

# Researchers revolutionize closed captioning

March 22 2012, by Lisa Zyga

---



Examples of different captioning styles: (a) scroll-up captioning; (b) pop-up captioning; (c) paint-on captioning; (d) cinematic captioning; and (e) dynamic captioning. The first four techniques can be categorized as static captioning, and different from them, dynamic captioning in (e) benefits hearing impaired audience by presenting scripts in suitable regions, synchronously highlighting them word-by-word and illustrating the variation of voice volume. Image credit: Hong, et al. ©ACM

(PhysOrg.com) -- Ever since closed video captioning was developed in the 1970s, it hasn't changed much. The words spoken by the characters or narrators scroll along at the bottom of the screen, enabling hearing impaired viewers - or all viewers when the sound is off - to follow along. Now a team of researchers from China and Singapore has developed a new closed captioning approach in which the text appears in translucent

talk bubbles next to the speaker. The new approach offers several advantages for improving the viewing experience for the more than 66 million people around the world who have hearing impairments.

The researchers, Meng Wang from the Hefei University of Technology in [China](#) and colleagues, won the Best Paper Award for their work on the new closed captioning method from the Association of Computing Machinery (ACM) Multimedia Conference in October 2010.

“The whole technique was motivated by solving the difficulties of hearing-impaired viewers in watching videos,” Wang told *PhysOrg.com*. “These viewers have difficulty in recognizing who is speaking, so we put scripts around the speaker's face; they have difficulty in tracking scripts, so we synchronously highlight the scripts.”

As the researchers explain, conventional closed captioning can be considered static captioning, since all spoken words are represented in the same way at the bottom of the screen, regardless of who said them or the vocal dynamics. In contrast, the researchers describe their new technique as dynamic captioning, since the text appears in different locations and styles to better reflect the speaker's identity and vocal dynamics. For example, the text is highlighted word by word in synchrony with the speech signals. In addition, a small indicator next to the talk bubble shows the variation of vocal volume.

Moreover, all of these features can be automatically implemented without any manual intervention. The engineers developed algorithms to automatically identify the speaker using the video's script file along with lip motion detection. Using a technique called visual saliency analysis, the technology can automatically find an optimal position for the talk bubble so that it interferes minimally with the visual scene. Professionals can also further adjust the generated captions, such as moving the talk [bubbles](#). When the speaker is off-screen, or a narrator is speaking, the

words appear at the bottom of the screen as in static closed captioning. The system estimates vocal volume of words and phrases by computing the power of the audio signal in 30-millisecond windows.

Processing a video for dynamic captioning takes approximately the same amount of time as the video duration itself (videos cannot be processed while running). However, processing time can vary depending on complexity. The researchers predict that the processing time can be significantly reduced by speeding up some of the individual processes.

In a user study with 60 hearing impaired individuals aged 11 to 22, the researchers found that 53 of the 60 individuals preferred dynamic captioning over static captioning. The seven individuals who chose static captioning mainly did so due to their familiarity with that method. On average, the users rated dynamic captioning higher than static in terms of enjoyment, and about the same in terms of naturalness, mainly due to some instances when the text position changes abruptly. The researchers hope to solve this problem by smoothing the variation in text position.

“In the technical papers, we have mentioned that there are several failure cases, such as putting scripts around an incorrect faces,” Wang said. “This is the main bottleneck for commercialization. In order to be commercialized, a better way is to further incorporate human intervention. For example, a professional user can quickly check the generated dynamic captions and then manually correct or edit some failure cases. It will cost much less time and effort than the pure manual generation of the whole captions as the user only needs to process those incorrect cases. We have already been studying it.”

Since this work is the first to help hearing impaired individuals enjoy an improved video experience, the researchers note that there is a lot of potential future work in this area. In addition to improving dynamic captioning, they hope to apply the technique to videos without script

files, as well as to perform more comprehensive user studies.

**More information:** Richang Hong, Meng Wang, et al. “Dynamic Captioning: Video Accessibility Enhancement for Hearing Impairment.” Proceedings of the International Conference on Multimedia. [DOI: 10.1145/1873951.1874013](https://doi.org/10.1145/1873951.1874013)

Richang Hong, Meng Wang, et al. “Video Accessibility Enhancement for Hearing Impaired Users.” ACM Transactions on Multimedia Computing, Communications, and Applications. [DOI: 10.1145/2037676.2037681](https://doi.org/10.1145/2037676.2037681)

© 2011 PhysOrg.com

Citation: Researchers revolutionize closed captioning (2012, March 22) retrieved 26 April 2024 from <https://phys.org/news/2012-03-revolutionize-captioning.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.