summary acts as a guide for where to find the most important parts of the source text. Using these definitions, the summaries we are creating in this current research can serve as either type depending on the use case. The summaries are incorporated into a meeting browser, and a time-constrained user can either read the summary in place of the entire transcript and/or use the summary as an efficient way of indexing into the meeting record.

Another division is between *multiple-document* and *single-document* summaries. In the latter case, information is gleaned from several source documents and summarized in a single output document; in these cases, redundancy is much more of an issue than with single-document summarization. In this research, we focus on summaries of individual meetings, but many of the methods are easily extendable to the task of summarizing and linking multiple archived meetings.

Similarly, this work focuses on *generic* summaries rather than *query-dependant* summaries, but the methods could be extended to query-dependent summarization. In generic summarization, each summary is created without regard to any specific information need, based on the inherent informativeness of the document. For query-dependent summarization, units are extracted based partly on how similar they are to a user-supplied query or information need.

It is possible to divide between *text* and *speech* summarization, or *text* and *multi-media* summarization, in the sense that the fields of research have separate but overlapping histories and use different types of data as input (and potentially as output as well), but of course the simplest way to approach speech summarization is to treat is as a text summarization problem, using a noisy text source. Speech summarization and text summarization approaches often use many of the same features or types of features. However, a central thesis of this work is that it is advantageous to use speech-specific features at various steps of the summarization process, compared with simply treating the problem as a text summarization task.

2 Previous Work

2.1 Text Summarization

Among the earliest work on automatic text summarization is the research by Luhn [25], who particularly focused on recognizing keywords in text. Luhn was among the first to recognize that the words with highest resolving power are words with medium or moderately high frequency in a given document.

A decade later, Edmundson [6] began to look beyond keywords for the summarization of scientific articles. He focused on four particular areas of interest: cue phrases, keywords, title words, and location. While keyword detection had been the subject of previous research the other areas were novel. Cue phrases are phrases that are very likely to signal an important sentence, and could include phrases such as “significantly”, “in conclusion” or “impossible” in the scientific articles domain. On the other hand, there are so-called Stigma phrases that may signal unimportance: specifically, these might be hedging or belittling expressions. Also particular to the type of academic articles Edmundson was working with is the Title feature, which weights each sentence according to how many times its constituent words occur in section or article titles. And finally, the Location feature weights sentences more highly if they occur under a section heading or occur very early or late in the article. Edmundson’s summarization system then works by scoring and extracting sentences based on a linear combination of these four features. These categories of features are still used today, though more often in machine-learning frameworks.

The ADAM system of the 1970s [30] relied heavily on cue phrases, but also strove to maximize coherence by analyzing whether a candidate sentence contained anaphoric references [7]. In the case that a candidate did, the system tried to either extract the preceding sentences as well or to re-write the candidate sentence so that it could stand alone. If neither of these were possible, the candidate was not chosen.

In the 1980s, several summarization methods arose that were inspired by findings in psychology and cognitive science [5, 9, 18]. These methods generally use human processing and understanding of text as a model for automatic abstraction. The source is interpreted and inferences are made based on prior knowledge. For an automatic summarization method, a schemata might be created relating to the domain of the data being summarized. What differentiates these methods from the earlier summarization methods described above is that the input is *interpreted* and *represented* more deeply than before. For example, the FRUMP system [5] uses “sketchy scripts” to model events in the real-world for the purpose of summarizing news articles. One example would be a sketchy script relating to earthquakes. We have prior knowledge about earthquakes, such as the magnitude on the Richter scale, the location of the epicenter, the number of deaths and the amount of damage inflicted. When a particular sketchy script is activated, these pieces of information are sought in the source data. An interesting overview of such approaches can be found in [7].

Summarization research underwent a major resurgence in the late 1980s and 1990s, primarily due to the explosion of data available from sources such as the web and newswire services. Because of the volume and variety of data to be summarized, the summarization techniques were more often extractive than abstractive, as the former is domain-independent,requires little or no prior knowledge, and can process a large amount of data efficiently. The field therefore tended to move away from the schema-based, cognition-inspired approaches of the 1980s.

Much of the work of this period revisited the seminal work of Edmundson [6] and his investigation of cue phrases, keywords, title words, and location features. The newer work incorporated these same features into machine-learning frameworks where classifiers are trained on human gold-standard extracts [22, 43], rather than manually tuning the weights of these features as Edmundson did. For the tasks of summarizing engineering papers [22] and computational linguistics papers [43], the most useful features were found to be cue phrases and locational features.

During this same period, other researchers investigated the use of rhetorical relations for the purpose of text summarization, particularly in the framework of Rhetorical Structure Theory (RST) [26]. A hypothesis of RST is that a given document can be represented as a single binary-branching rhetorical tree comprised of nuclei-satellite pairs, where a particular rhetorical relation exists between each nuclei-satellite pair. By pruning such a rhetorical tree, a summary of the entire text can be generated [37, 27, 28].

Contemporary work utilized linguistics resources such as WordNet, a database of lexical semantics, in order to derive relations between terms or phrases in a document. In work by Barzilay and Elhadad [1] lexical chains were detected according to the relatedness of document terms, and sentences corresponding to the strongest chains were extracted. The SUMMARIST system [15] utilized WordNet for concept detection in the summarization of news articles.

Also in the late 1990s, interest in multi-document summarization was growing. Creating a single summary of multiple documents presented, and still presents, and interesting challenge, as the summarizer must determine which documents are relevant to a given query and/or related to one another and must not extract the same information from multiple sources. In other words, the problem of *redundancy* is paramount. Carbonell and Goldstein [2] introducted the Maximal Marginal Relevance (MMR) algorithm, which scores a candidate sentence according to how relevant it is to a query (or how generally relevant, for a generic summary) and how similar it is to sentences that have already been extracted. The latter scores is used to penalize the former, thereby reducing redundancy in the resultant summary. MMR remains popular both as a stand-alone algorithm in its own right as well as a feature score in more complex summarization methods. Work by Radev et. al [39, 38] addressed single- and multi-document summarization via a centroid-method. A centroid is a pseudo-document consisting of important terms and their associated term-weight scores, representing the source document(s) as a whole. The authors address the redundancy probelem via the idea of cross-sentence information subsumption, whereby sentences that are too similar to other sentences are penalized, similar to the MMR method.

The work of Maybury [31] extended summarization work from merely processing and summarizing text to summarizing multi-modal event data. In the domain of battle simulation, the researchers took as input battle events such as missile fire, refueling, radar sweeps and movement and generated summaries based on the frequencies of such events and relations between such events. Not only are the inputs multi-modal events, but the output can be a combination of textual and graphical summaries in order to expedite perception and comprehension of the battle scene. The researchers also take into account that such summaries should be tailored to the user: for example, an intelligence officer might care more about enemy size and position whereas a logistician will care about refueling and supplies.

Since 2001, the Document Understanding Conference [1] has encouraged research in the area of multi-document, query-dependent summarization. For the text summarization community, this annual conference provides the benchmark tasks for comparing and evaluating state-of-the-art summarization systems. While the data used has primarily been newswire data, DUC has recently added tracks relating to the summarization of weblog opinions. Though a wide variety of systems have been entered in DUC, one finding is that the most competitive systems have extensive query-expansion modules. In fact, query-expansion forms the core of many of the systems [23, 16].

2.2 Speech Summarization

Chen and Withgott [3] identified areas of emphasis in speech data in order to create audio summaries, reporting results on two types of data: a recorded interview and telephone speech. The emphasis detection was carried out by training a hidden markov model on training data in which words had been manually labelled for varying degrees of emphasis. The features used in the model were purely prosodic, namely F0 and energy features. The authors reported near-human performance in selecting informative excerpts.

Rohlicek et. al [41] created brief summaries, or gists, of conversations in the air-traffic control domain. The basic summarization goals were to identify flight numbers and classify the type of flight, e.g. *takeoff* or *landing*. Such a system required components of speaker segmentation, speech recognition, natural language parsing and topic classification. The authors reported that the system achieved 98% precision of flight classification with 68% recall.

1. http://duc.nist.gov

One of the early projects on *speech* summarization was VERBMOBIL [40], a speech-to-speech translation system for the domain of travel planning. The system was capable of translating between English, Japanese and German. Though the focus of the project was on speech-to-speech translation, an abstractive summarization facility was added that exploited the information present in the translation module's knowledge sources. A user could therefore be provided with a summary of the dialogue, so that they can confirm the main points of the dialogue were translated correctly, for example. The fact that VERBMOBIL was able to incorporate abstractive summarization is due to the fact that the speech was limited to a very narrrow domain of travel planning and hotel reservation; normally it would be very difficult to create such structured abstracts in unrestricted domains.

Simultaneously work was being carried out on the MIMI dialogue summarizer [20], which was used for the summarization of spontaneous conversations in Japanese. Like VERBMOBIL, these dialogues were in a limited domain; in this case, negotiations for booking meetings rooms. The system creates a running transcript of the transactions so far, by recognizing domain-specific patterns and merging redundant information.

2.2.1 Summarization of Newscasts

One of the domains of speech summarization that has received the most attention and perhaps has the longest history is the domain of broadcast news summarization. Summarizing broadcast news is an interesting task, as the data consists of both spontaneous and read segments and so represents a middle-ground between text and spontaneous speech summarization. In Hirschberg et. al [12], a user interface tool is provided for browsing and information retrieval of spoken audio - in this case, National Public Radio broadcasts. The browser adds audio paragraphs, or *paratones*, to the speech transcript, using intonational information. This is a good example of how structure can be added to unstructured speech data in order make it more readable as well as more amenable to subsequent analysis incorporating structural features. Their browser also highlights keywords in the transcript based on acoustic and lexical information.

In Valenza et. al [44], summarization of the American Broadcast News corpus is carried out by weighting terms according to an acoustic confidence measure and a term-weighting metric from information retrieval called inverse frequency (described in detail in a later chapter). The units of extraction are n-grams, utterances and keywords, which are scored according to the normalized sums of their constituent words in the case of n-grams and utterances. When a user desires a low word-error rate (WER) above all else, a weighting parameter can be changed to favor the acoustic confidence score over the lexical score. One of the most interesting results of this work is that the WER of summaries portions are typically much lower than the overall WER of the source data, a finding that has since been attested in other work [33]. Valenza et. al also provide a simple but intuitive interface for browing the recognizer output.

In work by Hori and Furui [13] on Japanese broadcast news summarization, each sentence has a subset of its words extracted based on each word's topic score – a measure of its significance – and a concatenation likelihood, the likelihood of the word being concatenated to the previously extracted segment. Using this method, they report that 86% of the important words in the test set are extracted.

More recently in the broadcast news domain, Maskey and Hirschberg [29] found that the best summarization results utilized prosodic, lexical and structural features, but that prosodic features alone resulted in good-quality summarization. The prosodic features they investigated were broadly features of pitch, energy, speaking rate and sentence duration. Work by Ohtake et. al [36] explored using *only* prosodic features for speech-to-speech summarization of Japanese newscasts, finding that such summaries rated comparably with a system relying on speech recognition output.

2.2.2 Summarization of Meetings

In the domain of meetings, Waibel et. al [45] implemented a modified version of maximal marginal relevance applied to speech transcripts, presenting the user with the *n* best sentences in a meeting browser interface. The browser contained several information streams for efficient meeting access, such as topic-tracking, speaker activity, audio/video recordings and automatically-generated summaries. However, the authors did not research any speech-specific information for summarization; this work was purely text summarization applied to speech transcripts.

Zechner [46] investigated summarizing several genres of speech, including spontaneous meeting speech. Though relevance detection in his work relied largely on *tf.idf* scores, Zechner also explored cross-speaker information linking and question/answer detection, so that utterances could be extracted not only according to high *tf.idf* scores, but also if they were linked to other informative utterances.

On the ICSI corpus, Galley [10] used skip-chain Conditional Random Fields to model pragmatic dependencies such as QUESTION-ANSWER between paired meeting utterances, and used a combination of lexical, prosodic, structural and discourse features to rank utterances by importance. The types of features used were classified as *lexical features*, *information retrieval features*, *acoustic features*, *structural and durational featuers* and *discourse features*. Galley found that while the most useful single feature class was *lexical* features, a combination of acoustic, durational and structural features exhibited comparable performance according to Pyramid evaluation.

Simpson and Gotoh [42], also working with the ICSI meeting corpus, investigated speaker-independent prosodic features for meeting summarization. A problem of working with features relying on absolute measurements of pitch and energy is that these features vary greatly depending on the speaker and the meeting conditions, and thus require normalization. The authors therefore investigated the usefulness of speaker-indepedent features such as pauses, pitch and energy changes across pauses, and pitch and energy changes across units. They found that pause durations and pitch changes across units were the most consistent features across multiple speakers and multiple meetings.

Liu et. al [24] reported the results of a pilot study on the the effect of disfluencies on automatic speech summarization, using the ICSI corpus. They found that the manual removal of disfluencies did not improve summarization performance according to the ROUGE metric.

In our own work on the ICSI corpus, Murray et al. [33, 34] compared text summarization approaches with feature-based approaches incorporating prosodic features, with human judges favoring the feature-based approaches. In subsequent work [35], we began to look at additional speech-specific characteristics such as speaker and discourse features. One significant finding of these papers was that the ROUGE evaluation metric did not correlate well with human judgments on this test data.

2.2.3 Summarization of Lectures

Hori et al. [14] have developed an integrated speech summarization approach, based on finite state transducers, in which the recognition and summarization components are composed into a single finite state transducer, reporting results on a lecture summarization task.

Also in the lectures domain, Fujii et. al [8] attempted to label cue phrases and use cue phrase features in order to supplement lexical and prosodic features in extractive summarization. They reported that the use of cue phrases for summarization improved the summaries according to both f-scores and ROUGE scores.

Zhang et. al [47] compared feature types for summarization across domains, concentrating on lecture speech and broadcast news speech in Mandarin. They found that acoustic and structural features are more important for broadcast news than for the lecture task, and that the quality of broadcast news summaries is less dependent on ASR performance.

2.2.4 Voicemail Summarization

The SCANMail system [11] was developed to allow a user to navigate their voicemail messages in a graphical user interface. The system incorporated information retrieval and information extraction components, allowing a user to query the voicemail messages, and automatically extracting relevant information such as phone numbers. Huang et. al [17] and Jansche and Abbey [19] also described techniques for extracting phone numbers from voicemails.

Koumpis and Renals [21] investigated prosodic features for summarizing voicemail messages in order to send voicemail summaries to mobile devices. They reported that while the optimal feature subset for classification was the lexical subset, an advantage could be had by augmenting those lexical features with prosodic features, especially pitch range and pause information.

2.3 From Text to Speech

McKeown et. al [32] provided an overview of text summarization approaches and discussed how text-based methods might be extended to speech data. The authors described the challenges in summarizing differing speech genres such as Broadcast News and meeting speech and which features are useful in each of those domains. Their summarization work involved components of speaker segmentation, topic segmentation, detection of agreement/disagreement, and prosodic modelling, among others.

Christensen et. al [4] investigated how well text summarization techniques for newswire data could be extended to broadcast news summarization. In analyzing feature subsets, they found that positional features were more useful for text summarization than for broadcast news summarization and that positional features alone provided very good results for text. In contrast, no single feature set in their speech summarization experiments was as dominant, and all of the features involving position, length, term-weights and named entities made significant contributions to classification. They also found that increased word-error rate only caused slight degradation according to their automatic metrics, but that human judges rated the error-filled summaries much more severely.

3 Conclusion

In this document we have provided an overview of summarization types and a literature review of both text and speech summarization, looking particularly at speech domains of broadcast news, meetings and lectures. While there are certainly other interesting speech genres, speech summarization research has been focused largely on these few domains to date. It is hoped that by reviewing text summarization and speech summarization together, the best ideas of one community can inform the other and increase links between the parallel fields of research.

References

[1] R. Barzilay and M. Elhadad. Using lexical chains for summarisation. In *Proc. of ACL 1997, Madrid, Spain*, pages 10–18, 1997.

[2] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proc. of ACM SIGIR Conference on Research and Development in Information Retrieval 1998, Melbourne, Australia*, pages 335–336, 1998.

[3] F. Chen and M. Withgott. The use of emphasis to automatically summarize a spoken discourse. In *Proc. of ICASSP 1992, San Francisco, USA*, pages 229–232, 1992.

[4] H. Christensen, Y. Gotoh, B. Kolluru, and S. Renals. Are extractive text summarisation techniques portable to broadcast news? In *Proc. of IEEE Speech Recognition and Understanding Workshop, St. Thomas, USVI*, pages 489–494, 2003.

[5] G. DeJong. An overview of the FRUMP system. In W. G. Lehnert and M. H. Ringle, editors, *Strategies for Natural Language Processing*, pages 149–176. Lawrence Erlbaum Associates, Publishers, 1982.

[6] H. P. Edmundson. New methods in automatic extracting. *J. ACM*, 16(2):264–285, 1969.

[7] B. Endres-Niggemeyer. *Summarizing Information.* Springer, Berlin, 1998.

[8] Y. Fujii, N. Kitaoka, and S. Nakagawa. Automatic extraction of cue phrases for important sentences in lecture speech and automatic lecture speech summarization. In *Proc. of Interspeech 2007, Antwerp, Belgium*, pages 2801–2804, 2007.

[9] D. Fum, G. Guida, and C. Tasso. Forward and backward reasoning in automatic abstracting. In *Proc. of the (COLING '82), Prague, Czech Republic*, pages 83–88, 1982.

[10] M. Galley. A skip-chain conditional random field for ranking meeting utterances by importance. In *Proc. of EMNLP 2006, Sydney, Australia*, pages 364–372, 2006.

[11] J. Hirschberg, M. Bacchiani, D. Hindle, P. Isenhour, A. Rosenberg, L. Stark, L. Stead, S. Whittaker, and G. Zamchick. Scanmail: Browsing and searching speech data by content. In *Proc. of Interspeech 2001, Aalborg, Denmark*, pages 1299–1302, 2001.

[12] J. Hirschberg, S. Whittaker, D. Hindle, F. Pereira, and A. Singhal. Finding information in audio: A new paradigm for audio browsing and retrieval, 1999.

[13] C. Hori and S. Furui. Automatic speech summarization based on word significance and linguistic likelihood. In *Proc. of ICASSP 2000, Istanbul, Turkey*, pages 1579–1582, 2000.

[14] T. Hori, C. Hori, and Y. Minami. Speech summarization using weighted finite-state transducers. In *Proc. of Interspeech 2003, Geneva, Switzerland*, pages 2817–2820, 2003.

[15] E. Hovy and C.-Y. Lin. Automated text summarization in summarist. In I. Mani and M. T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 81–94. MITP, 1999.

[16] E. Hovy, C.-Y. Lin, and L. Zhou. A BE-based multi-document summarizer with query interpretation. In *Proc. of DUC 2005, Vancouver, CA*, 2005.

[17] J. Huang, G. Zweig, and M. Padmanabhan. Information extraction from voicemail. In *Proc. of ACL 2001, Toulouse, France*, pages 290–297, 2001.

[18] P. Jacobs and L. Rau. SCISOR: Extracting information from on-line news. *CACM*, 33(11):88–97, 1990.

[19] M. Jansche and S. Abney. Information extraction from voicemail transcripts. In *Proc. of EMNLP 2002, Philadelphia, USA*, pages 320–327, 2002.

[20] M. Kameyama and I. Arima. Coping with aboutness complexity in information extraction from spoken dialogues. pages 87–90, 1994.

[21] K. Koumpis and S. Renals. Automatic summarization of voicemail messages using lexical and prosodic features. *ACM Transactions on Speech and Language Processing*, 2:1–24, 2005.

[22] J. Kupiec, J. Pederson, and F. Chen. A trainable document summarizer. In *Proc. of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington, USA*, pages 68–73, 1995.

[23] F. Lacatusu, A. Hickl, P. Aarseth, and L. Taylor. Lite-GISTexter at DUC 2005. In *Proc. of DUC 2005, Vancouver, CA*, 2005.

[24] Y. Liu, F. Liu, B. Li, and S. Xie. Do disfluencies affect meeting summarization: A pilot study on the impact of disfluencies. In *Proc. of MLMI 2007, Brno, Czech Republic*, page poster, 2007.

[25] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research Development*, 2(2):159–165, 1958.

[26] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

[27] D. Marcu. Discourse trees are good indicators of importance in text. In I. Mani and M. T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 123–136. MITP, Cambridge, MA, 1995.

[28] D. Marcu. From discourse structures to text summaries. In *Proc. of ACL 1997*

Workshop on Intelligent Scalable Text Summarization, Madrid, Spain, pages 82–88, 1997.

[29] S. Maskey and J. Hirschberg. Comparing lexial, acoustic/prosodic, discourse and structural features for speech summarization. In *Proc. of Interspeech 2005, Lisbon, Portugal*, pages 621–624, 2005.

[30] B. Mathis. *Techniques for the evaluation and improvement of computer-produced abstracts.* Ohio State University Technical Report OSU-CISRC-TR-72-15, Ohio, USA, 1972.

[31] M. Maybury. Generating summaries from event data. *IPM*, 31(5):735–751, September 1995.

[32] K. McKeown, J. Hirschberg, M. Galley, and S. Maskey. From text to speech summarization. In *Proc. of ICASSP 2005, Philadelphia, USA*, pages 997–1000, 2005.

[33] G. Murray, S. Renals, and J. Carletta. Extractive summarization of meeting recordings. In *Proc. of Interspeech 2005, Lisbon, Portugal*, pages 593–596, 2005.

[34] G. Murray, S. Renals, J. Carletta, and J. Moore. Evaluating automatic summaries of meeting recordings. In *Proc. of the ACL 2005 MTSE Workshop, Ann Arbor, MI, USA*, pages 33–40, 2005.

[35] G. Murray, S. Renals, J. Moore, and J. Carletta. Incorporating speaker and discourse features into speech summarization. In *Proc. of the HLT-NAACL 2006, New York City, USA*, pages 367–374, 2006.

[36] K. Ohtake, K. Yamamoto, Y. Toma, S. Sado, S. Masuyama, and S. Nakagawa. Newscast speech summarization via sentence shortening based on prosodic features. In *Proc. of the ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition, Tokyo, Japan*, pages 167–170, 2003.

[37] K. Ono, K. Sumita, and S. Miike. Abstract generation based on rhetorical structure extraction. In *Proc. of COLING 1994, Kyoto, Japan*, pages 344–348, 1994.

[38] D. Radev, S. Blair-Goldensohn, and Z. Zhang. Experiments in single and multi-document summarization using mead. In *Proc. of DUC 2001, New Orleans, LA, USA*, 2001.

[39] D. Radev, H. Jing, and M. Budzikowska. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation and user studies. In *Proc. of the ANLP/NAACL 2000 Workshop, Seattle, WA*, pages 21–29, 2000.

[40] N. Reithinger, M. Kipp, R. Engel, and J. Alexandersson. Summarizing multilingual spoken negotiation dialogues. In *Proc. of ACL 2000, Hong Kong*, pages 310–317, Morristown, NJ, USA, 2000. Association for Computational Linguistics.

[41] J. R. Rohlicek. Gisting continuous speech. In *Proc. of ICASSP 1992, San Francisco, USA*, pages 384–384, 1992.

[42] S. Simpson and Y. Gotoh. Towards speaker independent features for information extraction from meeting audio data. In *Proc. of MLMI 2005, Edinburgh, UK*, page poster, 2005.

[43] S. Teufel and M. Moens. Sentence extraction as a classification task. In *Proc. of ACL 1997, Workshop on Intelligent and Scalable Text Summarization, Madrid, Spain*, pages 58–65, 1997.

[44] R. Valenza, T. Robinson, M. Hickey, and R. Tucker. Summarization of spoken audio through information extraction. In *Proc. of the ESCA Workshop on Accessing Information in Spoken Audio, Cambridge UK*, pages 111–116, 1999.

[45] A. Waibel, M. Bett, M. Finke, and R. Stiefelhagen. Meeting browser: Tracking and summarizing meetings. In D. E. M. Penrose, editor, *Proc. of the Broadcast News Transcription and Understanding Workshop, Lansdowne, VA, USA*, pages 281–286, 1998.

[46] K. Zechner. Automatic summarization of open-domain multiparty dialogues in diverse genres. *Computational Linguistics*, 28(4):447–485, 2002.

[47] J. Zhang, H. Chan, P. Fung, and L. Cao. Comparative study on speech summarization of broadcast news and lecture speech. In *Proc. of Interspeech 2007, Antwerp, Belgium*, pages 2781–2784, 2007.