USER EFFICIENT SPEECH RECOGNITION-BASED CAPTION EDITING SYSTEM

Ali Sobhi, Cole Miyamoto
March, 2009
Session ID: DHH-1010
Wednesday, March 18, 2009
3:10 PM – 3:40 PM
Marriott – Saddle Brook Rm

Speaker:
Ali Sobhi
Agenda

IBM Digital Multimedia Captioning and Editing System

Caption Editing System

Demo
Objectives & Deliverables

Define an appropriate captioning / transcribing and editing solution for HA & AC
   What we think is appropriate and makes sense
   What we help deliver through existing components and new work

Deliver a captioning system that can process media in the IBM media library
   Align with CIO requirements for accessible media

Create asset that can be commercialized
   Assist HA & AC service delivery organization
## Matrix of Media and Technology Components

<table>
<thead>
<tr>
<th>Media Source</th>
<th>Digital Recording</th>
<th>Media Transformation</th>
<th>Transcription</th>
<th>Edit</th>
<th>Translation</th>
<th>Edit</th>
<th>Media Generation / Synchronization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadcast - TV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Live</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Recorded</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broadcast – Radio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Webcast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Podcast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Audio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Speaker notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Charts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VoIP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Teleconference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• One-to-one</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• IM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• A / V Chat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Complex

### Many components
Products / Components Consideration

Caption Editing System (CES from TRL)
  Editing (master / client), Controlling GUI

Broadcast Captioning (BCC from CRL)
  Closed-caption merge and media generation

HLT’s new ASR engine controller
  API control for Automatic Speech Recognition (ASR) web service
High-Level Architecture

Requirements:
- Media upload
- Multi-user editing
- Speaker-independent
- High accuracy
- Batch processing
- Closed-captioned media
- Web-based & SMIL
- Transcription text

CES Components:
- Controlling GUI
- Editing – Master / Client

ASR Components:
- Web Service – SOA
- ASR
- Timing and Files

BCC Components:
- Closed-Caption
- Media generation
- Sync & highlighting

Future
Usage Scenario

Who will use this?
• PwD – deaf / hard of hearing
• Captioning of media archives
• Search engines
• Data mining
• Examples:
  • IBM
  • NCSU
  • NASA
  • National Archives
  • Broadcasting companies

How to use this?
• Install / configure client GUI
• Directory and file structure for I/O
• Define the DigiCapE server URL

• Upload appropriate media format to server

• Client GUI will be notified when the transcription is complete
  • The transcript with timing marks is sent back to the client and stored on local disk

• The user can then replay the media and view the transcript in the editing pane
  • User can edit the transcripts and adjust the speed and characteristics of the transcription

• The user uploads the modified transcription bundle back to DigiCapE server
  • Request type of output; Web, single media file

• The user is notified of the completion and the requested output media is stored locally on client
Possible Business Models

Client & Web Services Model

**Client:**
- Direct sale
- Updates
- Upgrades
- Training

**SOA Services:**
- Services-based
- Registration fee
- Usage fee
- Can be upgraded w/o any service interruption

Enterprise Solution Model

**Enterprise Sale:**
Complete sale and service inside
Agenda

IBM Digital Multimedia Captioning and Editing System

Caption Editing System

Demo
Caption Editing System: Overview

Captioning for Digital Archive of Lectures, Videos on the Web

Social Merit of Captioning
Accessibility needs for Seniors, Deaf or Hard of Hearing, Foreigners

Regulations and Guidelines
Section 508, WCAG 2.0, National Standards, Cooperate Instructions

There are still many technical issues to captioning

Speech

Speech Recognition Engine

Captioning System

Captioned Contents

Master Caption Editing System

Client Caption Editing System
The Big Picture: Collaborative Editing Community

- CES with Network Function
  - Wave Data ➔ Caption File

- Classes (Internal)
  - Open Contents, Community Classes, Key Note Lectures

- CES
  - CES Recorder ViaVoice ➔ Caption File
  - Student Editors ➔ Movie

- Server
  - Class Contents
  - Community Class Contents ➔ Caption File

- Collaboration with NPO
  - Volunteers

- Students w/ or w/o handicap

Provide Information to Local Citizens
Collaborative Editing by Role

One Skillful and Responsible Master Editor just flags right (O) or wrong(?) or delete (X) by listening to the audio

Novice Client Editors makes the actual correction

- Aligned buttons allow easy flag ops
- Complete Audio Synchronization
- Target Line Emphasis Audio Play
- Mouse-free Operation
- Automatic Jump

FTP or LAN
Complete Audio Synchronization

Ondemand Business

5 2.715 Onde And
6 3.393 Business

<Initially, Focus on the Audio>

I got it!!
<Audio Stops>

5 2.715 Onde And
6 3.393 Business

<Key Input>
Cursor to the right
Delete

“What was it?”

“Onde …”

<Audio restarts automatically>

Ondemand Business

5 2.715 Ondemand
6 3.393 Business

<Key Input>
Insert M, correct A to a
<Fix Complete>
<Jump to next error line>
**Media Coverage**

**Newspaper**
- **National**
  - 2005/03/15 Asahi Shinbun
  - 2005/03/23 Nikkei Sangyou Shinbun
  - 2005/03/30 Kagaku Kougyo Nippo
- **Local**
  - 2005/03/23 Shinano Mainichi
  - 2005/03/23 Yomiuri Shinbun
  - 2005/03/23 Nihon Keizai Shinbun
  - 2005/06/07 Nihon Keizai Shinbun
  - 2005/06/07 Shinano Mainichi
  - 2005/06/07 Toushin Journal
  - 2005/06/07 Shinshu Minpo

**Magazine Article**
- 2005/03/22 Weekly ASCII

**Papers**
- HCI International-Las Vegas 2005
- Rehabilitation Kougaku
- 2005 Education System Information CSUN 2006
- Closing the Gap 2006
- IEICE 2007
- HCI International-Beijing 2007

**TV**
- **National**
  - 2005/06/06 JNN CS (News Bird)
- **Local**
  - 2005/03/22 Nagano Asahi Housou
  - 2005/03/22 Shinetu Housou
  - 2005/06/06 Ueda Cable TV
  - 2005/06/06 NHK Nagano
  - 2005/06/06 TV Shinshu
  - 2005/06/06 Nagano Asahi Housou
  - 2005/06/06 Ueda Cable TV

**Internet Articles**
- 2007/10/26 Yahoo! Japan News
- 2007/10/26 Enterprise Watch
- 2007/10/26 PC Watch
Agenda

IBM Digital Multimedia Captioning and Editing System

Caption Editing System

Demo
Demo

Demonstrations:

1) Transcription by the latest ASR engine
2) Edited result of the transcription
Any Questions ??
Visit us
@
IBM Accessibility Booth # 270
Marriot
Accessing the Future:
A global collaborative exploration for accessibility in the next decade

July 20-21, 2009
Boston, MA

Conference tracks:
• Universal Design and Accessibility Standards
• Patient-centered Collaborative Care
• Accessible Online Workplaces and Communities
• Travel & Transportation