Spoken Content Retrieval: Challenges, Techniques and Applications

(Part 4: Beyond ASR Transcripts)

Gareth J. F. Jones

Centre for Next Generation Localisation
School of Computing, Dublin City University, Dublin, Ireland

Overview

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.
  - ASR may not be the best source of a transcript.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.
  - ASR may not be the best source of a transcript.
  - The actual words spoken may not be sufficient to support retrieval.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.
  - ASR may not be the best source of a transcript.
  - The actual words spoken may not be sufficient to support retrieval.
  - There may be data available which can be exploited to improve retrieval effectiveness.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.
  - ASR may not be the best source of a transcript.
  - The actual words spoken may not be sufficient to support retrieval.
  - There may be data available which can be exploited to improve retrieval effectiveness.
  - Techniques from text IR may be applied to improve retrieval effectiveness.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.
  - ASR may not be the best source of a transcript.
  - The actual words spoken may not be sufficient to support retrieval.
  - There may be data available which can be exploited to improve retrieval effectiveness.
  - Techniques from text IR may be applied to improve retrieval effectiveness.
  - It may not be clear what the retrieval units should be.

Introduction

- When thinking about SCR it is natural to think in terms of basing the retrieval on ASR transcripts of each recorded media file.
- While IR algorithms are robust to quite high levels of ASR errors, there are many reasons to look beyond them to improve the retrieval effectiveness.
  - ASR may not be the best source of a transcript.
  - The actual words spoken may not be sufficient to support retrieval.
  - There may be data available which can be exploited to improve retrieval effectiveness.
  - Techniques from text IR may be applied to improve retrieval effectiveness.
  - It may not be clear what the retrieval units should be.
- It is important to consider how to evaluate retrieval effectiveness for SCR tasks.

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?
  - Is there an alternative and better way of creating a transcript?

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?
  - Is there an alternative and better way of creating a transcript?
- Existing transcripts: close-captions, e,g, for broadcast TV; manual transcript, e.g. parliamentary proceedings.

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?
  - Is there an alternative and better way of creating a transcript?
- Existing transcripts: close-captions, e,g, for broadcast TV; manual transcript, e.g. parliamentary proceedings.
- Manually transcribe the content, e.g. via crowdsourcing.

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?
  - Is there an alternative and better way of creating a transcript?
- Existing transcripts: close-captions, e,g, for broadcast TV; manual transcript, e.g. parliamentary proceedings.
- Manually transcribe the content, e.g. via crowdsourcing.
  - How to ensure accuracy?

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?
  - Is there an alternative and better way of creating a transcript?
- Existing transcripts: close-captions, e,g, for broadcast TV; manual transcript, e.g. parliamentary proceedings.
- Manually transcribe the content, e.g. via crowdsourcing.
  - How to ensure accuracy?
  - Who should the workers be? -

Is ASR the best source of speech transcripts?

- Before using an ASR system to create a transcript, and especially before developing or training a new ASR to make a transcript in a particular domain, ask:
  - Is there already a transcript available?
  - Is there an alternative and better way of creating a transcript?
- Existing transcripts: close-captions, e,g, for broadcast TV; manual transcript, e.g. parliamentary proceedings.
- Manually transcribe the content, e.g. via crowdsourcing.
  - How to ensure accuracy?
  - Who should the workers be? - issues of confidentiality of the material, e,g, medical or enterprise content.

The Content of Speech Media

- Depending on the source of the speech stream, it can be underspecified in terms of the information content it contains.

The Content of Speech Media

- Depending on the source of the speech stream, it can be underspecified in terms of the information content it contains.
- When SCR moves away from domains involving planned speech such as broadcast news, speech is frequently informal or spontaneously produced.

The Content of Speech Media

- Depending on the source of the speech stream, it can be underspecified in terms of the information content it contains.
- When SCR moves away from domains involving planned speech such as broadcast news, speech is frequently informal or spontaneously produced.
- In these latter cases, the meaning conveyed by the spoken content is strongly dependent on the context:

The Content of Speech Media

- Depending on the source of the speech stream, it can be underspecified in terms of the information content it contains.
- When SCR moves away from domains involving planned speech such as broadcast news, speech is frequently informal or spontaneously produced.
- In these latter cases, the meaning conveyed by the spoken content is strongly dependent on the context:
  - knowledge that is shared between the speakers, e.g. they may use pronouns without specifying the entity to which they are referring.

The Content of Speech Media

- Depending on the source of the speech stream, it can be underspecified in terms of the information content it contains.
- When SCR moves away from domains involving planned speech such as broadcast news, speech is frequently informal or spontaneously produced.
- In these latter cases, the meaning conveyed by the spoken content is strongly dependent on the context:
  - knowledge that is shared between the speakers, e.g. they may use pronouns without specifying the entity to which they are referring.
  - the setting in which the speech is produced, e.g. long multi-topic discussion or short voice message.

The Content of Speech Media

- Informal, spontaneous speech generally has a high WER, often well above the 20% level to which SCR is generally considered robust.

The Content of Speech Media

- Informal, spontaneous speech generally has a high WER, often well above the 20% level to which SCR is generally considered robust.
- News sources often have multiple documents containing the same information, e.g. stories on a specific topic.

The Content of Speech Media

- Informal, spontaneous speech generally has a high WER, often well above the 20% level to which SCR is generally considered robust.
- News sources often have multiple documents containing the same information, e.g. stories on a specific topic.
  - Retrieval of any one of these may satisfy the user's information need.

The Content of Speech Media

- Informal, spontaneous speech generally has a high WER, often well above the 20% level to which SCR is generally considered robust.
- News sources often have multiple documents containing the same information, e.g. stories on a specific topic.
  - Retrieval of any one of these may satisfy the user's information need.
  - In other content there may only be single relevant recording, e.g. the details of a discussion in a meeting.

The Content of Speech Media

- Informal, spontaneous speech generally has a high WER, often well above the 20% level to which SCR is generally considered robust.
- News sources often have multiple documents containing the same information, e.g. stories on a specific topic.
  - Retrieval of any one of these may satisfy the user's information need.
  - In other content there may only be single relevant recording, e.g. the details of a discussion in a meeting.
    - The SCR must retrieve **this** recording, nothing else will do.

The Content of Speech Media

- Informal, spontaneous speech generally has a high WER, often well above the 20% level to which SCR is generally considered robust.
- News sources often have multiple documents containing the same information, e.g. stories on a specific topic.
  - Retrieval of any one of these may satisfy the user's information need.
  - In other content there may only be single relevant recording, e.g. the details of a discussion in a meeting.
    - The SCR must retrieve **this** recording, nothing else will do.
    - It is often not clear what retrieval unit to use in this setting.

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview
  - casual - oral testimony

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview
  - casual - oral testimony
- Structured presentation - lecture

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview
  - casual - oral testimony
- Structured presentation - lecture
- Focussed exchange:

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview
  - casual - oral testimony
- Structured presentation - lecture
- Focussed exchange:
  - professional - business meeting

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview
  - casual - oral testimony
- Structured presentation - lecture
- Focussed exchange:
  - professional - business meeting
  - casual - conversation between friends

From “spoken documents” to “spontaneous speech”

- Segmented read speech documents - news articles
- Structured conversation:
  - professional - news interview
  - casual - oral testimony
- Structured presentation - lecture
- Focussed exchange:
  - professional - business meeting
  - casual - conversation between friends
- Everything - life log archive.

Exploiting Metadata

- Much spoken content is accompanied by textual metadata.

Exploiting Metadata

- Much spoken content is accompanied by textual metadata.
- This may be about the content such as: title, creator, source, names of speakers, date of recording, language spoken.

Exploiting Metadata

- Much spoken content is accompanied by textual metadata.
- This may be about the content such as: title, creator, source, names of speakers, date of recording, language spoken.
- or it may summarize the content:

Exploiting Metadata

- Much spoken content is accompanied by textual metadata.
- This may be about the content such as: title, creator, source, names of speakers, date of recording, language spoken.
- or it may summarize the content:
  - assigned keywords: manually or automatically selected.

Exploiting Metadata

- Much spoken content is accompanied by textual metadata.
- This may be about the content such as: title, creator, source, names of speakers, date of recording, language spoken.
- or it may summarize the content:
  - assigned keywords: manually or automatically selected.
  - content summary: either extracted from the transcript, or created to describe the content in some way.

Exploiting Metadata

- In search of voice mail, dates, sender, etc can be used as simple filters to limited the search space.

Exploiting Metadata

- In search of voice mail, dates, sender, etc can be used as simple filters to limited the search space.
- Professional interviews may be accompanied by descriptive metadata assigned by domain experts and/or from a domain specific ontology.

Exploiting Metadata

- In search of voice mail, dates, sender, etc can be used as simple filters to limited the search space.
- Professional interviews may be accompanied by descriptive metadata assigned by domain experts and/or from a domain specific ontology.
  - e.g. oral history recordings of the Shoah Foundaion Institute used in the Malach collection are accompanied by short text summaries created by and domain keywords assigned by domain experts.

Exploiting Metadata

- In search of voice mail, dates, sender, etc can be used as simple filters to limited the search space.
- Professional interviews may be accompanied by descriptive metadata assigned by domain experts and/or from a domain specific ontology.
  - e.g. oral history recordings of the Shoah Foundaion Institute used in the Malach collection are accompanied by short text summaries created by and domain keywords assigned by domain experts.
- Recordings of lectures often accompanied by slide presentations, and possibly written notes or a textbook.

Exploiting Metadata

- In search of voice mail, dates, sender, etc can be used as simple filters to limited the search space.
- Professional interviews may be accompanied by descriptive metadata assigned by domain experts and/or from a domain specific ontology.
  - e.g. oral history recordings of the Shoah Foundaion Institute used in the Malach collection are accompanied by short text summaries created by and domain keywords assigned by domain experts.
- Recordings of lectures often accompanied by slide presentations, and possibly written notes or a textbook.
- Meetings may have associated minutes, and relate to a number of institutional or professional documents.

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.
- Results showed clearly that SCR based only on the ASR transcript was very poor.

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.
- Results showed clearly that SCR based only on the ASR transcript was very poor.
  - WER $\approx$ 25% - not particularly high

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.
- Results showed clearly that SCR based only on the ASR transcript was very poor.
  - WER $\approx$ 25% - not particularly high
  - very little content information spoken, little for queries to match against

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.
- Results showed clearly that SCR based only on the ASR transcript was very poor.
  - WER $\approx$ 25% - not particularly high
  - very little content information spoken, little for queries to match against
- Using manual summaries or manually assigned keywords very effective for search.

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.
- Results showed clearly that SCR based only on the ASR transcript was very poor.
  - WER $\approx 25\%$ - not particularly high
  - very little content information spoken, little for queries to match against
- Using manual summaries or manually assigned keywords very effective for search.
- Automatically assigned keywords much less effective for search.

Searching the Malach Collection

- The Malach collection was used as the focus of the CLEF Cross-Langauge Speech task 2005-2007.
- Results showed clearly that SCR based only on the ASR transcript was very poor.
  - WER $\approx$ 25% - not particularly high
  - very little content information spoken, little for queries to match against
- Using manual summaries or manually assigned keywords very effective for search.
- Automatically assigned keywords much less effective for search.
- Combining metadata with ASR transcripts generally produced a small overall improvement over metadata only retrieval.

Searching Lecture Recordings

- Basic metadata can be used to support search: name of lecture, name of course with which the lecture if associated, name of lecturer, venue at which lecture was delivred, etc.

Searching Lecture Recordings

- Basic metadata can be used to support search: name of lecture, name of course with which the lecture if associated, name of lecturer, venue at which lecture was delivred, etc.
- Slides used to deliver lecture can be aligned with a (noisy) ASR transcript.

Searching Lecture Recordings

- Basic metadata can be used to support search: name of lecture, name of course with which the lecture if associated, name of lecturer, venue at which lecture was delivred, etc.
- Slides used to deliver lecture can be aligned with a (noisy) ASR transcript.
  - Words on the slides can compensate for errors in the ASR transcript.

Searching Lecture Recordings

- Basic metadata can be used to support search: name of lecture, name of course with which the lecture if associated, name of lecturer, venue at which lecture was delivred, etc.
- Slides used to deliver lecture can be aligned with a (noisy) ASR transcript.
  - Words on the slides can compensate for errors in the ASR transcript.
  - Domain specific words which are OOV of the ASR system, are likely to appear in the slides, will thus be available to match with user queries.

Other Metadata

- Recognised features such as: speaker change points, identification of speakers, silence points, music or speech, non-verbal features (e.g. applause, laughter), channel (e.g. telephone vs desktop microphone.

Other Metadata

- Recognised features such as: speaker change points, identification of speakers, silence points, music or speech, non-verbal features (e.g. applause, laughter), channel (e.g. telephone vs desktop microphone.
- Extraction of affective (emotional) features - areas of emotional intensity, etc.

Other Metadata

- Recognised features such as: speaker change points, identification of speakers, silence points, music or speech, non-verbal features (e.g. applause, laughter), channel (e.g. telephone vs desktop microphone.
- Extraction of affective (emotional) features - areas of emotional intensity, etc.
- Such features can be used for filtering or displayed in a user interface to show structure of complex audio, e.g. the structure of a discussion in a meeting.

Expansion Techniques

- Query Expansion:

Expansion Techniques

- Query Expansion:
  - Compensates for ASR errors and can help address problems arising from OOV.

Expansion Techniques

- Query Expansion:
  - Compensates for ASR errors and can help address problems arising from OOV.
  - Consider Robertson expansion offer weight $ow(i) \approx idf(i) \times r(i)$ where $r(i)$ is the number of known relevant documents.

Expansion Techniques

- Query Expansion:
  - Compensates for ASR errors and can help address problems arising from OOV.
  - Consider Robertson expansion offer weight $ow(i) \approx idf(i) \times r(i)$ where $r(i)$ is the number of known relevant documents.
    $idf(i)$ values effectively smoothed by behaviour of LVCSR system.

Expansion Techniques

- Query Expansion:
  - Compensates for ASR errors and can help address problems arising from OOV.
  - Consider Robertson expansion offer weight $ow(i) \approx idf(i) \times r(i)$ where $r(i)$ is the number of known relevant documents.
    $idf(i)$ values effectively smoothed by behaviour of LVCSR system.
    Query expansion is robust in SCR. (Lam-adesina & Jones, 2006)
- Document Expansion:

Expansion Techniques

- Query Expansion:
  - Compensates for ASR errors and can help address problems arising from OOV.
  - Consider Robertson expansion offer weight $ow(i) \approx idf(i) \times r(i)$ where $r(i)$ is the number of known relevant documents.
    $idf(i)$ values effectively smoothed by behaviour of LVCSR system.
    Query expansion is robust in SCR. (Lam-adesina & Jones, 2006)
- Document Expansion:
  - Use document as a query to ASR collection or a related text collection.

Expansion Techniques

- Query Expansion:
  - Compensates for ASR errors and can help address problems arising from OOV.
  - Consider Robertson expansion offer weight $ow(i) \approx idf(i) \times r(i)$ where $r(i)$ is the number of known relevant documents.
    $idf(i)$ values effectively smoothed by behaviour of LVCSR system.
    Query expansion is robust in SCR. (Lam-adesina & Jones, 2006)
- Document Expansion:
  - Use document as a query to ASR collection or a related text collection.
  - Expand document to include words that are topically related, potentially spoken but misrecognised or OOV.

Extracting Retrieval Units

- Some spoken content forms natural retrieval units, e.g. voice messages.

Extracting Retrieval Units

- Some spoken content forms natural retrieval units, e.g. voice messages.
- Other content is easily segmented into clear single topic units retrieval, e.g. broadcast news.

Extracting Retrieval Units

- Some spoken content forms natural retrieval units, e.g. voice messages.
- Other content is easily segmented into clear single topic units retrieval, e.g. broadcast news.
  - Segmentation can be manual or automatic - automatic can be noisy.

Extracting Retrieval Units

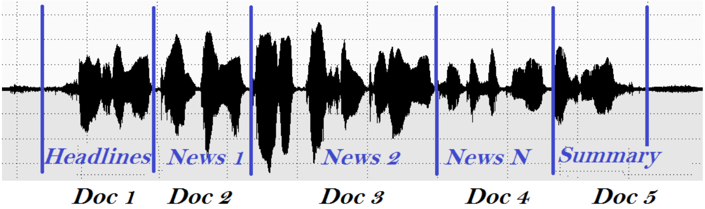
- Some spoken content forms natural retrieval units, e.g. voice messages.
- Other content is easily segmented into clear single topic units retrieval, e.g. broadcast news.
  - Segmentation can be manual or automatic - automatic can be noisy.
  - Automatic techniques borrowed from text segmentation, applied to ASR transcripts, ASR errors can affect segmentation behaviour.
- But other content cannot easily be segmented umambiguously into obvious topical units, e.g. meetings, where segmentation may be subjective or be query-dependent.

Extracting Retrieval Units

- Broadcast news:

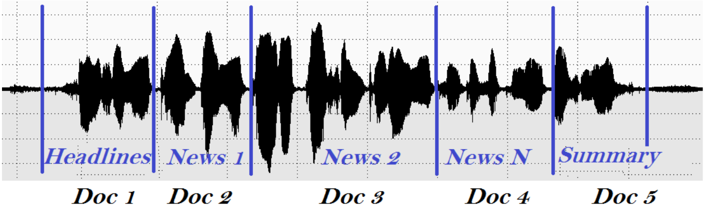
Extracting Retrieval Units

- Broadcast news:



Extracting Retrieval Units

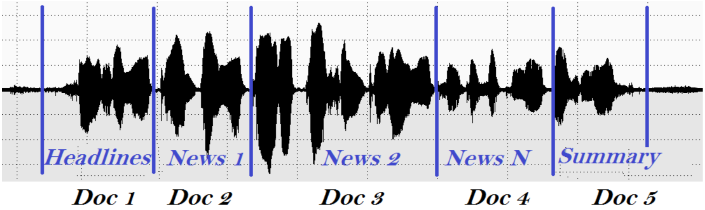
- Broadcast news:



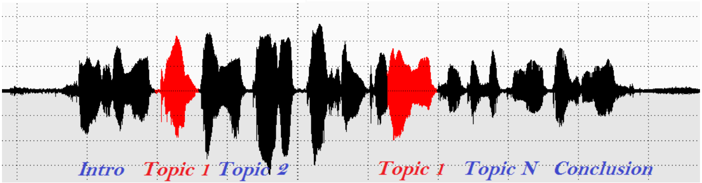
- Meetings:

Extracting Retrieval Units

- Broadcast news:



- Meetings:



Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:

Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:
  - fixed length, potentially overlapping segments;

Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:
  - fixed length, potentially overlapping segments;
  - based on automatic segmentation algorithms.

Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:
  - fixed length, potentially overlapping segments;
  - based on automatic segmentation algorithms.
- Extracted segments will generally only partially overlap with relevant content in spoken content.

Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:
  - fixed length, potentially overlapping segments;
  - based on automatic segmentation algorithms.
- Extracted segments will generally only partially overlap with relevant content in spoken content.
- Relevant content may be split between multiple segments.

Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:
  - fixed length, potentially overlapping segments;
  - based on automatic segmentation algorithms.
- Extracted segments will generally only partially overlap with relevant content in spoken content.
- Relevant content may be split between multiple segments.
  - Improved segmentation of content with respect to relevant content is a research challenge.

Extracting Retrieval Units

- Whether or not the content can be umambigously segmented by a human listener, for SCR, retrieval units can be extracted using various methods, including:
  - fixed length, potentially overlapping segments;
  - based on automatic segmentation algorithms.
- Extracted segments will generally only partially overlap with relevant content in spoken content.
- Relevant content may be split between multiple segments.
  - Improved segmentation of content with respect to relevant content is a research challenge.
- Ideally user should receive segments containing relevant content, with suggested "jump-in" point where they should start listening to the content.

Evaluation

SCR can be evaluated using various standard and new metrics.

Evaluation

SCR can be evaluated using various standard and new metrics.
Standard text retrieval ranking MAP metric.

Evaluation

SCR can be evaluated using various standard and new metrics.
Standard text retrieval ranking MAP metric.
Average Precision

$$AP = \frac{1}{n}.\sum_{r=1}^{N} P[r].rel(r)$$

where:
n - no of relevant documents
N = total number of documents retrieved
$P[r] =$ precision at rank r
$rel(r)$ - relevance at rank r, $rel(r) = 1$ if document is relevant,
$rel(r) = 0$ if it is non-relevant
MAP = AP averaged over a set of test queries

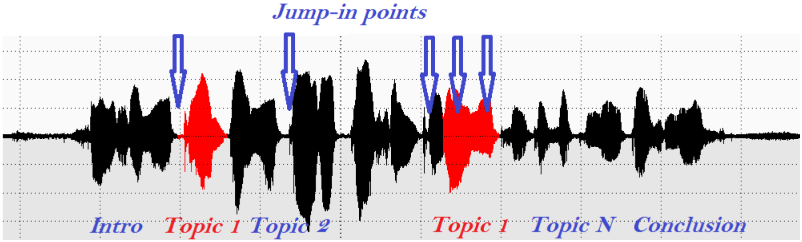
Evaluation

Note: MAP gives no indication of how much time user must spend listening to locate relevant content.

Evaluation

Note: MAP gives no indication of how much time user must spend listening to locate relevant content.

Alternative, measure effectiveness of retrieval of jump-in points in time for relevant content.



Evaluation

Generalized Average Precision (GAP):

$$GAP = \frac{1}{n} \cdot \sum_{r=1}^{N} P[r] \cdot \left(1 - \frac{Distance}{Granularity} \cdot 0.1\right)$$

where:
Distance = distance from start of segment to start of relevant content part.
Granularity = step for penalty function (granularity = 15 seconds at CLEF) - segments where relevant content starts after more then 150 second are considered non-relevant.
(Kekalainen & Jarvelin, 2002)(Liu & Oard, 2006)(Galuscáková, Pecina, & Hajic, 2012)

Evaluation

Generalized Average Precision (GAP):

$$GAP = \frac{1}{n} . \sum_{r=1}^{N} P[r] \cdot \left(1 - \frac{Distance}{Granularity} \cdot 0.1\right)$$

where:
Distance = distance from start of segment to start of relevant content part.
Granularity = step for penalty function (granularity = 15 seconds at CLEF) - segments where relevant content starts after more then 150 second are considered non-relevant.
(Kekalainen & Jarvelin, 2002)(Liu & Oard, 2006)(Galuscáková, Pecina, & Hajic, 2012)
Note: mGAP does not take into account how much time the user needs to spend listening to access the relevant content.

Evaluation

Segment Precision ($SP[r]$) at rank r:

$$SP = \frac{Rel + Rel}{S1 + S2}$$

Evaluation

Segment Precision ($SP[r]$) at rank r:

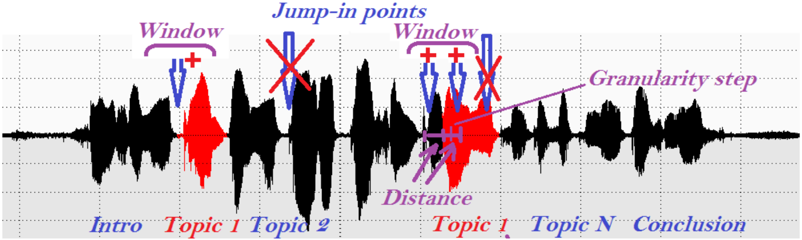
Average Segment Precision:

$$ASP = \frac{1}{n} . \sum_{r=1}^{N} SP[r] \cdot rel(s_r)$$

where:
$SP[r]$ = Segment Precision at rank $[r]$
$rel(s_r) = 1$, if relevant content is present, otherwise $rel(s_r) = 0$
(Eskevich, Magdy & Jones, 2012)

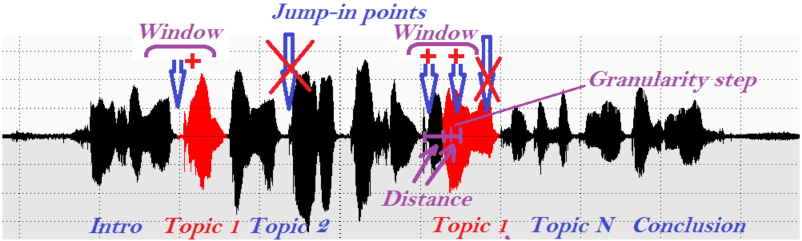
Evaluation

Penalize ASP results in the same way as mGAP



Evaluation

Penalize ASP results in the same way as mGAP



Mean Average Segment Distance-Weighted Precision (MASDWP):

$$ASDWP = \frac{1}{n}.\sum_{r=1}^{N} SP[r] \cdot rel(s_r) \cdot \left(1 - \frac{Distance}{Granularity} \cdot 0.1\right)$$

Evaluation

References:

- G.J.F.Jones and R.J.Edens, Automated Alignment and Annotation of AudioVisual Presentations, Proceedings of ECDK 2002, Rome, Italy. pp276-291, September 2002.
- A.M.Lam-Adesina and G.J.F.Jones. Using String Comparison in Context for Improved Relevance feedback in Different Text Media. In Proceedings of SPIRE 2006, Glasgow, Scotland, pp229-241, October 2006.
- J.Kekalainen and K.Jarvelin, Using Graded Relevance Assessments in IR Evaluation, Journal of the American Society for Information Science and Technology. 53(13):1120–1129, November 2002.
- B. Liu and D.W.Oard, One-Sided Measures for Evaluating Ranked Retrieval Effectiveness with Spontaneous Conversational Speech, In Proceedings of ACM SIGIR 2006, Seattle, U.S.A., August 2006
- M.Eskevich, W.Magdy and G.J.F.Jones. New Metrics for Meaningful Evaluation of Informally Structured Speech Retrieval, In Proceedings of ECIR 2012, Barcelona, April 2012.
- P.Galuscáková, P.Pecina, and J.Hajic, Penalty Functions for Evaluation Measures of Unsegmented Speech Retrieval. In Proceedings of CLEF 2012, volume, pages 100-111. Rome, Italy, September 2012.