Title: SPATIAL CONFERENCING SYSTEM AND METHOD

Abstract: The present invention relates to an enhanced method and an arrangement in a multi-party conferencing system having ability for spatial positioning of participant voices. The arrangement being configured to: process at least each received signal corresponding to a voice of a participant in a multi-party conferencing and extract at least one characteristic parameter for said voice of each participant (402), compare results (403) of said at least one characteristic parameters of at least each participant to find a similarity in said at least one characteristic parameter (404), and generate a virtual position for each participant voice through spatial positioning (405), in which a position of voices having similar characteristics is arranged distanced from each other in a virtual space.
SPATIAL CONFERENCING SYSTEM AND METHOD

TECHNICAL FIELD

The present invention relates to an arrangement and a method in a multi-party conferencing system

BACKGROUND

A human being can, using their two ears generally audibly preserve the direction and distance of a sound-source. Two cues are primarily used in the human auditory system to achieve this perception. These cues are the inter-aural time difference (ITD) and the inter-aural level difference (ILD) which result from the distance between the human's two ears and shadowing by the human's head. In addition to the ITD and ILD cues, a head-related transfer function (HRTF) is used to localize the sound-source in 3D space. The HRTF is the frequency response from a sound-source to each ear, which can be affected by diffractions and reflections of the sound waves as they propagate in space and pass around the human's torso, shoulders, head and pinna. Therefore, the HRTF for a sound-source generally differs from person to person.

In an environment where a plurality of people are talking at the same time, the human auditory system generally exploits information in the ITD cue, ILD cue and HRTF, and the ability to selectively focus one's listening attention on the voice of a particular talker. In addition, the human auditory system generally rejects sounds that are uncorrelated at the two ears, thus allowing the listener to focus on a particular talker and disregard sounds due to venue reverberation.

The ability to discern or separate apparent sound sources in 3D space is known as sound spatialization. The human auditory system has sound spatialization abilities which generally allow a human being to separate a plurality of simultaneously occurring sounds into different auditory objects and selectively focus on (i.e., primarily listen to) one particular sound.
For modern distance conferencing, one key component is a 3-dimensional audio spatial separation. This is used to distribute voice conference participants at different virtual positions around the listener. The spatial positioning helps the user identify different voices, even if they are unknown to the listener.

A wide range of methods for placing users in the virtual space can be perceived, and the easiest one is a random positioning. Random positioning, however, carries the risk that two voices similar sounding will be placed right next to each other. The benefit of spatial separation will be lost in those cases.

Spatial audio separation is well known. For example US 7,505,601 relates to a method and device for adding spatial audio capabilities by producing a digitally filtered copy of each input signal to represent a contra-lateral-ear signal with each desired talker location and treating each of a listener’s ears as separate end users.

SUMMARY

This summary is provided to introduce one or several selection of concepts, in a simplified form, that are further described hereafter in the detailed description. This summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

One of the objectives achieved by the present invention is to provide a conferencing system by spatial positioning of the participants in a manner that allows voices similar to each other are positioned in such a way that a user (listener) easily can distinguish different participants.

For this reason an arrangement in a multi-party conferencing system is provided. The arrangement comprises a processing unit and the arrangement is configured to: process at least each received signal corresponding to a voice of a participant in a multi-party conferencing and extract at least one characteristic parameter for the voice of each participant, compare results of the at least one characteristic parameters of at least each participant to find a similarity in the at least one characteristic parameter, and generate a virtual position for each participant voice through spatial positioning, in which a position of voices having similar characteristics is arranged distanced from each other in a virtual space. In the arrangement the spatializing is one or several of a virtual sound-source
positioning (VSP) method and a sound-field capture (SFC) method. The arrangement may further comprise a memory unit for storing sound characteristics and relating them to a participant profile.

5 The invention also relates to a computer for handling a multi-party conferencing. The computer comprises: a unit for receiving signals corresponding to a voice of a participant of the conferencing, a unit configured to analyze the signal, a unit configured to extract at least one characteristic parameter for the voice, a unit configured to compare the at least one characteristic parameter of at least each participant to find a similarity in the at least one characteristic parameter, a unit configured to generate a virtual position for each participant voice through spatial positioning, in which a position of voices having similar characteristics is arranged distanced from each other in a virtual space. The computer may further comprise a communication interface to a communication network.

10 The invention also relates to a communication device able of handling a multi-party conferencing. The communication device comprises: a communication portion, a sound input unit, a sound output unit, a unit configured to analyze a signal received from the communication network, the signal corresponding to voice of a party is the multi-party conferencing, a unit configured to extract at least one characteristic parameter for the voice, a unit configured to compare the at least one characteristic parameter of at least each participant to find a similarity in the at least one characteristic parameter, a unit configured to generate a virtual position for each participant voice through spatial positioning, in which a position of voices having similar characteristics is arranged distanced from each other in a virtual space and output through the sound output unit.

15 The invention also relates to a method in a multi-party conferencing system. The method comprises: analysing signal relating to one or several participant voices, processing at least each received signal and extracting at least one characteristic parameter for voice of each participant based on the signal, comparing result of the characteristic parameters to find similarity in the characteristic parameters, and generating a virtual position for each participant voice through spatial positioning, in which position of voices having similar characteristics is arranged distanced from each other in a virtual space.
BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will hereinafter be further explained by means of non-limiting examples with reference to the appended figures where:

5 Fig. 1 shows a schematic communication system according to the present invention,

Fig. 2 is block diagram of participant positioning in a system according to fig. 1,

Fig. 3 shows a schematic computer unit according to the present invention, and

Fig. 4 is a flow diagram according to one embodiment of the invention,

10 Fig. 5 is schematic communication device according to the present invention.

DETAILED DESCRIPTION

According to one aspect of the invention, the voice characteristics of the participants of a voice conference system are used to intelligently position similar voices far from each other when using spatial positioning.

Fig. 1 illustrates a conferencing system 100 according to one embodiment of the invention. The conferencing system 100 comprises a computing unit or conference server 110 which receives incoming calls from a number of user communications devices 120a-120c through one or several types of communication networks 130, such as public land mobile network, or public switched land network etc. The computer unit 110 communicates via one or several speakers 140a-140c to produce spatial positioning of the audio information. The speakers may also be substituted with a headphone(s).

25 With reference to Figs. 1 and 4, according to one aspect of the invention, when a participant of the communication device 120a-120c connects to the conference server 110, the received voice of the participant is analyzed 401 by an analyzing portion 111, which may be realized as a server component or a processing unit of the server. The voice is analyzed and one or several parameters characterizing each voice are extracted 402. The particular information that is extracted is beyond this application, but is considered common knowledge for a skilled person within voice recognition. This data may be retained and stored with information for recognition of the participant with a participant profile for future use. A storing unit 160 may be used for this purpose. The
voice characteristics as defined herein may comprise one or several of vocal range (registers), resonance, pitch, amplitude etc.

As mentioned above voice/speech recognition systems are well known for skilled persons. For example, some speech recognition systems make use of a Hidden Markov Model (HMM). A Hidden Markov Model outputs, for example, a sequence of n-dimensional real-valued vectors of coefficients (referred to as "cepstral" coefficients), which can be obtained by performing a Fourier transform of a predetermined window of speech, de-correlating the spectrum, and taking the first (most significant) coefficients. The Hidden Markov Model may have, in each state, a statistical distribution of diagonal covariance Gaussians which will give a likelihood for each observed vector. Each word, or each phoneme, will have a different output distribution; a hidden Markov model for a sequence of words or phonemes is made by concatenating the individual trained Hidden Markov Models for the separate words and phonemes. Decoding can make use of, for example, the Viterbi algorithm to find the most likely path.

One embodiment of the present invention may include an encoder to provide, e.g., the coefficients, or even the output distribution as the pre-processed voice recognition data. It is noted, however, that other speech models may be used and thus the encoder may function to extract other speech features.

When a participant joins a multi-party conference chat, the associated voice characteristics will be compared with the other participants' voice characteristics, and if participants are determined with similar voice patterns, that is with similar voices, are be positioned as far apart as possible. This helps all participants to build a distinct and accurate mental image of where participants are positioned.

Fig. 2 shows an example of the invention illustrating a "Listener" and a number of "Participants A-D". At the time of joining, the system concludes that, for example participant D has a voice pattern very similar to participant A. The system therefore places participant D to the far right, relative to the listener, to facilitate separation of the voices.

Fig. 3 illustrates a diagram of an exemplary embodiment of a suitable computing system (conferencing server) environment according to the present technique. The environment
illustrated in Fig 3 is only one example of a suitable computing system environment and
is not intended to suggest any limitation as to the scope of use or functionality of the
present technique. Neither should the computing system environment be interpreted as
having any dependency or requirement relating to any one or combination of components
exemplified in Fig 3.

As illustrated in Fig 3, an exemplary system for implementing the present technique
includes one or more computing devices, such as computing device 300 in its simplest
configuration, computing device 300 typically includes at least one processing unit 302
and memory 304.

Depending on the specific configuration and type of computing device, the memory 304
may be volatile (such as RAM), non-volatile (such as ROM and flash memory, among
others) or some combination of the two.

As exemplified in Fig 3 computing device 300 can also have additional features and
functionality. By way of example, computing device 300 can include additional storage
310 such as removable storage and/or non-removable storage. This additional storage
includes, but is not limited to, magnetic disks, optical disks and tape. Computer storage
media includes volatile and non-volatile media, as well as removable and non-removable
media implemented in any method or technology. The computer storage media provides
for storage of various information required to operate the device 300 such as computer
readable instructions associated with an operating system, application programs and
other program modules, and data structures, among other things. Memory 304, storage
310 are all examples of computer storage media. Computer storage media includes, but is
not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-
ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes,
magnetic tape, magnetic disk storage or other magnetic storage devices, or any other
medium which can be used to store the desired information and which can be accessed
by computing device 300. Any such computer storage media can be part of computing
device 300.

As exemplified in Fig 3, computing device 300 also includes a communications
interface(s) 312 that allows the device to operate in a networked environment and
communicate with a remote computing device(s), such as remote computing device(s).
Remote computing device can be a PC, a server, a router, a peer device or other common network node, and typically includes many or all of the elements described herein relative to computing device 300. Communication between computing devices takes place over a network, which provides a logical connection(s) between the computing devices. The logical connection(s) can include one or more different types of networks including, but not limited to, a local area network(s) and wide area network(s).

Such networking environments are commonplace in conventional offices, enterprise-wide computer networks, intranets and the Internet. It will be appreciated that the communications connection(s) and related network(s) described herein are exemplary and other means of establishing communication between the computing devices can be used.

As exemplified in Fig 3, communications connection and related network(s) are an example of communication media. Communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, but not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. The term "computer readable media" as used herein includes both storage media and communication media.

As exemplified in Fig 3, computing device 300 also includes an input device(s) 314 and output device(s) 316. Exemplary input devices 314 include, but are not limited to, a keyboard, mouse, pen, touch input device, audio input devices, and cameras, among others. A user can enter commands and various types of information into the computing device 300 through the input device(s) 314. Exemplary audio input devices (not illustrated) include, but are not limited to, a single microphone, a plurality of microphones in an array, a single audio/video (A/V) camera, and a plurality of cameras in an array. These audio input devices are used to capture a user's, or co-situated group of users', voice(s) and other audio information. Exemplary output devices 316 include, but are not limited to, a display device(s), a printer, and audio output devices, among others.
Exemplary audio output devices (not illustrated) include, but are not limited to, a single loudspeaker, a plurality of loudspeakers, and headphones.

These audio output devices are used to audibly play audio information to a user or co-situated group of users. With the exception of microphones, loudspeakers and headphones which are discussed in more detail hereafter, the rest of these input and output devices are well known and need not be discussed at length here.

The present technique can be described in the general context of computer-executable instructions, such as program modules, which are executed by computing device 300. Generally, program modules include routines, programs, objects, components, and data structures, among other things, that perform particular tasks or implement particular abstract data types. The present technique can also be practiced in a distributed computing environment where tasks are performed by one or more remote computing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote computer storage media including, but not limited to, memory 304 and storage device 310.

The present technique generally spatializes the audio in an audio conference between a plurality of parties situated remotely from one another. This is in contrast to conventional audio conferencing systems which generally provide for an audio conference that is monaural in nature due to the fact that they generally support only one audio stream (herein also referred to as an audio channel) from an end-to-end system perspective (i.e. between the parties). The present technique generally may involve one or several different methods for spatializing the audio in an audio conference, a virtual sound-source positioning (VSP) method and a sound-field capture (SFC) method. Both of these methods are assumed to be known to a person skilled in the art and not detailed herein.

The present technique generally results in each participant being more completely immersed in the audio conference and each conferences experiencing the collaboration that transpires as if all the conferences were situated together in the same venue.

The processing unit receives audio signals belonging to different participants, e.g. through communication network or input portions and analyze the voice characteristics. It may
also, upon recognition of a voice through analyzes fetch necessary information from the storage unit

When the voices are characterized, one or several spatialization methods, as mentioned earlier, are used to place different participants in the virtual room. The processing unit compares different characteristics and voices having most similar characteristics are placed as far apart as possible

The terms distance and far used in this description relate to a virtual room or space generated using sound reproducing means, such as speakers or headphones. The term participant as mentioned in this description relates to a user of the system of the invention and may be one of a listener or a talker.

It should be noted that the voice of one person may be influenced by, for example communication device/network quality and although if a profile is stored it may be analyzed each time a conference is set up.

The invention may also be used in a communication device as illustrated in one exemplary embodiment in Fig 5.

As shown in Fig 5, an exemplary device 500 may include a housing 510, a display 511, control buttons 512, a keypad 513, communication portion 514, a power source 515, a microprocessor 516 (or data processing unit), a memory unit 517, a microphone 518 and a speaker 520. The housing 510 may protect the components of device 500 from outside elements. Display 511 may provide visual information to the user. For example, display 511 may provide information regarding incoming or outgoing calls, media, games, phone books, the current time, a web browser etc. Control buttons 512 may permit the user to interact with device to cause device to perform one or more operations. Keypad 513 may include a standard telephone keypad. The microphone 518 is used to receive ambient sound, such as the voice of the user.

The communication portion comprises parts (not shown) such as a receiver, a transmitter, (or a transceiver), an antenna 519 etc, for establishing and performing communication with one or several communication networks 540.
The microphone and the speaker can be substituted with a headset comprising microphone and earphones.

Thus, when the communication device is used as a receiver in a conferencing application, the processing unit is configured to execute the instructions, which generate a spatial positioning of the participants' voices as described earlier.

It should be noted that the word "comprising" does not exclude the presence of other elements or steps than those listed and the words "a" or "an" preceding an element do not exclude the presence of a plurality of such elements. It should further be noted that any reference signs do not limit the scope of the claims, that the invention may be implemented at least in part by means of both hardware and software, and that several "means", "units" or "devices" may be represented by the same item of hardware.

A "device," as the term is used herein, is to be broadly interpreted to include a radiotelephone having ability for Internet/intranet access, web browser, organizer, calendar, a camera (e.g., video and/or still image camera), a sound recorder (e.g., a microphone), and/or global positioning system (GPS) receiver, a personal communications system (PCS) terminal that may combine a cellular radiotelephone with data processing, a personal digital assistant (PDA) that can include a radiotelephone or wireless communication system, a laptop, a camera (e.g., video and/or still image camera) having communication ability, and any other computation or communication device capable of transceiving, such as a personal computer, a home entertainment system, a television, etc.

The above mentioned and described embodiments are only given as examples and should not be limiting to the present invention. Other solutions, uses, objectives, and functions within the scope of the invention as claimed in the below described patent claims should be apparent for the person skilled in the art.
CLAIMS

1. An arrangement in a multi-party conferencing system, the arrangement comprising a processing unit, said processing unit being configured to:
   - process at least each received signal corresponding to a voice of a participant in a multi-party conferencing,
   - extract at least one characteristic parameter for said voice of each particular participant,
   - compare results of the at least one characteristic parameter of at least each particular participant to determine a similarity in said at least one characteristic parameter, and
   - generate a virtual position for each participant voice, using spatial positioning, where a position of voices having similar characteristics is arranged distanced from each other in a virtual space.

2. The arrangement of claim 1, wherein said spatializing is one or several of a virtual sound-source positioning (VSP) method and a sound-field capture (SFC) method.

3. The arrangement of claim 1, further comprising a memory unit for storing sound characteristics associated with a particular participant profile.

4. A computer for handling a multi-party conferencing, said computer comprising:
   - a unit for receiving signals corresponding to a particular conferee voices,
   - a unit configured to analyze said signals,
   - a unit configured to extract at least one characteristic parameter each signal,
   - a unit configured to compare said at least one characteristic parameter of at least each conferee to determine a degree of similarity in said at least one characteristic parameter, and
   - a unit configured to generate a virtual position for each conferee voice using spatial positioning, where a audible position of voices having similar characteristics is arranged distanced from each other in a virtual space.

5. The computer of claim 4, further comprising a communication interface to a communication network.
6. A communication device able of handling a multi-party conferencing, said communication device comprising:
   • a communication portion,
   » a sound input unit,
   • a sound output unit,
   • a unit configured to analyze a signal received from said communication network, said signal corresponding to voices of a plurality of conferee;
   β a unit configured to extract at least one characteristic parameter for said signal,
   • a unit configured to compare said at least one characteristic parameter to find a similarity in said at least one characteristic parameter, and
   • a unit configured to generate a virtual position for each conferee voice using spatial positioning, where an audible position of voices having similar characteristics is arranged distanced from each other in a virtual space for output using said sound output unit.

7. A method in a multi-party conferencing system, the method comprising:
   • analysing signal relating to one or several participant voices,
   • processing at least each received signal and extracting at least one characteristic parameter for voice of each participant based on said signal,
   • comparing result of the characteristic parameters to find similarity in said characteristic parameters, and
   • generating a virtual position for each participant voice through spatial positioning, in which position of voices having similar characteristics is arranged distanced from each other in a virtual space.
Fig. 1
Fig. 3
Fig. 4

401 Analyze participant voices

402 Determine characteristics

403 Compare characteristics

404 Determine similarity

405 Spatial position voices
A. CLASSIFICATION OF SUBJECT MATTER

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)
H04M

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

Electronic database consulted during the international search (name of database and, where practical, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>US 7 489 773 B1 (MORAN THOMAS [IE] ET AL) 10 February 2009 (2009-02-10) column 1, line 14 - line 22, column 2, line 40 - line 65, column 3, line 47 - line 54, column 5, line 39 - line 44, column 6, line 65 - column 7, line 62 figures 1,3-6</td>
<td>1-7</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C

See patent family annex

* Document defining the general state of the art which is not considered to be of particular relevance

† Document reporting on or after the International filing date

‡ 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)

Ω Document referring to an oral disclosure, use exhibition or other means

† Document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search

18 January 2010

Date of mailing of the international search report

25/01/2010

Authorized officer

Nash, Michael
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>US 7489773 B1 10-02-2009</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td>US 2009080632 A1 26-03-2009</td>
<td>NONE</td>
<td></td>
</tr>
</tbody>
</table>