Enhanced Web-Based Captioning System to Create Multi-Language Media Captions Effectively

Speaker: Kaoru Shinkawa

kaoruma@jp.ibm.com

Authors:
Kaoru Shinkawa – IBM Research – Tokyo, Japan
Ali Sobhi – IBM Research, Austin, US

28th Annual International Technology and Persons with Disabilities Conference,
Manchester Grand Hyatt, San Diego, CA
February 27th, 2013
Agenda

• Who will use captions?
• Closed Caption vs. Open Caption
• Accessibility Laws & Standards
• Current Status & Issues
• Media Captioner & Editor
• Collaborative Community Editing
  – Crowd-Sourcing
  – Speech Recognition and Manual Editing
• Translation Editing Feature
• Caption Formats
• Caption in Learning Management System (LMS)
• Demonstration
Who will use captions?

- Deaf or Hard of Hearing
- Learning Disabilities
- Online Education
- Captioning of media archives
- Search Engines
- Data Mining
Closed Caption
Turned on/off by the viewer
Open Caption

Open caption
Burned into the video Cannot be turned off

Captioned & Powered by DigiCapE

HR leader for the People with Disabilities Global Diversity Council.
Accessibility Laws & Standards

- US Section 508
- EU Mandate 376 (M376)
- 21\textsuperscript{st} Century Communications and Video Accessibility Act (CVAA)
- WCAG 2.0
- JIS X 8341-3
Current Status and Issues

IBM Media Library

Great portion of all media content on Media Library does not have captions or transcriptions

Total media count = 78,546
Media w/o transcript = 93%

Internet media with Captions in Japan (2009)

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Course Ware (Educational Contents)</td>
<td>6.3%</td>
</tr>
<tr>
<td>NHK and Commercial Broadcasting News site (*)</td>
<td>0%</td>
</tr>
<tr>
<td>Government Internet TV</td>
<td>45.3%</td>
</tr>
</tbody>
</table>

(+) Some programs are captioned and provided on YouTube
Media Captioner and Editor (MCE)

- Complete solution to caption, edit and create Web-ready digital media
- Helping people who are deaf or hard of hearing
- Utilizing state-of-the-art advanced, speaker-independent IBM Automatic Speech Recognition
- Optimized for manual editing process also (non-ASR)
- Fully accessible

Web-based GUI for complete job management

Intuitive and easy-to-use caption editor with synchronized audio and editing

Web-ready media with closed-captioning enabled
Collaborative Captioning with Crowd-Sourcing

• Many people can contribute captioning in a short period of time

• Use of Speech Recognition Engine can improve the efficiency of the captioning

Repeat for each caption line

- Most time consuming
- Requires concentration

Transcribe → Review → Adjustment

Transcribe

Review

Adjustment

Transcribe

Review

Adjustment
Three Groups to perform captioning

1) **Editors Group**
   - Listen to the audio and transcribe from scratch, or correct the transcribed text returned from speech recognition.

2) **Reviewers Group**
   - Review the caption entered by Editors and make correction. Reviewers should have some experience of editing caption.

3) **Verifiers Group**
   - Verify the caption edited and reviewed. Verify the layout/format of the caption and make correction. Verifier should have knowledge of captioning.

*Volunteers can be assigned to group 1) and 2).*
Working time of a person skilled in captioning reduced 4 times by distributing editing work to non-skilled volunteers with crowd-sourcing.
Speech Recognition & Manual Editing

• ASR availability for our customers’ Web media might be limited due to
  • Various audio conditions
  • Various speaking styles, dialects, multiple speakers
  • Language coverage of ASR

  ➔ These factors cannot be negligible on the Internet.

• Motivation of Manual Editing Feature
  – Explore a captioning method to cover the wide variety of Web media
  – Need a way of captioning independent of ASR
In order to understand as it is required **more than 90%** rate
For practical application, or in order to understand better, required for **more than 95%** rate

English broad casting is more than 90%, and also Japanese. Japanese is used the re-speak method.

(*1)The history of ASR Evaluations at NIST
(*2) Report of the study for the digital broadcasting for the visual and hearing impaired person, Ministry of Internal Affairs and Communications, Japan
http://www.soumu.go.jp/main_sosiki/kenkyu/digital/02ryutsu05_03000030.html
(*3) AC Japan original investigation
3 Stages of Collaborative Manual Editing

Audio decomposition
- Decompose into short speech blocks
- BGM are fairly OK
- Basically language independent

Collaborative captioning
- Each editor inputs captions only for a one minute or less voice chunk
- Editing line length is appropriate for typing text by listening and memorizing

Text formatting
- Unify punctuations
- Unify wording fluctuations
- Detect unnecessary words
- Format captions with video constraints
Featuring translation support and editing

Captioned media → Translation Engine → Editors make correction → Translated caption

Translation Screen

Source language  Translated language

Edit incorrectly translated text

Under development
Supporting Multiple Languages

Multi-language Caption

Language Selection

English

Spanish
Supported Caption format

- **SRT** (SubRip Text format) : DVD
- **DFXP** (The Distribution Format Exchange Profile) : Flash player
- **SMI** (Synchronized Accessible Media Interchange): Windows Media Player
- **CSV** (Comma Separated Values)

1
00:00:04,940 --> 00:00:09,850
over the past century but accelerating over the past couple

2
00:00:09,850 --> 00:00:14,750
of decades we've seen the emergence of the kind of global

**SRT**

**SMI**

<SYNC Start=4940><P Class=ENUSCC>over the past century but accelerating over the past couple</SYNC>
<SYNC Start=9850><P Class=ENUSCC></SYNC>
<SYNC Start=9850><P Class=ENUSCC>of decades we've seen the emergence of the kind of global</SYNC>

**DFXP**

<p begin="00:00:04.940" end="00:00:09.850"><captioning:status linestate="corrected" />
</captioning:comment value="" />
<captioning:situation value="on" />
over the past century but accelerating over the past couple</p>
<p begin="00:00:09.850" end="00:00:14.750"><captioning:status linestate="corrected" />
</captioning:comment value="" />
<captioning:situation value="off" />
of decades we've seen the emergence of the kind of global</p>
Demonstration – IBM AbilityLab Media Captioner & Editor (Collaborative Caption Editing System)
• Launched in October 2012
• All submitted videos are required to have captions
• Media Library is transforming from repository to include presentation also

More than 6 million students, nearly one-third of all students in higher education are learning on-line by using distant learning systems in U.S. (2011)

Accessible learning is difficult to provide...
- Text-to-speech (TTS)-based reading-aloud is rarely supported.
- It is possible to add captions to movies but effective authoring environment (e.g. crowd-sourcing) is not supported yet.

Difficulty of diverse contents types
- Texts, videos, presentation files, Q&A forums, tests / quizzes, games, etc.
Easy Web Learning (Overview)

Provides user interface improvement features (front end: EWB) and authoring helper (back end: CCES) as add-on for LMS.

Learning Management Systems
(e.g. Desire2Learn, etc.)

Easy Web Learning
*Frontend*

- Magnification
- Color settings
- Reading aloud
- Captions
- Audio Descriptions

Easy Web Learning
*Backend*

- Learner portal
- Course management
- Learning material repository
- Authoring tools
- Educator dashboard

Students with learning difficulties (PwDs, dyslexia, learning disabilities, etc.)

Teachers and faculties, and assistants

© 2013 IBM Corporation
EWB-Cloud demonstration

You tube web page for IBM AbilityLab Voice Transcription Manager
http://www.youtube.com/watch?v=NPNgQYyCFKc

EWB obtains caption from repository and shows it in the caption window

CCES Caption repository
Thank You!!