September 20, 2019

Universal accessible closed captioning and sign language

Daniel V. Klein

Joey Gutierrez

Universal accessible closed captioning and sign language

ABSTRACT

Visual performances often include audio portions, including spoken words. The verbal content can be difficult to hear or otherwise understand for individuals that have hearing impairments or are not fluent in the language of the verbal content. This disclosure provides a solutions to make verbal content of visual performances accessible to audience members. The described assistive techniques utilize a smartphone to present to audience members, closed captions or American Sign Language (ASL) interpreters of the verbal content portion, simultaneously as the audience members view the performance. The translation can be presented as an overlay with a view of the performance (AR), as an overlay on a projected image of the performance, (VR), or on a display screen of the smartphone (side-channel).

KEYWORDS

- Closed caption
- Subtitle
- American Sign Language (ASL)
- Augmented reality (AR)
- Virtual reality (VR)
- Wearable display

BACKGROUND

Visual performances, such as performing arts, informative presentations, on-screen shows, and the like, typically include an audio portion. An audience member is sometimes unable to hear sounds or understand the verbal content. For example, this may occur for any reason e.g., the audience member having a hearing impairment; voices of presenters being too

soft; accompanying music overpowering the words; a character's accent being unfamiliar to the audience member; verbal content being in a language that the audience member does not understand; etc. An audience member that is not able to clearly hear or comprehend verbal content relies on assistive solutions to receive the message being conveyed in the verbal content. Current methods of assisting with verbal content are suitable only for a subset of the audience.

Such solutions that make audio materials accessible include closed captioning and ASL interpreters. To rely on ASL, an individual needs to possess knowledge of sign language. Also, an audience member that uses an ASL interpreter needs to focus solely on the interpreter and needs to be positioned not be too far away from the interpreter. Assistive hearing devices, such as those provided at some venues may not support individuals who are completely deaf. Subtitles and supertitles provided at some venues, such as movie theatres and lecture halls, can be distracting and undesirable to people with sufficient hearing and understanding of the spoken language. There are no universal solutions that fit all audience members.

DESCRIPTION

Techniques are described for providing assistive relief to hearing-impaired individuals and individuals who experience language barriers. A smartphone or other device is used to present closed captioning text or an ASL interpreter (hereinafter, referred to as "translation"). The described techniques can be utilized at any venue that presents a visual performances to audience members, e.g., live theaters, movie theaters, lecture halls, music halls, etc. The translation can be displayed via augmented reality, virtual reality, or as a side channel.

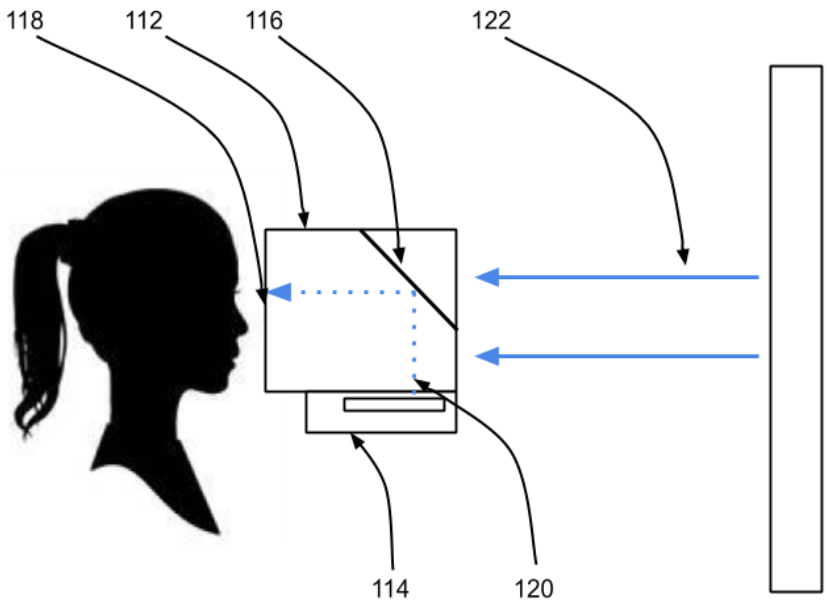

Fig. 1: Displaying translation via augmented reality

A user device, e.g., a smartphone of the audience member that wishes to view a translation of a performance (122), is used in augmented reality mode as follows. The user wears a headset (112) with a smartphone (114) mounted therein and a half-silvered mirror (116) positioned at a 45°, which provides a viewing portion (118). The translation (120) is provided as an augmented overlay with the view of the performance through the viewing portion.

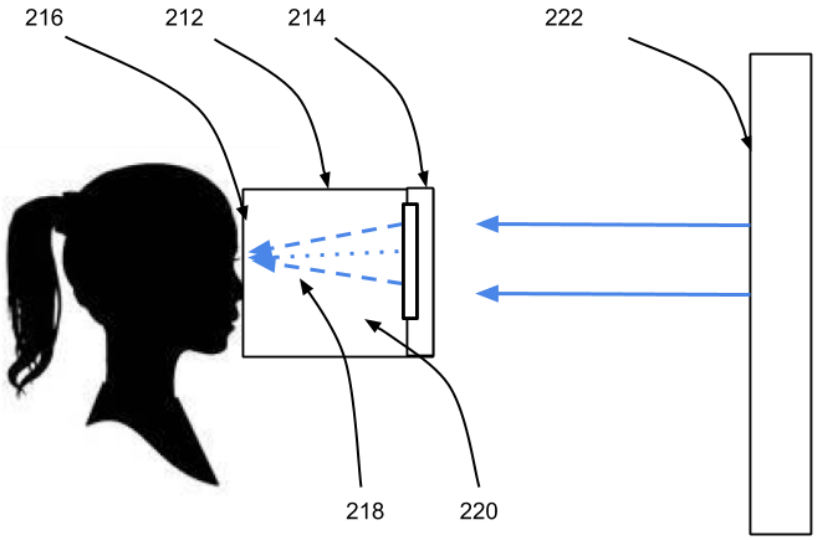

Fig. 2: Displaying translation via virtual reality

A user device, e.g., a smartphone of the audience member that wishes to view a translation of a performance (222), is used in virtual reality mode as follows. The user wears a headset (212) with a smartphone (214) mounted therein. The performance is reproduced in the view portion (216) with the translation (218) overlaid on a projected image (220) of the performance.

A locking feature is optionally provided in the augmented reality or virtual reality modalities. The lock is used to maintain the location of the translation relative to the user's view of the performance. An accelerometer or inclinometer of the smartphone device detects user movement. For example, the user turning their head away from a movie screen is detected and accordingly, the translation is updated to appear in the same location as before the turning of the head.

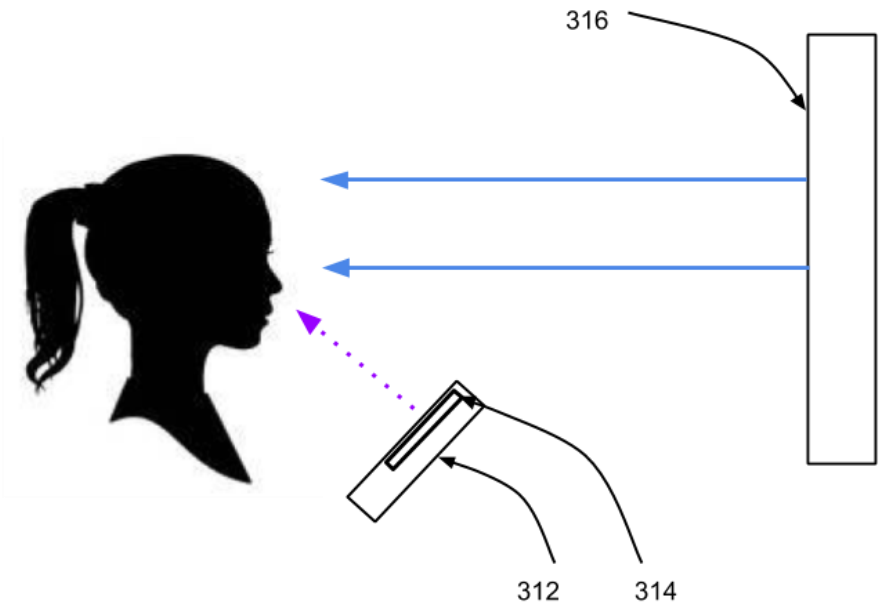

Fig. 3: Displaying translation as a side channel

A user device, e.g., a smartphone of the audience member that wishes to view a translation of a performance (316), is used in side channel mode as follows. The user holds a

smartphone or other device (312) while viewing the performance. The translation is displayed on a screen (314) of the device. The user can look away from the performance to view the translation on the display screen.

The translation for display in the various modalities can be obtained in a variety of ways. For example, the smartphone can communicate with a source of translation data at the venue. For example, translation data may be transmitted via infrared communication, WiFi, Bluetooth, or a translation device connected to a data port of the smartphone device. Alternatively, or in addition, a data stream that includes the translation can be received from a computing device, e.g. cloud server. Transmission of the data stream may be through WiFi, cellular, or a translation device connected to a data port of the smartphone device.

Speech-to-text applications, e.g., that include natural language parsing capabilities, can be included on the smartphone to translate verbal content into closed captioning. For example, a speech-to-text feature of a messaging application can be used to translate a live audio stream that is being broadcast to assistive listening devices provided at a venue.

The translation received at such a device is synchronized with the verbal content of the performance. For example, the venue can direct the timing of the transmission to coincide with the time of the performance. Alternatively, or in addition, the timing of the transmission can be synchronized with the performance by referencing the performance title and start time. An audio signal, e.g. a clapper sound that indicates the start of the show, can also be embedded in an on-screen show.

Translations are paired with corresponding performances when multiple performances occur at a particular venue, such as a movie theater. With user permission, the user location can be identified, e.g., by fine-grained GPS, location beacons such as Bluetooth low energy beacon

(BLEB), etc. and be used to determine the corresponding performance that an audience member is attending. If the user permits, a different source of information, e.g., a ticket receipt, such as an emailed receipt, can be used to identify the performance. Users can directly identify the performance, e.g., by scanning a QR code or a barcode as printed on a ticket stub. Optical character recognition (OCR) techniques can also be used to identify the show from scanned tickets.

A rewind or instant replay feature can be provided by storing the translation in non-persistent memory of the device. If an audience member misses a portion of the verbal content, a portion of the stored translation can be replayed. The audience member can request that the translation be played again from a specific earlier time, e.g., previous N seconds, to a previous phrase, etc.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure provides solutions to make verbal content of visual performances accessible to audience members. The described assistive techniques utilize a smartphone to present to audience members, closed captions or American Sign Language (ASL) interpreters of the verbal content portion, simultaneously as the audience members view the performance. The translation can be presented as an overlay with a view of the performance (AR), as an overlay on a projected image of the performance, (VR), or on a display screen of the smartphone (side-channel).

REFERENCES

1. Teofilo, Mauro Ricardo da Silva, and Vicente Lucena Jr. “Exploring Virtual Reality to Enable Deaf or Hard of Hearing Accessibility in Live Theaters: A Case Study” available online at https://tede.ufam.edu.br/bitstream/tede/7204/12/Tese_MauroTe%C3%B3filo_PPGI.pdf
2. De Angeli, Daniela, and Eamonn O'Neill. "A smartphone headset for augmented-reality experiences in museums." MW2015: Museums and the Web (2015).