

DEVICE AND METHOD FOR FACILITATING A PROVISION OF A CONTENT

Field of Invention

The present invention relates to a device and method for facilitating a provision of
5 a content.

Background Art

The following discussion of the background to the invention is intended to facilitate
an understanding of the present invention only. It may be appreciated that the
discussion is not an acknowledgement or admission that any of the material
10 referred to was published, known or part of the common general knowledge of the
person skilled in the art in any jurisdiction as at the priority date of the invention.

Deaf, Hard of hearing and Visually Impaired (DH&VI) have accessibility
challenges. Many DH&VI may be socially excluded from theatrical events such as
movies, plays, musicals or concerts, because there may be no means or rare
15 means for them to access such entertainment content. Even though subtitles may
be overlaid on a screen, the subtitles may be foreign subtitles for foreign content,
and not be closed captions for the DH. Some theatres have attempted to embed
inset sign language interpretation video, but this may distract the DH from the
actual content and as such the adoption of embedded insets has been limited.

20 Developing countries may encounter more technical challenges in implementing
captions. For example, captions embedded in the content may sometimes be lost
in the process of retransmission, in particular when the contents are transcoded
using bandwidth-saving data compression schemes. On the receiving end, many
viewers in the developing countries still have legacy TV sets that are incapable of
25 decoding and displaying the captions embedded in the contents.

Similar problems may be experienced by the VI. There may be no means or rare
means for them to access audio description tracks.

Even though some contents-sharing website provides captions, the captions may
be only for the contents provided by the website. In this regard, using the website
30 may not be applicable to the theatrical events.

Considering the above, there exists a need to provide a solution that meets the mentioned needs or alleviates the challenges at least in part.

Summary

Throughout the specification, unless the context requires otherwise, the word
5 “comprise” or variations such as “comprises” or “comprising”, will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Furthermore, throughout the specification, unless the context requires otherwise, the word “include” or variations such as “includes” or “including”, will be
10 understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

There are various types of Augmented Reality (AR) technologies and uses. The invention or disclosure described herein can augment a first content (also referred to as a primary/source/reference content) (e.g. movie, TV show, video, etc.) with a
15 second content (also referred to as a companion content) (e.g. captions, sign language video, audio description, etc.), delivered on-demand, and synchronised to the first content. The invention or disclosure seeks to provide a synchronised second content with a first content for the DH&VI.

The technical solution is provided in the form of a device and a method for
20 facilitating a provision of a content. In particular, the device can output a second content, for example the secondary content such as captions, associated with a first content, for example the primary content such as a video, to be synchronised with the first content outputted, by matching a first identifier associated with a temporal location of the first content outputted against one or more identifiers of
25 the second content and using the matched identifier of the second content.

In one aspect, there is a device for facilitating a provision of a content comprising: a processor operable to detect a temporal location of a first content outputted; and an output unit operable to output a second content associated with the first content, to be synchronised with the first content outputted, wherein the processor
30 is operable to extract a first identifier associated with the temporal location of the first content outputted, match the first identifier against one or more identifiers of

the second content, and use a temporal location associated with a second identifier matched with the first identifier to synchronise the output of the second content with the first content outputted.

5 In some embodiments, the processor is operable to control the output unit to output the second content from the temporal location associated with the second identifier.

In some embodiments, the processor is operable to extract a fragment of the first content which is associated with the temporal location of the first content outputted.

10 In some embodiments, the processor is operable to digitise the fragment of the first content.

In some embodiments, the device further comprises a database storing the digitised fragment of the first content.

15 In some embodiments, the processor is operable to extract the first identifier associated with the temporal location of the first content outputted, upon the digitised fragment of the first content is stored in the database.

In some embodiments, the processor is operable to identify the first content from the extracted first identifier.

20 In some embodiments, the processor is operable to search the second content associated with the first content from a reference database.

In some embodiments, the processor is operable to receive the second content from the reference database.

In some embodiments, the processor is operable to receive an audio input of the first content to detect the temporal location of the first content.

25 In some embodiments, the device further comprises a display unit operable to display the second content associated with the first content.

In some embodiments, the device further comprises an audio output unit operable to output the second content associated with the first content.

In some embodiments, the identifier is a fingerprint of the content.

In some embodiments, the temporal location is a time-related data. The time-related data may be a time stamp, the time stamp may have a data format comprising HH:MM:SS.FF wherein HH is the hour; mm is the minutes; ss is the seconds, and FF is the video/film frame number associated with at least one particular event.

In another aspect, there is a method for facilitating a provision of a content comprising: detecting a temporal location of a first content outputted; extracting a first identifier associated with the temporal location of the first content outputted; matching the first identifier against one or more identifiers of a second content associated with the first content; using a temporal location associated with a second identifier matched with the first identifier to synchronise an output of the second content with the first content outputted; and outputting the second content associated with the first content, to be synchronised with the first content outputted.

Other aspects of the invention will become apparent to those of ordinary skilled in the art upon review of the following description of specific embodiments of the invention in conjunction with the accompanying drawings.

Brief Description of the Drawings

The present invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Fig. 1 shows a flow diagram in accordance with some embodiments of the present invention.

Fig. 2 shows another flow diagram in accordance with some embodiments of the present invention.

Fig. 3 shows an exemplary device of some embodiments of the present invention.

Fig. 4 shows another exemplary device of some embodiments of the present invention.

Fig. 5 shows another exemplary device of some embodiments of the present invention.

Fig. 6 shows another exemplary device of some embodiments of the present invention.

Fig. 7 shows another exemplary device of some embodiments of the present invention.

- 5 Fig. 8 shows another exemplary device of some embodiments of the present invention.

Other arrangements of the invention are possible and, consequently, the accompanying drawings are not to be understood as superseding the generality of the preceding description of the invention.

10 **Description of Embodiments**

Throughout the description, the term 'content' is to be construed broadly to include various information such as data, text, music, multimedia, and/or combinations of the aforementioned.

- 15 Fig. 1 shows a flow diagram in accordance with some embodiments of the present invention.

The device may comprise a processor and an output unit. The output unit may include, but not be limited to a display unit and an audio output unit.

- The processor is operable to detect a temporal location of a first content (S10). For example, the processor may comprise an audio input unit such as a
20 microphone, and receive an audio input of the first content to detect the temporal location of the first content.

- The processor is operable to extract a part or portion of the first content which is associated with the temporal location of the first content outputted (S20). In some embodiments, the processor is operable to digitise the part or portion of the first
25 content to generate a digital signal for the first content. The part or portion may also be referred to as a fragment or segment.

It may be appreciated that the device may further comprise a database storing the digitised fragment of the first content. The processor is operable to extract the first

identifier associated with the temporal location of the first content, upon the digitised fragment of the first content is stored in the database.

In some embodiments, the processor is operable to identify the first content from the extracted first identifier. In some other embodiments, the processor may not
5 identify the first content, for example in case the first content is a predetermined content.

Thereafter, the processor is operable to match the first identifier against one or more identifiers of a second content associated with the first content (S30), and use a temporal location associated with a second identifier matched with the first
10 identifier to synchronise the output of the second content with the first content outputted (S40). For example, the processor is operable to control the output unit to output the second content from the temporal location associated with the second identifier.

The output unit may then output the second content associated with the first
15 content, to be synchronised with the first content (S50).

In this manner, the invention or disclosure can augment the first content which a user is watching or listening to with the second content (e.g. sign language video, caption, audio description, etc.), delivered in a synchronisation with the first content.

20 Fig. 2 shows another flow diagram in accordance with some embodiments of the present invention.

Content acquisition

a. The first content may be ingested in a digital format for an identification processing. In some embodiments, the content identification can be performed
25 using an audio fingerprint matching. As such, at least a fragment of the first content's audio may be ingested first. The first content's audio may be accessible in an analogue format (e.g. sound waves from a TV or movie theatre loudspeaker) or a digital format (e.g. audio component of an HDMI signal, internet radio stream, etc.)

b. The content acquisition step may be initiated and/or triggered manually by the user, typically via a control pad and/or panel (e.g. via actuating a start button). When triggered, the processor activates the microphone or line input, which feeds analogue audio into an analogue-to-digital converter (ADC) module. The ADC module digitises the first content's audio, and stores it in the database for the identification processing. If the audio source is available in a digital format (e.g. audio component of an HDMI signal), then the content audio acquisition can be more direct, without the need for analogue-to-digital conversion via the ADC module. In such a case, the digital audio can be stored directly into the database for the identification processing. Reliance on analogue or digital audio may depend on the embodiment. For instance, a smart glass embodiment (Fig. 3) relies on analogue audio, whereas a set-top box embodiment (Fig. 5) can extract digital audio directly from the HDMI signal. Regardless, a fragment of first content's audio information may be acquired for the purpose of the content identification. In this regard, the first content's entire audio information is not required to be captured.

c. Fingerprint extraction. Upon storage of a digital audio fragment into the database, the processor then extracts the fragment's audio fingerprint, for example using an audio fingerprint extraction algorithm.

Content and temporal location identification

a. Content identification. The processor is configured to search and match the fingerprint of the acquired audio fragment against a reference database containing one or more fingerprints. The reference database may exist internally in the device, or externally such as a remote database (cloud-based). If the fingerprint database is remote, the processor may comprise a data communications module to communicate with the remote data resources (e.g. internet/web server, database, etc.) Each fingerprint in the reference database may be an extract from other content titles (e.g. movie, TV show, etc.) The reference fingerprints may be extracted from the first content's entire audio track, not just fragments. Each reference fingerprint may also be logically linked to related metadata (e.g. title,

year, cast, summary, etc.) The search/matching algorithm generates a list of most likely matched fingerprints. The database returns a closest match to the processor.

b. The fingerprint matching query also returns the audio fragment's temporal location, in relation to the start of the matched reference content. For instance, if the audio fragment acquired when the start-button was actuated contained a lyrical line "Tonight, we will dine in hell", then the temporal location is the exact time from the start of the matched movie, when that line was output. This temporal location data is used to synchronise the delivery of the second content with the first content.

c. Related content. In addition to returning closest content matches, the query may also return related information, such as the nearest matches' content title, release year, cast, summary, etc. The query result may also include a list of available second content (e.g. sign language videos, captions, audio description, etc.) related to the matched content.

Second content selection

The processor received an audio signal of the first content being consumed, extracted the first content's audio fingerprint, matched it against a reference fingerprint database, identified the closest match, identified the temporal location, and retrieved related information about the closest match(es). The second content and the respective file locations of the same is also identified. At this juncture, the user may be given an option (via a user interface) to select from the list of available second content, or the device may automatically select one, based on the user's preference setting. For instance, if the user previously set that his/her preferred second content is American Sign Language, then that, if available, may be automatically selected for the said user.

Synchronised content delivery

a. Content buffering. Once the desired second content has been selected, the processor prepare/readies the second content for playback/delivery. Readying

may entail accessing the second content file and/or buffering of the same. The selected second content may be accessed locally, within the device's storage, or remotely, typically an external storage source, such as a media streaming server.

b. Synchronized playback. Once ready, the device's processor then plays the selected second content—synchronised to what the viewer is watching, based on the previously identified temporal location.

c. Output. The device then delivers the second content to audio/video interface(s), accordingly—sign language video to a visual display or HDMI port, audio description to an audio output unit (e.g. headphone jack, line out, speaker jack, etc.)

There may be various devices which the invention or disclosure may be applied, as follows:

a. Smart glass for sign language video display (referred to as “GlassSign 100”);

b. Smart glass for caption display (referred to as “GlassCap 200”);

c. Set-top Box for overlaying subtitles/captions over an existing video signal (referred to as “STB 400”);

d. Desktop Caption Display (referred to as “DeskCap 500”) for displaying captions/subtitles on a table-top character display;

e. Audio Description Sound Bar (referred to as “ADBar 600”) for delivering audio description track for the visually impaired;

f. Audio Description Line Device (referred to as “ADLine 700”) for delivering audio description track to other audio devices with line input capability (headphones, amplifier, etc.)

Each device may be a standalone embodiment; each may be self-contained, with all the necessary components to carry out the invention or disclosure, without the need for other external devices, such as smartphone, computer, etc.

Each device may comprise at least one of the following components:

- a. Processor (also referred to as “Computer Processing Unit (CPU)”) – microcontroller, FPGA, GPU, etc.
- b. Database (also referred to as “Memory”) – internal and/or provision for external memory; for OS, program code, data, media, etc.
- c. Power source – battery and/or line power
- d. Media input/output interfaces – video I/O (e.g. HDMI, video display, character display, etc.), audio I/O (e.g. 3.5mm, 1/4", RCA, XLR, built-in speaker, etc.)
- e. Data connectivity mechanisms – USB, Wi-Fi, Bluetooth, Ethernet, etc.
- f. Software programs – Operating system, program code, etc.
- g. Controls – switches, buttons, sliders, etc.
- h. Apparatus-specific components – microphone, analogue-to-digital converter (ADC), display driver, etc.

15

Fig. 3 shows an exemplary device of some embodiments of the present invention.

The device, hereinafter referred to as GlassSign 100, may be a smart glass that detects what the viewer is watching and then delivers a synchronized sign language video (and/or captions) on a video display in the smartglass.

- GlassSign 100 is primarily targeted for the use of DH patrons in a movie theatre, where the GlassSign's 100 private viewing feature does not distract neighbouring viewers. As such, the GlassSign 100 may be arranged to have a sufficient power (e.g. battery charge) for the duration of at least one full-length movie, and sufficient memory to locally store a full-length companion sign language video.
- The sign language video may be pre-loaded (e.g. via USB data transfer) by the movie theatre operator, and be updated when the movie schedule changes.

25

GlassSign 100 may comprise a video-capable display 101 such as a micro-OLED display, and a smartglass circuit board 102 which includes a microcontroller 103,

a memory 104, a battery 105, a power/data port (USB) 106, a control switch 107, a microphone 108, an ADC module 109 and/or a software 110. The GlassSign 100 may further comprise an enclosure 111 and an eyeglass mount 112.

The GlassSign 100 may be powered on with a prolonged press of the control switch 107. When switched on, the microcontroller 103 boots up the firmware software 110. Upon boot-up completion, the unit is ready, on standby mode, awaiting further control. When the control switch 107 is pressed briefly, the content detection process is initiated—the microcontroller 103 activates the microphone 108, which then receive the ambient audio of the first content that the user is watching; then the ADC 109 converts the analogue audio from the microphone into a digital audio file. The microcontroller 103 then extracts the fingerprint of the digitized audio fragment and stores the content fragment's audio fingerprint temporarily in the memory 104.

In some embodiments, since the GlassSign 100 for cinema is for specific companion use (e.g. the first content or movie is known), content identification may not be necessary while the temporal location is needed. The microcontroller 103 attempts a temporal location identification of the fragment's fingerprint against the movie's entire audio fingerprint, so as to effect the synchronised playback.

Once the temporal location is identified, the microcontroller 103 plays the companion sign language video (stored and/or pre-loaded in memory 104) and displays on the video display 101, using the temporal location as a timing reference for the synchronised companion content playback. If the sign language video is out of synchronisation, the process can be reinitiated by pressing the control switch 107.

The eyeglass mount 112 may be used to attach and/or adjust the circuitry enclosure 111 either to the viewer's own eyeglass, or to a default grade/refraction-free eyeglass, which serves to make GlassSign 100 a hands-free wearable implement.

Optionally, GlassSign 100 can also display text (captions, subtitles) on the video display. However, this may not be ideal because captions/subtitles are usually displayed across both visual fields (both eyes). In some embodiments, GlassSign

100 may be suited as a one-eye complement, akin to a picture-in-picture/inset video of a sign language interpreter in one corner of the screen. For a dedicated caption smartglass, there is GlassCap 200 to be described.

5 Fig. 4 shows another exemplary device of some embodiments of the present invention.

The device, GlassCap 200, may be a smartglass that detects what the viewer is watching, then displays synchronised captions/subtitles on the smartglass. GlassCap 200 is similar in functionality and general operation to GlassSign 100.

10 However GlassCap 200 may employ a text/character display that is projected and/or reflected onto a transparent/translucent glass in the user's field of view.

As shown in Figure 4, GlassCap 200 is a heads-up display (HUD) similar to conventional Virtual Reality / Augmented Reality HUDs. The HUD is worn on the head, typically its cushion 201 is pressed against the forehead, like a visor, held in
15 place by an adjustable/elastic band 202 wrapped around the head. A compartment 203 is an enclosure which holds most of the electronics, including components such as a microcontroller 204, a memory 205, a battery 206, a power/data port (USB) 207, an ADC module 208 and a software 209. Attached to the compartment is a graphical display 210, facing the transparent/translucent
20 glass 211, which may be adjusted and/or tilted to suit the user's preference.

The GlassCap 200 is powered on with a prolonged press of the control switch 212. When switched on, the microcontroller 204 boots up the firmware software. Upon boot-up completion, the unit is ready, on standby mode, awaiting further control. When the control switch 212 is pressed briefly, the content detection process is
25 initiated. The microcontroller 204 activates the microphone 213, which then receives the audio signal of the first content that the user is watching, and then the ADC module 208 converts the analogue audio from the microphone 204 into a digital audio file. The microcontroller 204 then extracts the fingerprint of the digitized audio fragment and stores the content fragment's audio fingerprint
30 temporarily in the memory 205.

Fig. 5 shows another exemplary device of some embodiments of the present invention.

The device, STB 400, may be a set-top box that overlays the synchronised second content over an incoming video feed. STB 400 uses the input video stream to identify the content and temporal location of the content fragment acquired. It then gives the viewer the option to select at least one secondary content, for example a caption/subtitle in a particular language. Upon selection, the STB 400 accesses the second content and overlays it over the incoming video, then outputs the composite (video input and second content overlay) onto the video output port.

As shown in Fig. 5(a) and (b), STB 400 has two video ports, for example an HDMI Video Input port 401 and an HDMI Video Output port 402. STB 400 also has an infrared sensor 403, a power/data port (USB) 404, an infrared sensor extension port 405, a control switch (Set-Up/Menu button) 406 and a STB enclosure 407.

The internal circuit board 413 is shown in Fig. 5(c). The internal circuit board 413 may comprise a microcontroller 408, a memory 409, a power/data port 404 and a control switch/button 406. There may also be new set-top boxes that are specific to the needs/usage context of the STB 400. Like other set-top boxes, there is a video output port 402. Because of the STB's 400 likely use with a TV monitor, there is a provision for infrared remote control (not shown). As such, there is a provision for an infrared sensor 403, as well as an infrared extension port 405, which may be useful should the STB 400 be set up in a location without a direct line-of-sight to the user (e.g. STB 400 behind the TV, or inside a media cabinet/drawer). There is also a Wi-Fi module 411, which is the STB's 400 means for accessing the second content (e.g. subtitles/captions), in addition to updating its firmware and/or communicating with a smart device/app. There is a Field Programmable Gate Array (FPGA) chip 407, which is used for the computationally intensive task of digital video overlays. It may be appreciated that the first set-top box to detect the content being fed into the video input port 401, overlay a synchronised second content to the same video, then output the composite video.

Employing a FPGA chip 407 enables the option to detect and/or identify the content by analysing the video. In some embodiments, the STB 400 may use the

same identification audio fingerprint matching method as the other embodiments. The audio fragment in this case is extracted directly from the HDMI input port 401, by programmatically separating the audio and video streams of the HDMI input signal.

5 Once an audio fragment from the HDMI input port 401 is acquired, the microcontroller 408 extracts the fingerprint and the matching query is initiated, to identify the first content, and the fragment's temporal location, in relation to the closest matched content. The matching second content options are displayed and/or overlaid onto the video output, which the user can see. The user selects
10 using the infrared remote control; the microcontroller 408 then fetches the second content (e.g. captions/subtitles), then overlays them onto the video, synchronised using the temporal location as a basis. The composite video (e.g. input video, plus synchronized captions) is crunched by the FPGA, and is fed to the HDMI video output port 402.

15

Fig. 6 shows another exemplary device of some embodiments of the present invention.

The device, DeskCap 500, may be a desktop character display that detects what the viewer is watching, then displays synchronised captions and/or subtitles on
20 the character display. DeskCap 500 is a simpler (no plugging of HDMI In and Out cables) and cheaper (no pricey video processing components, such as FPGA) alternative to the STB 400. DeskCap 500 is ideally suited for placement in front of a TV monitor and/or screen, serving as an external caption and/or subtitle display. DeskCap 500 is particularly useful when a user's TV set is old and incapable of
25 displaying closed captions and/or subtitles, and/or when the broadcaster does not provide captions/subtitles.

DeskCap 500 is nearly identical in functionality and operation to GlassCap 200. Both are operable to receive an audio signal (e.g. via an on-board microphone) to what the viewer is watching, identify the first content and its temporal location, and
30 then display the selected second content (e.g. caption and/or subtitles), in

synchronisation. DeskCap 500 is desktop display, while GlassCap 200 is a wearable device.

DeskCap 500 may be a digital character display (e.g. LED, LCD, or other digital display mechanisms), similar to a simple digital signage. DeskCap 500 is equipped with core circuitry for content identification and synchronised second content delivery to the display.

Fig. 7 shows another exemplary device of some embodiments of the present invention.

The device, ADBar 600, may be a sound bar (a desktop loudspeaker) that detects what the viewer is watching and then outputs a synchronised Audio Description track. Like conventional soundbars, ADBar 600 is arranged to be placed in front of a TV display or video screen. The conventional soundbar is arranged to improve audio, over the built-in TV speakers. ADBar 600 is similar, but goes the extra mile for the blind and/or visually impaired, by augmenting their experience with Audio Description track. ADBar 600 can also be used to replace a program's default audio track with a dub (foreign language) track or commentaries.

As with the previous apparatus/embodiments, ADBar 600 has a circuitry for receiving the audio signal to what the viewer is watching, identifying the content and temporal location, presenting audio track options, remotely accessing/retrieving the selected content, and then playing the second content (e.g. audio description, dub, commentaries, etc.) through the ADBar's 600 loudspeakers.

Fig. 8 shows another exemplary device of some embodiments of the present invention.

The device, ADLine 700, is an intermediating audio device that has two audio line (or headphone) ports, Line In and Line Out. ADLine's 700 internal circuitry detects the audio content being fed into the Line In input jack, then outputs the selected synchronized second content (e.g. Audio Description, dub tracks, audio

commentaries, etc.) to the Line Out port. The internals are similar to previous embodiments, except ADLine 700 does not employ a microphone. ADLine 700 acquires an analogue audio fragment directly from the Line In for digital conversion, fingerprint extraction/matching and identification.

5

It may be appreciated that the invention or disclosure can be used in other situations where synchronised second content may be useful. The invention or disclosure can be used for non-entertainment purposes, such as listening to live translations, watching live captions/sign language interpretation of an event (e.g. conference presentation, news, etc.). The invention or disclosure can also be used for educational purposes, in school or corporate training. In general, the disclosure may be applicable for synchronization of related content.

It may be appreciated that the invention or disclosure can also be used for musical accompaniment needs. For instance, the first content may be a rhythm section (e.g. drum and bass) of a musical piece, whereas the second content may be an audio accompaniment containing other musical parts, such as guitar, keyboards, and other instruments in an arrangement. The invention or disclosure can be used for an on-demand karaoke service, whereby backing tracks are delivered based on a song that was detected or was sung/hummed by the user. The smart glass embodiment may display the musical score, synchronised to the first content. The set-top box embodiment may serve overlays of lyrics, synchronised to the song.

It may be appreciated that the audio watermarking or video fingerprinting/watermarking systems can be used in lieu of the preferred audio fingerprint-matching methods. Such audio watermarking or video fingerprinting/watermarking systems may use the at least one content identification mechanism/method.

It may be appreciated that the devices may also be used for other purposes. ADBar 600 and ADLine 700 can be used for dub tracks, commentaries, etc.

It may be appreciated that some embodiments can be outfitted with Bluetooth™ and/or Wi-Fi™ capability that would allow additional controls and processing via

external devices, such as smartphones, tablets, computers, etc. Some embodiments can be conceived to be standalone, self-contained apparatus.

As mentioned above, GlassSign 100 can be enhanced to display captions on the video display as well, even if this set-up is not ideal because the GlassSign 100 video display is affixed to one side of an eyeglass. Captions are generally better displayed across both eyes' field of view.

GlassCap 200 can be enhanced (e.g. more powerful processor, more memory, and higher resolution display) to project and/or reflect sign language video, in addition to its original purpose of synchronised caption/subtitle display.

STB 400 can also be enhanced (e.g. more powerful processor and memory) to overlay sign language video. It can also be enhanced to overlay and/or replace the incoming audio track with a companion audio content (e.g. audio description, dub tracks, commentaries, etc.)

DeskCap 500 can be upgraded to a higher resolution screen, making it possible to display both sign language video and captions.

DeskCap's 500 core components can be repurposed and combined with a mini projector, to augment videos with projected captions/subtitles.

ADBar 600 and DeskCap 500 can be combined into one multimedia device, capable of delivering captions/subtitles and/or companion audio.

In the various described embodiments, the temporal location may include, but is not limited to, a time-related data. The time-related data may be a time stamp, the time stamp may have a data format expressed in the following manner: - HH:MM:SS.FF wherein HH is the hour; mm is the minutes; ss is the seconds, and FF is the video/film frame number associated with at least one particular event.

As an example, at a time-data indicating 01:00:01.05, or an hour, one second and five frames from the start of the reference multimedia content such as a movie/video may indicate an actor clapping.

As described above, the invention or disclosure has at least the following advantages:

- a. Stand-alone devices. The methods may be manifested in stand-alone, application-specific, Augmented Reality (AR) embodiments/apparatus: an AR smartglass embodiment, an AR Heads-Up Display (HUD) embodiment, a set-top box, a smart desktop character display, a smart loudspeaker, an AR audio intermediating device, as described above. The above reduce or omit the need to use expensive and complicated devices such as smartphones, tablets, computers, etc. for operating these embodiments.
- 10 b. Automatic content recognition/identification. Some embodiments can be self-contained to automatically identify the content that the user is watching or listening to. Upon identification, the invention provides available second content options, for example captions, subtitles, audio description, sign language video, foreign dub tracks, audio commentary, etc. beyond what was made available and
15 accessible by the first content provider (e.g. broadcaster, VOD service provider, etc.)
- c. Synchronized. Upon selection, the second content may be automatically synchronised to the program that the user is watching or listening to. For instance, an Audio Description track can be delivered, synchronized to what the Visually
20 Impaired (VI) is watching.
- d. Dynamic content. As mentioned, the second content, such as captions, provided by broadcasters/VOD operators are tightly coupled with their content. Their second content options are often limited depending on the service provider. The invention can augment and extend the content options.
- 25 e. Accessibility. Captions embedded by broadcasters can become corrupted, if not lost, as signals get transcoded in the video transmission chain, making the captions inaccessible. Similarly, in emerging markets, there may be many legacy TV sets that are incapable of displaying captions and/or subtitles. The invention can augment and extend that.
- 30 f. Crowdsourced content companion. The second content portal provides a self-service internet/web interface that allows content owners to self-manage their

second content. Better yet, viewers/fans can even contribute second content too, such as fan-generated subtitles. As such, additional second content can be added dynamically, even crowdsourced, unlike broadcast and VOD content, where second content options are typically preset and/or limited by the service provider.

- 5 It may be appreciated by the person skilled in the art that variations and combinations of features described above, not being alternatives or substitutes, may be combined to form yet further embodiments falling within the intended scope of the invention.

CLAIMS

1. A device for facilitating a provision of a content comprising:

a processor operable to detect a temporal location of a first content; and

an output unit operable to output a second content associated with the first

5 content, the second content synchronisable with the first content,

wherein the processor is operable to extract a first identifier associated with the first temporal location, match the first identifier against one or more identifiers of the second content, and use the temporal location to associate with a second identifier matched with the first identifier to synchronise the output of the second
10 content with the first content.

2. The device according to claim 1, wherein the processor is operable to control the output unit to output the second content from the temporal location associated with the second identifier.

15

3. The device according to claim 2, wherein the processor is operable to extract a fragment of the first content which is associated with the temporal location of the first content outputted.

20 4. The device according to claim 3, wherein the processor is operable to digitise the fragment of the first content.

5. The device according to claim 4 further comprising a database storing the digitised fragment of the first content.

25

6. The device according to claim 5, wherein the processor is operable to extract the first identifier associated with the temporal location of the first content outputted, upon the digitised fragment of the first content is stored in the database.

7. The device according to claim 6, wherein the processor is operable to identify the first content from the extracted first identifier.

5 8. The device according to claim 7, wherein the processor is operable to search the second content associated with the first content from a reference database.

9. The device according to claim 8, wherein the processor is operable to
10 receive the second content from the reference database.

10. The device according to claim 1, wherein the processor is operable to receive an audio input of the first content to detect the temporal location of the first content.

15

11. The device according to claim 1 further comprising a display unit operable to display the second content associated with the first content.

12. The device according to claim 1 further comprising an audio output unit
20 operable to output the second content associated with the first content.

13. The device according to claim 1, wherein the identifier is a fingerprint of the content.

25 14. The device according to claim 1, wherein the temporal location is a time-related data.

15. The device according to claim 14, wherein the time-related data is a time stamp.

16. The device according to claim 15, wherein the time stamp has a data
5 format comprising HH:MM:SS.FF wherein HH is the hour; mm is the minutes; ss is the seconds, and FF is the video/film frame number associated with at least one particular event.

17. A method for facilitating a provision of a content comprising:

10 detecting a temporal location of a first content outputted;

extracting a first identifier associated with the temporal location of the first content outputted;

matching the first identifier against one or more identifiers of a second content associated with the first content;

15 using a temporal location associated with a second identifier matched with the first identifier to synchronise an output of the second content with the first content outputted; and

outputting the second content associated with the first content, to be synchronised with the first content outputted.

20