

New Transcription System using Automatic Speech Recognition (ASR) in the Japanese Parliament (Diet)

-- The House of Representatives --

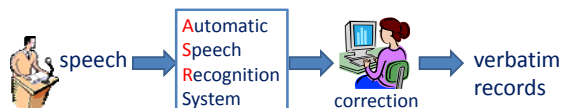
Tatsuya Kawahara
(Kyoto University, Japan)
kawahara@i.kyoto-u.ac.jp

Brief History

- 1890: Foundation of Japanese "Imperial" Parliament
 - Verbatim records have been made by manual shorthand since the first session
- 2005: terminated recruiting stenographers investigated ASR for a new system
- 2007: a prototype system & preliminary evaluation
- 2008: system design
- 2009: system implementation Intersteno 2009
- 2010: system deployment and trials
- 2011: official operation Intersteno 2011

System Overview

- **ALL** plenary sessions and committee meetings
- Speech captured by the stand microphones
 - Separate channels for interpellator & (minister + speaker)
- ASR system generates an initial draft
 - System's recognition errors to be corrected ~10%
 - Disfluencies & colloquial expressions to be corrected ~10%
 - **Reporters still play an important role!**



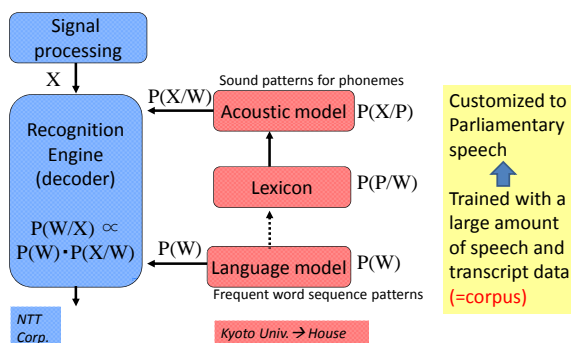
Japanese Language-specific Issues

- Need to convert *kana* (phonetic symbol) to *kanji* (Chinese characters)
- Conversion often ambiguous ← many homonyms
(ex.) KAWAHARA → 河原 (not 川原)
 - **Very hard to type in real-time**
 - Only limited stenographers using special keyboard can
- Difference between spoken-style and transcript-style
(ex.) *じゃ、これいいですか* → *では、これはいいですか*
 - need to rephrase in many cases
 - **Re-speaking is not so simple!**

Requirements for ASR System

- High accuracy → technically most difficult
 - Over 90% preferred
 - No problem in plenary sessions
 - Difficult in committee meetings (spontaneous, interactive)
- Fast turn-around → feasible with current PC
 - Each reporter assigned 5-minute segment
 - ASR should be performed almost in real-time, so reporters can start working promptly even during the session
- Compliance to orthographic transcript guideline
 - Only official meeting records used for training ASR
 - Electric dictionary of 60K lexical entries proofed

ASR System: Kyoto Univ. Model integrated to NTT System



Data of Parliamentary Meetings

- Huge archive of official meeting records (text)
 - 15M words per year...comparable to newspapers
- Huge archive of meeting speech
 - 1200 hours per year

However,

- Official meeting records are different from actual utterances due to editing process by reporters
 - Difference between spoken-style and written-style
 - Disfluency phenomena (fillers, repairs)
 - Redundancy (discourse markers)
 - Grammatical correction

→ More in Japanese
→ More in English (EU PPS)

Corpus of Parliamentary Meetings

- Faithful transcripts of utterances including fillers, which are aligned with official records
- 200 hours, 2.4M words

{えー} それでは少し、今 {そのー} 最初に大臣からも、{そのー} 貯蓄から投資へという流れの中に {ま} 資するんじゃないだろうかとかいうような話もありましたけれども、{だけど}、{まあ} あなたが言うとも本当にならなくなる(んで) {で}、{えー} もう少し(ですね、あのー) これは {あー} 財務大臣に {えー} お尋ねをしたいと思います(が)。
{ま} その {あの} 見通しはどうかということでもありますけれども、これについては、{あのー} 委員御承知の {その} 「改革と展望」の中で {ですね}、我々の今 {あのー} 予測可能な範囲で {えー} 見通せるものについてはかなりはっきりと書かせていただいているつもりでございます。

- Vital for satisfactory performance
- Very costly and need to be updated

Faithful transcript (what was uttered)

target of ASR

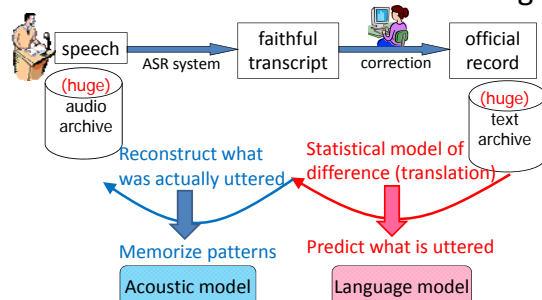
{えー} それでは少し、今 {そのー} 最初に大臣からも、{そのー} 貯蓄から投資へという流れの中に {ま} 資するんじゃないだろうかとかいうような話もありましたけれども、{だけど}、{まあ} あなたが言うとも本当にならなくなる(んで) {で}、{えー} もう少し(ですね、あのー) これは {あー} 財務大臣に {えー} お尋ねをしたいと思います(が)。
{ま} その {あの} 見通しはどうかということでもありますけれども、これについては、{あのー} 委員御承知の {その} 「改革と展望」の中で {ですね}、我々の今 {あのー} 予測可能な範囲で {えー} 見通せるものについてはかなりはっきりと書かせていただいているつもりでございます。

13% difference 93% are simple edits

Official record

それでは少し、今 最初に大臣からも、貯蓄から投資へという流れの中に 資するんじゃないだろうかとかいうような話もありましたけれども、{だけども}、あなたが言うとも本当にならなくなる(んで) {ので}、もう少し これは 財務大臣に お尋ねをしたいと思います。その 見通しはどうかということでもありますけれども、これについては、委員御承知の 「改革と展望」の中で、我々の今 予測可能な範囲で 見通せるものについてはかなりはっきりと書かせていただいているつもりでございます。

Innovative Approach for Corpus Generation and ASR Model Training



- Precise modeling of spontaneous speech in Parliament
- Evolve in time, reflecting change of MPs and topics

Evaluation of ASR System

- Accuracy
 - Character Correct compared against official record
 - 89.4% for 108 meetings in 2010 & 2011
 - Over 95% when limited to plenary sessions
 - No meetings got less than 85%
 - Update of models gives improvement of 0.7%
- Processing Time
 - 0.5 in Real-Time Factor
 - 2.5 min. for 5-min. segment
- Post-processing
 - Fillers are automatically annotated & removed
 - Automation of other edits is difficult... research ongoing

Post-Editor used by Reporters

- For efficient correction of ASR errors and cleaning transcripts
- Screen editor (word-processor interface); not line editor
 - so that reporters can concentrate on making correct sentences
 - designed by reporters, not by engineers!
- Easy reference to original speech (+video)
 - by time, by utterance, by character (cursor)
 - can speed up & down replay of speech
- Re-speaking function is not incorporated, though technically feasible

Side Effect of ASR-based System

- Everything (text/speech/video) digitized and hyper-linked
 - by speaker turn, by utterance
 - good platform even if ASR result is not usable
 - efficient search & retrieval

Demonstration

Summary and Future Perspectives

- Highest-standard ASR system dedicated to Parliamentary meetings
- 89% Character (85% Word) Correct
- will improve (evolve) with more data accumulated
- Drastic change from manual short-hand to fully ICT-based system
 - Need time for reporters to get accustomed
 - Need to develop a new training methodology
- Reporters play a central role in making verbatim records!