
Effective Video Captioning using Collaborative Editing

Reiko Nagatsuma, Kohtaroh Miyamoto and Takashi Saitoh
IBM Research- Tokyo, IBM Japan

Effective Video Captioning Using Collaborative Editing

Prepared by:

IBM Research Lab, Yamato, Japan

Reiko Nagatsuma, reiko@jp.ibm.com

Kohtaroh Miyamoto, kmiya@jp.ibm.com

Takashi Saitoh, saito@jp.ibm.com

Presented by:

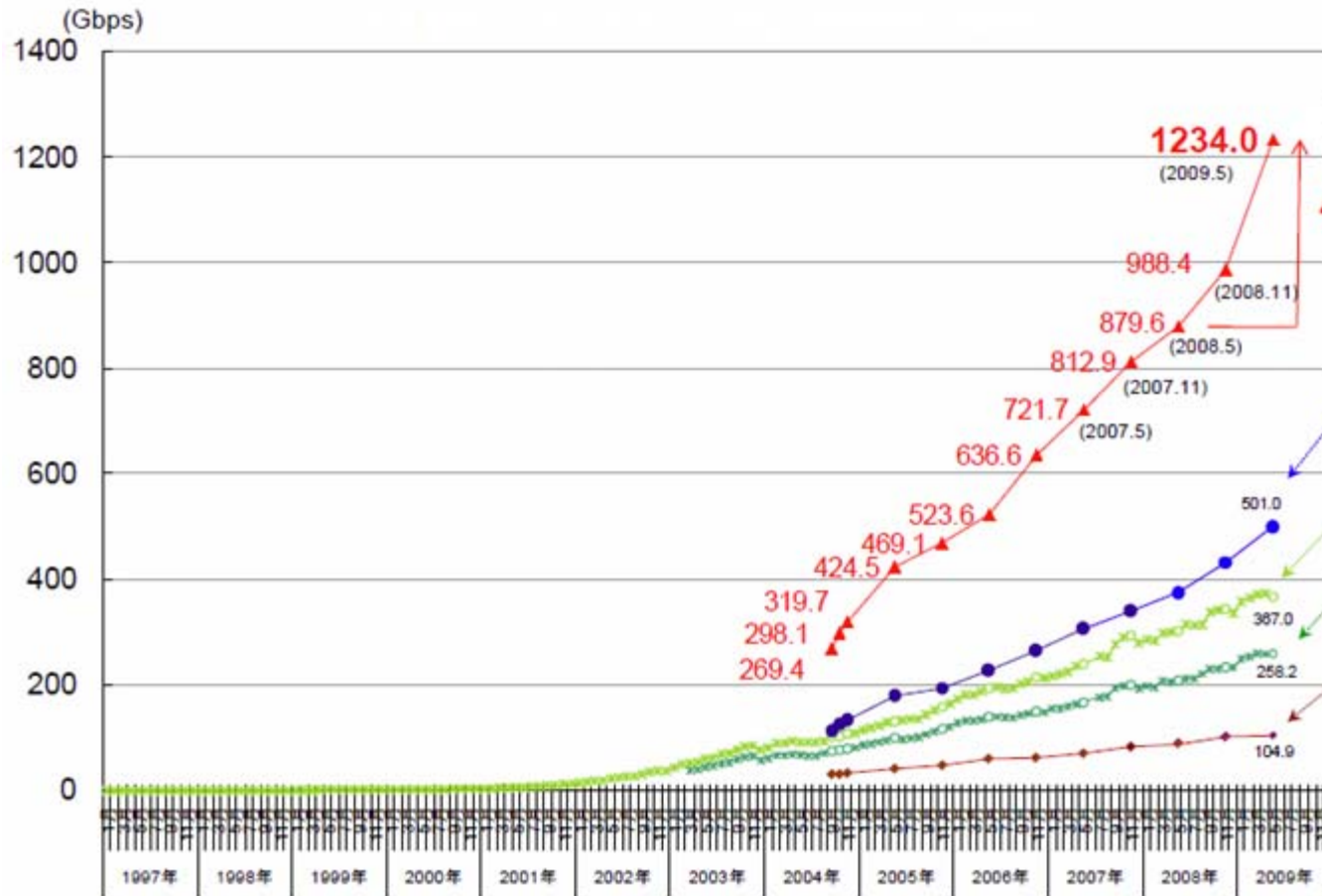
Ali Sobhi – IBM Research, HA & AC

sobhi@us.ibm.com

Session ID:	WEB-1012
Date:	Thursday, March 17, 2011
Time:	12:00 PM
Duration:	30 Minutes
Room:	Emma B

Needs & field of the multimedia contents including movie & audio is rapidly expanding

Variation of the network traffic in Japan (average)

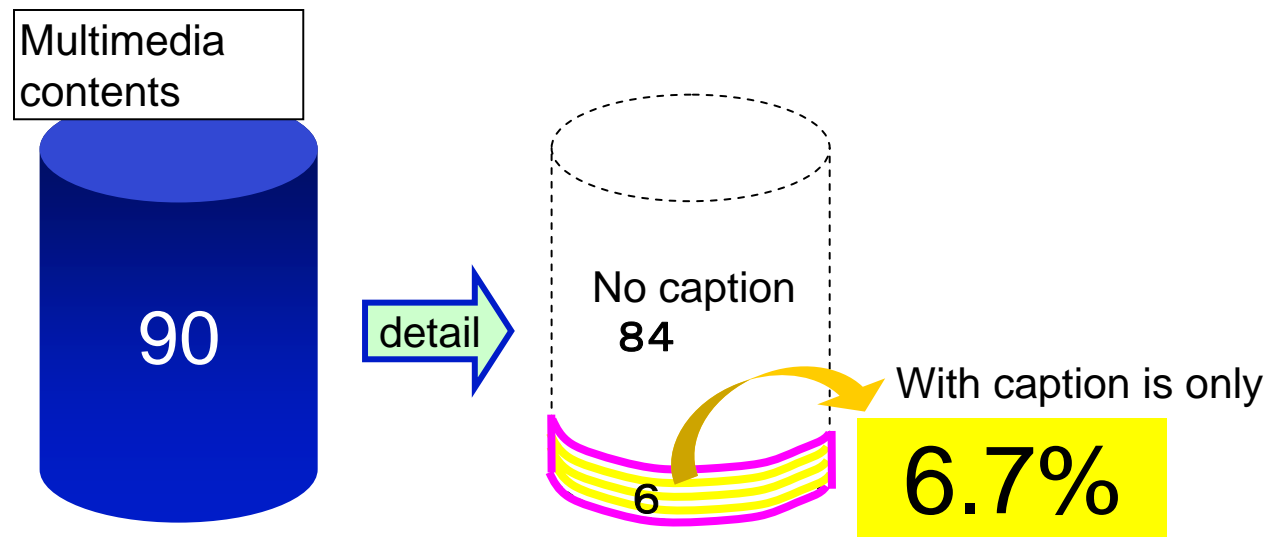


The data stream on the internet increased about 40 % from 2008 to 2009.

34% increase per person

This is due to the dissemination of multimedia contents

WCAG2.0 suggests that the multimedia on the internet should have captions. But, the accessibility of the internet multimedia contents is hardly progressing.



- Ratio of the caption of the interview with the governor in Japan(*1)
 - With Caption : 6/90 (6.7%)
 - Without Caption : 84/90 (93.3%)

*1) August 2009

The problems of multimedia web content

Actual state of speech recognition

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 neglected anymore on the Internet

Problems in editing and listening to the voice

Need to raise the efficiency of the type with listening to the voice:

- Captioning is the VDT work, VDT work incurs semi-acute fatigue
- Long durations of caption editing are not effective
- Break time of several minutes (10 min or more) is required to heal this fatigue

One way to prevent fatigue and maintain productivity is to divide up the long contents into short segments

Overview of the collaborative caption editing system (CCES) (1)

Basic concept

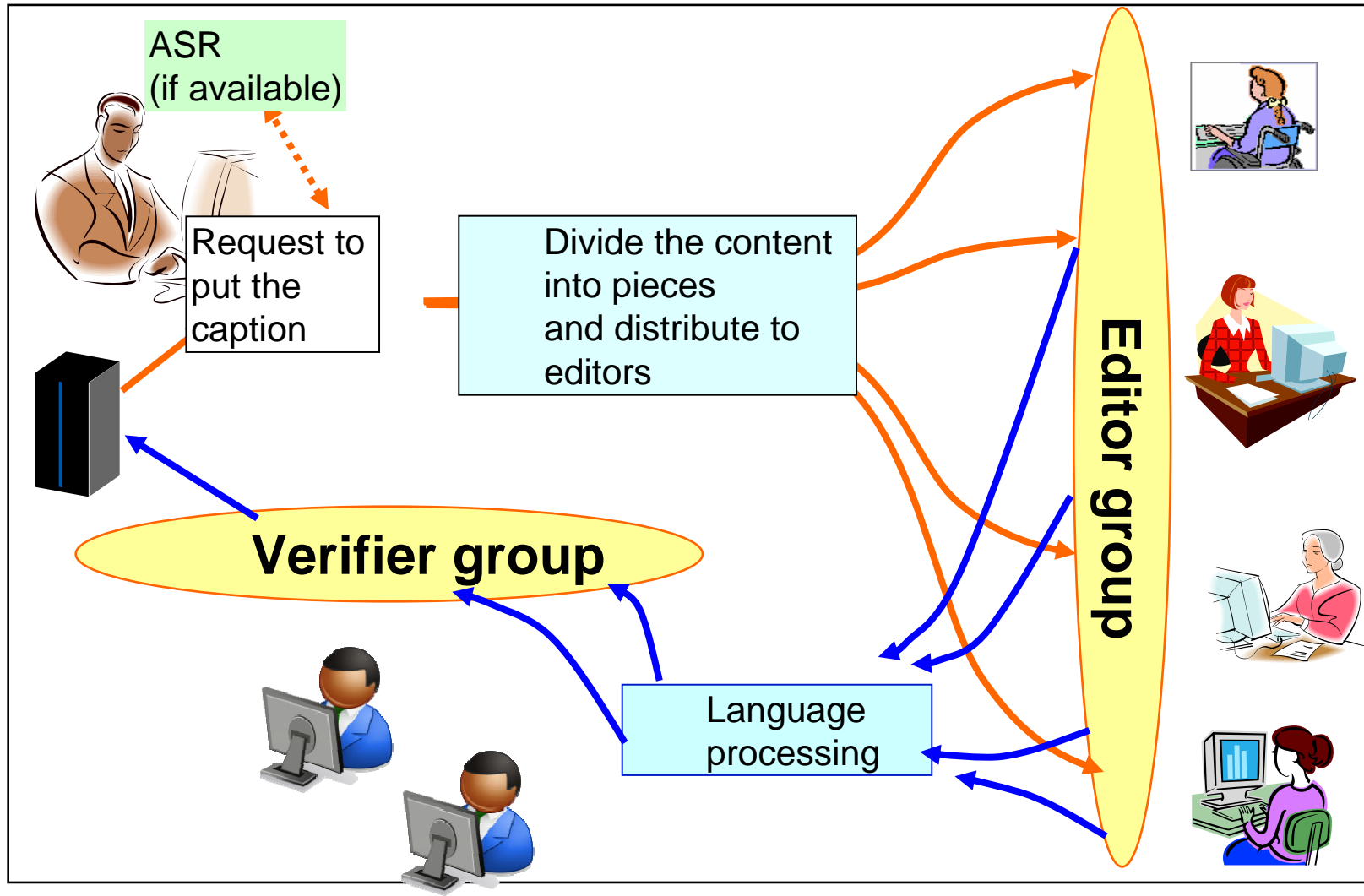
1: shorten the work unit

2: captioning with many people

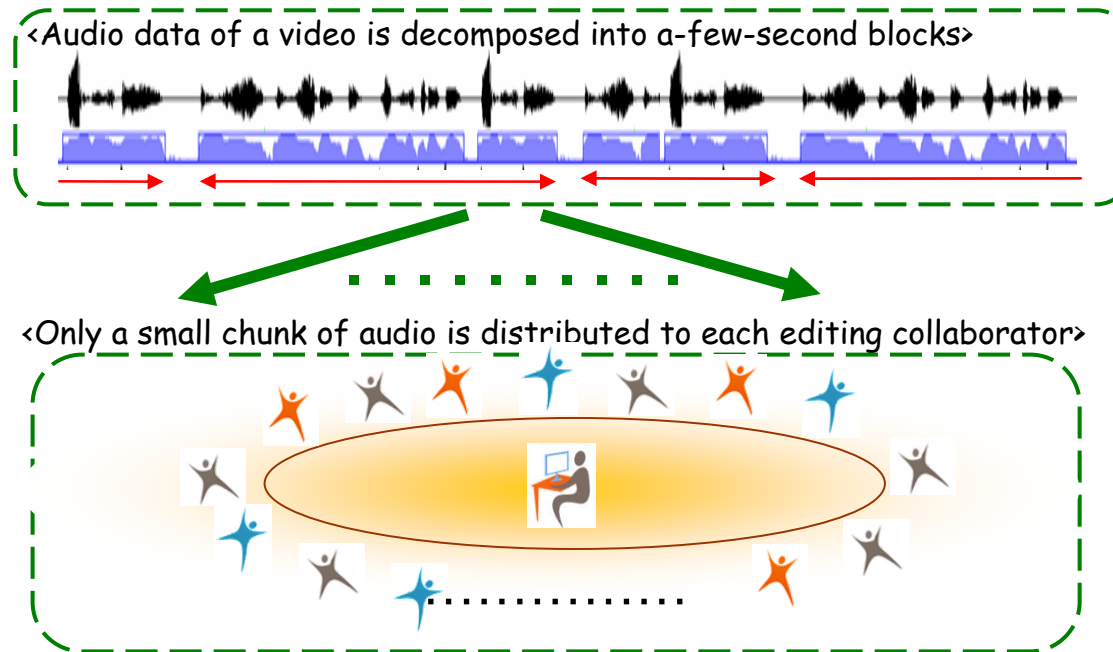
3: then shorten the total time to add the caption to the contents

- Separate the work items to collaborative editing with several persons
- Design the editor for easy input of the captions and listening to the voice
- After adding the caption, the system proceeds to the language processing phase

Overview of the CCES

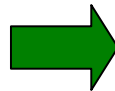


Overview of the CCES (2) – working unit



Audio decomposition

- Decompose into short speech blocks
- BGM are fairly OK
- Basically language independent
- Preliminary experiments have been done



Collaborative captioning

- Use the same CES-Web of DigiCapE
- Each editor inputs captions only for a 1-minute or less voice chunk
- Editing line length is appropriate for typing text by listening and memorizing

Overview of the CCES (3) – Editing view

- Input the caption data with audio playing repeatedly in a loop
 - System finds the low power point (like a breath) and cuts the voice data around 3 to 6 sec.
 - Move to next line with <Enter> key
 - Input the data, both audio and video can be played back
-
- If ASR is available, the recognition rate should be higher than 70%
 - If not, the efficiency of the editing is lower than without ASR

test

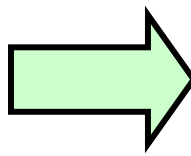
	Start Time	Status	Selection	Caption	Audio
1	00:00:00.109			We just talk about the Caption Editing System with collaboration	
2	00:00:18.444			It is not a	
3	00:00:19.113			editor but	
4	00:00:22.566				
5	00:00:25.241				

Overview of the CCES (4) – language process

- Corrected the line with specified character length which required by contents owner
- Indicated the candidate of the interjection
- Detecting the spelling inconsistency is in the future investigation plan

Caption

More major network channels are setting up video players on their sites..and the good news is, the players show captions! More and more captioned programming is now available through Fox.com ... read the review at Disabled in the Digital Age, and others. Plus there is a new online tv broadcaster, Hulu.com, that makes some captioned programming available.

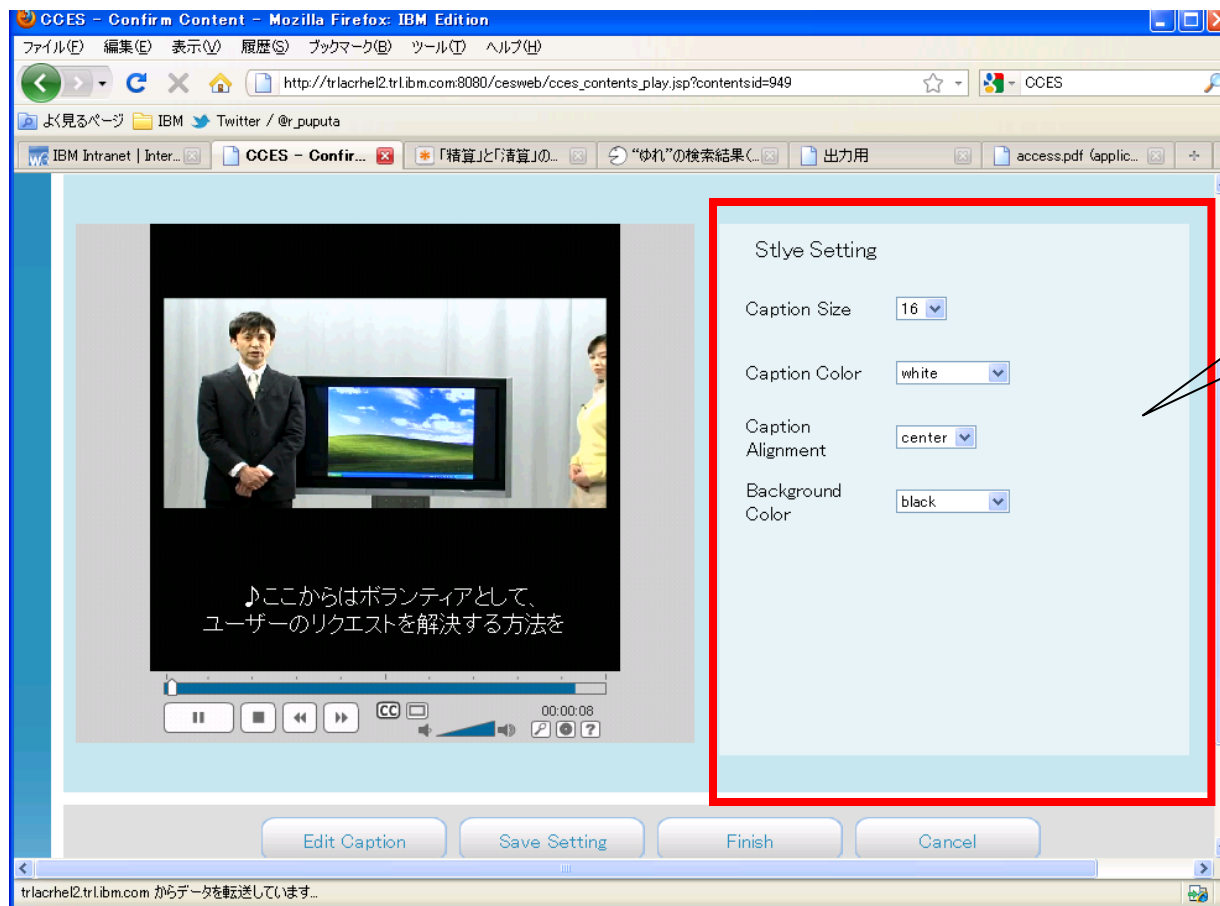


1 Caption

More major network channels are setting up video players on their sites..and the good news is, the players show captions! More and more captioned programming is now available through Fox.com ... read the review at Disabled in the Digital Age, and others. Plus there is a new online tv broadcaster, Hulu.com, that makes some captioned programming available.

Overview of the CCES (5) – contents owner

- Input some policy for the requested contents
 - character length, verifier task, and so on
- Check the contents after adding the caption



It is able to change the attribute of the caption

Functions and Characteristics of Captioning

- **Collaborative editing by minimizing workload per person**
 - Eliminate the work burden because of the short work unit per person (1 min per person)
 - One can join as an editor from normal web browser
 - Suitable for home office, because of minimum data transfer during editing.
- **User-friendly editor even for the aged and the visually impairment person**
 - Captioning can be controlled by a keyboard
 - User interface designed for input efficiency and avoiding the difficult setup and controls
 - Allows registering profile of the editors for the area of specialty
- **Language processing for compensating the editorial error by large-scale collaborative editing**
 - Spelling inconsistencies can be detected by the system automatically
 - Pre-format the line length and ray-out the text suitable to the contents automatically
 - The results of voice recognition can be edited as initial data of captioning.

Advantages

Creation of work style based on new social networks

Wisdom of clouds about accessibility and captioning itself

Drastically extending its coverage of Web media

Summary

- **CCES is the system that can work with many people.**
 - **It can work with/without ASR**
- **More features are in plan**
 - **Predictive text input**
 - **Inquiry to other editor**
 - **Additional language process support**