

A SYSTEM AND METHOD FOR REAL-TIME TRANSLATION IN COMMUNICATIONS

[001] The invention relates to the fields of wireless communication and artificial intelligence. More specifically, a system and method for real-time translation in communications, in which the translations are controllable and the audio and/or video stream are mixed and synchronized.

BACKGROUND OF THE INVENTION

[002] Currently, users accessing real-time translation services need to interface with an application software, having software modules that perform translation, which is installed on hardware devices. Moreover, users can typically only view one type of audio-visual media or translated text within one media stream of the user interface (UI) of the application software. This limitation prevents the simultaneous playback of audio-visual media streams on the UI of one single device. This makes it challenging to implement communication scenarios where real-time communication between one or more users speaking in different native languages is required. Thus, this impacts the ability of users from having a comprehensive and rich communication experience.

[003] Furthermore, in view of communication frameworks that deliver multimedia services, such those based on the Internet Protocol Multimedia Core Network Subsystem (IMS) architecture, the existing communication frameworks, which include Voice over Long-Term Evolution (VoLTE) and/or Video over Long-Term Evolution (ViLTE), are only used for end-to-end transmission of an audio stream for communication between users without being further leveraged to provide real-time translated communication between users.

[004] There are disclosed technologies over the prior art relating to systems and methods for real-time translated communication. Among them include United States Patent Application US20110246172A1, which discloses a real-time audio translation system for videoconferencing. It includes a controller that examines multiple audio streams and selects specific ones for translation. The system has multiple translator resources to handle the translation of the selected audio streams. A translator resource selector works with the

controller to direct the chosen audio streams to the appropriate translator resources. The system can translate speech into text for subtitles for allowing different participants to receive translations based on their preferred languages.

[005] However, the teachings derived from the aforementioned prior art are insufficient in providing real-time translated audio and/or visual communication between users, especially over cellular networks. Furthermore, the aforementioned prior art also fails to inform means to provide translations that are controllable for a synchronized audio and/or video communication in real-time. Accordingly, it is desirable for a system and method for real-time translation in communications that offers such features, which may further leverage on artificial intelligence (AI) technology for translations to be continuously optimized to meet the communication needs of people from different countries or regions.

SUMMARY OF INVENTION

[006] The main objective of the present invention is to provide a system and method for real-time translation in communications, in which the translations are controllable and the audio and/or video stream are mixed and synchronized, in real-time, for mutual communication between users or parties.

[007] In particular, the present invention aims to leverage on VoLTE or ViLTE audio/video calling capabilities to [i] introduce media streams to realize real-time translation of languages and/or dialects, [ii] provide streaming of text and video between two or more devices, and [iii] provide a simultaneous output of audio streams. The usage of VoLTE or ViLTE audio/video calling capabilities advantageously allows the present invention to accommodate high-demand scenarios where a large bandwidth is required for communication between terminal(s) or device(s). This shall enable barrier-free communication between users of different languages, and moreover, enhance auditory and visual interaction and perception of both parties in a call scenario, thereby providing parties a high-quality communication experience.

[008] Furthermore, the present invention provides synchronization and integration of translated audio-video media streams and translated audio streams through networks having

communication protocols with VoLTE or ViLTE channels. More specifically, media streams are to be displayed on the same screen according to business subscription requirements as set by the user(s) on terminal(s) or their device(s). The audio calls are configured to for them be received by the terminal(s) or device(s) at the same rate, thereby allowing users to view various types of translated audio-video content on terminal(s) or their device(s). The translated audio-video content may a combination of speech, text and, image animations, which are combined in a single media stream. Moreover, the users may also be able to simultaneously hear corresponding translated audio content. This improves communication barriers caused by language differences and adds elements of social interaction, further enhancing the quality of the communication experience in call scenarios.

[009] The present application intends to provide a system for real-time translation in communications, comprising a plurality of terminals, at least one network, which is configured to allow media content to be exchanged or streamed between the terminals in real-time or near real-time via a service route at a substantially synchronized or near-synchronized manner, wherein the media content includes original media content and translated media content, and at least one device that is either part of, or in communication with, the network, having at least one processor that operates a plurality of modules, with the device configured to receive the original media content from at least one transmitting terminal for generating the translated media content therefrom that is sent to at least one corresponding receiving terminal via the service route. The original media content includes at least one original audio stream, at least one original video stream, or a combination thereof, and the translated media content includes the original audio stream, the original video stream, at least one translated audio stream, at least one data file having text transcribed and/or translated from the original audio stream that is embeddable within the original video stream as one or more subtitle tracks, or a combination thereof, which are composited and synchronized by the device based on specifications of the service route of the network.

[0010] Preferably, the network has a communication framework that is based on the Internet Protocol Multimedia Core Network Subsystem (IMS) architecture, wherein the framework is either a Voice over Long-Term Evolution (VoLTE) channel, a Video over Long-Term Evolution (ViLTE) channel, or a combination thereof.

[0011] Preferably, the modules include a business logic module that is configured to either or both authenticate user identity information of at least one corresponding user of one corresponding terminal, and determine service subscription information of at least one corresponding user of one corresponding terminal that relates to any one or a combination of input language settings of the terminals, output language settings of the terminals, and an output format of the translated media content that are being exchanged or streamed in real-time or near real-time at the terminals.

[0012] Preferably, regarding the system, the output format of the translated media content of one receiving terminal include any one of a first type of translated media content, which is composited and synchronized media stream from the transmitting terminal having the original audio stream, and a subtitled video stream having the original video stream and any one or a combination of one subtitle track transcribed from the original audio stream and translated subtitle tracks transcribed and translated from the original audio stream, or a second type of translated media content, which is composited and synchronized media stream from the transmitting terminal having the translated audio stream, and the subtitled video stream having the original video stream and any one or a combination of one subtitle track transcribed from the original audio stream and the translated subtitle tracks transcribed and translated from the original audio stream. The translated subtitle track, the translated audio stream, or a combination thereof, are of languages according to the output language settings of the receiving terminal.

[0013] Preferably, regarding the system, the modules include an automated speech recognition module that is configured to recognize an original language from the original audio stream, transcribe a first set of text in the original language from the original audio stream, and generate at least one audio transcription data file having the transcribed first set of text, which is embeddable within the original video stream.

[0014] Preferably, regarding the system, the modules include a comprehension and translation module that is configured to receive audio transcription file having the transcribed first set of text, translate the transcribed first set of text into a second set of text that is of at least one translated language, and generate at least one translated transcription data file having the second set of text, which is embeddable within the original video stream.

[0015] Preferably, regarding the system, the modules include a text-to-speech module that is configured to generate the translated audio stream based on the second set of text of the translated transcription data file.

[0016] Preferably, regarding the system, the text-to-speech module is further configured identify and/or sample timbre(s) within the original audio stream to generate the translated audio stream that mimics the voice of one user at the transmitting terminal in which the original audio stream is transmitted therefrom.

[0017] Preferably, regarding the system, the modules include a service control module that is configured to determine receipt of at least one media translation service request message from at least one terminal within a pre-set time period, for the service control module, or in conjunction with the other modules, to either initiate or continue exchange or streaming of translated media content between the terminals in real-time or near real-time, stop the exchange or streaming of translated media content between the terminals in real-time or near real-time, and optionally, delete the translated media content that is stored within the device.

[0018] Preferably, regarding the system, the service control module is further configured to determine receipt of at least one audio segmentation service request message from at least one terminal, within a pre-set period, for the service control module, or in conjunction with the other modules, to either correct or update the translated media content that are being exchanged or streamed between the terminals in real-time or near real-time within the pre-set time period, or delete the translated media content that is stored within the device.

[0019] Preferably, regarding the system, the modules include a signalling module and a media processing module which are configured to enable the original media content and the translated media content to be exchanged between corresponding terminals via the service route over the network in a synchronized manner in real-time or near real-time.

[0020] Preferably, the system further comprises a cache unit configured to store any one or a combination of the original media stream, the translated media stream, and their associated media stream resources, for a pre-determined time period, as they are being exchanged or streamed between the terminals in real-time or near real-time via the service route.

[0021] The present application further intends to provide a method for real-time translation in communications, comprising the steps of allowing media content to be exchanged or streamed between a plurality of terminals in real-time or near real-time, via a service route at a substantially synchronized or near-synchronized manner, by at least one network, wherein the media content includes original media content and translated media content, and receiving the original media content from at least one transmitting terminal, by at least one device that is either part of, or in communication with, the network, wherein the device has at least one processor that operates a plurality of modules, generating the translated media content from the original media content, by the device, and sending the translated media content to at least one corresponding receiving terminal via the service route, by the device. The original media content includes at least one original audio stream, at least one original video stream, or a combination thereof. The translated media content includes the original audio stream, the original video stream, at least one translated audio stream, at least one data file having text transcribed and/or translated from the original audio stream that is embeddable within the original video stream as one or more subtitle tracks, or a combination thereof, which are composited and synchronized by the device based on specifications of the service route of the network.

[0022] Preferably, regarding the method, the network has a communication framework that is based on the Internet Protocol Multimedia Core Network Subsystem (IMS) architecture, wherein the framework is either a Voice over Long-Term Evolution (VoLTE) channel, a Video over Long-Term Evolution (ViLTE) channel, or a combination thereof.

[0023] Preferably, the method further comprises the steps of authenticating user identity information of at least one corresponding user of one corresponding terminal, by a business logic module, determining service subscription information of at least one corresponding user of one corresponding terminal that relates to any one or a combination of input language settings of the terminals, output language settings of the terminals, and an output format of the translated media content that are being exchanged or streamed in real-time or near real-time at the terminals, by the business logic module, or a combination thereof.

[0024] Preferably, the method further comprises steps carried out by an automated speech recognition module, which include recognising an original language from the original audio stream, transcribing a first set of text in the original language from the original audio stream, and generating at least one audio transcription data file having the transcribed first set of text, which is embeddable within the original video stream.

[0025] Preferably, the method further comprises steps carried out by a comprehension and translation module, which include translating the transcribed first set of text into a second set of text that is of at least one translated language, and generating at least one translated transcription data file having the second set of text, which is embeddable within the original video stream.

[0026] Preferably, the method further comprises the step of generate the translated audio stream based on the second set of text of the translated transcription data file by a text-to-speech module.

[0027] Preferably, the method further comprises steps carried out by a service control module, which includes determining receipt of at least one media translation service request message from at least one terminal within a pre-set time period, for initiating or continuing exchange or streaming of translated media content between the terminals in real-time or near real-time, determining absence of at least one media translation service request message from at least one terminal within a pre-set time period, for stopping the exchange or streaming of translated media content between the terminals in real-time or near real-time, determining receipt of at least one audio segmentation service request message from at least one terminal, for correcting or updating the translated media content that are being exchanged or streamed between the terminals in real-time or near real-time, within the pre-set time period, and determining absence of at least one audio segmentation service request message from at least one terminal, for deleting the translated media content that is stored within the device.

[0028] The present application also intends to provide a computer-readable medium for real-time translation in communications, being part of a device that is either part of, or in communication with, a network that is configured to allow media content to be exchanged

or streamed between a plurality of terminals in real-time or near real-time, via a service route, wherein the media content includes original media content and translated media content, wherein the computer-readable medium stores modules including their instructions that, when executed by a processor of the device, causes the processor to enable the device to receive the original media content from at least one transmitting terminal, generate translated media content from the original media content, send the translated media content to corresponding receiving terminals in real-time or near real-time, via the service route. The original media content includes at least one original audio stream, at least one original video stream, or a combination thereof. The translated media content includes the original audio stream, the original video stream, at least one translated audio stream, at least one data file having text transcribed and/or translated from the original audio stream that is embeddable within the original video stream as one or more subtitle tracks, or a combination thereof, which are composited and synchronized by the device based on specifications of the service route of the network.

[0029] One skilled in the art will readily appreciate that the invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The embodiments described herein are not intended as limitations on the scope of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0030] To facilitate an understanding of the invention, there is illustrated in the accompanying drawings the preferred embodiments from an inspection of which when considered in connection with the following description, the invention, its construction and operation and many of its advantages would be readily understood and appreciated.

[0031] **FIG. 1** illustrates a block diagram of a system for real-time translated communication, according to one example embodiment of the present invention.

[0032] **FIG. 2** illustrates a sequence diagram that pertains to a method for real-time translated communication, according to one example embodiment of the present invention.

[0033] **FIG. 3** illustrates a flowchart that pertains to the method for real-time translated communication, which may be regarded as a further extension of **FIG. 2**, according to one example embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0034] The present invention relates to a system and method for real-time translation in communications, in which the translations are controllable and the audio and/or video stream are mixed and synchronized, in real-time or near real-time. The invention may also be presented in a number of different embodiments with common elements.

[0035] According to the concept of the invention, there is provided at least one device that facilitates communication between two or more users via their corresponding terminals. More specifically, the aforementioned device includes at least one computer-readable medium and at least one processor. The computer-readable medium has one or more modules and their related instructions, and the processor operates these modules based on their instructions for enabling real-time translated communication between the users.

[0036] The invention will now be described in greater detail, by way of example, with reference to the figures. For ease of reference, common reference numerals or series of numerals will be used throughout the figures when referring to the same or similar features common to the figures.

[0037] **FIG. 1** illustrates a block diagram of a system for real-time translated communication, according to one example embodiment of the present invention. As shown, the system may comprise at least one device **10** that is located within or part of a network **20**, and one or more terminals **31, 32** that are each operated by a corresponding user **A, B**.

[0038] With reference to **FIG. 1**, the device **10** may be configured to facilitate communication between the users **A, B** via their terminals **31, 32**. Furthermore, the device **10** may be configured to perform and provide translations of media content that is being exchanged or streamed between terminals **31, 32** in real-time or near real-time. The device

10, may be, by way of example, a computing device, a server, a network device, or the like. In alternative embodiments, there may be a plurality of devices **10** that are involved.

[0039] With reference to **FIG. 1**, the network **20** relates to a communication network, and it is preferable that the device **10** is part of the network. More specifically, the network **20** may have a plurality of devices besides device **10**, and device **10** is to substantively interact with other devices within network **20** for exchange or streaming of data. In alternative embodiments, the network **20** may not include device **10**, i.e. the device **10** is a device that is separate from the network **20** but in communication therewith.

[0040] With reference to **FIG. 1**, the network **20** is configured to be based on framework(s) that is on the Internet Protocol Multimedia Core Network Subsystem (IMS) architecture, more specifically, the framework(s) may relate to Voice over Long-Term Evolution (VoLTE), Video over Long-Term Evolution (ViLTE), a combination of them, their derivatives, and/or their technological equivalent counterparts. Besides that, the network **20** may be in the form of a local area network (LAN), a wide area network (WAN) such as the internet, or any other form of network based on standards provided by the 3rd Generation Partnership Project 1 / 2 (3GPP 1 / 2), the Global System for Mobile Communications (GSM), or the like. Furthermore, the IMS architecture, which may be based on the Internet, may support telephony-like and/or videotelephony-like applications. Through such applications, online translation video mixing and synchronous audio streaming communication can be achieved.

[0041] With reference to **FIG. 1**, the terminals **31, 32** may include a first terminal **31** that is operated by a first user **A**, and a second terminal **32** that is operated by a second user **B**. In particular, each terminal **31, 32** is configured to be at least able to [i] receive audio and/or graphical inputs from their corresponding user **A, B**, and [ii] provide audio and/or graphical outputs to their corresponding user **A, B**. Furthermore, the terminals **31, 32** are also configured to be capable of being in communication with the network **20**. With that, exchange or streaming of data, in the form of media content, may occur between the terminals **31, 32**, in real-time or near real-time, with the device **10** preferably being within the communication pathway between the terminals **31, 32**. It is to be noted that the number of terminals may not be limited to as described or depicted, and in certain embodiments,

there may be a plurality of terminals that communicate with each other in real-time or near real-time, with the device **10**, or a plurality thereof, providing translated media content to the terminals.

[0042] With reference to **FIG. 1**, the terminals **31**, **32** each may be conventional commercially available end-user devices, which may include smartphones, personal device assistants, or any devices capable of interacting with a network. Furthermore, it is preferred that these devices have means, or are capable of being coupled with means, to capture or record audio and/or visual data/or information, such as image sensors, microphones, etc.

[0043] With reference to **FIG. 1**, the device **10** may comprise at least one processor **11**, at least one data storage medium **12**, and at least one cache unit **13**. In particular, the processor **11** may be substantially interfaced with the data storage medium **12** and the cache unit **13**. Furthermore, the processor **11** may run at least one application software **11a** for operating one or more modules that carry out functions that relate to real-time or near real-time translation of media content that is being exchanged or streamed between the terminals **31**, **32**.

[0044] With reference to **FIG. 1**, the processor **11** is configured to perform the scheduling and the execution of software instructions of the software modules of that may be part of the application software **11a** running thereon. The processor **11** may be, but shall not be limited to, a conventional processor, application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA), a graphics processing unit (GPU), or a combination thereof.

[0045] With reference to **FIG. 1**, the data storage medium **12** may store data therewithin in a permanent or a semi-permanent manner. More specifically, the data storage medium **12** may have non-volatile memory, for it to substantially act as a non-transitory computer-readable medium. The data storage medium **12** may store computer-executable instructions related to the modules, which may be in the form of computer-readable codes, for them to be read, executed and operated by the processor **11**. With that, the data storage medium **12** may be one or a combination of digital storage mediums such as a flash memory unit, a read-only memory (ROM) unit, hard-disk drives, solid-state drives, hybrid drives, or the like.

[0046] With reference to **FIG. 1**, the cache unit **13** may store data therewithin in a temporary manner. More specifically, the cache unit **12** may have volatile memory, i.e. or e.g., be a transitory or transient computer-readable medium, and it may temporarily store data that were from the runtime of the modules operated by the processor **11**. In certain embodiments, the cache unit **13** is configured to store data or usage data of any one or a combination of the translated and composited media content, the translated audio stream, data files related to transcribed texts or translated texts, an original video stream, or an original audio stream. With that, the cache unit **13** may be one or a combination of digital storage mediums such as a random-access memory (RAM) devices and their synchronous / asynchronous variants, cache memory that is part of the processor **11**, or the like.

[0047] It is to be noted that, in certain embodiments, the system may instead include or further include a unit having substantially volatile memory or is one that may undergo data loss, i.e. or e.g., a transitory or transient computer-readable medium, which may store computer-executable instructions related to the modules that may be in the form of computer-readable codes, for them to be read, executed, and operated by the processor **11**. By way of example, embodiments of this aforementioned unit may be similar to those of the cache unit **13**. In certain embodiments, the cache unit **13** may be the entity that stores the aforementioned computer-executable instructions.

[0048] With reference to **FIG. 1**, the processor **11** is configured to operate a plurality of modules based on their corresponding instructions provided by the data storage medium **12**. The modules may include a service control module **1101**, an audio and video (AV) transcoding module **1102**, an audio and video (AV) fusion module **1103**, a signalling module **1104**, a media processing module **1105**, a business logic module **1106**, an audio and signal (AS) module **1107**, an automated speech recognition (ASR) module **1108**, a comprehension and translation module **1109**, a text-to-speech (TTS) module **1110**, an automatic call distribution (ACD) module **1111**, and a session border controller (SBC) module **1112**.

[0049] It is to be noted that the aforementioned modules may substantively be in communication with each other through signals and/or interfaces known in the art. It is noted that while the aforementioned modules may be in a software embodiment, they may also be a hardware embodiment where they are directly connected to the processor **11**, or they are

their own independent computer system. It should also be noted that some of the aforementioned modules may not necessarily be part operated by the processor **11**, and they may be operated by other devices or processors within the network **20** in which the device **10** is in communication therewith. While not shown, one or more ancillary software modules may be operated by the processor **11** to support the operations of the device **10**.

[0050] Regarding the service control module **1101**, it is configured to process messages received from the terminals **31**, **32** and provides control instructions to the other modules of the device **1101**. In particular, the service control module **1101** may be involved in pulling (i.e. receiving), pushing (i.e. sending), and/or analysing, media content and/or user information that are being streamed and/or exchanged between the terminals **31**, **32**. More specifically, the service control module **1101** may facilitate the conversion of a received media stream from either terminal **31**, **32** into a translated media stream based on the user information in real-time or near real-time.

[0051] Regarding the audio and video transcoding module **1102**, it is a module that may be part of the service control module **1101**, or a separate module in communication therewith as shown in **FIG. 1**. In particular, the audio and video transcoding module **1102** is configured to transcode media content that is being exchanged or streamed between the terminals **31**, **32**. More specifically, the audio and video transcoding module **1102** may carry out transcoding upon audio data in the media content and/or video data within the video content. Moreover, the media content that is transcoded may be the translated media content that is to be pushed to either one or both terminals **31**, **32** from the device **10**. Besides transcoding, the audio and video transcoding module **1102** may be configured to carry out any one or a combination of audio data compression, video data compression, and/or bandwidth requirement reduction, for improving transmission efficiency of media content that is being streamed or exchanged between the terminals **31**, **32** in real-time or near real-time.

[0052] In particular, the audio and video transcoding module **1102** may optionally providing encoding of the input format and/or output format of the media content. More specifically, it may support video encoding of one or more input/output formats, which include, but shall not be limited to, H.264, VP8, or the like. Furthermore, it may support audio encoding of

one or more input/output formats, which include, but shall not be limited to, G.711, G.722, G.729, or the like.

[0053] In particular, the audio and video transcoding module **1102** may be configured to control the complexity of the media content in an adaptive manner, upon receiving relevant instructions from the service control module **1101**. More specifically, for a translated media content in the form of an audio and/or video stream that are to be transmitted on channels such as the VoLTE or ViLTE channel, its encoding bit rate for its video and/or audio may be adaptively determined according to the texture complexity, motion complexity, audio presentation speed and pre-coding results of the original audio stream and/or the original video stream of the original media content.

[0054] In particular, the audio and video transcoding module **1102** may be instructed by the service control module **1101** to perform the aforementioned adaptive control by changing factors such as the network environment and terminal adaptation conditions, so that bit rate control can be achieved based on the communication scenario.

[0055] In particular, the audio and video transcoding module **1102** may be configured to perform intelligent machine learning-based adaptive control upon audio and video coding through trained machine learning module(s) operated by the device **10**, upon receiving relevant instructions from the service control module **1101**. In particular, pre-processing and mixing of the media content is performed to improve the synchronization quality of the audio and video. The pre-processing may include adaptive picture restoration processing and/or sharpening of picture, audio, text and digital image of the media streams, noise reduction, and decompression distortion, etc. With that, a multi-resolution output may be provided, wherein the output may be of the same resolution, an up-sampled resolution, and/or a down-sampled resolution. The supported resolutions may include 480p, 720p, 1080p, 4K, 8K, and/or the like, while the supported frame rates may include 12fps, 24fps, 25fps, 30fps, 50fps, 60fps, and/or the like.

[0056] Regarding the audio and video fusion module **1103**, it is configured to perform an integration of any one or a combination of audio, video, and text for a relevant media content. In particular, the audio and video fusion module **1103** may receive any one or a combination

of translated texts, an original video stream of at least one terminal **31**, **32**, and a translated audio stream, as input, and subsequently, it may provide a composited media content that combines or mixes the aforementioned input(s) into its output. More specifically, the integrated media content may correspond to the translated media content that is to be received by either terminal **31**, **32** in real-time or near real-time.

[0057] Regarding the signalling module **1104**, it is configured to receive media content that is being streamed or exchanged between the terminals **31**, **32** from a corresponding service route for communication between the terminals **31**, **32** to be established. Moreover, it is also configured to push or transmit the original media content and/or the translated media content to a corresponding service route via relevant communication protocols for them to be streamed or exchanged between the terminals **31**, **32** in real-time in a synchronized or in near real-time in a near-synchronized manner. In particular, the signalling module **1104** may provide the media content to the service control module **1101**, and it may receive instructions from the service control module **1101** to push or transmit the media content to corresponding terminals **31**, **32**. In certain embodiments, the signalling module **1104** is involved in established a connection for between the device **10** and the network **20**, and the network **20** may have audio/video IMS communication architecture. In certain embodiments, the signalling module **1104** may further be configured to convert the media content streamed or exchanged between the terminals **31**, **32** into appropriate transmission signals that may correspond to audio/video specifications for the network **20** (e.g. network with IMS communication architecture). In certain embodiments, the signalling module **1104** may be part of an application server that is separate from the device **10** in which the device **10** is connected thereto.

[0058] Regarding the media processing module **1105**, it is configured to operate in tandem with the signalling module **1104**. More specifically, it is configured to reserve, or facilitate reservation of, media content resources for the terminals **31**, **32** prior to and/or after media content is streamed or exchanged between them. In particular, the media processing module **1105** may also be configured to receive and/or identify a user identity and their corresponding service subscription information. Subsequently, it may store an authenticated user identity and their corresponding service subscription in the cache unit **13**. Furthermore, the media processing module **1105** may be configured to facilitate the transmission of the

translated media content between the terminals **31**, **32** in a synchronous manner through corresponding reserved media resources. The media processing module **1105** may collect the audio streams of the corresponding terminals **31**, **32** through the media transmission protocol. In certain embodiments, the media processing module **1105** may be part of an application server that is separate from the device **10** in which the device **10** is connected thereto.

[0059] In certain embodiments, the functions and/or the operations signalling module **1104** and the media processing module **1105** may be combined. By way of example, the signalling module **1104** may be configured to also carry out certain or all function and/or operations of the media processing module **1105**. By way of example, the media processing module **1105** may be configured to also carry out certain or all function and/or operations of the signalling module **1104**.

[0060] Regarding the business logic module **1106**, it is configured to handle user information and service subscription information received from the terminals **31**, **32**. More specifically, it may be configured to authenticate a user identity of the users **A**, **B** of terminals **31**, **32**, and their corresponding service subscription information. Subsequently, it may store the authenticated identity and their corresponding service subscription the cache unit **13**, or it may instruct other relevant modules to do so. Moreover, the business logic module **1106** may be configured to determine language settings and video/audio output format settings as set by the users **A**, **B** via the terminals **31**, **32**. Furthermore, the business logic module **1106** may also be configured to determine an allocation of media stream resources within the cache unit **31**.

[0061] In particular, the service subscription information handled by the business logic module **1106** may relate to any one or a combination of [i] input language settings from the terminals **31**, **32**, [ii] output languages settings, in the form of audio and/or text, for the terminals **31**, **32**, and [iii] the types of output format for the media content that is synchronously streamed or exchanged between the terminals **31**, **32** in real-time or near real-time. Furthermore, the user information handled by the handled by the business logic module **1106** may relate to any one or a combination of [i] an active service subscription, and [ii] an inactive service subscription.

[0062] Regarding the AS module **1107**, it is configured to perform operations upon the media content that is exchanged between the terminals **31, 32**. By way of example, it may be configured to separate the video stream and/or the audio stream within the media content, and it may subsequently pass the audio stream to other relevant modules for further processing. The AS module **1107** may also be part of the business logic module **1106**.

[0063] Regarding the ASR module **1108**, it is configured to carry out audio identification and/or audio analysis upon the media content that is synchronously streamed or exchanged between in terminals **31, 32** in real-time or near real-time. In particular, the ASR module **1108** may receive information about the input languages of the terminals **31, 32** from the business logic module **1106**, and it may be further configured to recognize text and generate one or more sets of text that is transcribed from one input language received by one terminal. The ASR module **1108** may be further configured to store the sets or transcribed text in the cache unit **13**. The ASR module **1108** may also be part of the business logic module **1106**.

[0064] With reference to the example system of **FIG. 1**, during communication between the terminals **31, 32**, the ASR module **1108** may, in real-time or near real-time, recognize and generate a first set of text transcribed from a first input language received from the first terminal **31**, which corresponds to an audio stream of a speech of a first language spoken by the first user **A** to the first terminal **31**. Moreover, during communication between the terminals **31, 32**, during communication between the terminals **31, 32**, the ASR module **1108** may, in real-time or near real-time, recognize and generate a second set of text transcribed from a second input language received from the second terminal **32**, which corresponds to a real-time audio stream of a speech of a second language spoken by the second user **B** to the second terminal **32**.

[0065] Regarding comprehension and translation module **1109**, it is configured to translate the sets of transcribed text into language(s) that corresponds to configured output settings of each terminal **31, 32**. In particular, the comprehension and translation module **1109** may translate one set of transcribed text, corresponding to one terminal, into one set of translated text for another different terminal. It is to be noted that the communication language between the terminals **31, 32** is uncertain, and translations may be performed by the comprehension

and translation module **1109** is done according to the service subscription information set by each terminal **31, 32**. It is to be noted that the translations performed by this module, or by extension, the device **10**, is not implemented on the terminals **31, 32**, but is implemented during the communication process as both terminals **31, 32** communicate via the network **20**.

[0066] By way of example, with reference to the example system of **FIG. 1**, during communication between the terminals **31, 32**, the comprehension and translation module **1109** may, in real-time or near real-time, retrieve the first set of transcribed text that is in the first language as spoken by the first user **A** from the cache unit **13**. With that, the comprehension and translation module **1109** may, in real-time or near real-time, translate the first set of transcribed text into the second language, which corresponds to a first set of translated text.

[0067] By way of example, with reference to the example system of **FIG. 1**, during communication between the terminals **31, 32**, the comprehension and translation module **1109** may, in real-time or near real-time, retrieve the second set of transcribed text that is in the second language as spoken by the second user **B** from the cache unit **13**. With that, the comprehension and translation module **1109** may, in real-time or near real-time, translate the second set of transcribed text into the first language, which corresponds to a second set of translated text.

[0068] Regarding the TTS module **1110**, it is configured to generate at least one translated audio stream that corresponds to one set of translated text. The TTS module **1110** is further configured to be in communication with the audio and video fusion module **1103**, and under instructions from the service control module **1101**, for the translated audio stream to replace an original audio stream of a video stream for the video stream to become an edited video stream, which is exchanged or streamed between the terminals **31, 32** in real-time or near real-time.

[0069] Regarding the TTS module **1110**, it may be configured to have one or more voice settings for the translated audio stream. The TTS module **1110** may refer to the service subscription information of the user, and will generate the translated audio stream in

accordance thereto. In particular, the translated audio stream generated by the TTS module **1110** may have characteristic of the users' voice.

[0070] By way of example, with reference to the example system of **FIG. 1**, during communication between the terminals **31**, **32**, the TTS module **1110** may, in real-time or near real-time, retrieve the first set of translated text. With that, the TTS module **1110** may, in real-time or or near real-time, generate a first translated audio stream, which corresponds to a speech spoken in the second language. Based on the settings of the business logic module **1106**, media content from the first terminal **31**, which may be in the form of an audio and/or video stream, may have its original audio stream replaced with the audio translated into the second language so that the second user **B** at the second terminal **32** may perceive the media stream in a language that is understood by them.

[0071] By way of example, with reference to the example system of **FIG. 1**, during communication between the terminals **31**, **32**, the TTS module **1110** may, in real-time or near real-time, retrieve the second set of translated text. With that, the TTS module **1110** may, in real-time or near real-time, generate a second translated audio stream, which corresponds to a speech spoken in the second language. Based on the settings of the business logic module **1106**, media content from the second terminal **32**, which may be in the form of an audio and/or video stream, may have its original audio stream replaced with the translated audio of the first language so that the first user **A** at the first terminal **32** may perceive the media stream in a language that is understood by them.

[0072] In particular, the device **10** may also be configured to further perform automatic real-time or or near real-time correction of text interpretation of audio stream synchronization calls. More specifically, the service control module **1101**, in conjunction with the ASR module **1108**, comprehension and translation module **1109**, and TTS module **1110**, may selectively correct text within the text translation, in an intelligent manner, based on complete sentence segments, for the corrected text to be displayed in real-time or near real-time, according to pre-set audio rules. For this, there may one or more designated trained machine learning model(s) operating on the device **10** for performing text correction through artificial intelligence.

[0073] Regarding the ACD media module **1111** and SBC module **1112**, they are configured to perform call management between the terminals **31**, **32** to ensure that media content is exchanged or streamed between the intended terminals during the communication session. The operations of these modules are as known in the art and hence not described.

[0074] **FIG. 2** illustrates a sequence diagram that pertains to a method for real-time translation in communications, according to one example embodiment of the present invention. **FIG. 3** illustrates a flowchart that pertains to the method for real-time translation in communications, which may be regarded as a further extension of **FIG. 2**, according to one example embodiment of the present invention. It is noted that the steps described in the sequence diagram, the flowchart, and within this detailed description, are not to be interpreted as non-limiting, and minor modifications to the steps (e.g. repetitions, additions, omissions, or swaps) are permissible by a skilled person without substantial deviation from as described.

[0075] The operation of the present invention shall now be described with reference to **FIGS 1 – 3**. By way of example, the first user **A** may interact with the first terminal **31** to initiate a call, and hence, the first terminal **31** may also be referred to as a “calling terminal”. The call may be transmitted through the network **20** and device **10** to reach the second terminal **32**. With that, the second user **B** may interact with the second terminal **32** to receive the call, and hence, the first terminal **31** may also be referred to as a “called terminal”. As the call takes place, real-time media content that is being exchanged or streamed between the first terminal **31** and second terminal may be translated by the device **10**, until the call is terminated by either terminal **31**, **32**. Whilst the steps may be described according to a unidirectional communication setting (e.g. a simplex communication), it is to be noted that the steps described are also applicable in a multidirectional communication setting (e.g. a full duplex communication), and the steps that may be performed or initiated by the first terminal **31** may be similarly performed or initiated by the second terminal **32**.

[0076] The first set of steps related to establishment of a call between the terminals **31**, **32** through a service route shall now be described.

[0077] For call establishment, there may be a first precursor step that involves adjusting the settings of the first terminal **31** by configuring any one or a combination of [i] input language settings of the first terminal **31**, [ii] output language settings of the first terminal **31**, and [iii] an output format for the media content that is to be outputted onto the first terminal **31**. This step may be performed by the first user **A**. In particular, the input language settings may relate to a first language may be spoken by the first user **A**, the output language settings may relate to the first language and/or other languages that are understood by the first user **A**, and the output format of the media content relates to the output audio and/or video stream displayed or outputted from the first terminal **31**. In alternative embodiments, the first terminal **31** may automatically identify the language spoken by the first user **A**, and as such, the configuration of input language settings of the first terminal **31** may not be required.

[0078] For call establishment, there may be a second precursor step that involves adjusting the settings of the second terminal **32** by configuring any one or a combination of [i] input language settings of the second terminal **32**, [ii] output language settings of the second terminal **32**, and [iii] an output format for the media content that is to be outputted onto the second terminal **32**. This step may be performed by the second user **B**. In particular, the input language settings may relate to a second language may be spoken by the second user **B**, the output language settings may relate to the second language and/or other languages that are understood by the second user **B**, and the output format of the media content relates to the output audio and/or video stream displayed or outputted from the second terminal **32**. In alternative embodiments, the second terminal **32** may automatically identify the language spoken by the second user **B**, and as such, the configuration of input language settings of the second terminal **32** may not be required.

[0079] The steps for the first set of steps may begin with a first step that involves sending a call entry request (i.e. an INVITE signal) to the network **20** (e.g. the IMS communication network). This step may be initiated by the first user **A** and performed by the first terminal **31**. In particular, call entry request is made in accordance to the VoLTE and/or ViLTE communication framework(s). Furthermore, the call entry request may include identifier information pertaining to the intended call recipient (i.e. the second terminal **32**).

[0080] Next, there is a second step that involves forwarding the call entry request (i.e. the INVITE signal) to the ACD media processing module **1111**. This step may be carried out by the SBC signalling module **1112**.

[0081] Next, there is a third step that involves receiving the call entry request (i.e. the INVITE signal) by the device **10**. More specifically, its service control module **1101** receives this request.

[0082] Next, there is a fourth step that involves establishing a service identifier of the first terminal **31**. This step may be carried out by the signalling module **1104** under instructions from the service control module **1101**. In particular, this service identifier may be established by the signalling module **1104** based on the network **20** and its related signalling instruction requirements.

[0083] Next, there is a fifth step that involves providing a user identity information and a service subscription information to the device **10**. This step may be carried out by the first terminal **31**.

[0084] Next, there is a sixth step that involves determining a user identity information and a service subscription information that corresponds to the first user **A** of the first terminal **31**. This step may be carried out by any one or both the signalling module **1104** and the business logic module **1104**.

[0085] Next, there is a seventh step that involves storing the user identity information and the service subscription information related to the first user **A** of the first terminal **31** in the cache unit **13**. This step may be carried out by the business logic module **1104**. The user identity information may further include authentication information of the first user **31**.

[0086] Next, there is an eighth step that involves generating a service logic message according to the user identity and the service subscription information of the first user **A** of the first terminal **31**. This step may be carried out by the business logic module **1104**. The service logic message may relate to an instruction for the device **10** to translate media content

transmitted from any terminal that is to be received by the first terminal **31** into the first language based on the service subscription information.

[0087] Next, there is a ninth step that involves reserving or allocating media content resources for the first terminal **31**. This step may be carried out by the media processing module **1105** under instructions from the service control module **1101**.

[0088] Next, there is a tenth step, the signalling module **1104** may be configured to carry out the step of establishing a service route together with the media processing module **1105**. In particular, this step may be carried out based on authenticated user identity information of the first terminal **31**.

[0089] Next, there is an eleventh step that involves performing protocol conversion on the call entry request (i.e. the INVITE signal). This step may be carried out by the signalling module **1104**.

[0090] Next, there is a twelfth step that involves transmitting the converted INVITE signal through the SBC module **1112** to the network **20** of the second terminal **32** for this signal to reach the second terminal **32** through the service route.

[0091] Next, there is a thirteenth step that involves receiving the INVITE signal, by the second terminal **32**.

[0092] Next, there is a fourteenth step that involves sending a call receipt message (i.e. an ACKNOWLEDGEMENT(ACK) signal) to the network **20** (e.g. the IMS communication network). This step may be initiated by the second user **B** and performed by the second terminal **32**, which may be done when the second user **B** decides to answer the call via their second terminal **32**. Similarly, the call receipt message may be made in accordance to the VoLTE and/or ViLTE communication framework(s).

[0093] Next, there is a fifteenth step that involves forwarding the call receipt message (i.e. the ACK signal) to the ACD media processing module **1111** of the network **20**. This step may be carried out by the SBC signalling module **1112** of the network **20**.

[0094] Next, there is a sixteenth step that involves receiving the call receipt message (i.e. the ACK signal) by the device **10**. More specifically, its service control module **1101** receives this request.

[0095] Next, there is a seventeenth step that involves establishing a service identifier of the second terminal **32**. This step may be carried out by the signalling module **1104** under instructions from the service control module **1101**. In particular, this service identifier may be established by the signalling module **1104** based on the network **20** and its related signalling instruction requirements.

[0096] Next, there is an eighteenth step that involves providing a user identity information and a service subscription information to the device **10**. This step may be carried out by the second terminal **32**.

[0097] Next, there is a nineteenth step that involves determining a user identity information and a service subscription information that corresponds to the second user **B** of the second terminal **32**. This step may be carried out by any one or both the signalling module **1104** and the business logic module **1104**.

[0098] Next, there is a twentieth step that involves storing the user identity information and the service subscription information related to the second user **B** of the second terminal **32** in the cache unit **13**. This step may be carried out by the business logic module **1104**. The user identity information may further include authentication information of the second user **32**.

[0099] Next, there is a twenty-first step that involves generating a service logic message according to the user identity and the service subscription information of the second user **B** of the second terminal **32**. This step may be carried out by the business logic module **1104**. The service logic message may relate to an instruction for the device **10** to translate media content transmitted from any terminal that is to be received by the second terminal **32** into the second language based on the service subscription information.

[00100] Next, there is a twenty-second step that involves reserving or allocating media content resources for the second terminal **32**. This step may be carried out by the media processing module **1105** under instructions from the service control module **1101**.

[00101] Next, there is a twenty-third step that involves performing protocol conversion on the call receipt message (i.e. the ACK signal). This step may be carried out by the signalling module **1104**.

[00102] Next, there is a twenty-fourth step that involves transmitting the converted call receipt message (i.e. the ACK signal) through the SBC module **1112** to the network **20** of the first terminal **31** for this signal to reach the first terminal **31**.

[00103] Next, there is a twenty-fifth step that involves receiving the call receipt message (i.e. the ACK signal), by the first terminal **31**.

[00104] Finally, there is a twenty-sixth step that involves establishing the call between the terminals **31**, **32**. In particular, the call is connected and media content may now be exchanged or streamed between the first terminal **31** and the second terminal **32** in real-time or near real-time through the call service route, with the device **10** being a communication intermediary between the terminals **31**, **32**. More specifically, the device **10** may be involved in pulling and/or pushing media content between the terminals **31**, **32**.

[00105] It is to be noted that in certain embodiments, the signalling protocol of the signal processing module **1104** may be the same as the signalling protocol of the network (i.e. the IMS communication network). In this case, the signal processing module **1104** does not need to perform protocol conversion on the INVITE signal and/or the ACK signal, and it only needs to forward the INVITE signal and/or the ACK signal to the network **20**.

[00106] The second set of steps related to media content being exchanged or streamed between the terminals **31**, **32** through the call route shall now be described.

[00107] The steps for the second steps of steps may begin with a first step that involves providing an input of media content to the first terminal **31**. This step may be carried out by

the first user **A**. In particular, the media content that is inputted into the first terminal **31** may be regarded as the original media content from the first terminal **31**, and it may be in the form of an original audio stream, an original video stream, or a combination thereof. Furthermore, the original media content from the first terminal **31** may have an audio stream in a first language, which is a language that is spoken by the first user **A** in the audio stream.

[00108] Next, there is a second step that involves transmitting the original media content from the first terminal **31** over the network **20**.

[00109] Next, there is a third step that involves receiving the original media content from the first terminal **31** by the media processing module **1105**. In particular, the media processing module **1105** may obtain the any one or both the original video stream and the original audio stream of the first terminal **31** through the corresponding media transmission protocol.

[00110] Next, there is a fourth step that involves receiving the original media content from the first terminal **31** by the service control module **1101**.

[00111] Next, there is a fifth step that involves analysing the original media content from the first terminal **31**. This step may be carried out by the service control module **1101**. In particular, the service control module **1101** may identify a correspondence between [i] the original media content from the first terminal **31** and [ii] the input language settings of the first terminal **31** as per the user information and service subscription information of the first terminal **31** that may have been stored in the cache unit **13**. More specifically, the input language settings of the first terminal **31** is determined to identify the first language that was spoken in the original media content of the first terminal **31**. In certain embodiments, the service control module **1101** may instruct relevant modules to separate the original media content into an original audio stream and/or an original video stream (if applicable).

[00112] Next, there is a sixth step that involves identifying the service subscription information of the second terminal **32**, which may have been stored in the cache unit **13**. This step may be carried out by the service control module **1101**. More specifically, the

service control module **1101** may determine the output language settings of the second terminal **32** and the output format settings of the second terminal.

[00113] Next, there is a seventh step that involves providing the original media content of the first terminal **31**, or more specifically, the original audio stream, to the ASR module **1108**. This step may be carried out by the service control module **1101**.

[00114] Next, there is an eighth step that involves performing real-time or near real-time audio identification analysis upon the original audio stream of the first terminal **31**. This step may be carried out by the ASR module **1108** under instructions of the service control module **1101**, in which the service control module **1101** may inform the ASR module **1108** to perform its analysis with reference to the first language. With that, the ASR module **1108** may identify any one or a combination of words, phrases, slangs, dialects, or the like, which are spoken in the first language by the user **A** at the first terminal **31**. In certain embodiments, the ASR module **1108** may automatically identify the first language within the original audio stream of the first terminal **31** without prior knowledge of the language spoken in the said audio stream.

[00115] Next, there is a ninth step that involves transcribing a first set of text that is in the first language based on the original audio stream of the first terminal **31**. This step may be carried out by the ASR module **1108**.

[00116] Next, there is a tenth step that involves generating at least one audio transcription data file having the transcribed first set of text, which may be embeddable within the original video stream of the first terminal **31** as a subtitle track. This step may be carried out by the ASR module **1108**. The generated audio transcription data file(s) may be continuously updated in real-time or near real-time for transcribed text of the first language to be added into the first set of text. Furthermore, the audio transcription data file(s) may include timestamps to indicate a temporal relationship between the transcribed text and the instance when the corresponding speech was spoken in the original audio stream of the first terminal **31** in the first language.

[00117] Next, there is an eleventh step that involves retrieving the audio transcription data file(s) by the comprehension and translation module **1109** under instructions of the service control module **1101**. In particular, the service control module **1101** may inform the comprehension and translation module **1109** of the output language settings of the second terminal **32** (i.e. the second language).

[00118] Next, there is a twelfth step that involves translating the audio transcription data file(s). This step may be performed by the comprehension and translation module **1109**. More specifically, the comprehension and translation module **1109** may translate the transcribed first set of text, which is in the first language, into at least one second set of text, which is of the second language. In particular, besides the second language, the comprehension and translation module **1109** may also be configured to translate the transcribed first set of text into other languages besides the first language.

[00119] Next, there is a thirteenth step that involves generating at least one translated transcription data file having the second set of text, which is embeddable within the original video stream of the first terminal **31** as a subtitle track. This step may be carried out by the comprehension and translation module **1109**. The translated transcription data file(s) may be continuously updated in real-time or near real-time for translated transcribed text of the second language to be added into the second set of text. Furthermore, the translated transcription data file(s) may include timestamps to indicate a temporal relationship between the translated transcribed text on the second language and the instance when the corresponding speech was spoken in the original audio stream of the first terminal **31** in the first language.

[00120] Next, there is a fourteenth step that involves retrieving the translated transcription data file(s) by the TTS module **1110** under instructions of the service control module **1101**.

[00121] Next, there is a fifteenth step that involves generating a translated audio stream based on the second set of text of the translated transcription data file(s). In particular, the translated audio stream may be of the second language or any other language understood by the second user **B** at the second terminal **32**.

[00122] Next, there is a sixteenth step that involves compiling any one or a combination of the original video stream of the first terminal **31**, the audio transcription data file(s) of the original audio stream of the first terminal in the first language, the translated transcription data file(s) of the original audio stream of the first terminal in the second language, and the translated audio stream, by the service control module **1101**.

[00123] Next, there is a seventeenth step that involves synthesising the translated media content based on the output format settings of the second terminal **32**. This step may be carried out by the service control module **1101** together with the audio-video fusion module **1102** by mixing and/or compositing any one or a combination of the streams and/or data files compiled in the previous step in a synchronized manner device based on specifications of the service route of the network **20**. In particular, the output format of the translated media content may be any one of a first type of translated media content or a second type of translated media content.

[00124] In particular, the first type of translated media content may be a composited and synchronized media stream having the original audio stream of the first terminal **31**, and a subtitled video stream having the original video stream of the first terminal **31** and any one or a combination of one or more subtitle tracks. In particular, the audio transcription data file(s) may be embedded within the original video stream of the first terminal **31** for it to have a subtitle track displaying text in the first language. Moreover, or alternatively, the translated transcription data file(s) may be embedded within the original video stream of the first terminal **31** for it to have a subtitle track displaying text in the second language or any other language understood by the second user **B** at the second terminal **32**. The subtitle track(s) allow subtitles to be displayed alongside with the graphics of the original video stream based on their timestamps.

[00125] In particular, the second type of translated media content may be a composited and synchronized media stream having the translated audio stream of the first terminal **31** in the second language or any other language understood by the second user **32** at the second terminal, and a subtitled video stream having the original video stream of the first terminal **31** and any one or a combination of one or more subtitle tracks. In particular, the audio transcription data file(s) may be embedded within the original video stream of the first

terminal **31** for it to have a subtitle track displaying text in the first language. Moreover, or alternatively, the translated transcription data file(s) may be embedded within the original video stream of the first terminal **31** for it to have a subtitle track displaying text in the second language or any other language understood by the second user **32** at the second terminal. The subtitle track(s) allow subtitles to be displayed alongside with the graphics of the original video stream based on their timestamps.

[00126] Next, there is an eighteenth step that involves retrieving the translated media content in its intended output format, by the service control module **1101**.

[00127] Next, there is a nineteenth step that involves pushing the translated media content in its intended output format, to the media processing module **1105** and/or the signalling module **1104**. The media processing module **1105** and/or the signalling module **1104** may perform their operations upon the translated media content sequentially in any desired order.

[00128] In particular, the media processing module **1105** may carry out the step of performing technical processing that includes any one of a combination of as transcoding and compression of audio and video media streams, bandwidth requirement reduction, and transmission efficiency improvement. It may further attempt to synchronize the translated media content.

[00129] In particular, the signalling module **1104** may carry out the step of converting the translated media content into appropriate transmission signals that may correspond to audio and/or video specifications for the network **20** (i.e. the IMS communication network).

[00130] Next, there is a twentieth step that involves sending the translated media content to the second terminal **32** from the network **20**. This step may be carried out by any one of the signalling module **1104** or the media processing module **1105**.

[00131] Finally, there is a twenty-first step that involves receiving the translated media content by the second terminal **32** through the established call route. More specifically, the second terminal **32** may be able display the translated media content in the intended format as per the output format settings of the second terminal **32** as set by the second user **B**. In

particular, the translated media content may be present the original video stream and/or the original audio stream of the first terminal **31** having subtitles and/or audio in the second language or any other language understood by the second user **B**.

[00132] Optionally, within the second set of steps, there is further included a third set of steps related to updating and/or correcting the translated media content in real-time or near real-time.

[00133] The third set of steps may begin with a first step that involves determining a receipt of at least one audio segmentation service request message was sent from at least one of the terminals **31**, **32** in real-time or near real-time. This step may be carried out by the service control module **1101**. In particular, the audio segmentation service request message may be sent when either user **A**, **B** interacts with their corresponding terminal **31**, **32** to send this request therefrom.

[00134] Next, there may be a second step that involves determining that the audio segmentation service request message is received within a pre-set time period. This step may be carried out by the service control module **1101**.

[00135] Next, there may be a third step that involves instructing any one or a combination of the ASR module **1108**, the comprehension and translation module **1109**, and the TTS module **1110** to correct or update translations made for the media content that is exchanged or streamed between the terminals **31**, **32** in real-time or near real-time, within the pre-set time period. More specifically, for the translated media content, its translated audio stream and/or the translated text of its subtitle data files may be corrected or updated within the pre-set time period.

[00136] Next, there may be a third step that involves pushing the corrected or updated translated media content, to the corresponding terminal that had sent the audio segmentation service request message. With that, the output of the said terminal may reflect changes in its received translated media content. By way of example, a subtitle track that is displayed on the translated media content being streamed may instantaneously change to a language different from previous. By way of another example, a translated audio stream that is

vocalized from the terminal may instantaneously change to a language different from previous.

[00137] In particular, should the service control module determine that the audio segmentation service request message is absent within the pre-set time period, the service control module **1101** may subsequently carry out the step of deleting media content, more specifically, translated media content, which may be stored in the cache unit **13**.

[00138] Optionally, within the second set of steps, there is further included a fourth set of steps related to enabling the translated audio stream to have voiceprints of the users **A, B**.

[00139] The fourth set of steps may begin with a first step that involves determining a receipt of a user-voiceprint audio and video translation service request message in real-time or near real-time. This step may be carried out by the service control module **1101**. In particular, the user-voiceprint audio and video translation service request message may be sent when either user **A, B** interacts with their corresponding terminal **31, 32** to send this request therefrom.

[00140] The fourth set of steps may include a step involving determining the user-voiceprint audio and video translation service request message is received, by the service control module **1101**. Next, the step of instructing the relevant modules to generate the translated transcription data files may be carried out by the service control module **1101**. Next, the step of identifying and/or sampling timbre(s) within the original audio stream may be carried out by the TTS module **1110**. This allows for information of the original voiceprint to be retained. Next, the step of generating the translated audio stream based on the identified and/or sampled timbre(s) may be carried out by the TTS module **1110**. With that, the translated audio stream may be as if one user in the call is speaking in the translated language having their corresponding voice pitch, tone, intonation, etc., even though the aforementioned user does not speak the translated language. It is to be noted that the aforementioned steps mentioned in this paragraph may be extensions of the steps previously described for the second set of steps.

[00141] The fourth set of steps may also include a step involving determining the user-voiceprint audio and video translation service request message is absent, by the service

control module **1101**. Next, the step of instructing the relevant modules to generate the translated transcription data files may be carried out by the service control module **1101**. Next, the step of generating the translated audio stream based using audio of default or pre-loaded timbre(s) may be carried out by the TTS module **1110**. With that, the translated audio stream has no correspondence to the speech made by the user within the original audio stream, i.e. it may have a voice produced from artificial intelligence that may have a different voice pitch, tone, intonation, or the like. It is to be noted that the aforementioned steps mentioned in this paragraph may be extensions of the steps previously described for the second set of steps.

[**00142**] Optionally, within the first set of steps or second set of steps, there is further included a fifth set of steps related to ending the communication between the users **A**, **B**.

[**00143**] The fifth set of steps may include a step involving determining receipt of a call termination or disconnection service request message from at least one terminal **31**, **32**. Upon the service control module **1101** determining that such a service request message is received within a pre-set time, the step of disconnect the corresponding terminal from the service route is carried out by the media processing module **1105** to terminate real-time or near real-time translated communication for the disconnected terminal. With that, the translated media content that may be stored in the cache unit **13** may be deleted.

[**00144**]The fifth set of steps may include a step involving determining absence of a call termination or disconnection service request message from at least one terminal **31**, **32**. Upon the service control module **1101** determining that such a service request message is absent within a pre-set time, the service control module **1101** may initiate or continue exchange or streaming of translated media content between the terminals **31**, **32** in real-time or near real-time, which may be in accordance to the any one of the previously described set of steps.

[**00145**] It is to be noted that for the aforementioned sets of steps may allow media content to be exchanged or streamed between the terminals **31**, **32** in real-time, or near real-time with minimal to no latency.

[00146] It is to be noted that for the aforementioned sets of steps, the media content may be stored in the cache unit **13** for a pre-determined period of time. More specifically, any one or a combination of the original audio stream from a transmitting terminal, the original video stream from the transmitting terminal, the audio transcription data file(s), the translated transcription data file(s), the translated audio stream for the receiving terminal, the composited and/or synchronized audio-video stream for the receiving terminal, may be stored in the cache unit **13** after they are generated from a corresponding step. Moreover, the media content stored in the cache unit **13** may be retrieved by other modules for relevant steps to be carried out.

[00147] From hereon, one or more example implementations of the present invention shall be described.

[00148] In a first example implementation, for the first terminal **31**, it may have settings in which its input language settings is French and its output language settings is English. Whereas, for the second terminal **32**, it may have settings in which its input language settings is Italian and its output language settings is English.

[00149] Hence, for the first example implementation, during communication (i.e. a call) between the first user **A** via the first terminal **31** and the second user **B** via the second terminal, as the first user **A** speaks French via the first terminal **31**, the second user **B** via the second terminal **32** perceives English audio and/or English subtitles, which are translated from the French spoken by the first user **A**, and optionally, they may further perceive French subtitles of the French spoken by the first user **A**. Conversely, as the second user **B** speaks Italian via the second terminal **31**, the first user **A** via the first terminal **32** perceives an English audio and/or English subtitles, which are translated from the Italian spoken by the second user **B**, and optionally, they may further perceive Italian subtitles of the Italian spoken by the second user **B**. The present application may handle the translation and composition of the media stream exchanged between the terminals **31**, **32** in real-time or near real-time, as previously described. Optionally, the screen of the interface of each terminal **31,32** may also display a digital image or video (e.g. a picture or video stream of a corresponding user) as its background.

[00150] In a second example implementation, for the first terminal **31**, it may have settings in which its input language settings is Spanish and its output language settings is English. Whereas, for the second terminal **32**, it may have settings in which its input language settings is Italian and its output language settings is English.

[00151] Hence, for the second example implementation, during communication (i.e. a call) between the first user **A** via the first terminal **31** and the second user **B** via the second terminal, as the first user **A** speaks Spanish via the first terminal **31**, the second user **B** via the second terminal **32** perceives an English audio and/or English subtitles, which are translated from the Spanish spoken by the first user **A**, and optionally, they may further perceive Spanish subtitles of the French spoken by the first user **A**. Conversely, as the second user **B** speaks Italian via the second terminal **31**, the first user **A** via the first terminal **31** perceives an English audio and/or English subtitles, which are translated from the Italian spoken by the second user **B**, and optionally, they may further perceive Italian subtitles of the Italian spoken by the second user **B**. The present application may handle the translation and composition of the media stream exchanged between the terminals **31**, **32** in real-time or near real-time, as previously described. Optionally, the subtitles may be displayed in a timed manner. Optionally as well, the screen of the interface of each terminal **31,32** may also display a digital image or video (e.g. a picture or video stream of a corresponding user) as its background.

[00152] In a second example implementation, during the call, should the first user **A** interact with the first terminal **31** may to change its input language settings to French and its output language settings to Japanese, and should the second user **B** interact with second terminal **32** to change its input language settings to Korean and its output language settings to Japanese, the output formats of the translated media content displayed or vocalized by the terminals **31**, **32** may instantaneously change in real-time or near real-time, without the call being terminated or disconnected. More specifically, during communication (i.e. a call) between the first user **A** via the first terminal **31** and the second user **B** via the second terminal, as the first user **A** speaks French via the first terminal **31**, the second user **B** via the second terminal **32** perceives Japanese audio and/or Japanese subtitles, which are translated from the French spoken by the first user **A**, and optionally, they may further perceive

Japanese subtitles of the French spoken by the first user **A**. Conversely, as the second user **B** speaks Korean via the second terminal **32**, the first user **A** via the first terminal **32** perceives a Japanese audio and/or Japanese subtitles, which are translated from the Korean spoken by the second user **B**, and optionally, they may further perceive Korean subtitles of the Korean spoken by the second user **B**. The present application may handle the translation and composition of the media stream exchanged between the terminals **31**, **32** in real-time or near real-time, as previously described. Optionally, the subtitles may be displayed in a timed manner. Optionally as well, the screen of the interface of each terminal **31,32** may also display a digital image or video (e.g. a picture or video stream of a corresponding user) as its background.

[00153] It is to be noted that in further or alternative embodiments of the invention, the exchange of translated media content may occur between the terminals **31**, **32** after a period of time following establishment of the call. By way of example, when one user (i.e. the recipient) answers a call via their terminal, they may not immediately understand the language spoken by the other user (i.e. the caller), hence, the recipient may prompt their terminal to send a media translation service request message over the network **20** to reach the device **10**. Upon receipt of the media translation service request message over the network **20** by the service control module **1101**, the device **10** may then perform relevant steps to initiate or continue exchange or streaming of translated media content to the terminal of the recipient based on any one or a combination of the sets of steps as previously described. In certain embodiments, the input language settings, the output language settings, and the output format settings may only be set by the recipient and/or the caller, and provided to the device **10** after a period of time following establishment of the call. Furthermore, should the service control module **1101** determine that the media translation service request message is absent (i.e. not received) within a pre-set time, the service control module **1101** may carry out the steps of stopping exchange or streaming of the media content translation operations for either one or a combination of the terminals **31**, **32**, and deleting the translated media content that were stored in the cache unit **13**.

[00154] It is to be noted that in alternative embodiments, the business logic module **110**, the AS module **1107**, the ASR module **1108**, and the comprehension and translation module **1109** may be provided by third-party service provider(s). In one possible implementation,

these modules may be provided by different third-party service providers or by one third-party service provider of the same.

[00155] It is to be noted that in further or alternative embodiments of the invention, the device **10** may be configured to temporarily store the translated media content in the data storage medium **12** and/or the cache unit **13** prior to them being sent to the terminals **31, 32**. More specifically, in scenarios where the call network quality is poor, the translated media streams may be delivered in a delayed manner to ensure viewing and/or hearing smoothness between the terminals **31, 32**. In particular, the users **A, B** may be informed that the media stream is being delayed and shall be notified when the stream is loaded. Preferably, either the service control module **1101** and/or the business logic module **1106** may read from the data storage medium **12** and/or the cache unit **13** to load these audio-video media streams. With such an implementation, the wait time for loading the media content is reduced.

[00156] It is to be noted that in further or alternative embodiments of the invention, the device **10** may further operate a multimedia streaming module (not shown) that is configured to manage or facilitate the streaming of media content exchanged or streamed between the terminals **31, 32** in real-time or near real-time.

[00157] It is to be noted that the functionalities described in this invention can be implemented using hardware, software, firmware, or any combination of these. When implemented via software, these functions can be stored on a computer storage medium and executed through one or more instructions or codes. The storage medium can be any accessible medium, usable by either a general-purpose or a special-purpose computer.

[00158] It is to be noted that those skilled in the art should also understand that the invention provides examples focused on various functional modules to maintain clarity and conciseness. In actual applications, these modular functionalities will be tailored and integrated as necessary to meet specific service requests.

[00159] It is to be noted that for the various embodiments described for the present invention, the control methods, storage media, and systems can be implemented through various alternative means. The division of modules or units is a logical organization of

functionalities, but actual implementations may use different combinations or divisions. Multiple modules can be combined, integrated into other device applications, or certain features may be omitted if not needed. The interface coupling, direct coupling, or communication connections shown or discussed can be selected either fully or partially, depending on practical needs to achieve the objectives of these embodiments.

[00160] With this, the details pertaining to a system and method for real-time translation in communications are described. This invention shall enable for mutual communication between users that speak in different languages. The present disclosure includes as contained in the appended claims, as well as that of the foregoing description. Although this invention has been described in its preferred form with a degree of particularity, it is understood that the present disclosure of the preferred form has been made only by way of example and that numerous changes in the details of construction and the combination and arrangements of parts may be resorted to without departing from the scope of the invention.

CLAIMS

1. A system for real-time translation in communications, comprising
 - a plurality of terminals that include a first terminal and a second terminal;
 - at least one network, which is configured to allow media content to be exchanged or streamed between the terminals in real-time or near real-time, via a service route at a substantially synchronized or near-synchronized manner, wherein the media content includes original media content of at least one language that is inputted to each terminal, and translated media content of at least one language that is outputted at each terminal; and
 - at least one device that is either part of, or in communication with, the network, having at least one processor that operates a plurality of modules, with the device configured to receive the original media content from at least one terminal for generating the translated media content therefrom that is sent to at least one corresponding terminal via the service route;
 - wherein the original media content includes at least one original audio stream, at least one original video stream, or a combination thereof;
 - wherein the translated media content includes the original audio stream, the original video stream, at least one translated audio stream, at least one data file having text transcribed and/or translated from the original audio stream that is embeddable within the original video stream as one or more subtitle tracks, or a combination thereof, which are composited and synchronized by the device based on specifications of the service route of the network;
 - wherein the device receives a signal from the first terminal that is to be passed to the second terminal for receipt by the second terminal, during which the device determines the language of the translated media content for the first terminal, and as the device receives another signal from the second terminal that is to be passed to the first terminal that indicates call receipt by the second terminal, the device determines the language of the translated media content for the second terminal, in which communications are established between the terminals thereafter where the device then receives the original media content from the terminals for translation into translated media content based on the determined languages;
 - wherein the device is further configured to allow a change of both the language of the original media content and the language of the translated media content translated thereinto for one terminal, upon receipt of a message from the aforementioned terminal that relates to a change in its languages, during exchange or streaming of the media content

between the terminals in real-time or near real-time, in which the languages for the original media content and the translated media content are both different from previous; and

wherein the translated text is selectively corrected by the device in an automated manner, based on complete sentence segments, as the translated media content is exchanged or streamed between the terminals.

2. The system according to claim 1, wherein the network has a communication framework that is based on the Internet Protocol Multimedia Core Network Subsystem (IMS) architecture, wherein the framework is either a Voice over Long-Term Evolution (VoLTE) channel, a Video over Long-Term Evolution (ViLTE) channel, or a combination thereof.

3. The system according to claim 1 or 2, wherein the modules include a business logic module that is configured to either or both

authenticate user identity information of at least one corresponding user of one corresponding terminal; and

determine service subscription information of at least one corresponding user of one corresponding terminal that relates to any one or a combination of input language settings of the terminals, output language settings of the terminals, and an output format of the translated media content that are being exchanged or streamed in real-time or near real-time, at the terminals.

4. The system according to claim 3, wherein the output format of the translated media content include any one of

a first type of translated media content, which is composited and synchronized media stream from one terminal having the original audio stream, and a subtitled video stream having the original video stream and any one or a combination of one subtitle track transcribed from the original audio stream and translated subtitle tracks transcribed and translated from the original audio stream; or

a second type of translated media content, which is composited and synchronized media stream from the one terminal having the translated audio stream, and the subtitled video stream having the original video stream and any one or a combination of one subtitle track transcribed from the original audio stream and the translated subtitle tracks transcribed and translated from the original audio stream.

wherein the translated subtitle track, the translated audio stream, or a combination thereof, are of languages according to the output language settings of the terminal that receives the translated media content.

5. The system according to any one of the preceding claims, wherein the modules include an automated speech recognition module that is configured to
recognize an original language from the original audio stream; and
transcribe a first set of text in the original language from the original audio stream;
and
generate at least one audio transcription data file having the transcribed first set of text, which is embeddable within the original video stream.

6. The system according to claim 5, wherein the modules include a comprehension and translation module that is configured to
receive audio transcription file having the transcribed first set of text;
translate the transcribed first set of text into a second set of text that is of at least one translated language; and
generate at least one translated transcription data file having the second set of text, which is embeddable within the original video stream.

7. The system according to claim 6, wherein the modules include a text-to-speech module that is configured to generate the translated audio stream based on the second set of text of the translated transcription data file.

8. The system according to claim 7, wherein the text-to-speech module is further configured identify and/or sample timbre(s) within the original audio stream to generate the translated audio stream that mimics the voice of one user at one terminal in which the original audio stream is transmitted therefrom.

9. The system according to any one of the preceding claims, wherein the modules include a service control module that is configured to determine receipt of at least one media translation service request message from at least one terminal within a pre-set time period, for the service control module, or in conjunction with the other modules, to either

initiate or continue exchange or streaming of translated media content between the terminals in real-time or near real-time; or

stop the exchange or streaming of translated media content between the terminals in real-time or near real-time, and optionally, delete the translated media content that is stored within the device.

10. The system according to claim 9, wherein the service control module is further configured to

delete the translated media content that is stored within the device.

11. The system according to any one of the preceding claims, wherein the modules include a signalling module and a media processing module which are configured to enable the original media content and the translated media content to be exchanged between corresponding terminals via the service route over the network in a synchronized manner in real-time or near real-time.

12. The system according to any one of the preceding claims, further comprising a cache unit configured to store any one or a combination of the original media stream, the translated media stream, and their associated media stream resources, for a pre-determined time period, as they are being exchanged or streamed between the terminals in real-time or near real-time, via the service route.

13. A method for real-time translation in communications, comprising the steps of
allowing media content to be exchanged or streamed between a plurality of terminals, which include a first terminal and a second terminal, in real-time or near real-time, via a service route at a substantially synchronized or near-synchronized manner, by at least one network, wherein the media content includes original media content of at least one language that is inputted to each terminal, and translated media content of at least one language that is outputted at each terminal, which involves

receiving the original media content from at least one terminal, by at least one device that is either part of, or in communication with, the network, wherein the device has at least one processor that operates a plurality of modules;

generating the translated media content from the original media content, by the device; and

sending the translated media content to at least one corresponding terminal via the service route, by the device;

wherein the original media content includes at least one original audio stream, at least one original video stream, or a combination thereof;

wherein the translated media content includes the original audio stream, the original video stream, at least one translated audio stream, at least one data file having text transcribed and/or translated from the original audio stream that is embeddable within the original video stream as one or more subtitle tracks, or a combination thereof, which are composited and synchronized by the device based on specifications of the service route of the network;

wherein the device receives a signal from the first terminal that is to be passed to the second terminal for receipt by the second terminal, during which the device determines the language of the translated media content for the first terminal, and as the device receives another signal from the second terminal that is to be passed to the first terminal that indicates call receipt by the second terminal, the device determines the language of the translated media content for the second terminal, in which communications are established between the terminals thereafter where the device then receives the original media content from the terminals for translation into translated media content based on the determined languages;

wherein the device is further configured to allow a change of both the language of the original media content and the language of the translated media content translated thereinto for one terminal, upon receipt of a message from the aforementioned terminal that relates to a change in its languages, during exchange or streaming of the media content between the terminals in real-time or near real-time, in which the languages for the original media content and the translated media content are both different from previous; and

wherein the translated text is selectively corrected by the device in an automated manner, based on complete sentence segments, as the translated media content is exchanged or streamed between the terminals.

14. The method according to claim 13, wherein the network has a communication framework that is based on the Internet Protocol Multimedia Core Network Subsystem (IMS)

architecture, wherein the framework is either a Voice over Long-Term Evolution (VoLTE) channel, a Video over Long-Term Evolution (ViLTE) channel, or a combination thereof.

15. The method according to claim 13 or 14, further comprising the steps of
authenticating user identity information of at least one corresponding user of one corresponding terminal, by a business logic module;

determining service subscription information of at least one corresponding user of one corresponding terminal that relates to any one or a combination of input language settings of the terminals, output language settings of the terminals, and an output format of the translated media content that are being exchanged or streamed in real-time or near real-time at the terminals, by the business logic module;

or a combination thereof.

16. The method according to any one of claims 13 to 15, further comprising steps carried out by an automated speech recognition module, which include

recognising an original language from the original audio stream;

transcribing a first set of text in the original language from the original audio stream;

and

generating at least one audio transcription data file having the transcribed first set of text, which is embeddable within the original video stream.

17. The method according to claim 16, further comprising steps carried out by a comprehension and translation module, which include

translating the transcribed first set of text into a second set of text that is of at least one translated language; and

generating at least one translated transcription data file having the second set of text, which is embeddable within the original video stream.

18. The method according to claim 17, further comprising the step of generate the translated audio stream based on the second set of text of the translated transcription data file by a text-to-speech module.

19. The method according to any one of claims 13 to 18, further comprising steps carried out by a service control module, which include

determining receipt of at least one media translation service request message from at least one terminal within a pre-set time period, for initiating or continuing exchange or streaming of translated media content between the terminals in real-time or near real-time;

determining absence of at least one media translation service request message from at least one terminal within a pre-set time period, for stopping the exchange or streaming of translated media content between the terminals in real-time or near real-time; and

deleting the translated media content that is stored within the device.

20. A computer-readable medium for real-time translation in communications, being part of a device that is either part of, or in communication with, a network that is configured to allow media content to be exchanged or streamed between a plurality of terminals, include a first terminal and a second terminal, in real-time or near real-time, via a service route, wherein the media content includes original media content of at least one language that is inputted to each terminal, and translated media content of at least one language that is outputted at each terminal, wherein the computer-readable medium stores modules including their instructions that, when executed by a processor of the device, causes the processor to

enable the device to receive the original media content from at least one terminal;

generate translated media content from the original media content; and

send the translated media content to corresponding terminals in real-time or near real-time, via the service route;

wherein the original media content includes at least one original audio stream, at least one original video stream, or a combination thereof; and

wherein the translated media content includes the original audio stream, the original video stream, at least one translated audio stream, at least one data file having text transcribed and/or translated from the original audio stream that is embeddable within the original video stream as one or more subtitle tracks, or a combination thereof, which are composited and synchronized by the device based on specifications of the service route of the network;

wherein the processor of the device, as instructed by the computer-readable medium, receives a signal from the first terminal that is to be passed to the second terminal for receipt by the second terminal, during which the device determines the language of the translated

media content for the first terminal, and as the device receives another signal from the second terminal that is to be passed to the first terminal that indicates call receipt by the second terminal, the device determines the language of the translated media content for the second terminal, in which communications are established between the terminals thereafter where the device then receives the original media content from the terminals for translation into translated media content based on the determined languages;

wherein processor of the device, as instructed by the computer-readable medium, is further configured to allow a change of both the language of the original media content and the language of the translated media content translated thereinto for one terminal, upon receipt of a message from the aforementioned terminal that relates to a change for its languages, during exchange or streaming of the media content between the terminals in real-time or near real-time, in which the languages for the original media content and the translated media content are both different from previous; and

wherein the translated text is selectively corrected by the processor of the device based on instructions from the computer-readable medium, in an automated manner, based on complete sentence segments, as the translated media content is exchanged or streamed between the terminals.