Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop

Tomoyosi Akiba (Toyohashi University of Technology)
Hiromitsu Nishizaki (Yamanashi University)
Kiyoaki Aikawa (Tokyo University of Technology)
Tatsuya Kawahara (Kyoto University)
Tomoko Matsui (The Institute of Statistical Mathematics)

Outline

- Background
- Task Definition
  - Document Collection
  - Transcriptions
  - Subtasks
- Evaluation Results
  - STD subtask
  - SDR subtask
- Introduction of Oral Presentations

Background (1/2)

- Increase of the multi-media contents through the Internet.
- Multi-media contents is much more difficult to be retrieved than texts.
  - Metadata is not always available.
  - Manual annotation is expensive.
- Spoken Document Retrieval
  - Retrieval using speech track
  - One of the promising methods for retrieving multi-media contents with speech.

Background (2/2)

- Previous evaluation frameworks related to Spoken Document Retrieval
  - TREC SDR Track (1996-2000), TREC Video Track (2001-2002), TRECVID (2003-2010)
  - CLEF CL-SDR (2003-2004), CLEF CL-SR (2005-2007), CLEF QAST (2007-2009), VideoCLEF (2008-2009), Mediaeval (2010-2011)
  - NIST STD evaluation (2006)
- NTCIR-9 SpokenDoc (2011)
  - Both STD and SDR task
  - First evaluation targeting Japanese and lecture speech
  - Investigate Boundary-free passage retrieval task

Outline

- ✓ Background
- Task Definition
  - Document Collection
  - Transcriptions
  - Subtasks
- Evaluation Results
  - STD subtask
  - SDR subtask
- Introduction of Oral Presentations

Document Collection

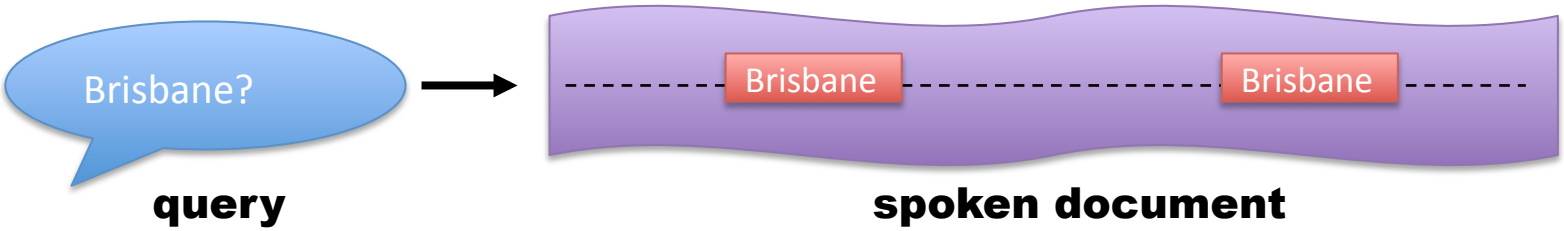
- Corpus of Spontaneous Japanese (CSJ)
  - provided from National Institute for Japanese Language and Linguistics (NINJAL)
- ALL set: 2702 lectures from CSJ
  - Lectures at academic societies
  - Simulated lectures on a given subject
  - 612 hours length
- CORE set: 177 lectures
  - Subset of ALL
  - 44 hours length
  - only for STD subtask

Reference Automatic Transcriptions

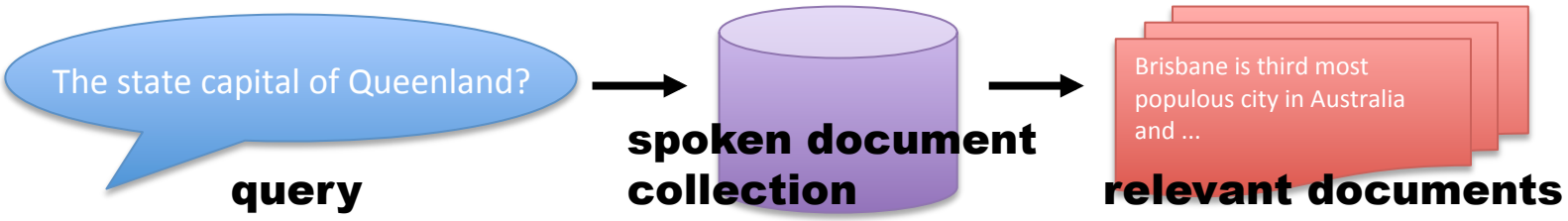
- The organizers provided two transcriptions of CSJ.
  - Word-based transcription
    - Produced by using Large Vocabulary Continuous Speech Recognition (LVCSR)
      - Vocabulary Size: 27K
      - using the word tri-gram language model
  - Syllable-based transcription
    - Produced by using the syllable-based speech recognition
      - Syllable Dictionary
      - using the syllable tri-gram language model
- The acoustic model and the language models were trained by using the CSJ itself, but by keeping the open condition.

Task Definition Overview: Two Subtasks

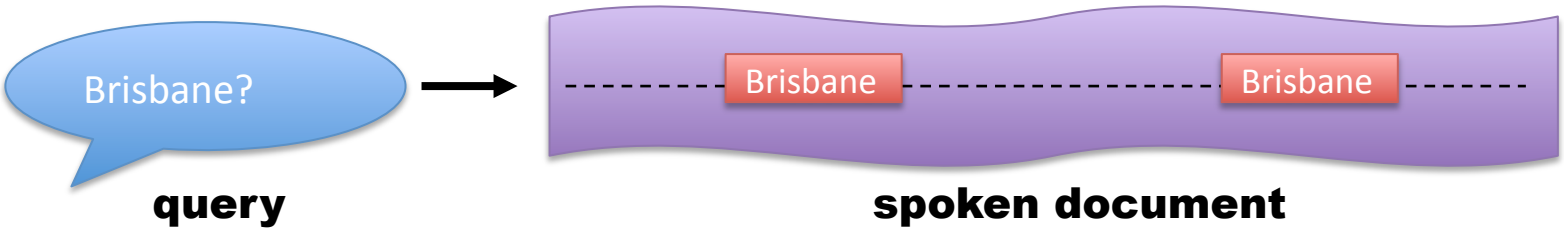
- Spoken Term Detection (STD) subtask
  - Find the occurrence of the given query term



- Spoken Document Retrieval (SDR) subtask
  - Find the segments related to the given query topic



Task Definition of STD Subtask



- From the target documents, the system is required to find the Inter Pausal Units (IPUs) in that the given query term is uttered at least one time.
  - Target: ALL or CORE set
  - Input: a query term
  - Output: a list of scored IPUs

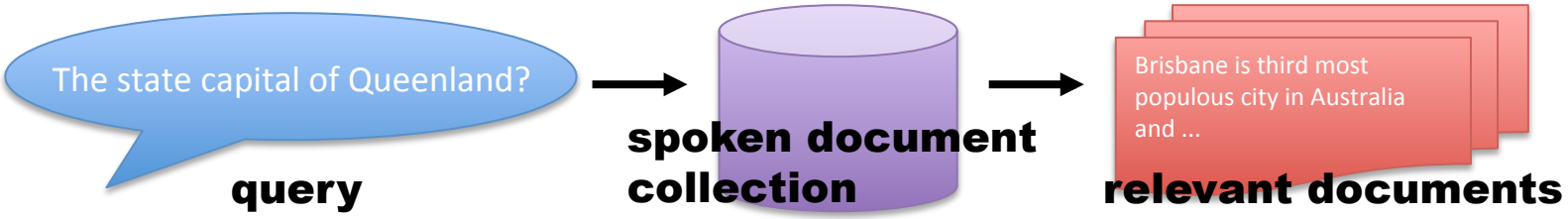
Inter Pausal Unit (IPU)

- Speech segment surrounded between two pauses no shorter than 200ms.
  - A spoken document = a sequence of IPUs
- Used as an atomic unit in our task definition.
  - Ignore time differences within IPU
- Enable us to apply conventional IR metrics based on discrete units.

Evaluation Metrics for STD subtask

- Using IPUs as the basic unit.
  - Precision, Recall, and F-measure (micro average)
    - Recall-precision curve
    - F-measure at the specified detection threshold (actual F-measure)
    - Max F-measure at the optimal detection threshold
  - Mean Average Precision (macro average)

Task Definition of SDR subtask



- From the target documents, the system is required to find the segments that are relevant to the given query topic.
  - Target: ALL set
  - Granularity of a segment: lecture or passage
    - lecture retrieval task
    - passage retrieval task
  - Input: a query topic
  - Output: an ordered list of segments

An Example Query Topic and its Relevant Passages

Query: 情報検索性能を評価するにはどのような方法があるか知りたい。
(How can we evaluate the performance of information retrieval?)

sequence of IPUs

0072: <雑音>
0073: (D ぷそ)
0074: (F えー)漏れなくという方に関係している
0075: で(F その)評価尺度としていわゆる再現率と呼ばれているものは(F その)どれだけ(D も)網羅的に
0076: (F えー)検索ができているかということを表わす尺度です
0077: <雑音>
0078: (F え)もう一つの
0079: (F え)スペシフィシティーというのは(F そのー)
0080: もう一方の特徴で(F あの)目的の重要な要素である(F その)正確に
0081: (F えー)検索するということに関係してますこれは(F あのー)評価尺度で言うと
0082: <雑音>
0083: (F え)精度
0084: (F えー)プリシジョンと呼ばれてるやつですね精度
0085: に関係するもんですけれど(F その)できるだけ(F その)文書の
0086: 内容
0087: を特徴的な要素を掴まえている
0088: という
0089: ことが(F ま)望ましい訳です
0090: で当然のことなんか(F ま)両者はある程度(D 排)

Relevant Passage

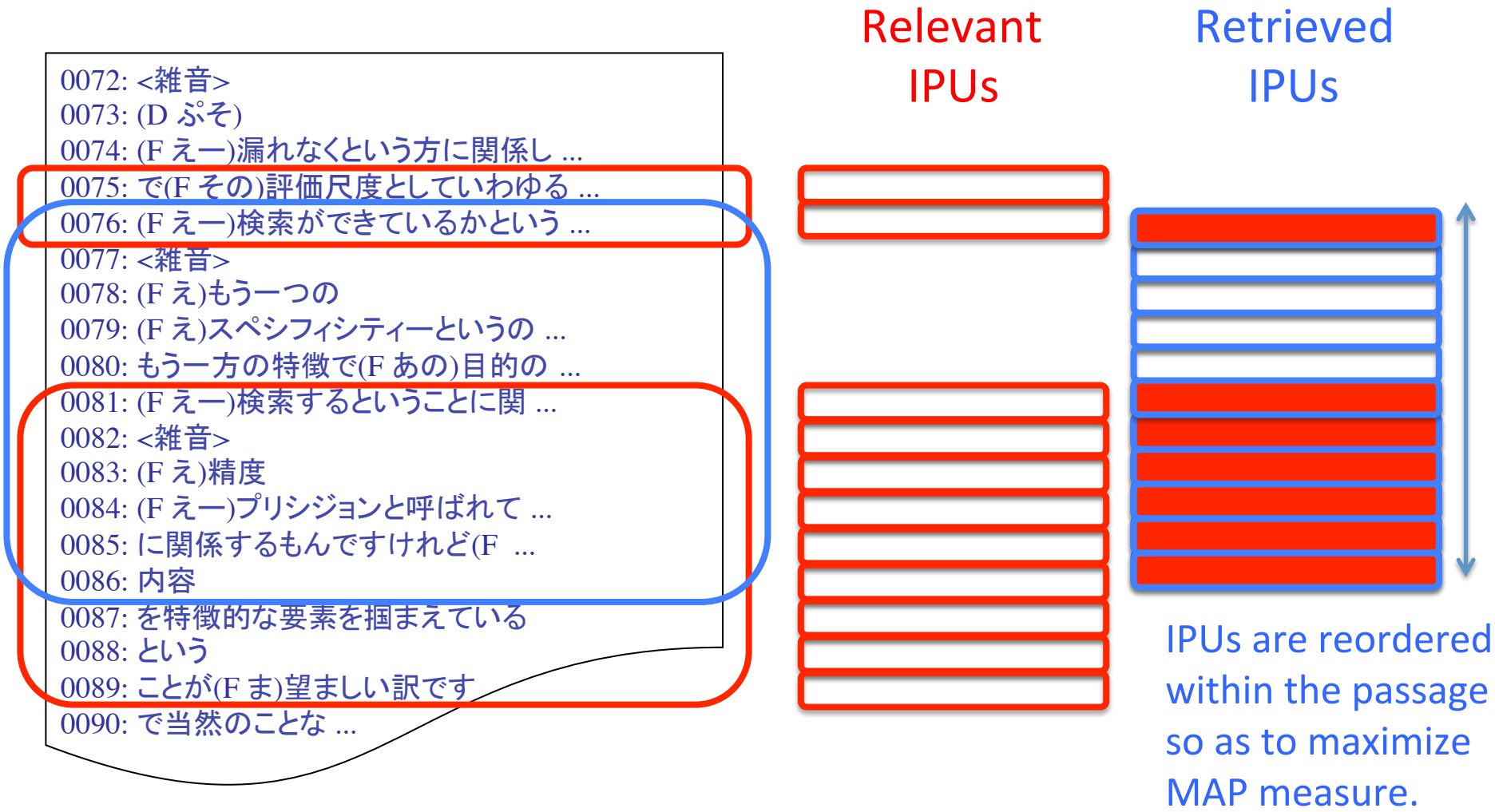
Relevant Passage

A01M0958

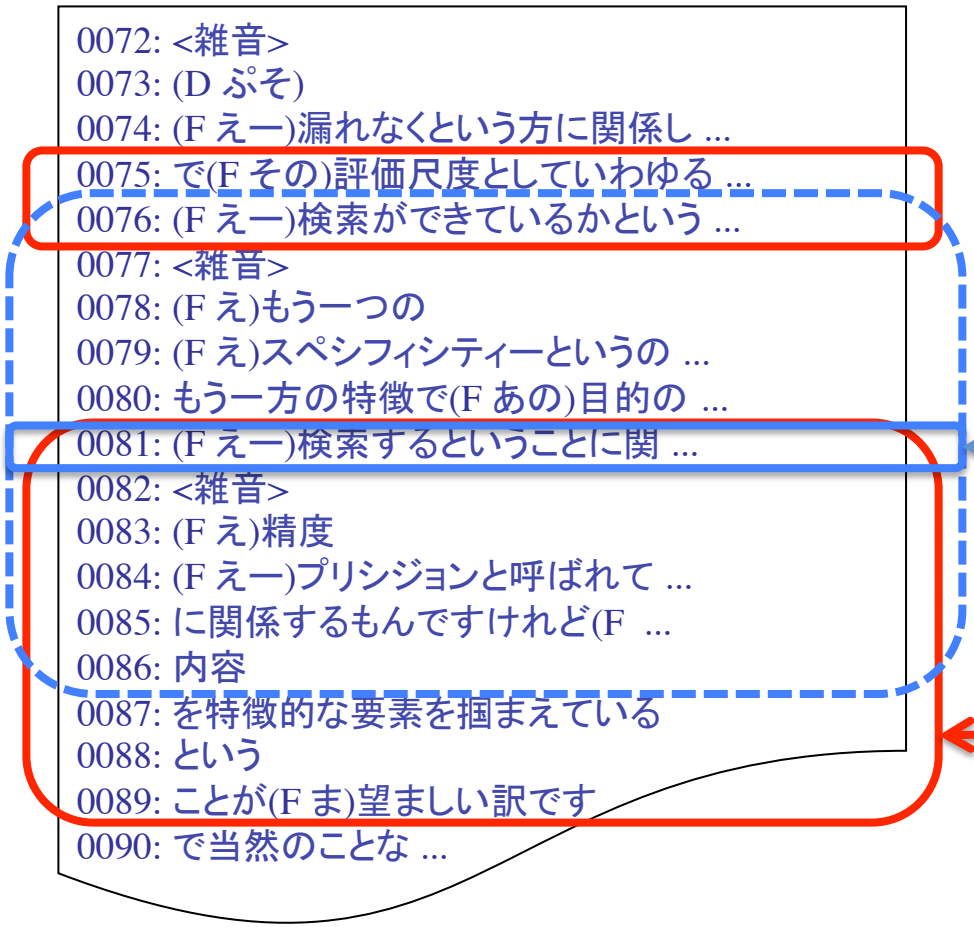
Evaluation Metrics for SDR subtask

- For lecture retrieval task
  - Mean Average Precision (MAP)
- For passage retrieval task
  - Utterance-based metric
    - utterance-based MAP (uMAP)
  - Passage-based metric
    - point-wise MAP (pwMAP)
    - fractional MAP (fMAP)

Utterance-based MAP (uMAP)



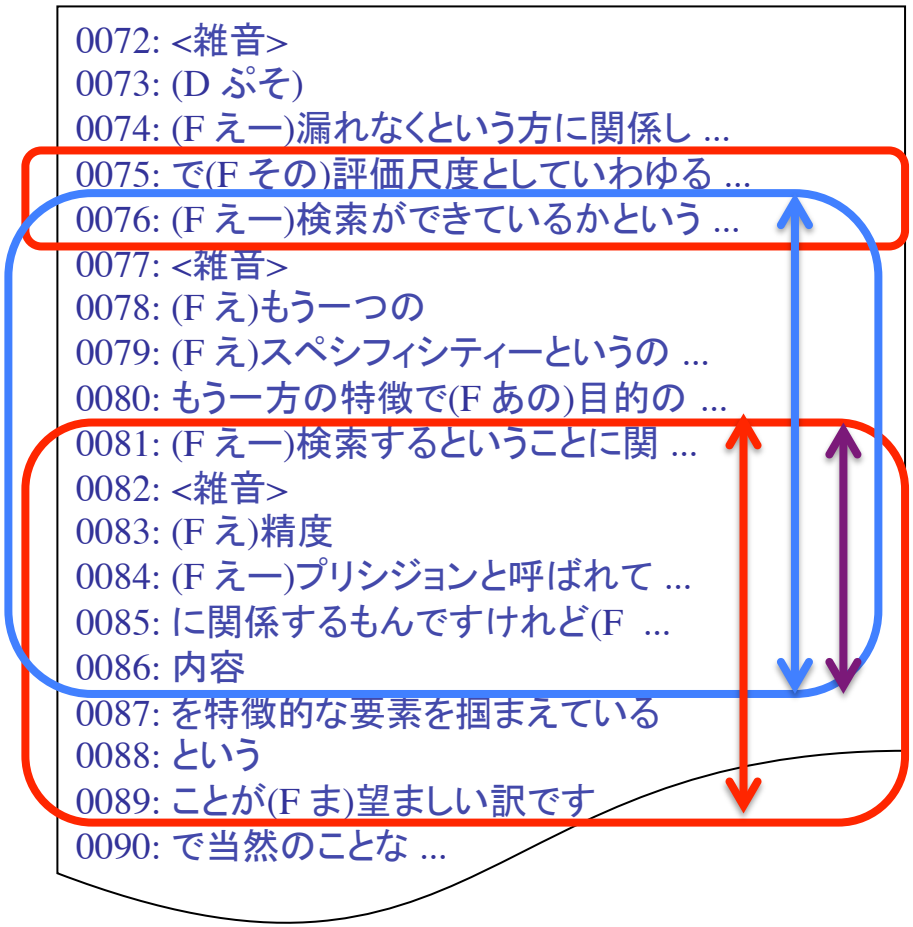
Point-wise MAP (pwMAP)



Center IPU of Retrieved Passage

This relevant passage as a whole is considered to be retrieved.

Fractional MAP (fMAP)



$$fAveP_q = \frac{1}{|R_q|} \sum_{i=1}^{|P_q|} rel(p_i, R_q) \frac{\sum_{j=1}^{i} prec(p_j, R_q)}{i}$$

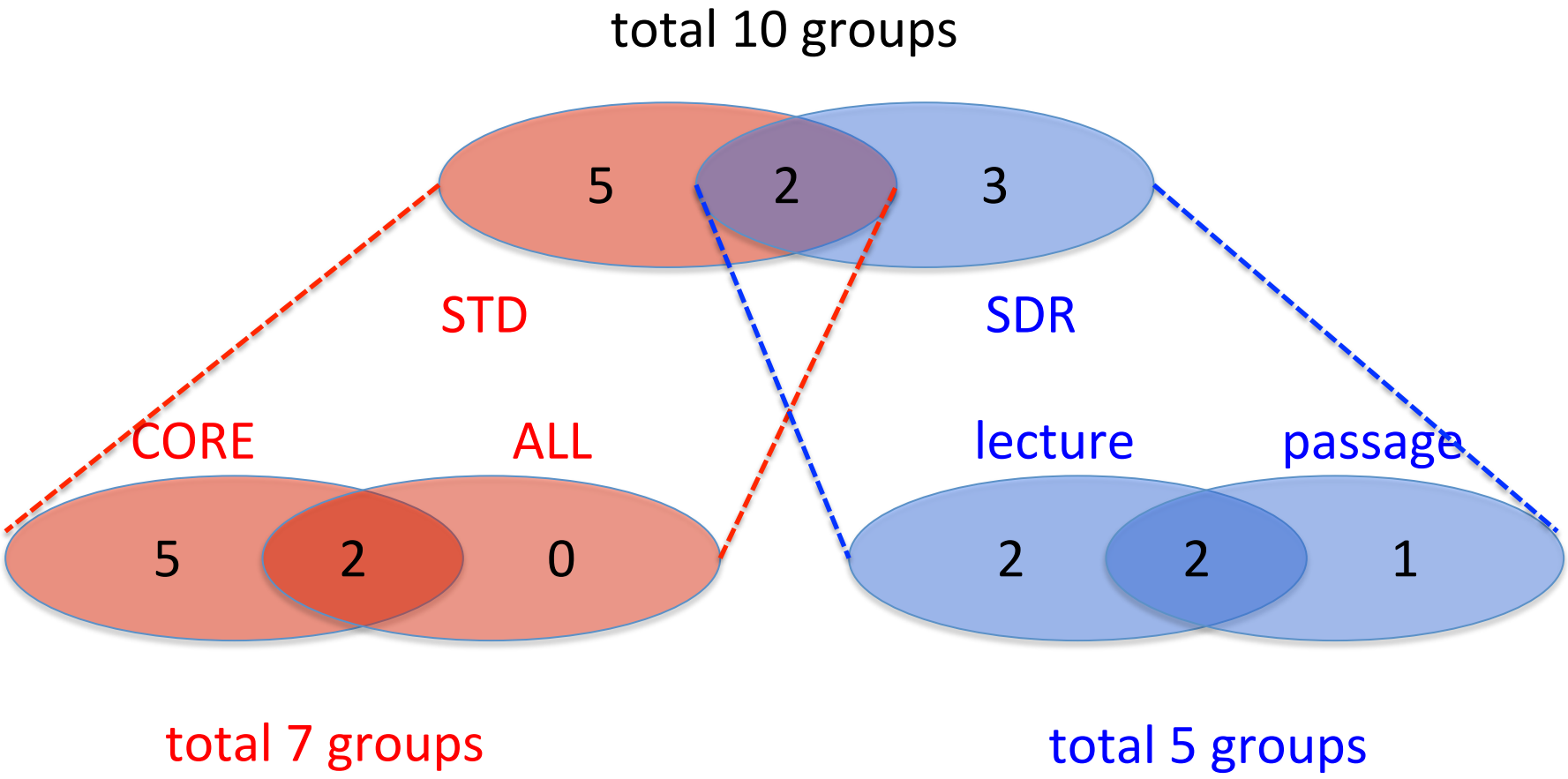
$$prec(p, R_q) = \max_{r \in R_q} \frac{|r \cap p|}{|p|}$$

$$rel(p, R_q) = \max_{r \in R_q} \frac{|r \cap p|}{|r|}$$

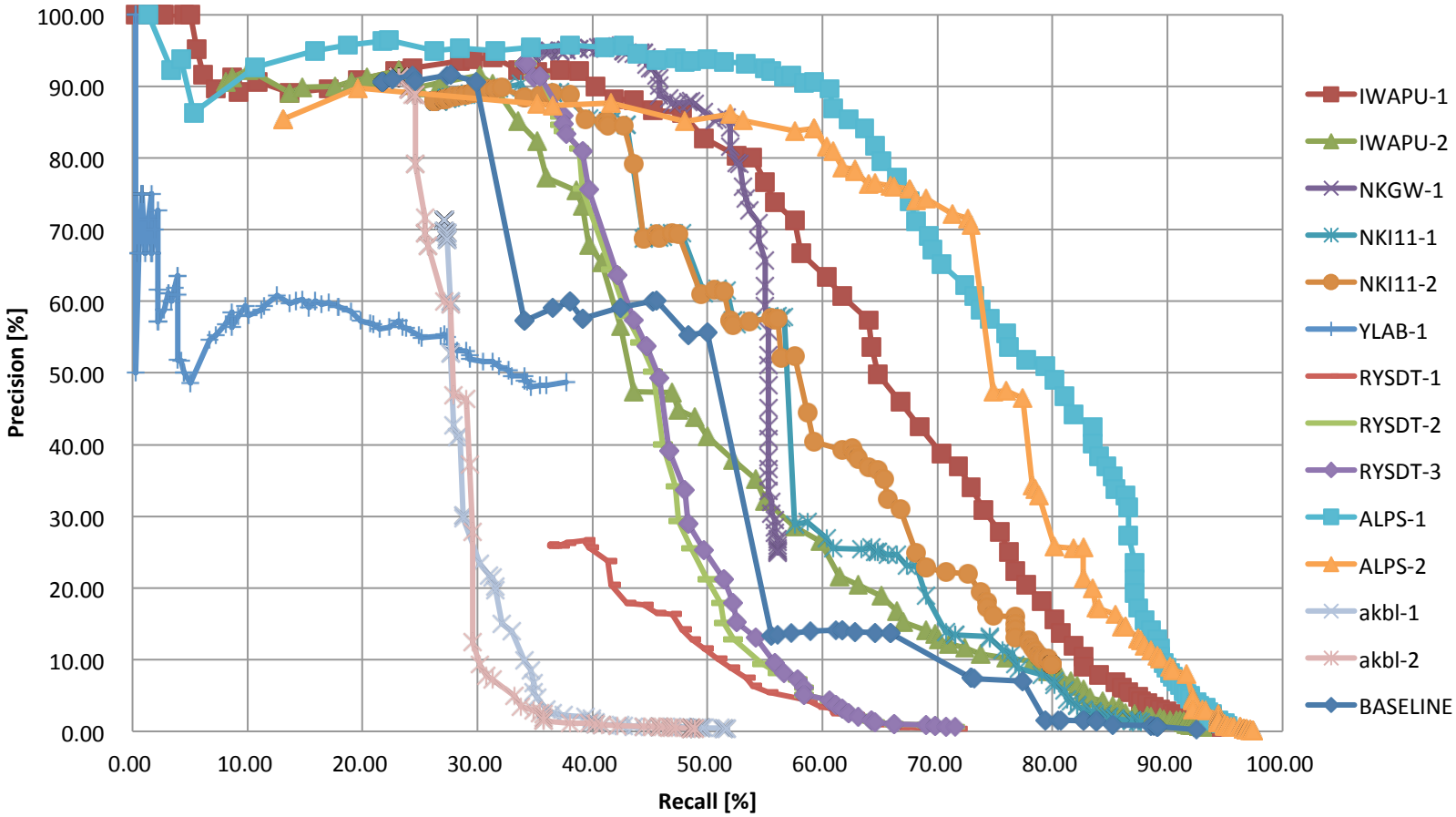
Outline

- ✓ Background
- ✓ Task Definition
  - ✓ Document Collection
  - ✓ Transcriptions
  - ✓ Subtasks
- Evaluation Results
  - STD subtask
  - SDR subtask
- Introduction of Oral Presentations

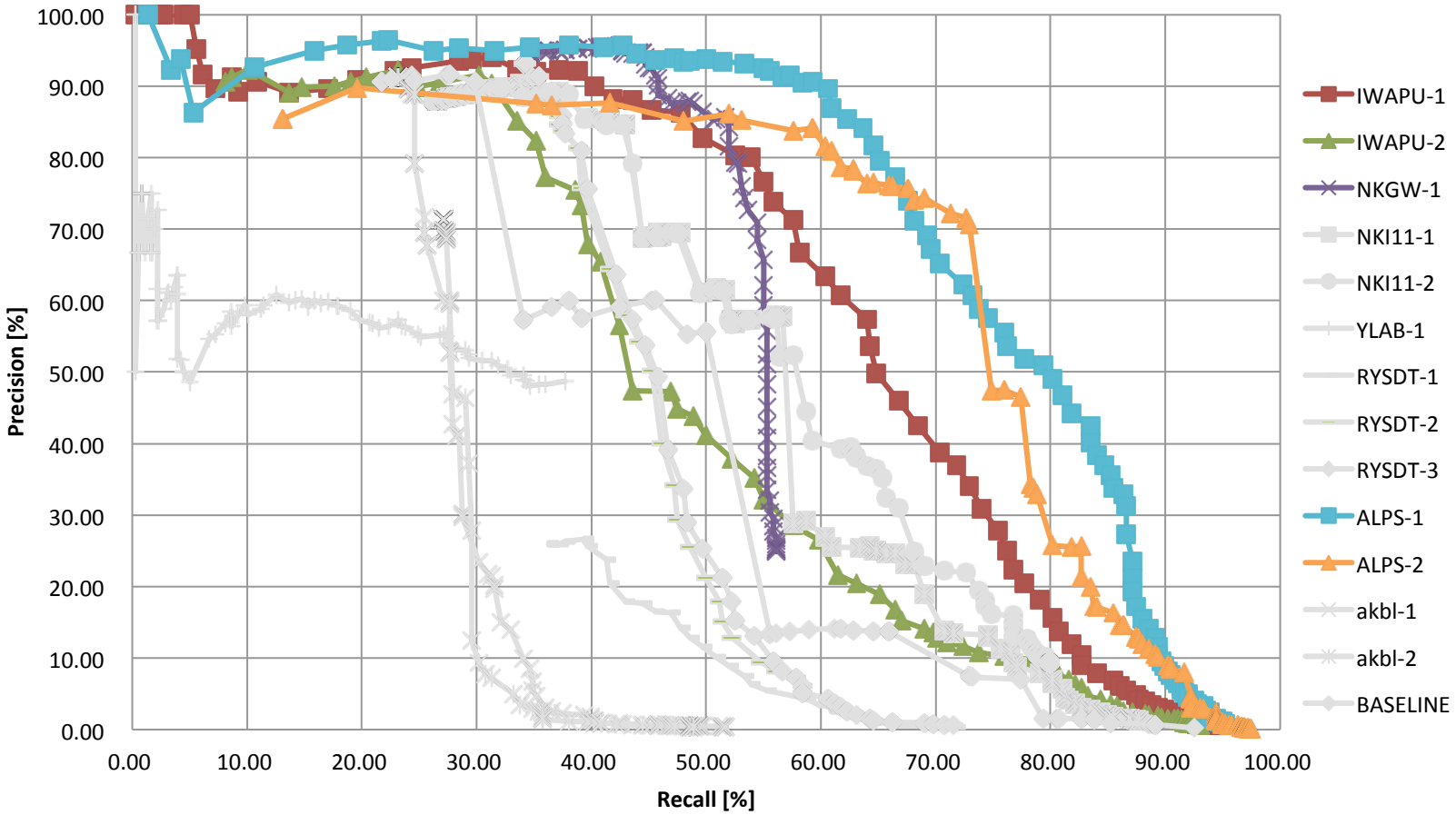
Participant Groups



Recall-Precision Curves for STD subtask for CORE set

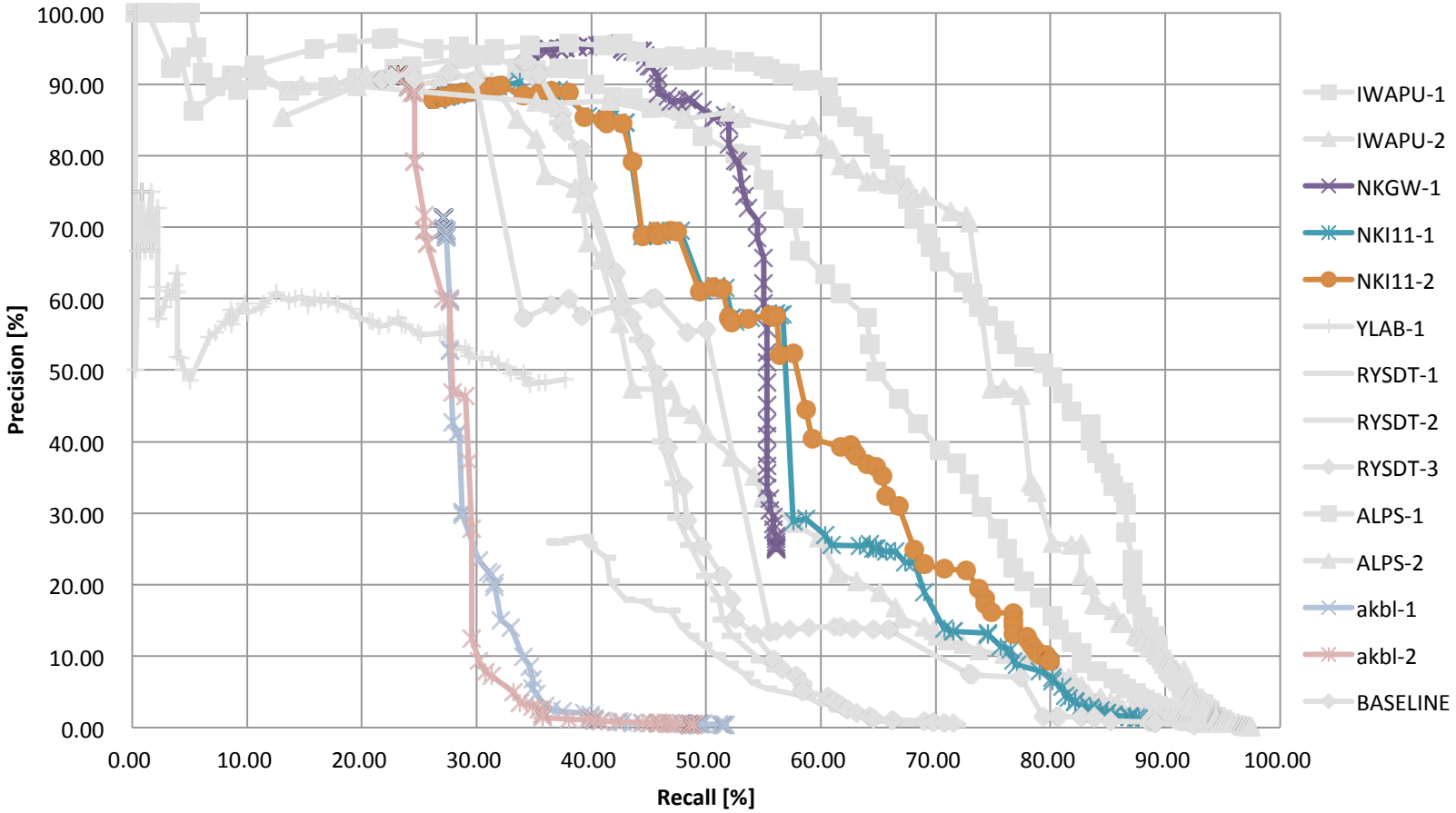


Recall-Precision Curves for STD subtask for CORE set



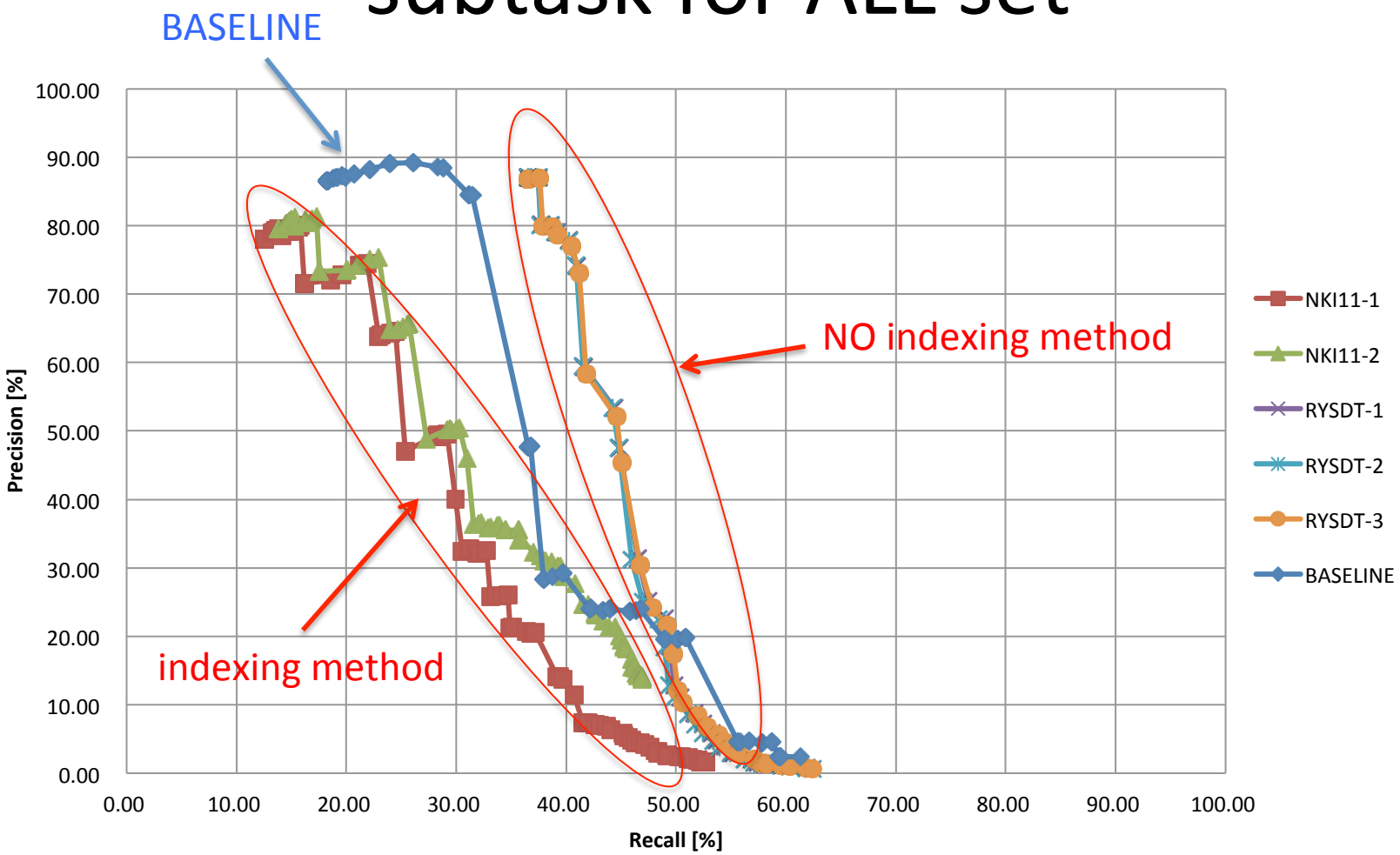
5 runs using multiple transcriptions

Recall-Precision Curves for STD subtask for CORE set

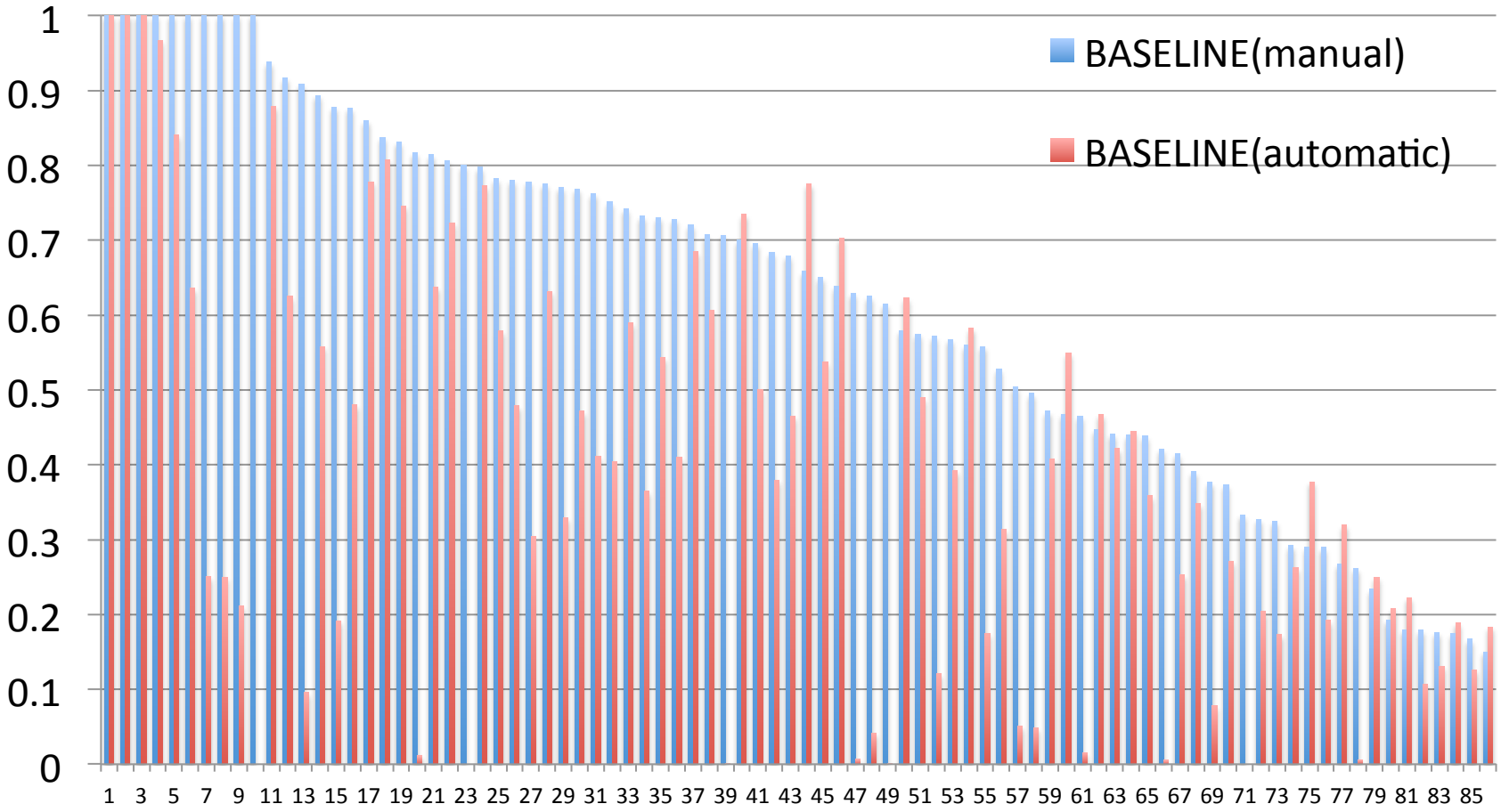


5 runs using indexing for efficiency

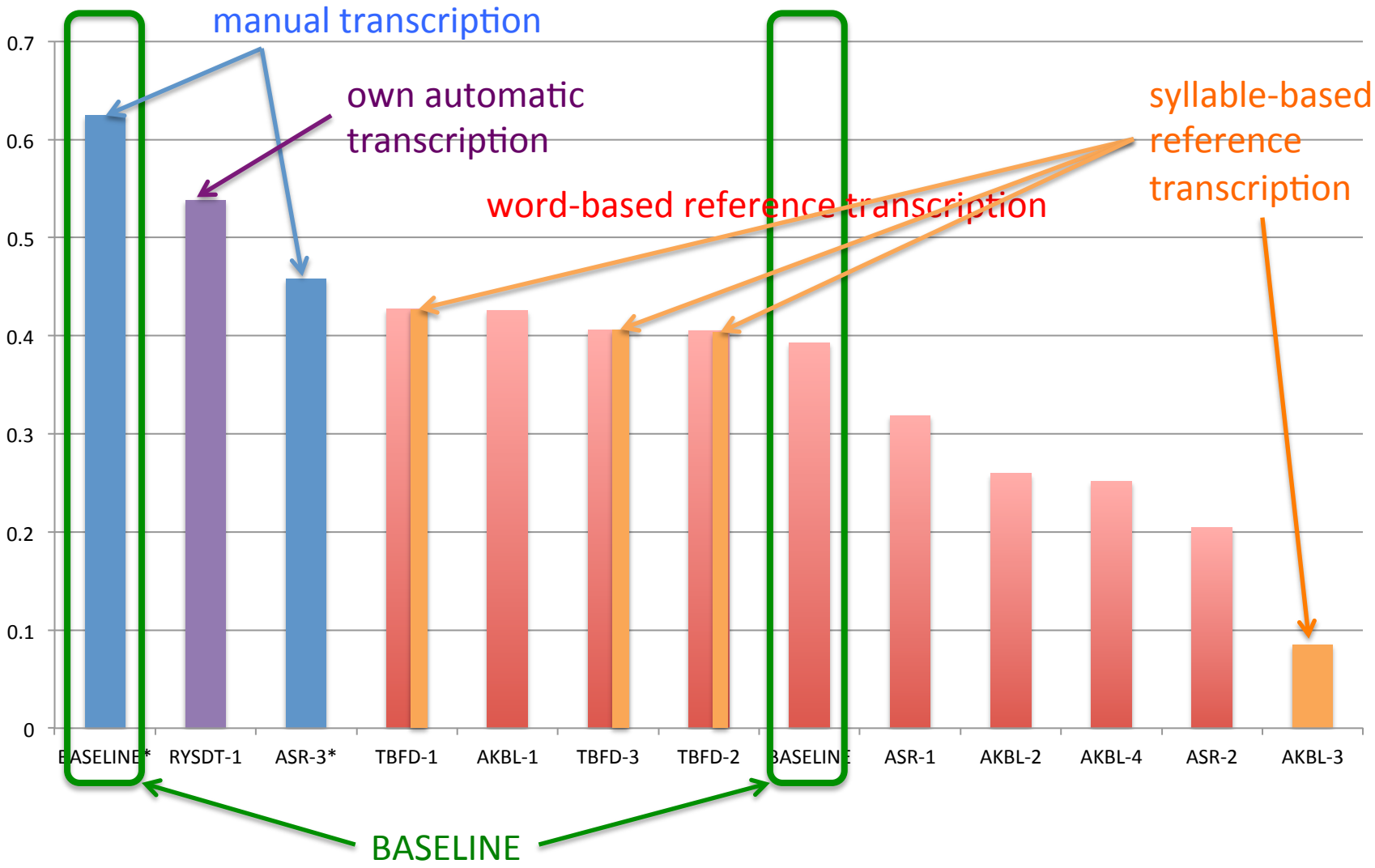
Recall-Precision Curves for STD subtask for ALL set



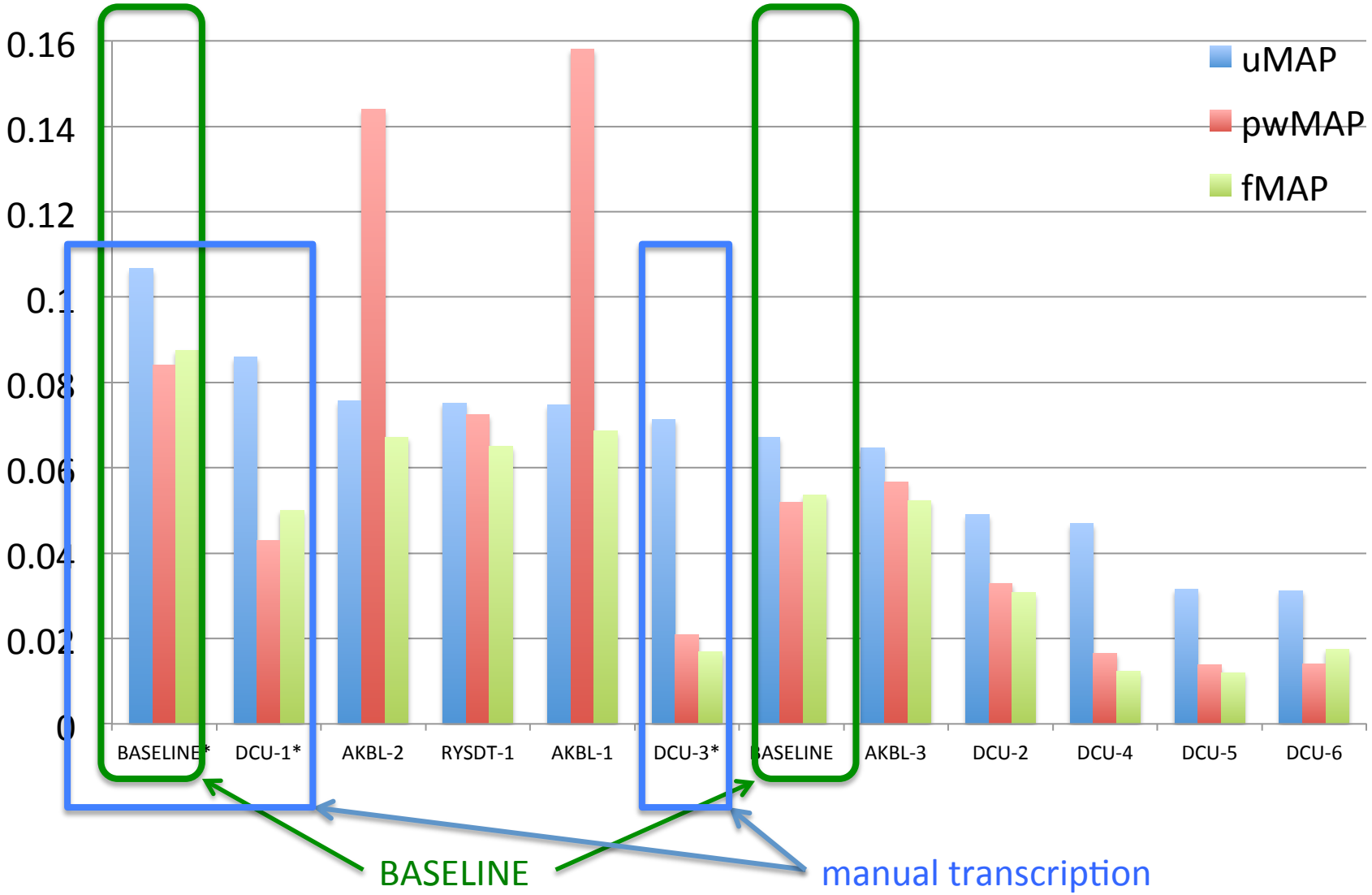
BASELINE Performance (MAP) per Query for SDR lecture retrieval task



Evaluation Result (MAP) for the lecture retrieval of SDR subtask



Evaluation Result for the passage retrieval of SDR subtask



Findings from STD subtask results

- Task participations were motivated from various research interests, e.g. some for performance, some for efficiency, etc.
  - It, in turn, made comparison among the methods difficult as they often did not share same bases.
    - using various kinds and numbers of transcriptions
    - indexing or not
- Use of multiple transcriptions was one of the most effective methods for improving the performance.
- Indexing methods made the detection thousands times faster than the basic DTW algorithm, while some of them did not degrade the performance much.

Findings from SDR subtask results

- Common techniques used for text-based IR could help SDR, e.g. pseudo relevance feedback, query and document expansion, TF-IDF weighting, language modeling, etc.
- Some techniques specific for SDR could also be effective, e.g. using multiple recognition candidates, sub-word based indexing, etc.
- Using good transcription (including less errors) consistently improved the IR performance.
- Our boundary-free passage retrieval task was much harder than what had been expected. Solving both the tasks, which are finding the relevant portions in the documents and determining their boundaries, have to be investigated further.

Oral Presentations

- STD subtask
  - ALPS lab at UY
    - H. Nishizaki, H. Furuya, S. Natori, and Y. Sekiguchi, "Spoken Term Detection using multiple speech recognizers' output at NTCIR-9 SpokenDoc STD subtask"
    - Best performer of STD subtask targeting CORE set
  - Nakagawa LAB
    - K. Iwami and S. Nakagawa, "High speech spoken term detection by combination of n-gram array of a syllable lattice and LVCSR result for NTCIR-SpokenDoc"
    - Achieved efficient detection without much performance loss
- SDR subtask
  - Team Big Four Dragon
    - S. Tsuge, H. Ohashi, N. Kitaoka, K. Takeda, and K. Kita, "Spoken document retrieval method combining query expansion with continuous syllable recognition for NTCIR-SpokenDoc"
    - Best performer among those using reference transcription in lecture retrieval
  - DCU
    - M. Eskevich and G. J. F. Jones, "DCU at the NTCIR-9 SpokenDoc passage retrieval task"
    - Applied text segmentation methods to find the arbitrary length passage