UB at the NTCIR-12 SpokenQuery&Doc-2: Spoken Content Retrieval Using Multiple ASR Hypotheses and Syllables

Jianqiang Wang

Department of Library and Information Studies, Graduate School of Education, University at Buffalo, U.S.A.
jw254@buffalo.edu, http://www.buffalo.edu/~jw254

Task: Spoken Content Retrieval (SCR) in Japanese
For each of the 80 search topics, retrieving relevant documents from 2,259 slide group segments that are created from 98 conference presentation lectures. Both documents and topics contain audio, manual transcription, and automatic speech recognition (ASR) words and syllables.

Challenges and motivation
ASR errors may affect the SCR effectiveness; Previous work on news broadcast and oral history; Increased volume of recorded class lectures; Users, information needs, and user-system interaction.

Areas of interest
Usefulness of multiple ASR hypotheses; Usefulness of ASR syllables; Performance of different ASR engines (Julius vs KALDI).

Techniques
ASR followed by text-based information retrieval (IR); Treating multiple ASR hypotheses as independent terms; Okapi BM25 term weighting and document tanking.

Official runs and results

Run id	Document term	Query term	MAP	Relative MAP
SQSCR-UB-SGS-TXT-1	Manual transcription words	Manual transcription words (verbose)	0.1953	Reference run
SQSCR-UB-SGS-TXT-2	1-best Julius words	1-best Julius words	0.1128	57.8%
SQSCR-UB-SGS-TXT-3	5-best Julius words	1-best Julius words	0.0994	50.9%
SQSCR-UB-SGS-TXT-4	1-best Julius words	5-best Julius words	0.1127	57.7%
SQSCR-UB-SGS-TXT-5	5-best Julius words	5-best Julius words	0.0966	49.5%
SQSCR-UB-SGS-TXT-6	1-best Julius syllables	1-best Julius syllables	0.0253	13.0%
SQSCR-UB-SGS-TXT-7	1-best KALDI words	1-best KALDI words	0.1946	**99.7%**

In terms of MAP
- Julius ASR significantly lower than manual transcription.
- KALDI ASR indistinguishable from manual transcription.
- ASR words significantly better than ASR syllables (Julius).
- Multiple ASR hypotheses (words) showing little effect.

Query-by-query comparison
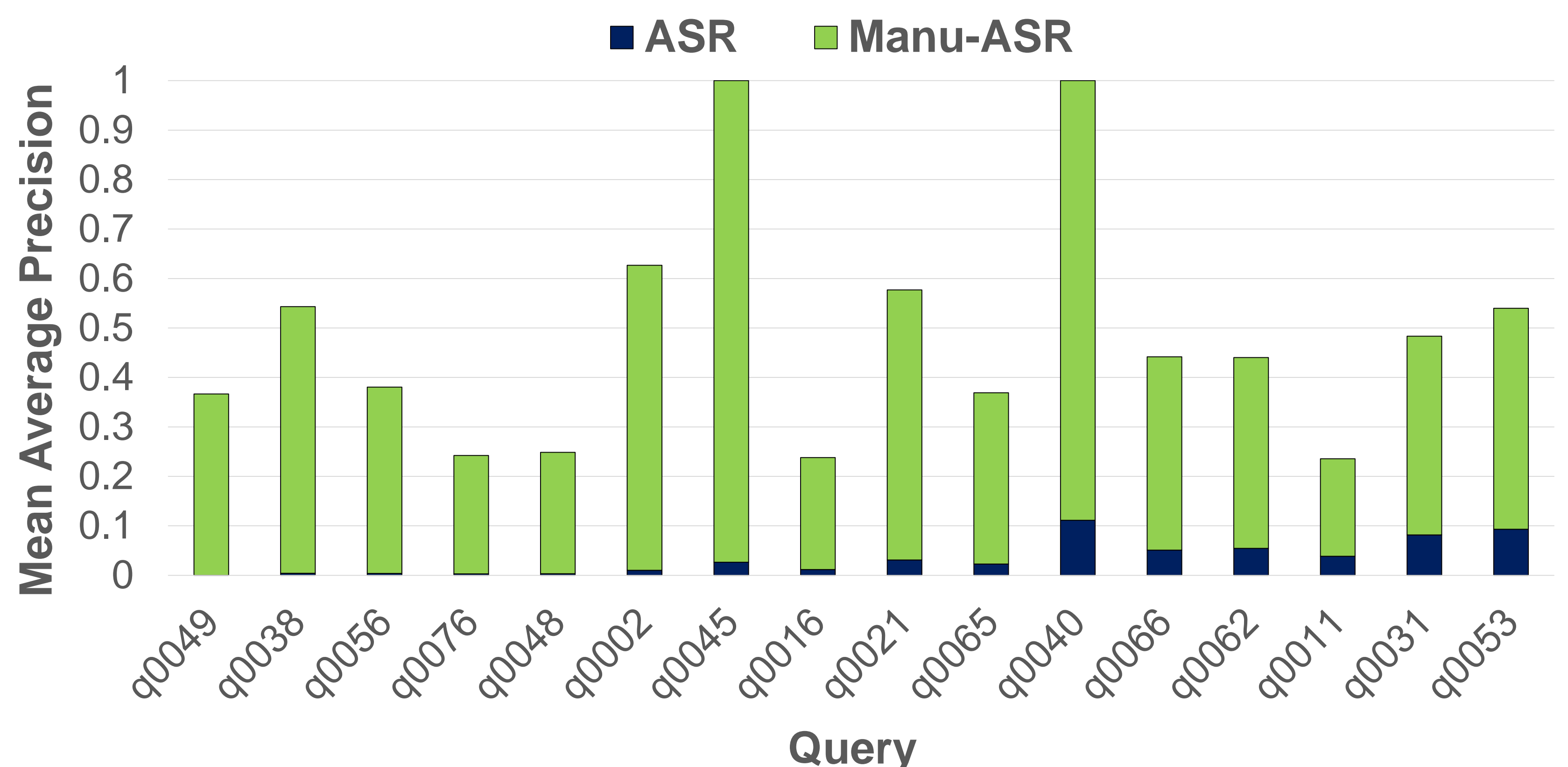
- Among 29 queries that have an AP of >=0.2 on Manu, 16 achieved only less than 20% of the Manu AP when 1-best Julius ASR words were used.
- These are the topics/queries deserving close examination.

Conclusions and future work
- ASR errors degrades SCR effectiveness, but some systems can produce ASR text leading to results comparable to manual transcription. Therefore, detailed analysis of different ASRs is needed.
- ASR syllables alone are not reliable for SCR. Work is underway on enhancing/compensating for ASR words with ASR syllables and other subword units.
- Multiple ASR hypotheses should be used more sophisticatedly, including taking into consideration their confidence scores.