May 15, 2017

Displaying Closed Captioning Using Speech-To-Text

Brant Candelore
SonyDp@convergenceiplaw.com

Mike Nejat

Peter Shintani

DISPLAYING CLOSED CAPTIONING USING SPEECH-TO-TEXT

IPD# 201605720

ABSTRACT

A system and method for providing synchronized Closed Captioning (CC) for live programming is disclosed herein. The method includes providing CC by speech-to-text conversion of the live content followed by synchronization of the CC with the dialogue to be displayed in a device. The speech-to-text conversion may be run in real-time or non-real time. The system provides annotations for different speakers and background music. The live content may also be stored on a video recorder, such as a set-top box, and processed offline. Similarly, a web stream may be recorded and processed offline. Alternatively, it may also be possible to enter the URL of the content into a processing website which will add CC to the video. The disclosed system and method provide for synchronized CC for live broadcasts and for user-generated content delivered over the internet for the hearing-impaired.

BACKGROUND

One of the biggest complaints with closed captioning (CC) is that it is not supported for use in live broadcasting and streaming of user-generated content delivered over the internet, and thereby unavailable to the hearing-impaired. CC, if available for live programs is delayed from the actual dialogue. There is a lack of useful solutions for users requiring CC for live broadcasts or content.

DESCRIPTION

A system and method for providing synchronized closed captioning (CC) for live programming is provided herein. The method as depicted in FIG. 1 includes providing CC, if unavailable, by speech-to-text conversion of the live content followed by synchronization of

the CC with the dialogue to be displayed in a device.

The speech-to-text conversion includes separating voice frequencies from background sounds prior to submitting it to an interpretation engine. The interpretation engine may be provided with data concerning the type of content e.g. football or soccer game, user generated content about automobile repair, etc. The speech-to-text conversion distinguishes between different speakers and also background music, which is annotated in the CC. The conversion of speech-to-text may be in real-time or non-real-time. The content may be stored on a video recorder, such as a set-top box, and processed offline. Similarly, a web stream might be recorded and processed offline. Alternatively, it may also be possible to enter the URL of the content into a processing website which will add CC to the video.

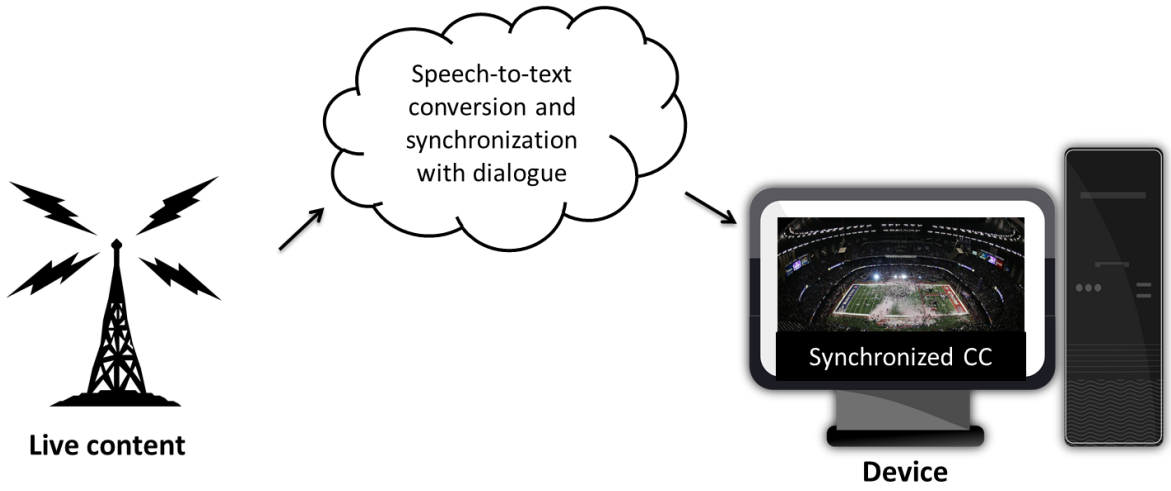

FIG. 1 – A method for providing synchronized CC for live content

The disclosed system and method provide for synchronized CC of live programming such as a football game or user generated web content for the hearing-impaired.