



- (51) **International Patent Classification:**
G06F 3/01 (2006.01) *G06N 3/08* (2023.01)
- (21) **International Application Number:**
PCT/EP2024/053796
- (22) **International Filing Date:**
15 February 2024 (15.02.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
23157288.4 17 February 2023 (17.02.2023) EP
- (71) **Applicant: SONY SEMICONDUCTOR SOLUTIONS CORPORATION** [JP/JP]; 4-14-1 Asahi-cho, Atsugi-shi, Kanagawa 243-0014 (JP).
- (71) **Applicant (for AL only): SONY EUROPE B.V.** [NL/GB]; The Heights, Brooklands, Weybridge Surrey KT13 0XW (GB).
- (72) **Inventors: MARKHASIN, Lev**; c/o IP Europe, Sony Europe B.V., Zweigniederlassung Deutschland, Stuttgart Technology Center, Hedelfinger Str. 61, 70327 Stuttgart (DE). **LIN, Shan**; c/o IP Europe, Sony Europe B.V.,

- Zweigniederlassung Deutschland, Stuttgart Technology Center, Hedelfinger Str. 61, 70327 Stuttgart (DE). **BEL-GACEM, Iheb**; c/o IP Europe, Sony Europe B.V., Zweigniederlassung Deutschland, Stuttgart Technology Center, Hedelfinger Str. 61, 70327 Stuttgart (DE).
- (74) **Agent: MFG PATENTANWAELTE MEYER-WILD-HAGEN, MEGGLE-FREUND, GERHARD PARTG MBB**; Amalienstrasse 62, 80799 Munich (DE).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,

(54) **Title:** DEVICE, METHOD, AND COMPUTER PROGRAM TO CONTROL AN AVATAR

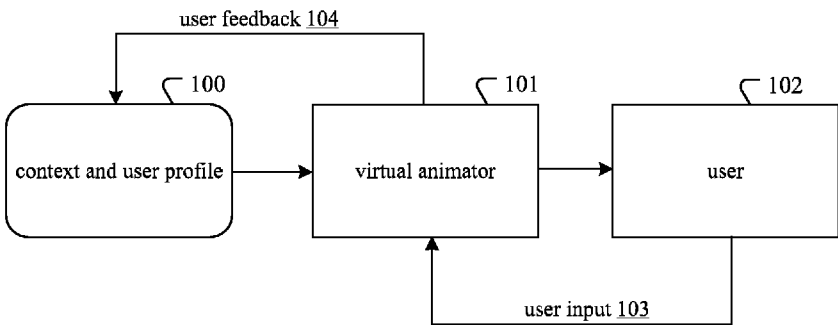


Fig. 1

(57) **Abstract:** An electronic device comprising circuitry configured to acquire (400, 402) user information (603, 702) related to a detected user; and generate (406) a personalized strategy (407) of an avatar (409) based on the user information (603, 702).

SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

DEVICE, METHOD, AND COMPUTER PROGRAM TO CONTROL AN AVATAR

TECHNICAL FIELD

The present disclosure generally pertains to the field of virtual reality, and in particular to method, device and computer program for performing image processing.

TECHNICAL BACKGROUND

Generally, neural networks, such as Deep Neural Network (DNN) and Convolutional Neural Network (CNN) are known, and they are used in a plurality of technical fields, for example in image processing. Known image processing devices may use DNN and CNN for image reconstruction, multispatial and multispectral image generation, object recognition and the like.

Moreover, DNN and CNN typically have an input layer, an output layer and multiple hidden layers between the input layer and the output layer. In image processing, a neural network may be trained to output images having high spectral resolution or high spatial resolution, using as an input to the neural network, a color channel image, such as an RGB image (having red, green and blue color channels).

Although there exist techniques for image processing, it is generally desirable to improve image processing devices and methods.

SUMMARY

According to a first aspect, the disclosure provides an electronic device comprising circuitry configured to acquire user information related to a detected user and generate a personalized strategy of an avatar based on the user information.

According to a second aspect, the disclosure provides a method comprising acquiring user information related to a detected user and generating a personalized strategy of an avatar based on the user information.

According to a third aspect, the disclosure provides a computer program comprising instructions, the instructions when executed on a processor causing the processor to perform acquiring user information related to a detected user and generating a personalized strategy of an avatar based on the user information.

Further aspects are set forth in the dependent claims, the following description and the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments are explained by way of example with respect to the accompanying drawings, in which:

Fig. 1 schematically shows a process of enhancing virtual reality experience by personalizing the virtual reality experience;

Fig. 2 schematically shows in more detail an embodiment of a context and user profile as described in Fig. 1;

Fig. 3 schematically shows an embodiment of a virtual club animator as described in Fig. 1;

Fig. 4 shows a flow-diagram of a process of generating a personalized avatar;

Fig. 5a schematically shows in more detail an embodiment of a process of user detection performed in the process of generating a personalized avatar as described in Fig. 4, wherein the virtual animator is projected as a hologram;

Fig. 5b schematically shows in more detail an embodiment of a process of user detection performed in the process of generating a personalized avatar as described in Fig. 4, wherein the virtual animator is displayed in mixed reality glasses;

Fig. 6 schematically shows in more detail an embodiment of a process of performing basic user information collection performed in the process of generating a personalized avatar as described in Fig. 4;

Fig. 7 schematically shows in more detail an embodiment of a process of performing user analysis performed in the process of generating a personalized avatar as described in Fig. 4;

Fig. 8 schematically shows in more detail an embodiment of a process of personalized strategy generation performed in the process of generating a personalized avatar as described in Fig. 4;

Fig. 9 schematically shows in more detail an embodiment of a process of personalized avatar generation, wherein a text prompt is used;

Fig. 10 shows a flow diagram visualizing a method for generating a personalized avatar for enhancing a virtual reality experience;

Fig. 11 illustrates an embodiment of a venue wherein personalized avatars and users interact with each other; and

Fig. 12 shows a block diagram depicting an embodiment of an electronic device that can implement the processes of generating a personalized avatar.

DETAILED DESCRIPTION OF EMBODIMENTS

Before a detailed description of the embodiments under reference of Fig. 1 to Fig. 12 are given, general explanations are made.

As mentioned in the outset, neural networks, such as Deep Neural Network (DNN) and Convolutional Neural Network (CNN) are known, and they have reached state-of-the-art level performance in many domains, such as of image processing, image reconstruction, multispatial and multispectral image generation, language processing and the like.

As it is generally known, deep learning models and neural networks are used in virtual reality scenarios. For example, for generating an avatar in a virtual reality scenario a neural network using a text-to-video model system may be used. Virtual reality is used not only in gaming scenarios but also in everyday life scenarios.

In some cases, there are people find it awkward, to go for example to a club or a restaurant alone. People who are shy/introverted/new in a city/solo tourists may want to go to a club to dance and to meet new people but may hesitate to go if they don't find someone that can accompany. And even in company many people just stick to the company they came with and have difficulties to approach strangers.

It has been recognized that, for example, shy people may mean boring atmosphere in a club and therefore it may be in the club's best interest to help its guests to get in contact with each other.

Consequently, some embodiments pertain to an electronic device comprising circuitry configured to acquire user information related to a detected user and generate a personalized strategy of an avatar based on the user information.

The electronic device may be a digital (video) camera, an edge computing enabled image sensor, such as smart sensor associated with smart speaker, or the like, a smartphone, a personal computer, a laptop computer, a personal computer, a wearable electronic device, electronic glasses, professional music equipment or the like, a circuitry, a processor, multiple processors, logic circuits or a mixture of those parts.

The circuitry may include one or more processors, logical circuits, memory (read only memory, random memory, etc., storage memory, i.e. hard disc, compact disc, flash drive, etc.), an interface for communication via a network, such as a wireless network, internet, local area network, or the like, a CMOS (Complementary Metal Oxide Semiconductor) image sensor, a CCD (Charge Coupled Device) image sensor, or the like.

The avatar may be for example, a virtual animator that greets people entering the club and starts a conversation, makes small talk, tries to create a fun and welcoming atmosphere with them trying to lighten the mood quickly. In this manner, the presence of a user is detected, i.e. a detected user and user information related to the detected user is acquired.

- 5 The personalized strategy may be a list of actions, or at least one action that is generated and being performed by the avatar. The personalized strategy may be adjusted to the acquired user information, to an acquired user feedback as well as to the venue style. In this manner, the virtual reality experience, e.g., clubbing experience, may be more enjoyable for all users. The avatar may be customized to reflect the values of the venue and may be adapted to the clubbers personalities. The avatar may be multilingual and may make the virtual reality experience more inclusive.

The personalized strategy generation may be an action planning as described herein. The personalized strategy may be an action plan as described herein.

- 15 In some embodiments, the circuitry may be further configured to generate the avatar based on the generated personalized strategy. The avatar generated based on the generated personalized strategy is a personalized avatar. For example, using the acquired user information a personalized avatar is generated, that is an avatar based on the user profile, user interest, user preferences and the like. Such an avatar is a personalized avatar, e.g., a virtual (human) club animator, that creates a welcoming atmosphere for shy and/or introverted people, helps them socialize in any place, such as a venue/club, connects them with compatible people based on their data and profiles and initiates dancing. For example, the personalized avatar may take the form of a hologram or be displayed by mixed reality glasses.

- 25 In some embodiments, the circuitry may be further configured to perform user analysis based on the user information to obtain a user analysis result. User analysis may include performing body scanning to obtain physical information, information regarding the age, the gender, the fitness level of a user and the like. A user analysis result may comprise all the above information as well as basic user information acquired from an interface to users.

- 30 The user information may be all kind of information acquired by the interface to users. For example, information regarding the profile of the user, the profile of the venue, information that the user actively inputs by talking or by any other means and information that can be acquired by microphone and camera.

The user analysis may be a user data analysis as described herein.

In some embodiments, the circuitry may be configured to perform personalized strategy generation based on the user analysis result to obtain the personalized strategy of the avatar.

In some embodiments, performing user analysis may comprise performing body scanning to obtain a physical analysis result included in the user analysis result.

5 In some embodiments, the generated personalized strategy may comprise at least one of performing small talk with a user, engage user to dance, engage user to drink, engage user to sing and user matching. The generated personalized strategy may comprise for example introducing guests to each other and leading the conversation until a chemistry has been established or stopping the conversation in a non-awkward way if no chemistry can be established.

10 The introduction of people to each other may be based on data collected during an initial conversation or based on certain profiles. For example, the electronic device and/or avatar may have access to the users' social media profiles and may use that data to optimize the matching.

Alternatively, or additionally, the electronic device and/or avatar may have access to the users' dating profiles and match them specifically with the intent of dating. For example, it may also
15 provide support by jokingly pushing the two users towards each other.

In this manner, shy or average people may have the same experience as very social, or extrovert people have by getting in contact with a lot of new people without feeling bad when starting a conversation. For example, the personalized avatar may initiate the dancing if users are too shy to start by themselves.

20 Since it is normal to go out without a friend, the personalized avatar may make sure that everyone is enjoying themselves/having a positive experience.

In some embodiments, the avatar may be a virtual animator presented to a user as a hologram. In this case there may be a need for one or more speakers, one or more microphones and one or more cameras in the walls, floor and ceiling. The device may activate the speakers that are best
25 for a specific person without enhancing the overall sound in the room. A person may then give a command like "louder" to get stronger sound. In this case all people see and hear the same avatar and speech. The behaviour of the avatar may be based on camera and microphone input. Lights in the place may be optimized for the hologram experience.

Alternatively, in some embodiments, the avatar may be a virtual animator presented to a user in
30 mixed reality glasses. In this case there may be a need for an earpiece with a microphone.

Glasses may be available for usage or may be private given that a specific application is installed and some kind of check-in is performed. In this case different users may see different avatars and

each avatar may give certain hidden hints like how to approach the other person to speed up the dating etc. like “offer her to buy a drink”.

In some embodiments, the circuitry may be further configured to perform human detection and human localization to obtain the detected user.

- 5 In some embodiments, the user information may be acquired by an interface to users implemented by one or more camera and/or one or more microphones.

In some embodiments, the user information may comprise user profile information and context information.

- 10 In some embodiments, the avatar may talk to the detected user to acquire basic user information included in the user information. Basic user information may comprise information whether the user is new or old, information about user drink/eat preferences, user experience history, and the like.

In some embodiments, generating a personalized strategy of an avatar may be implemented by a deep learning model.

- 15 In some embodiments, the deep learning model may be implemented by a neural network.

In some embodiments, performing user analysis may be implemented by a deep learning model.

In some embodiments, the personalized avatar generation may be implemented based on a transformer model.

In some embodiments, the transform model may use text-to-video conversion technology.

- 20 In some embodiments, the user detection, the basic user information collection and the personalized avatar generation may be implemented by an interface to users.

- The interface to users may acquire any possible information. For example, information regarding the profile of the user, the profile of the venue, information that the user actively inputs by talking or by any other means, information that can be acquired by microphone and camera, without
25 limiting the present disclosure in that regard.

Some embodiments pertain to a method comprising acquiring user information related to a detected user and generating a personalized strategy of an avatar based on the user information.

Some embodiments pertain to a computer program comprising instructions, the instructions when executed on a processor causing the processor to perform comprising acquiring user information related to a detected user and generating a personalized strategy of an avatar based on the user information.

- 5 The electronic device may be a Hardware-Software (HW-SW) product that simulates a human who keeps company, animates party behaviour and assists in meeting other people. It may provide more fun for most users, More fun may lead to more users and more drinks and food.

General concept of a virtual animator

Fig. 1 schematically shows a process of enhancing virtual reality experience by personalizing the virtual reality experience.

A context and user profile 100 is transferred to a personalized avatar, here a virtual animator 101. The virtual animator 101 processes data of the context and user profile 100 to project/display the personalized virtual animator to a user 102. The user 102 transmits to the virtual animator 101 a user input 103, such as a user feedback indicative of the interaction between the virtual animator 101 and the user 103. The virtual animator 101 updates the context and user profile 100 based on user feedback 104 comprising the user input 103. In this manner, the virtual animator 101 becomes a personalized virtual animator providing an enhanced virtual reality experience to the user 102. For example, the user feedback 104 may include feedback of the user regarding previous interaction with the virtual animator and feedback regarding the atmosphere of the place where the user and the virtual animator are met, e.g., a club, a restaurant, social events, conferences, universities, schools, or the like.

In the embodiment of Fig. 1, the user 102 is described as one person, without limiting the present embodiment in that regard. Alternatively, there may be a plurality of users interacting with one virtual animator, e.g., a virtual club animator 101, or with a plurality of avatars (virtual animators).

Fig. 2 schematically shows in more detail an embodiment of a context and user profile as described in Fig. 1. The context and user profile 100 comprises user information 200 and context information 201. The user information 200 includes a conversation history of the user with the virtual club animator or with other users, user voice, a user interaction history with the virtual club animator, the profile of the user and the user's preferences, and the like. The context information 201 includes values of venue (e.g., club), style of venue, environment information, e.g., information acquired by cameras which for example are suitable also for dark environments.

For example, the user profile may be acquired by the user's social media profile, dating user's profile, etc. The user interaction history with the virtual club animator may include for example, feedback of the user regarding previous interaction with the virtual club animator and feedback regarding the atmosphere of the place (see user feedback 104 in Fig. 1) where the user and the virtual club animator are met, e.g., a club. In this manner the virtual animator meets the user, talk with the user and learns the user's habits and characteristics.

The values of venue may for example include a customer category to which the venue refers to. The style of venue may for example include the kind of music that usually is played in the venue.

In the embodiment of Fig. 2, the avatar is described as virtual club animator which interacts with the user within a club, without limiting the present embodiment in that regard. The avatar may interact with the user in any possible place, e.g., in a restaurant, social events, conferences, universities, schools, or the like.

Fig. 3 schematically shows an embodiment of a virtual club animator as described in Fig. 1. The virtual club animator 101 receives as an input user information 200 and context information 201 via an interface to users 300, wherein the interface to users 300 comprises a user interface. In other words, the virtual club animator 101 collects information about the user and the venue via the interface to users 300. User data analysis 301 is performed on the collected information, here user information 200 and context information 201, to obtain a data analysis result 302. Action planning 303 is performed on the data analysis result 302 to obtain an action plan 304. The action plan 304 is received by the interface to users 300 and is output as a display of a personalized virtual club animator with motion and sound 305.

In the embodiment of Fig. 3, the interface to users is used to collect information about the users e.g., by an active input from the users such as by talking to them, and to collect information from one or more cameras and one or more microphones existing in a place, e.g. a venue, where the virtual animator and the user are also present.

Generating a personalized avatar

Fig. 4 shows a flow-diagram of a process of generating a personalized avatar. As described in Figs. 1 and 3, when a user enters a place, such a venue, a club, a restaurant, universities, schools, a place for social events, for conferences, or the like, an avatar meets the user and talks with the user.

At 400, user detection is performed to detect whether a user is present or not and to obtain presence information of a detected user (see 102 in Fig. 1). The user detection may be performed by

the interface to users (see 300 in Fig. 3) of the avatar (see virtual animator 101 in Figs. 1 and 3). At 402, basic user information collection is performed on the detected user (see 102 in Fig. 1) to obtain basic user information, such as if the user is new/old customer, what are the user drink/eat preferences, the user experience history, etc. The basic user information collection may be performed by the interface to users (see 300 in Fig. 3) of the avatar (see virtual animator 101 in Figs. 1 and 3). At 404, user analysis is performed to obtain a user analysis result. The user analysis is performed based on the basic user information and also based on further acquired information, as described in more detail in Fig. 7 below. The user analysis may be performed by the user data analysis (see 301 in Fig. 3) of the avatar (see virtual animator 101 in Figs. 1 and 3). The user analysis result may comprise the basic user information, a user profile, user matching, a dance type recommendation of the user, etc. At 406, personalized strategy generation is performed based on the user analysis result to obtain at least one personalized strategy. The personalized strategy generation may be performed by the action planning (see 302 in Fig. 3) of the avatar (see virtual animator 101 in Figs. 1 and 3), and the at least one personalized strategy corresponds to the action plan (see 304 in Fig. 3), as described in Fig. 3. At 408, personalized avatar generation is performed based on the at least one personalized strategy to generate a personalized avatar 409, such as a personalized virtual animator (see 305 in Fig. 3).

In the embodiment of Fig. 4, the avatar may be customized to reflect the values of the place, may adapt to the users' personalities, and may be multilingual. For example, if the avatar is within a club, the virtual animator may make the club more inclusive and may make the clubbing experience more enjoyable for everyone. In this manner the avatar changes its reactions based on the reactions of the user.

In the embodiment of Fig. 4, the personalized avatar 409 is generated to perform the at least one generated personalized strategy. For example, the personalized avatar 409 speaks to the user, dance with the user and the like. In other words, there is a sound/speech generation and also motion generation. The sound/speech generation of the avatar may be implemented as described by Long Ouyang et al, in the published paper "Training language models to follow instructions with human feedback", arXiv:2203.02155 [cs.CL]. The motion generation of the avatar may be implemented as described by Ahmed A, et al., in the published paper "STAR: Sparse Trained Articulated Human Body Regressor", arXiv:2008.08535 [cs.CV], and by Fangzhou Hong, et al. in the published paper "AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars".

In other words, the avatar is trained using, e.g., a deep learning model, such as a neural network, wherein a speech-to-motion/video or text-to-motion/video conversion technology is used.

In the embodiment of Fig. 4, the personalized avatar 409 is placed within the venue, so there may be performed a scene understanding and avatar localization. The visual scene understanding may be implemented as described by Cheng Zhang, et al., in the published paper “Holistic 3D Scene Understanding from a Single Image with Implicit Representation”. The avatar localization such that the 3D avatar is correctly placed in the scene may be implemented as described by Mohamed Hassan, et al. in the published paper “Populating 3D Scenes by Learning Human-Scene Interaction”.

User detection

Fig. 5a schematically shows in more detail an embodiment of a process of user detection performed in the process of generating a personalized avatar as described in Fig. 4, wherein the virtual animator is projected as a hologram. The virtual animator performs human detection 500 to obtain a detected human. Then, the virtual animator performs human localization 501 to detect the location of the human, and thus to detect the presence of a user, e.g. a detected user.

Fig. 5b schematically shows in more detail an embodiment of a process of user detection performed in the process of generating a personalized avatar as described in Fig. 4, wherein the virtual animator is displayed in mixed reality glasses. In the embodiment of Fig. 5b, the presence of the detected user is obtained by wearing glass and earpiece 502 and then by activating a user device and related application 503.

Basic user information collection

Fig. 6 schematically shows in more detail an embodiment of a process of performing basic user information collection performed in the process of generating a personalized avatar as described in Fig. 4 above.

An avatar projection/display 600 is performed to obtain an avatar projected as a hologram or displayed in mixed reality glasses. An avatar meeting 601 is performed with a detected user 401. A talk with the avatar 602 is performed with the detected user 401 to obtain user information 603, such as if the detected user 401 is a new or old costumer, what the detected user to wants to drink and/or eat, acquire information regarding the experience history like user feedback, etc. All these information named user information 603 are included in the basic user information 403.

User analysis

Fig. 7 schematically shoes in more detail an embodiment of a process of performing user analysis performed in the process of generating a personalized avatar as described in Fig. 4.

A body scanning 700 is performed to obtain physical information. A user information acquisition via user data analysis 701 is performed based on the basic user information 403 and the physical
5 to obtain detailed user information 702, such as a user profile, a user matching, a dance type recommendation, etc. All these information is included in the user analysis result 405.

Personalized strategy generation

Fig. 8 schematically shows in more detail an embodiment of a process of personalized strategy generation performed in the process of generating a personalized avatar as described in Fig. 4
10 above.

A data analysis 800 is performed on the user analysis result 405 to obtain a data analysis result. A user category determination 801 is performed based on the data analysis result to obtain a user category determination result which categorizes the user. A personalized strategy generation 802 is performed to obtain at least one personalized strategy 407.

The personalized strategy 407 may for example include a warmup talk, such as a small talk between the user and the avatar, performs user matching, introduce users to each other, start a dancing session and lead a user to the dancing floor, etc. Based on the warmup talk, an emotion analysis may be performed to obtain an emotion analysis result used for example as feedback for improving the generated personalized strategy. For example, starting a dancing session and leading
20 a user to the dancing floor may be followed by cheering, applause, and encouragement feedback used as feedback for improving the generated personalized strategy. A personalized avatar (see 409 in Fig. 4) is generated to perform the at least one generated personalized strategy.

Personalized avatar generation

Fig. 9 schematically shows in more detail an embodiment of a process of personalized avatar generation, wherein a text prompt indicates “Generate “Avatar showing to user dancing moves””. A text-to-video model neural network neural network 806 receives as input a text
25 prompt 805, here “Generate “Avatar showing to user dancing moves””, indicating generation of a personalized avatar that shows to the user some dancing moves. The text-to-video model neural network 806 output a personalized avatar that moves and shows to the user some dancing moves.

In the embodiment of Fig. 9 the text-to-video model neural network is an encoder-decoder neural network may be implemented as described by Fangzhou Hong, et al. in the published paper “AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars”.

30

Text-to-video models guide creation of a video through natural language. The conditioned text-to-video system, such as the personalized avatar generation 408 can be implemented based on a transformer model.

A transformer model is a deep learning model that uses natural language processing (NLP) and is designed to process sequential input data, for example natural language with translation and text summarization. The transformer has an encoder-decoder architecture and gets as input text a text prompt (see 805 in Fig. 9). The encoder-decoder architecture comprises an encoder and a decoder. The encoder consists of encoding layers which process the input iteratively one layer after another. The decoder consists of decoding layers that process the encoder's output iteratively one layer after another.

For example, the transform model uses a text to video conversion technology implemented by the encoder-decoder type of neural network, described above.

A text-to-video model is a machine learning model having as input text a natural language description, such as the text prompt 805 and output a sequence of images, such as a video, see 807 in Fig. 9 matching that description. Typically, text-to-video model uses a language model and a generative video model. The language model transforms the input text, e.g. the text prompt 805, into a latent representation. The generative video model produces a video conditioned on that representation, such as the 805 in Fig. 9.

The text-to-video model is trained on large amounts of image data and text data sets. The text encoding step of the text-to-video model may be implemented by a recurrent neural network, such as a long short-term memory (LSTM) network. The video generation step of the text-to-video model may be implemented by conditional generative adversarial network or by a diffusion model.

Method

Fig. 10 shows a flow diagram visualizing a method for generating a personalized avatar for enhancing a virtual reality experience.

At 900, a user detection (see 400 in Fig. 4) is performed to obtain a detected user. At 901, a basic user information collection (see 402 in Fig. 4) is performed to obtain basic user information (see 403 in Fig. 4). At 902, a user analysis (see 404 in Fig. 4) is performed to obtain a user analysis result (see 405 in Fig. 4). At 903, a personalized strategy generation (see 406 in Fig. 4) is performed based on the user analysis result (see 405 in Fig. 4) to obtain at least one personalized

strategy (see 407 in Fig. 4). At 904, a personalized avatar generation (see 408 in Fig. 4) is performed based on the at least one personalized strategy (see 407 in Fig. 4) to obtain a personalized avatar (see 409 in Fig. 4).

Virtual club animator

5 Fig. 11 illustrates an embodiment of a venue wherein personalized avatars and users interacts with each other, as described in the above figures. In the embodiment of Fig. 11, a venue, such as a club is presented. In the club a plurality of speakers, cameras and microphones are installed, which are configured to capture the environment of the club, the customers of the club, the avatars, here virtual club animators, the voices of the users and of the avatars, images capturing the
10 inside of the club, and the interactions between a plurality of users and a plurality of personalized avatars, here personalized virtual club animators. The plurality of speakers, cameras and microphones are installed in the walls, floor, and ceiling of the club. The users 102 interact by talking with the virtual club animators 101. As described in the figures above, the virtual club animators 101 talk with the users 102 to collect a plurality of user information (see 200 in Fig. 2) in order
15 to generate a personalized virtual club animator (see 409 in Fig. 4) that knows the user 102 preferences. The virtual animator 101 by greeting people entering the club and by starting a conversation with them, makes small talk with the users 102, tries to create a fun and welcoming atmosphere with them trying to lighten the mood quickly. Then the personalized virtual club animator introduces guests to each other and leads the conversation until a chemistry has been es-
20 tablished or stops the conversation in a non-awkward way if no chemistry can be established. For example, the personalized virtual club animator may initiate the dancing if people are too shy to start by themselves. In this manner, the personalized virtual club animator may provide to shy or average people the same experience as very social, or extrovert people have in a club by getting in contact with a lot of new people without feeling bad when starting a conversation.

25 In the embodiment of Fig. 11 the virtual club animator 101 is shown via a hologram. In this case there need to be many speakers, microphones and cameras in the walls, floor and ceiling of the club. The system activates the speakers that are best for a specific person without enhancing the overall sound in the room. A person can then give a command like “louder” to get stronger sound. In this case all people see and hear the same avatar and speech. The behaviour of the avatar is based on camera and microphone input. Lights in the club are optimized for the hologram
30 experience.

In the embodiment of Fig. 11, the virtual club animator is a hologram, without limiting the present embodiment in that regard. Alternatively, the virtual club animator may be displayed by

mixed/virtual reality glasses. In this case, an earpiece with a microphone is needed. Virtual reality glasses may be available in the club for usage or may be private given, for example, when a specific application (App) is installed, and a check-in process may be performed. In this case different people may see different personalized club animators and the personalized club animator
5 may give certain hidden hints like how to approach the other person to speed up the dating etc., e.g., “offer her to buy a drink”, or the like.

In the embodiment of Fig. 11, the place where the users and the avatars interact is a club, without limiting the present embodiment in that regard. Alternatively, the concept of a virtual club animator may take place to other social contexts and places, such as restaurants, social events, conferences, universities, schools, and the like. For example, if a user wants to go for dinner to a
10 restaurant alone but wants to have a conversation, a virtual animator may join the user at the table with a virtual meal.

Implementation

Fig. 12 shows a block diagram depicting an embodiment of an electronic device that can implement the processes of generating a personalized avatar for enhancing a virtual reality experience.
15 The electronic device 1200 comprises a CPU 1201 as processor. The electronic device 1200 further comprises a microphone array 1210, a loudspeaker array 1211 and a neural network unit (NN) 1220 that are connected to the processor 1201. The NN unit 1220 may for example be an artificial neural network in hardware, e.g., a neural network on GPUs or any other hardware specialized for the purpose of implementing an artificial neural network. The NN1220 may for example implement the process of performing personalized avatar generation, as described in
20 Fig. 4. Loudspeaker array 1211 consists of one or more loudspeakers that are distributed over a predefined space and is configured to render 3D audio. The electronic device 1200 further comprises a user interface 1212 that is connected to the processor 1201. This user interface 1212 acts
25 as a man-machine interface and enables a dialogue between an administrator and the electronic system. The user interface 1212 may be a graphical user interface (GUI). For example, the user may actively input information, e.g., by talking to the avatar, as described in Figs. 1, 2, 3, 4, and 6. Still further, an administrator may make configurations to the system using this user interface 1212. The electronic device 1200 further comprises a Bluetooth interface 1204, and a WLAN interface 1205. These units 1204, 1205 act as I/O interfaces for data communication with external
30 devices. For example, additional loudspeakers, microphones, and video cameras with Ethernet, WLAN or Bluetooth connection may be coupled to the processor 1201 via these interfaces 1204, and 1205.

The electronic system 1200 further comprises a data storage 1202 and a data memory 1203 (here a RAM). The data memory 1203 is arranged to temporarily store or cache data or computer instructions for processing by the processor 1201. The data storage 1202 is arranged as a long-term storage, e.g., for recording sensor data obtained from the microphone array 1210 and provided to
5 or retrieved from the NN unit 1220. The data storage 1202 may also store audio data that represents audio messages, which the public announcement system may transport to people moving in the predefined space.

It should be noted that the description above is only an example configuration. Alternative configurations may be implemented with additional or other sensors, storage devices, interfaces, or
10 the like.

It should be further noted that alternatively the electronic device 1200 may be implemented with a digital signal processor (DSP) or a graphics processing unit (GPU), without limiting the present disclosure in that regard.

It should also be noted that the division of the electronic device of Fig. 12 into units is only made
15 for illustration purposes and that the present disclosure is not limited to any specific division of functions in specific units. For instance, at least parts of the circuitry could be implemented by a respectively programmed processor, field programmable gate array (FPGA), dedicated circuits, and the like.

It should be recognized that the embodiments describe methods with an exemplary ordering of
20 method steps. The specific ordering of method steps is, however, given for illustrative purposes only and should not be construed as binding.

All units and entities described in this specification and claimed in the appended claims can, if not stated otherwise, be implemented as integrated circuit logic, for example, on a chip, and functionality provided by such units and entities can, if not stated otherwise, be implemented by
25 software.

In so far as the embodiments of the disclosure described above are implemented, at least in part, using software-controlled data processing apparatus, it will be appreciated that a computer program providing such software control and a transmission, storage or other medium by which such a computer program is provided are envisaged as aspects of the present disclosure.

Note that the present technology can also be configured as described below.

- (1) An electronic device comprising circuitry configured to acquire (400, 402) user information (603, 702) related to a detected user; and generate (406) a personalized strategy (407) of an avatar (409) based on the user information (603, 702).
- 5 (2) The electronic device of (1), wherein the circuitry is further configured to generate (408) the avatar (409) based on the generated personalized strategy (407).
- (3) The electronic device of (1) or (2), wherein the circuitry is further configured to perform user analysis (404) based on the user information (603, 702) to obtain a user analysis result (302).
- 10 (4) The electronic device of (3), wherein the circuitry is configured to perform personalized strategy generation (406) based on the user analysis result (302) to obtain the personalized strategy (407) of the avatar (409).
- (5) The electronic device of (3), wherein performing user analysis (404) comprises performing body scanning (700) to obtain a physical analysis result (701) included in the user analysis result (302).
- 15 (6) The electronic device of any one of (1) to (5), wherein the generated personalized strategy (407) comprises at least one of performing small talk with a user, engage user to dance, engage user to drink, engage user to sing and user matching.
- (7) The electronic device of any one of (1) to (6), wherein the avatar (409) is a virtual animator presented to a user as a hologram.
- 20 (8) The electronic device of any one of (1) to (7), wherein the avatar (409) is a virtual animator presented to a user in mixed reality glasses.
- (9) The electronic device of any one of (1) to (8), wherein the circuitry is further configured to perform human detection (500) and human localization (501) to obtain the detected user.
- 25 (10) The electronic device of any one of (1) to (9), wherein the user information (603, 702) is acquired by an interface to users (300) implemented by one or more camera and/or one or more microphones.
- (11) The electronic device of any one of (1) to (10), wherein the user information (603, 702) comprises user profile information (200) and context information (201).

(12) The electronic device of any one of (1) to (11), wherein the avatar (409) talks to the detected user to acquire basic user information (603) included in the user information (603, 702).

(13) The electronic device of any one of (1) to (12), wherein generating a personalized strategy (407) of an avatar (409) is implemented by a deep learning model.

5 (14) The electronic device of (13), wherein the deep learning model is implemented by a neural network.

(15) The electronic device of (5), wherein performing user analysis (404) is implemented by a deep learning model.

10 (16) The electronic device of (2), wherein the personalized avatar generation (408) is implemented based on a transformer model.

(17) The electronic device of (16), wherein the transform model uses text-to-video conversion technology.

15 (18) The electronic device of (2), wherein the user detection (400), the basic user information collection (402) and the personalized avatar generation (408) are implemented by an interface to users (300).

(19) A method comprising:

acquiring (400, 402) user information (603, 702) related to a detected user (401); and

generating (406) a personalized strategy (407) of an avatar (409) based on the user information (603, 702).

20 (20) A computer program comprising instructions, the instructions when executed on a processor causing the processor to perform the method of (19).

CLAIMS

1. An electronic device comprising circuitry configured to acquire user information related to a detected user; and generate a personalized strategy of an avatar based on the user information.
- 5 2. The electronic device of claim 1, wherein the circuitry is further configured to generate the avatar based on the generated personalized strategy.
3. The electronic device of claim 1, wherein the circuitry is further configured to perform user analysis based on the user information to obtain a user analysis result.
4. The electronic device of claim 3, wherein the circuitry is configured to perform personal-
10 ized strategy generation based on the user analysis result to obtain the personalized strategy of the avatar.
5. The electronic device of claim 3, wherein performing user analysis comprises performing body scanning to obtain a physical analysis result included in the user analysis result.
6. The electronic device of claim 1, wherein the generated personalized strategy comprises
15 at least one of performing small talk with a user, engage user to dance, engage user to drink, engage user to sing and user matching.
7. The electronic device of claim 1, wherein the avatar is a virtual animator presented to a user as a hologram.
8. The electronic device of claim 1, wherein the avatar is a virtual animator presented to a
20 user in mixed reality glasses.
9. The electronic device of claim 1, wherein the circuitry is further configured to perform human detection and human localization to obtain the detected user.
10. The electronic device of claim 1, wherein the user information is acquired by an interface to users implemented by one or more camera and/or one or more microphones.
- 25 11. The electronic device of claim 1, wherein the user information comprises user profile information and context information.
12. The electronic device of claim 1, wherein the avatar talks to the detected user to acquire basic user information included in the user information.

13. The electronic device of claim 1, wherein generating a personalized strategy of an avatar is implemented by a deep learning model.

14. The electronic device of claim 13, wherein the deep learning model is implemented by a neural network.

5 15. The electronic device of claim 5, wherein performing user analysis is implemented by a deep learning model.

16. The electronic device of claim 2, wherein the personalized avatar generation is implemented based on a transformer model.

10 17. The electronic device of claim 16, wherein the transform model uses text-to-video conversion technology.

18. The electronic device of claim 2, wherein the user detection, the basic user information collection and the personalized avatar generation are implemented by an interface to users.

19. A method comprising:

acquiring user information related to a detected user; and

15 generating a personalized strategy of an avatar based on the user information.

20. A computer program comprising instructions, the instructions when executed on a processor causing the processor to perform the method of claim 19.

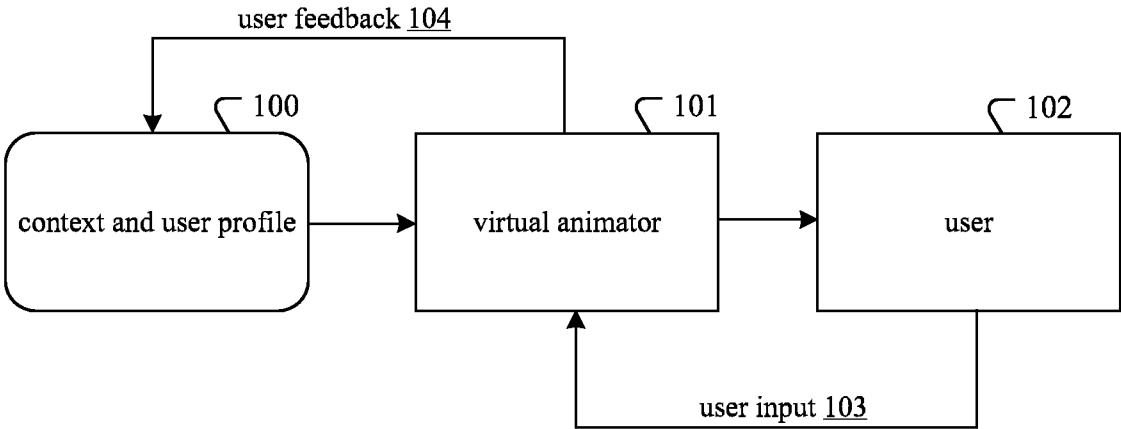


Fig. 1

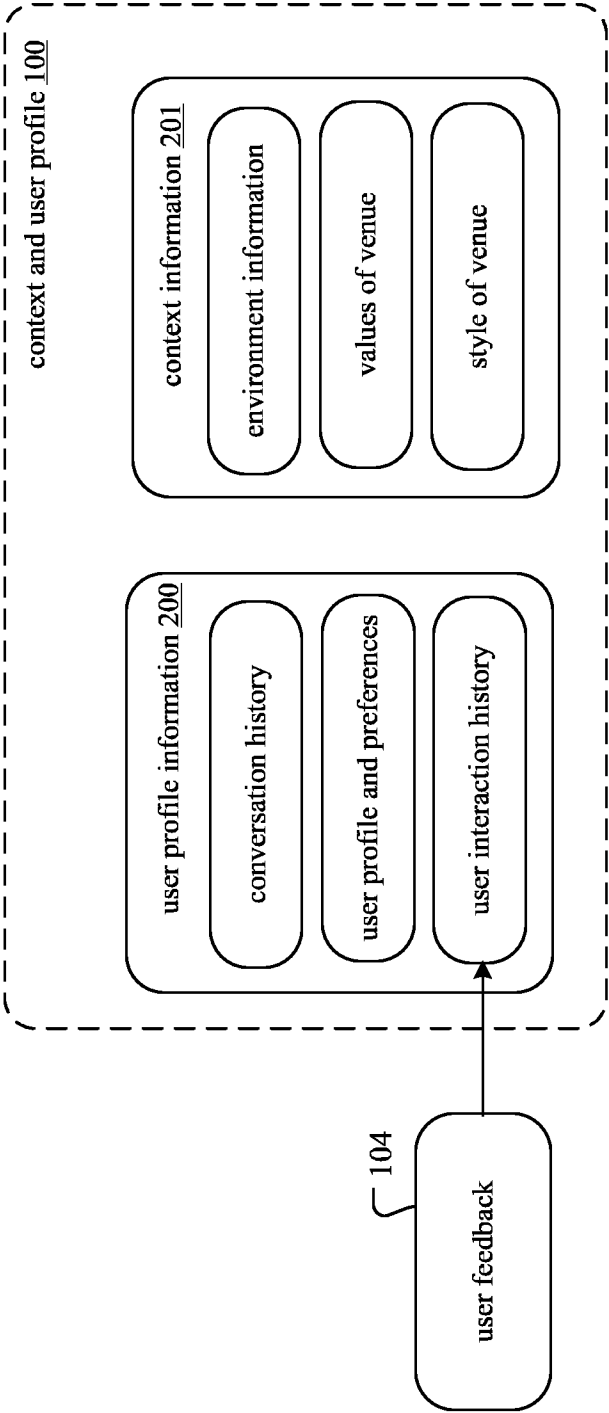


Fig. 2

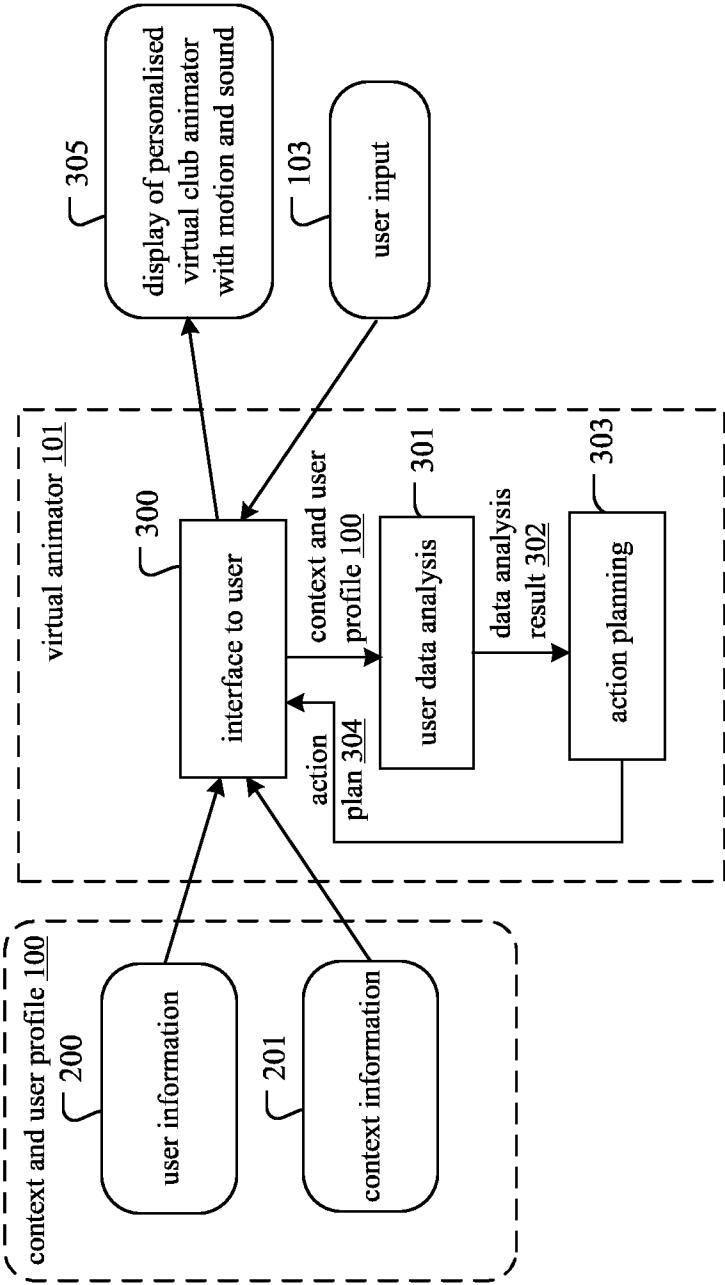


Fig. 3

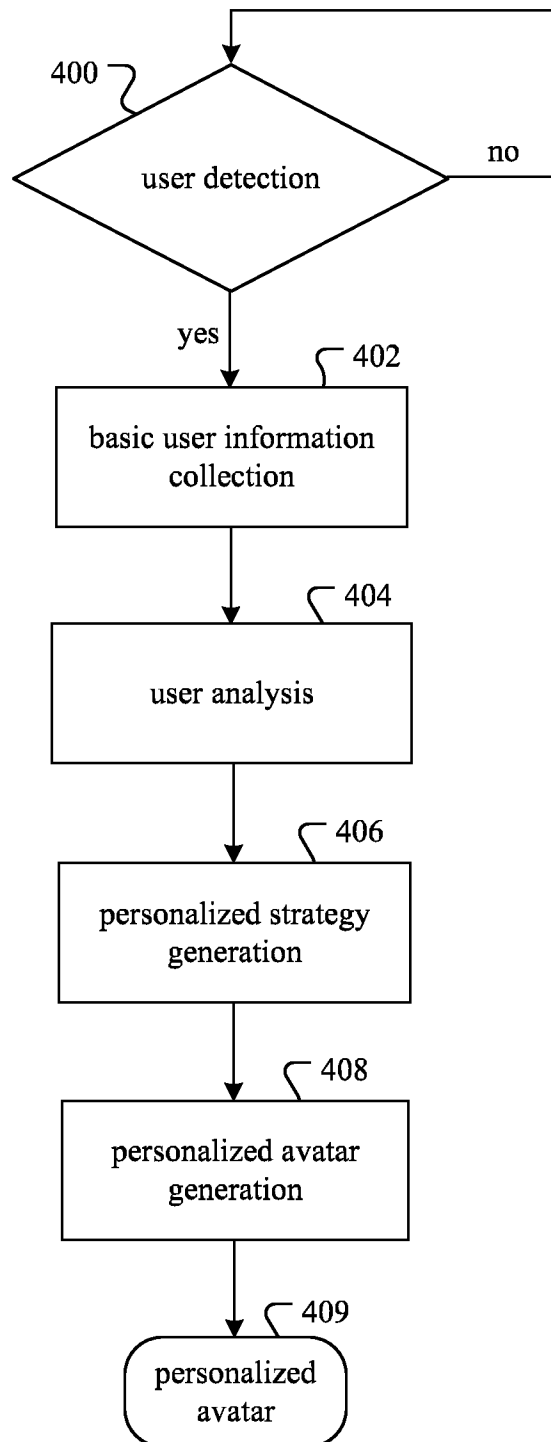
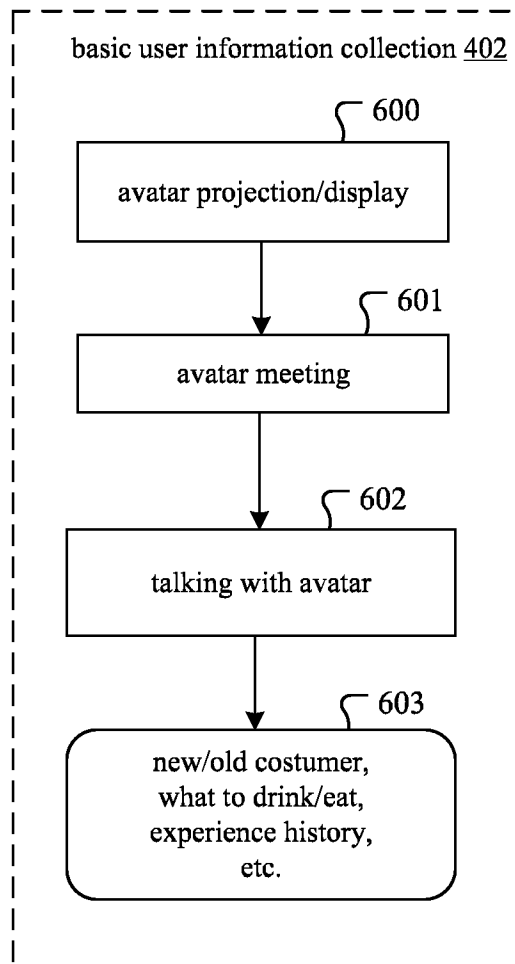
**Fig. 4**



Fig. 5a

Fig. 5b

**Fig. 6**

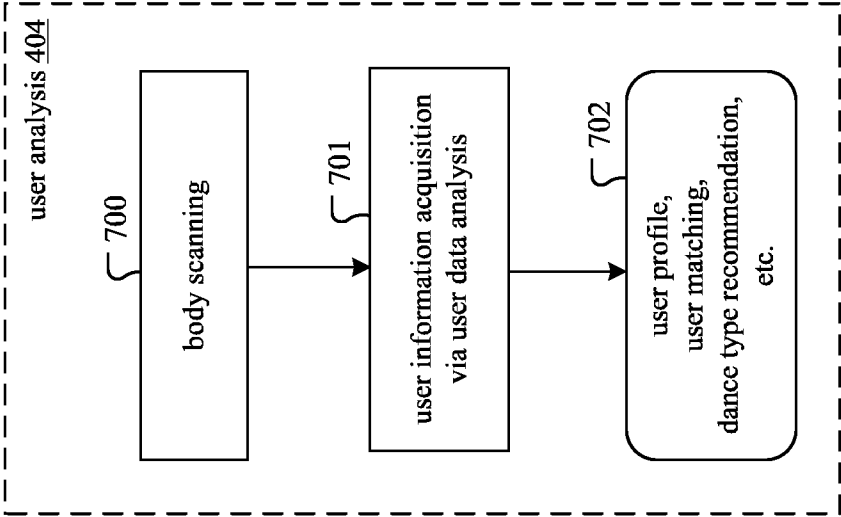


Fig. 7

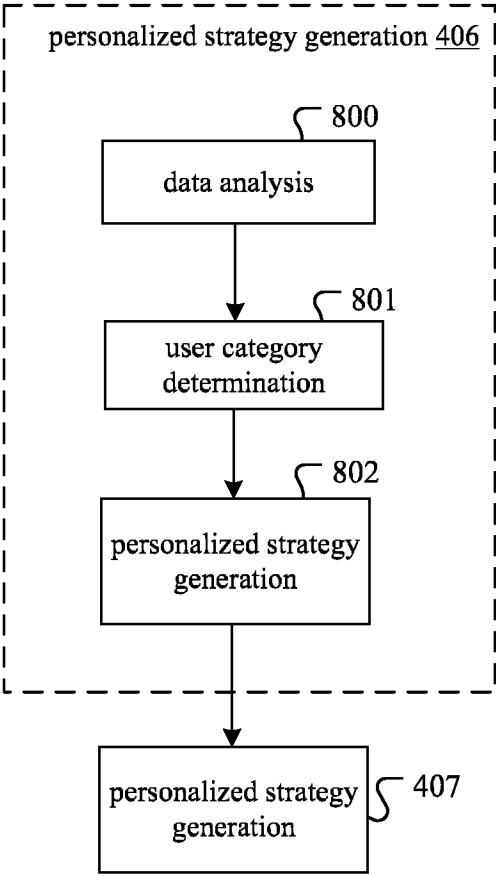


Fig. 8

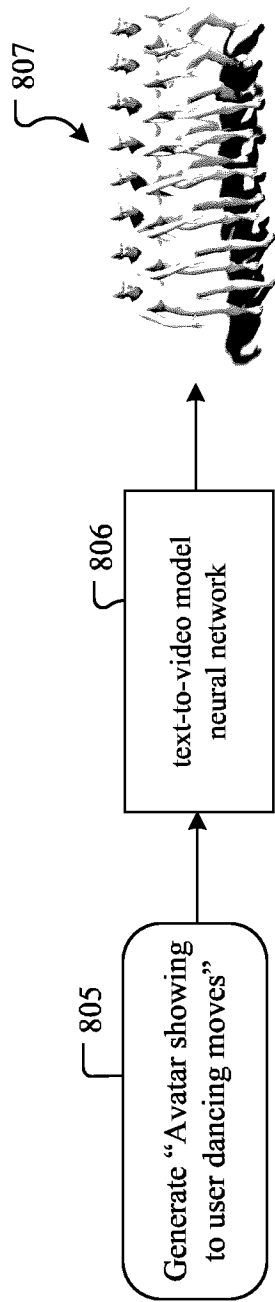
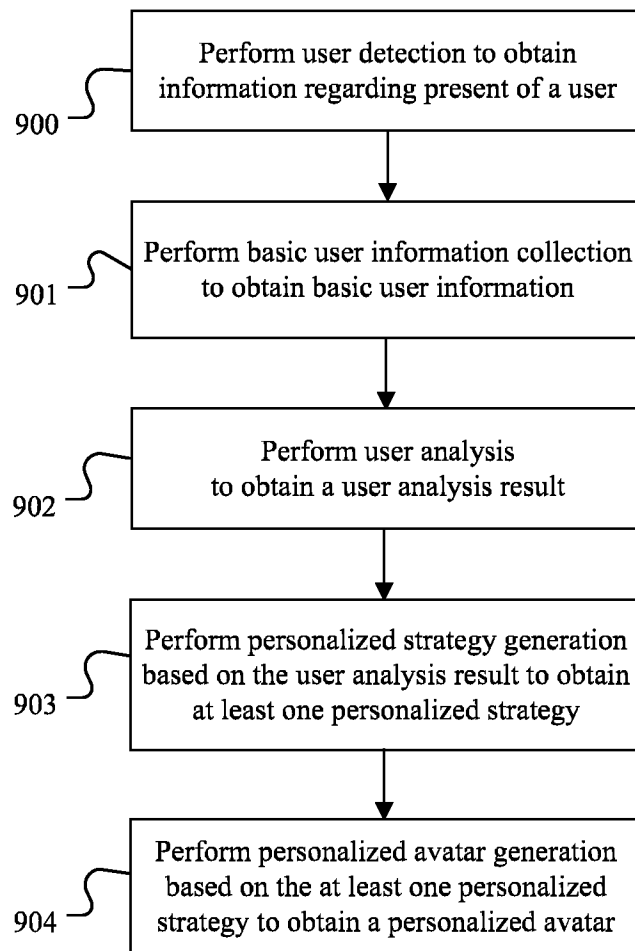


Fig. 9

**Fig. 10**

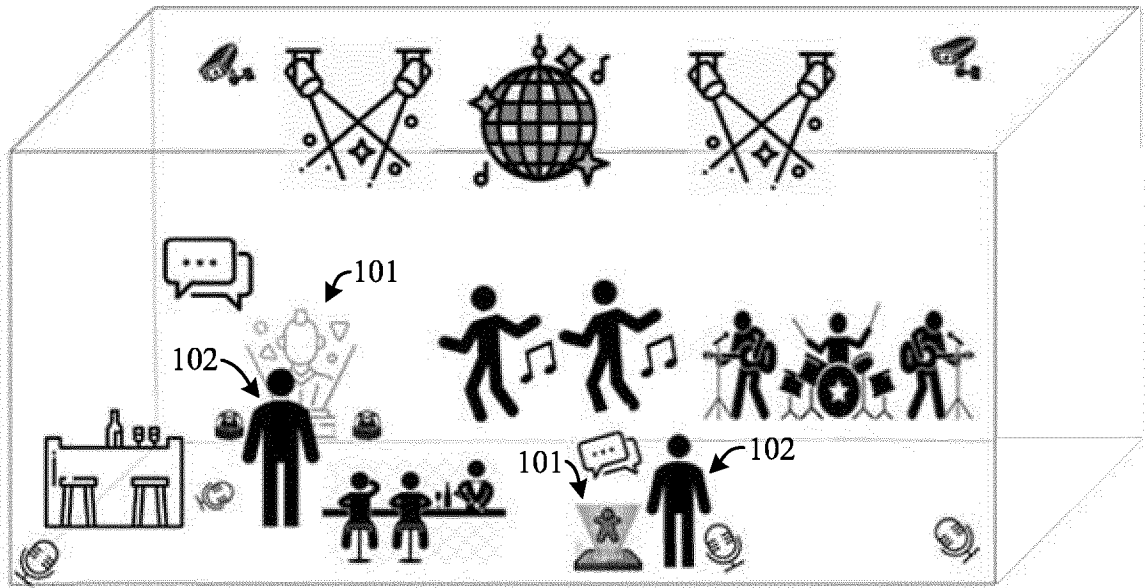


Fig. 11

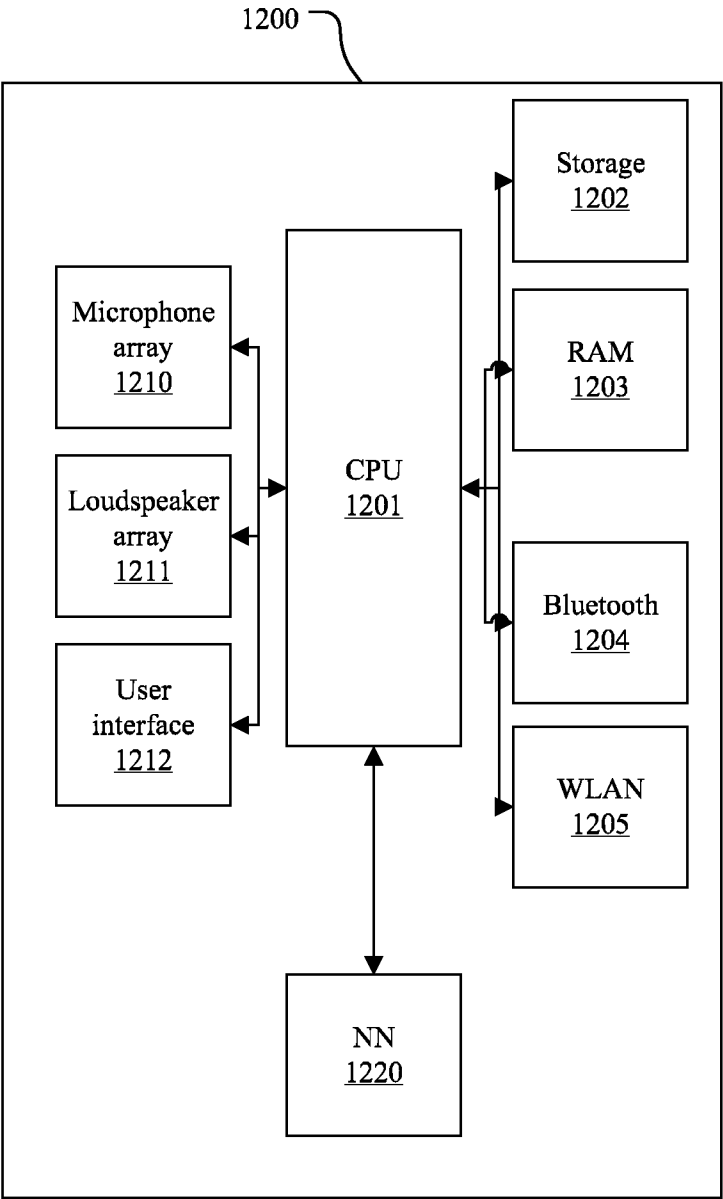


Fig. 12

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2024/053796

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F3/01 G06N3/08

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

A63F G06F G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2023/038213 A1 (KALEAL III ROBERT LOUIS	1-16,
Y	[US] ET AL) 9 February 2023 (2023-02-09)	18-20
	abstract	17
	paragraph [0024] - paragraph [0027]	
	paragraph [0030] - paragraph [0034]	
	paragraph [0038]	
	paragraph [0042]	
	paragraph [0045] - paragraph [0047]	
	paragraph [0060] - paragraph [0061]	
	paragraph [0110]	
	paragraph [0130] - paragraph [0132]	
	paragraph [0155]	

	-/--	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

18 April 2024

Date of mailing of the international search report

30/04/2024

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2

NL - 2280 HV Rijswijk

Tel. (+31-70) 340-2040,

Fax: (+31-70) 340-3016

Authorized officer

Fedrigo, Enrico

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2024/053796

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2022/148248 A1 (MCINTYRE-KIRWIN ARMANDO [US] ET AL) 12 May 2022 (2022-05-12) abstract paragraph [0003] paragraph [0031] - paragraph [0032] paragraph [0036] paragraph [0041] paragraph [0059] - paragraph [0069] paragraph [0093] - paragraph [0095] -----	1, 3, 4, 6, 8, 11, 19, 20
X	US 2012/130717 A1 (XU NING [CN] ET AL) 24 May 2012 (2012-05-24) abstract paragraph [0019] - paragraph [0021] paragraph [0025] paragraph [0038] - paragraph [0044] paragraph [0049] -----	1-5, 13-17, 19, 20
Y	FANGZHOU HONG ET AL: "AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 17 May 2022 (2022-05-17), XP091227613, abstract "Our approach" -----	17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2024/053796

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2023038213 A1	09-02-2023	US 2015037771 A1	05-02-2015
		US 2017027528 A1	02-02-2017
		US 2018116609 A1	03-05-2018
		US 2020245954 A1	06-08-2020
		US 2023038213 A1	09-02-2023

US 2022148248 A1	12-05-2022	AU 2020287622 A1	18-11-2021
		CA 3137927 A1	10-12-2020
		CN 114303116 A	08-04-2022
		EP 3980865 A1	13-04-2022
		JP 2022534708 A	03-08-2022
		KR 20220039702 A	29-03-2022
		US 2022148248 A1	12-05-2022
		US 2023145369 A1	11-05-2023
		WO 2020247590 A1	10-12-2020

US 2012130717 A1	24-05-2012	CN 102568023 A	11-07-2012
		US 2012130717 A1	24-05-2012
