

(19)



(11)

EP 2 984 763 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
21.02.2018 Bulletin 2018/08

(51) Int Cl.:
H04M 9/08 ^(2006.01) **G10L 15/20** ^(2006.01)
G10L 15/22 ^(2006.01) **G10L 21/0208** ^(2013.01)

(21) Application number: **13881816.6**

(86) International application number:
PCT/US2013/036088

(22) Date of filing: **11.04.2013**

(87) International publication number:
WO 2014/168618 (16.10.2014 Gazette 2014/42)

(54) **SYSTEM FOR AUTOMATIC SPEECH RECOGNITION AND AUDIO ENTERTAINMENT**

SYSTEM FÜR AUTOMATISCHE SPRACHERKENNUNG UND AUDIOUNTERHALTUNG

SYSTÈME POUR LA RECONNAISSANCE VOCALE AUTOMATIQUE ET LES DIVERTISSEMENTS AUDIO

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

- **WOLFF, Tobias**
89077 Ulm (DE)
- **SUHADI**
89077 Ulm (DE)

(43) Date of publication of application:
17.02.2016 Bulletin 2016/07

(74) Representative: **South, Nicholas Geoffrey et al**
A.A. Thornton & Co.
10 Old Bailey
London EC4M 7NG (GB)

(73) Proprietor: **Nuance Communications, Inc.**
Burlington, MA 01803 (US)

(72) Inventors:
• **BUCK, Markus**
88400 Biberach (DE)
• **HAULICK, Tim**
89143 Blaubeuren (DE)

(56) References cited:
US-A1- 2004 264 686 US-A1- 2006 002 547
US-A1- 2007 280 472 US-A1- 2007 280 486
US-A1- 2008 232 617 US-A1- 2009 022 330
US-A1- 2009 262 950 US-A1- 2009 316 923

EP 2 984 763 B1

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

TECHNICAL FIELD

[0001] The present invention relates to selective playback channel activation for an audio playback system in a media system.

BACKGROUND ART

[0002] Automatic speech recognition (ASR) is often used to accept user commands for operating a voice-controlled device. When the device includes an audio playback system, background noise, interfering sounds, and loudspeaker feedback can degrade the ASR's performance. Conventional methods for remedying these distortions are often computationally expensive. US 2007/280486 A1 discloses a vehicle communication system comprising: a plurality of microphones adapted to detect speech signals of different vehicle passengers, a mixer for combining the audio signal components of the different microphones to a resulting speech output signal, and a weighting unit for determining the weighting of the audio signal components for the resulting speech output signal, where the weighting unit determines the weighting of the signal components based upon non-acoustical information about the presence of a vehicle passenger.

[0003] US 2009/022330 A1 discloses a method for processing sound signals in a vehicle multimedia system. A sound signal is recorded by a microphone. A voice signal component resulting from reproducing a voice signal and an audio signal component resulting from reproducing an audio signal are compensated in the recorded sound signal. A system for processing sound signals and a vehicle multimedia system are also disclosed.

SUMMARY

[0004] Some embodiments of the present disclosure are directed to a device for providing different levels of sound quality in an audio entertainment system. The device includes a speech enhancement system with a reference signal modification unit and a plurality of acoustic echo cancellation filters. Each acoustic echo cancellation filter is coupled to a playback channel. The device includes an audio playback system with loudspeakers. Each loudspeaker is coupled to a playback channel. At least one of the speech enhancement system and the audio playback system operates according to a full sound quality mode and a reduced sound quality mode. In the full sound quality mode, all of the playback channels contain non-zero output signals. In the reduced sound quality mode, a first subset of the playback channels contains non-zero output signals and a second subset of the playback channels contains zero output signals.

[0005] In some embodiments, the audio playback system activates all the playback channels during the full sound quality mode and activates a subset of the play-

back channels during the reduced sound quality mode. The reduced sound quality mode can be a stereo mode or a mono mode. In some embodiments, the speech signal modification unit of the speech enhancement system outputs the non-zero output signals to all of the acoustic echo cancellation filters during the full sound quality mode, and outputs the non-zero output signals to a first subset of the acoustic echo cancellation filters and the zero output signals to a second subset of the acoustic echo cancellation filters during the reduced sound quality mode. The speech signal modification unit can output the zero output signals instead of non-zero output signals intended for the second subset of acoustic echo cancellation filters during the reduced sound quality mode.

[0006] In many embodiments, the speech enhancement system deactivates a subset of the acoustic echo cancellation filters during the reduced sound quality mode. The subset of the acoustic echo cancellation filters corresponds to the second subset of the playback channels that contain zero output signals.

[0007] In various embodiments, the device includes an automatic speech recognition (ASR) system that operates according to a broad listening mode and a selective listening mode. The ASR system can be coupled to the audio playback system so that i) the at least one of the speech enhancement system and the audio playback system operates in the full sound quality mode when the ASR system operates in the broad listening mode, and ii) the at least one of the speech enhancement system and the audio playback system operates in the reduced sound quality mode when the ASR operates in the selective listening mode. A subset of microphones coupled to the ASR system can be activated during the full sound quality mode and all of the microphones can be activated during the reduced sound quality mode.

[0008] In some embodiments, the ASR system sends the at least one of the speech enhancement system and the audio playback system a signal indicative of a mode to which the ASR system is switching. In various embodiments, the ASR system sends the at least one of the speech enhancement system and the audio playback system a signal indicative of a mode in which the ASR system is operating. In many embodiments, the at least one of the speech enhancement system and the audio playback system switches between the full and reduced sound quality modes in response to a switching cue. The switching cue can include one or more mode switching words from a speech input, one or more dialog states, or one or more visual cues from the possible speakers.

[0009] Some embodiments of the present disclosure are directed to a computer program product encoded in a non-transitory computer-readable medium for providing different levels of sound quality in an audio entertainment system. The product includes program code for operating at least one of a speech enhancement system coupled to playback channels and an audio playback system coupled to the playback channels. Execution of the program code causes receipt of a signal indicative of a

mode of the at least one of a speech enhancement system and an audio playback system, operation of the at least one of a speech enhancement system and the audio playback system so that all of the playback channels contain non-zero output signals in a full sound quality mode, and operation of the at least one of a speech enhancement system and the audio playback system so that a first subset of the playback channels contain non-zero output signals and a second subset of the playback channels contain zero output signals in a reduced sound quality mode.

[0010] In some embodiments, the product includes program code for deactivating one or more playback channels in the audio playback system when the mode is the reduced sound quality mode. In many embodiments, the product includes program code for switching the audio playback system to a stereo mode or a mono mode. In various embodiments, the product includes program code for transmitting non-zero output signals to all acoustic echo cancellation filters of the speech enhancement system in the full sound quality mode; and program code for transmitting non-zero output signals to a first subset of the acoustic echo cancellation filters and zero output signals to a second subset of the acoustic echo cancellation filters in the reduced sound quality mode.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The foregoing features of embodiments will be more readily understood by reference to the following detailed description, taken with reference to the accompanying drawings, in which:

Figure 1 shows an arrangement for using a voice controlled user interface with the system of the present invention.

Figure 2 shows a state diagram for a user dedicated voice controlled user interface according an embodiment of the present invention.

Figures 3 and 4 show systems with selective playback channel activation for an audio playback system coupled to an automatic speech recognition system.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

[0012] The present disclosure is directed towards optimizing the computational load of a media system that has an audio playback system and a speech enhancement system coupled to one another. In some embodiments, the media system has a user dedicated automatic speech recognition (ASR) system coupled to the speech enhancement system. The speech enhancement system is also coupled to multiple microphones, which enable the media system to detect sound over longer distances, a wider span, or both.

[0013] When the media system is not detecting

speech, a subset of the microphones may be active (e.g., one or two microphones). In this mode, the media system detects sound in any direction from the active microphone(s). When the media system detects speech, the speech enhancement system's algorithm is spatially steered towards the assumed direction of the speaking user (e.g., beamforming). The ASR system limits the voice control functionality to one selected user rather than to any user who happens to be in the vicinity. This may be based, for example, on a user speaking a special activation word that invokes the user limiting functionality. In another example, the limited voice control functionality is invoked when a user selects a push-to-talk (PTT) button, or any other method. The ASR system may then remain dedicated to the designated user until a specific dialog ends or some other mode switching event occurs. While operating in this mode, due to the beamforming, the system does not respond to any spoken inputs from other users (e.g., interfering speakers).

[0014] The audio playback system includes multiple loudspeakers, each of which is coupled to a playback channel. Each of the microphones also detects sound from the loudspeakers, and the acoustic echo degrades the performance of the ASR system. To alleviate or cancel this effect, the speech enhancement system applies acoustic echo cancellation (AEC) filters to the loudspeaker audio signals and subtracts the filtered signals from the microphone signals. For each sound signal received through a microphone, the speech enhancement system accounts for the audio signals from each of the loudspeakers. Because each microphone-loudspeaker pair defines a unique acoustic path, a corresponding AEC filter is needed to reduce the impact of that particular loudspeaker's audio signal. To account for all of the microphone-loudspeaker acoustic paths, the total number of AEC filters for the speech enhancement system to apply is the product of the number of microphones, M , and the number of independent playback channels, L , associated with the loudspeakers.

[0015] Although AEC filters can effectively reduce the impact of loudspeaker feedback on microphone signals, they are computationally complex and time consuming. When a media system has numerous microphones and/or loudspeakers, the filters' processing times may result in unacceptable latencies for a user issuing commands through a voice controlled interface.

[0016] In the present disclosure, to maintain acceptable latencies, the media system activates a subset of the playback channels in the audio playback system when it is selectively listening for voice commands such that the audio playback system outputs sound of reduced sound quality. In this reduced sound quality mode, all of the system's microphones are active and the speech enhancement system's algorithm is spatially steered towards the assumed direction of the speaker, for beamforming or other multi-microphone processing methods. Further, in this mode, since only a subset of the loudspeakers is active, fewer loudspeakers are adding

acoustic echo to the microphone sound signals. In light of the reduced number of acoustic sound paths associated with the subset of active loudspeakers, fewer AEC filters are needed to remove loudspeaker feedback from the microphone sound signals. During this mode, the reduced number of active loudspeakers diminishes the user's experience of the sound system. Nevertheless, because the user is operating the media system at this time, the reduced sound quality may well be acceptable to the user.

[0017] When the user finishes giving voice commands, the media system enters a different mode, e.g., a full sound quality mode. During this mode, the user is presumably enjoying the content of the media system instead of issuing voice commands. Thus, the media system devotes fewer resources to detecting voice commands, or potential voice commands. Only a subset of the microphones is active and the subset listens broadly for a voice command, such as an activation word. Further, the system activates all of the playback channels and their associated loudspeakers to maximize the user's sound experience.

[0018] Figure 1 shows an arrangement for a voice controlled user interface 100 used in a media system that has an audio playback system, a user dedicated automatic speech recognition (ASR) system, and a speech enhancement system coupled to one another. The audio playback system and/or speech enhancement system includes at least two different operating modes. In some embodiments, in full sound quality mode, all of the playback channels associated with the audio playback system's loudspeakers are active, thereby providing an enhanced user sound experience. In some embodiments, in full sound quality mode, the speech enhancement system outputs received sound signals to all of its acoustic echo cancellation filters. In some embodiments, the voice controlled user interface 100 also includes at least two different operating modes. A broad listening mode of the interface 100 can be coupled to the full sound quality mode of the audio playback system and/or speech enhancement system. Thus, in some embodiments, whenever the interface 100 operates in the broad listening mode, the audio playback system and/or speech enhancement system operates in the full sound quality mode.

[0019] In a broad listening mode, the voice controlled user interface 100 broadly accepts speech inputs via one or more of the microphones 103, without any spatial filtering, from any one of multiple speakers 102 in a room 101. In this mode, the voice controlled user interface 100 uses a limited broad mode recognition vocabulary that includes a selective mode activation word.

[0020] When the voice controlled user interface 100 detects the activation word, or receives an activation signal from a push-to-talk (PTT) button, it enters a selective listening mode that uses spatial filtering to limit speech inputs to a specific speaker 102 in the room 101 using an extended selective mode recognition vocabulary. For

example, the selected specific speaker may use the voice controlled user interface 100 in the selective listening mode following a dialog process to control one or more devices such as a television 105 and/or a computer gaming console 106.

[0021] The selective listening mode may be coupled to a reduced sound quality mode of the audio playback system and/or speech enhancement system. Thus, in some embodiments, whenever the interface 100 switches to the selective listening mode, the audio playback system and/or speech enhancement system switches to the reduced sound quality mode. In some embodiments, in the reduced sound quality mode, a subset of the playback channels associated with the audio playback system's loudspeakers is active. In some embodiments, in the reduced sound quality mode, the speech enhancement system outputs received zero signals to some of its acoustic echo cancellation filters in lieu of sound signals that were intended for the filters. Since the speech enhancement system thus does not apply all of its AEC filters, latency for the ASR system is reduced such that the media system responds to user voice commands.

[0022] In various embodiments, the components of the media system can switch modes based on different switching cues: dialog-state, certain activation words, visual gestures, activation of a push-to-talk (PTT) button. The different listening modes may also use different recognition vocabularies, for example, a limited vocabulary in broad listening mode and a larger recognition vocabulary in selective listening mode. To limit the speech inputs to a specific speaker, the system may use acoustic speaker localization and/or video processing means to determine speaker position.

[0023] In some embodiments, potential users are detected by means of image processing using images from one or more cameras. Image processing may rely on detection of one or more user cues to determine and select the dedicated user, for example, gesture recognition, facial recognition, etc. Based on the results of such user selection, the steering direction of the acoustic spatial filter can be controlled, continuing to rely on ongoing visual information. User feedback (via a GUI) can be given to identify the direction and/or identity of the selected dedicated user, for example, to indicate the spatial steering direction of the system. Further, activation gestures may be used to switch the media system from a broad listening mode to a selective listening mode, a full sound quality mode to a reduced sound quality mode, or vice versa.

[0024] Figure 2 shows a state diagram for the operation of the voice controlled user interface 100. Starting in broad listening mode 205 on the left side of Figure 2, the closed arc on the far left shows that the voice controlled user interface 100 stays in broad listening mode 205 so long as the activation word is not detected and a specific speaker has not been identified. The arc on top of Figure 2 shows the transition from broad listening mode 205 to selective listening mode 210 that occurs when the voice

controlled user interface 100 detects that the activation word has been spoken and a specific speaker is successfully identified. When in selective listening mode 210, the closed arc on the far right side of Figure 2 shows that the voice controlled user interface 100 stays in selective listening mode 210 so long as the location of the specific speaker is known, until either a dialog process with the speaker ends or some other event changes the mode back to broad listening mode 205 as shown by the arc across the bottom of Figure 2.

[0025] Figures 3 and 4 show systems with selective playback channel activation for an audio playback system coupled to an automatic speech recognition system. Figure 3 shows a media system 300 with a plurality of microphones 305, a speech enhancement system 310, an automatic speech recognition system 315, an audio playback system 320, and a plurality of loudspeakers 325. The speech enhancement system 310 includes a reference signal modification unit 350, a splitter 355, a plurality of AEC filters 360, a plurality of adders/subtractors 365, and a beamforming and noise reduction unit 370.

[0026] In many embodiments, the total number of playback channels, L , in the audio playback system 320 is equal to the number of loudspeakers 324. In various embodiments, the total number of AEC filters 360 is equal to the product of the number of independent playback channels, L , and the number of microphones 305, M , i.e., $M \cdot L$.

[0027] In operation, the audio playback system 320 receives audio signals to send to the loudspeakers 325. In various examples, the audio signals are for a television show, a video game being played on a media console (not shown), music being streamed from an Internet radio station, audio for a movie being watched on a home entertainment system, or any other audio, as would be understood by one of ordinary skill in the art.

[0028] In some embodiments, during the full sound quality mode, the audio playback system 320 transmits the audio signals over all of the playback channels. In particular, the audio playback system 320 sends the audio signals to the reference signal modification unit 350. In some embodiments, the reference signal modification unit 350 decorrelates the audio signals so that the audio signals resulting from application of the AEC filters 360 would converge. Then, the reference signal modification unit 350 sends the audio signals to the splitter 355. In various embodiments, the splitter 355 directs the audio signals to different loudspeakers 325, thereby attaining a "surround sound" effect for users of the media system 300.

[0029] In some embodiments, during the full sound quality mode, the reference signal modification unit 350 also sends all of the audio signals to the AEC filters 360. In the full sound quality mode, a subset of the microphones 305 is active. Although the operation herein is described with respect to one active microphone in the full sound quality mode, more than one microphone may

be active, so long as at least one microphone remains inactive. The media system 300 sends the identity of the active microphone 305 to the speech enhancement system 310. The speech enhancement system 310 identifies the AEC filters corresponding to the acoustic paths between the active microphone 305 and each of the loudspeakers 325.

[0030] When the speech enhancement system 310 receives an audio signal on a playback channel, the system 310 applies the AEC filter 360 corresponding to the active microphone 250 and that playback channel to the audio signal. The adder/subtractor(s) 365 subtract the outputs of the applied AEC filters 360 from the microphone sound signal to reduce or remove effects of acoustic echo. The adder/subtractor(s) 365 sends the resulting sound signal to the beamforming and noise reduction unit 370. The unit 370 applies its algorithms to the sound signal before sending the final sound signal to the automatic speech recognition unit 315. In some embodiments, during the full sound quality mode, the unit 370 only applies its noise reduction algorithms to the sound signal.

[0031] The automatic speech recognition unit 315 interprets the sound signal to determine the user's command. The automatic speech recognition unit 315 sends the command to the media system 300, which operates the system 300 accordingly. In some embodiments, the command includes one or more activation words. In other embodiments, the system 300 detects a different switching cue, such as dialog-state, visual gestures, or activation of a push-to-talk (PTT) button.

[0032] In response, the media system 300 switches the user interface 100 from the broad listening mode to the selective listening mode. The media system 300 can also switch the audio playback system 420 and/or the speech enhancement system 410 from the full sound quality mode to the reduced sound quality mode. In some embodiments, a speech dialogue system (not shown) of the system 300 notifies the audio playback system 320 of the change in mode, whereas in other embodiments, the speech dialogue system notifies the speech enhancement system 310. The speech enhancement system 310 may direct the notification to the reference signal modification unit 350, although in some embodiments, the reference signal modification unit 350 may receive the notification directly.

[0033] In some embodiments, during the reduced sound quality mode, some of the playback channels are deactivated. In some embodiments, the audio playback system 320 changes to a stereo or mono mode in response to the notification from the speech dialogue system of the change in mode. For example, instead of outputting audio signals on all of the playback channels, the audio playback system 320 outputs signals on one (e.g., mono mode) or two (e.g., stereo mode) channels.

[0034] In other embodiments, the audio playback system 320 continues outputting audio signals on all of the playback channels. When the reference signal modification unit 350 receives the audio signals, the unit 350 may

block the audio signals on some of the channels. For example, the unit 350 may process audio signals on active channels and passes those signals to the splitter 355, loudspeakers 325, and/or AEC filters 260. The unit 350 may output zero signals on the remaining playback channels. The loudspeakers 325 that do receive audio signals output those signals. Because a subset of the loudspeakers 325 outputs sound, the overall sound experience is diminished compared to the sound experience when all of the loudspeakers 325 are active. Nevertheless, since the users are focusing on operating the media system 300 during this time, the sound experience may still be acceptable.

[0035] In some embodiments, the reference signal modification unit 350 sends audio signals on the active channels to the AEC filters 360. During the reduced sound quality mode, all of the microphones are active. In some embodiments, the speech enhancement system 310 identifies the AEC filters corresponding to the acoustic paths between each of the microphones 305 and each of the active loudspeakers 325. The speech enhancement system 310 applies the identified AEC filters to the corresponding audio signals on the active playback channels. In some embodiments, the speech enhancement system 310 deactivates AEC filters corresponding to the inactive playback channels. In other embodiments, the speech enhancement system 310 applies all of the AEC filters to outputs on the playback channels. Since some of the playback channels have outputting zero signals, computations for AEC filters applied to these channels are completed quickly as zero signals.

[0036] The microphones 305 detect sound in the media system's 300 environment, and each microphone 305 sends its sound signal to the speech enhancement system 310. The adder/subtractor(s) 365 subtract the outputs of the applied AEC filters 360 from the microphone sound signals to reduce or remove effects of acoustic echo. The adder/subtractor(s) 365 sends the resulting sound signals to the beamforming and noise reduction unit 370. The unit 370 applies its beam forming and noise reduction algorithms to the sound signals before sending the final sound signals to the automatic speech recognition unit 315.

[0037] In some situations, the media system 300 detects or determines a switching cue for switching from the selective listening mode to the broad listening mode, or from the reduced sound quality mode to the full sound quality mode. In some examples, the media system 300 may detect a change in a dialog-state, a deactivation word, or a visual gesture. In further examples, the system 300 may determine that the push-to-talk (PTT) button is no longer active. In various examples, the system 300 may determine that no further verbal or gesture commands have been received for a predetermined period of time. For example, the automatic speech recognition system 315 may determine that the microphone sound signals have not included any words in the recognition vocabulary applicable to the selective listening mode, for

that period of time. After this period of time elapses, the system 300 determines that the user no longer intends to operate the system 300.

[0038] The media system 300 switches from the selective listening mode back to the broad listening mode, and/or from the reduced sound quality mode to the full sound quality mode. In some embodiments, a speech dialogue system (not shown) of the system 300 notifies the audio playback system 320 of the change in mode, whereas in other embodiments, the speech dialogue system notifies the speech enhancement system 310. The audio playback system 320 and/or the speech enhancement system 310 activates all of the playback channels. In some embodiments, the speech enhancement system 310 activates all of the AEC filters. The media system 300 deactivates at least some of the microphones. The media system 300 remains in this mode until it detects another switching cue to enter the reduced sound quality mode.

[0039] Figure 4 shows a media system 400 with a plurality of microphones 405, a speech enhancement system 410, an automatic speech recognition system 415, an audio playback system 420, and a plurality of loudspeakers 425. The speech enhancement system 410 includes a unit 452 for correlating audio signals on the playback channels, a unit 453 for selecting a subset of the playback channels, a plurality of AEC filters 460, a plurality of adders/subtractors 465, and a beamforming and noise reduction unit 470. Operation of the media system 400 is similar to operation of the media system 300. However, in system 400, the unit 452 processes the audio signals from the audio playback system 420 to produce correlated audio signals on all of the channels. Because the signals are correlated, only one audio signal needs to be used for the AEC filters 460. The unit 453 selects one of the channels and passes its audio signal to the AEC filters 460. In some embodiments, the unit 453 selects more than one playback channel.

[0040] Embodiments of the invention may be implemented in whole or in part in any conventional computer programming language such as VHDL, SystemC, Verilog, ASM, etc. Alternative embodiments of the invention may be implemented as pre-programmed hardware elements, other related components, or as a combination of hardware and software components.

[0041] Embodiments can be implemented in whole or in part as a computer program product for use with a media system. Such implementation may include a series of computer instructions fixed either on a tangible medium, such as a computer readable medium (e.g., a diskette, CD-ROM, ROM, or fixed disk) or transmittable to a media system, via a modem or other interface device, such as a communications adapter connected to a network over a medium. The medium may be either a tangible medium (e.g., optical or analog communications lines) or a medium implemented with wireless techniques (e.g., microwave, infrared or other transmission techniques). The series of computer instructions embodies

all or part of the functionality previously described herein with respect to the system. Those skilled in the art should appreciate that such computer instructions can be written in a number of programming languages for use with many computer architectures or operating systems. Furthermore, such instructions may be stored in any memory device, such as semiconductor, magnetic, optical or other memory devices, and may be transmitted using any communications technology, such as optical, infrared, microwave, or other transmission technologies. It is expected that such a computer program product may be distributed as a removable medium with accompanying printed or electronic documentation (e.g., shrink wrapped software), preloaded with a media system (e.g., on system ROM or fixed disk), or distributed from a server or electronic bulletin board over the network (e.g., the Internet or World Wide Web). Of course, some embodiments of the invention may be implemented as a combination of both software (e.g., a computer program product) and hardware. Still other embodiments of the invention are implemented as entirely hardware, or entirely software (e.g., a computer program product).

[0042] Although various exemplary embodiments of the invention have been disclosed, it should be apparent to those skilled in the art that various changes and modifications can be made which will achieve some of the advantages of the invention without departing from the true scope of the invention.

Claims

1. A device for providing different levels of sound quality in an audio entertainment system, the device comprising:

a speech enhancement system (310, 410) with a reference signal modification unit (350) and a plurality of acoustic echo cancellation filters (360, 460), each acoustic echo cancellation filter being coupled to a playback channel; and an audio playback system (320, 420) with loudspeakers (325, 425), each loudspeaker being coupled to a playback channel, wherein at least one of the speech enhancement system and the audio playback system is configured to operate according to:

- i) a full sound quality mode, during which all of the playback channels contain non-zero output signals, and
- ii) a reduced sound quality mode, during which a first subset of the playback channels contains non-zero output signals and a second subset of the playback channels contains zero output signals

wherein the device is **characterized by** at least

one of the speech enhancement system and the audio playback system being configured to operate in the reduced sound quality mode when the device is selectively listening for a voice command.

2. The device of claim 1, wherein the audio playback system (320, 420) is configured to activate all the playback channels during the full sound quality mode and activates a subset of the playback channels during the reduced sound quality mode.

3. The device according to claim 2, wherein the reduced sound quality mode is a stereo mode or a mono mode.

4. The device of claim 1, wherein the speech signal modification unit of the speech enhancement system (310, 410) is configured to output the non-zero output signals to all of the acoustic echo cancellation filters (360, 460) during the full sound quality mode, and to output the non-zero output signals to a first subset of the acoustic echo cancellation filters and the zero output signals to a second subset of the acoustic echo cancellation filters during the reduced sound quality mode.

5. The device according to claim 4, wherein the speech signal modification unit is configured to output the zero output signals instead of non-zero output signals intended for the second subset of acoustic echo cancellation filters (360, 460) during the reduced sound quality mode.

6. The device according to claim 1, wherein the speech enhancement system (310) is configured to deactivate a subset of the acoustic echo cancellation filters (360, 460) during the reduced sound quality mode, the subset of the acoustic echo cancellation filters corresponding to the second subset of the playback channels that contain zero output signals.

7. The device according to claim 1, further comprising an automatic speech recognition, ASR, system (315, 415) that is configured to operate according to a broad listening mode and a selective listening mode, wherein the ASR system is coupled to the audio playback system (320, 420) so that

- i) the at least one of the speech enhancement system (310, 410) and the audio playback system is adapted to operate in the full sound quality mode when the ASR system operates in the broad listening mode, and
- ii) the at least one of the speech enhancement system and the audio playback system is adapted to operate in the reduced sound quality mode when the ASR operates in the selective listening

mode.

8. The device according to claim 7, wherein a subset of microphones (103, 305, 405) coupled to the ASR system (315, 415) are configured to be activated during the full sound quality mode and all of the microphones are configured to be activated during the reduced sound quality mode.
9. The device according to claim 7, wherein the ASR system (315, 415) is configured to send the at least one of the speech enhancement system (310, 410) and the audio playback system (320, 420) a signal indicative of: (i) a mode to which the ASR system is switching; or (ii) a mode in which the ASR system is operating.
10. The device according to claim 1, wherein the at least one of the speech enhancement system (310, 410) and the audio playback system (320, 420) is configured to switch between the full and reduced sound quality modes in response to a switching cue.
11. The device according to claim 10, wherein the switching cue includes one or more mode switching words from a speech input, one or more dialog states, or one or more visual cues from the possible speakers.
12. A computer program product encoded in a non-transitory computer-readable medium for providing different levels of sound quality in an audio entertainment system, the product comprising:

program code for operating at least one of a speech enhancement system (310, 410) coupled to playback channels and an audio playback system (320, 420) coupled to the playback channels, the program code adapted to:

receive a signal indicative of a mode of the at least one of a speech enhancement system and an audio playback system;
 operate the at least one of a speech enhancement system and the audio playback system so that all of the playback channels contain non-zero output signals in a full sound quality mode; and
 operate the at least one of a speech enhancement system and the audio playback system so that a first subset of the playback channels contain non-zero output signals and a second subset of the playback channels contain zero output signals in a reduced sound quality mode;
 wherein the program code is further **characterized by** being adapted to operate at least one of the speech enhancement sys-

tem and the audio playback system in the reduced sound quality mode when the audio entertainment system is selectively listening for a voice command.

13. The computer program product of claim 12, further comprising:

program code for deactivating one or more playback channels in the audio playback system (320, 420) when the mode is the reduced sound quality mode.

14. The computer program product of claim 12, further comprising:

program code for switching the audio playback system (320, 420) to a stereo mode or a mono mode.

15. The computer program product of claim 12, further comprising:

program code for transmitting non-zero output signals to all acoustic echo cancellation filters (360, 460) of the speech enhancement system (310) in the full sound quality mode; and
 program code for transmitting non-zero output signals to a first subset of the acoustic echo cancellation filters and zero output signals to a second subset of the acoustic echo cancellation filters in the reduced sound quality mode.

Patentansprüche

1. Vorrichtung zum Bereitstellen verschiedener Stufen von Klangqualität in einem Audio-Entertainment-System, die Vorrichtung umfassend:

ein Sprachverbesserungssystem (310, 410) mit einer Referenzsignalmodifizierungseinheit (350) und einer Vielzahl von akustischen Echokompensationsfiltern (360, 460), wobei jeder akustische Echokompensationsfilter mit einem Wiedergabekanal gekoppelt ist; und
 ein Audio-Wiedergabesystem (320, 420) mit Lautsprechern (325, 425), wobei jeder Lautsprecher mit einem Wiedergabekanal gekoppelt ist,
 wobei mindestens eines aus dem Sprachverbesserungssystem und dem Audio-Wiedergabesystem konfiguriert ist, gemäß Folgendem zu arbeiten:

i) einem Modus mit voller Klangqualität, während dem sämtliche Wiedergabekanäle Ausgangssignale ungleich Null enthalten,

- und
 ii) einen Modus mit herabgesetzter Klangqualität, während dem eine erste Teilmenge der Wiedergabekanäle Ausgangssignale ungleich Null enthält und eine zweite Teilmenge der Wiedergabekanäle Ausgangssignale gleich Null enthält,
- wobei die Vorrichtung **dadurch gekennzeichnet ist, dass** mindestens eines aus dem Sprachverbesserungssystem und dem Audio-Wiedergabesystem konfiguriert ist, in dem Modus mit herabgesetzter Klangqualität zu arbeiten, wenn die Vorrichtung selektiv nach einem Sprachbefehl horcht.
2. Vorrichtung nach Anspruch 1, wobei das Audio-Wiedergabesystem (320, 420) konfiguriert ist, sämtliche Wiedergabekanäle während des Modus mit voller Klangqualität zu aktivieren, und eine Teilmenge der Wiedergabekanäle während des Modus mit herabgesetzter Klangqualität zu aktivieren.
 3. Vorrichtung nach Anspruch 2, wobei der Modus mit herabgesetzter Klangqualität ein Stereo-Modus oder ein Mono-Modus ist.
 4. Vorrichtung nach Anspruch 1, wobei die Sprachmodifizierungseinheit des Sprachverbesserungssystems (310, 410) konfiguriert ist, während des Modus mit voller Klangqualität die Ausgangssignale ungleich Null an sämtliche akustische Echokompensationsfilter (360, 460) auszugeben und während des Modus mit herabgesetzter Klangqualität die Ausgangssignale ungleich Null an eine erste Teilmenge der akustischen Echokompensationsfilter und die Ausgangssignale gleich Null an eine zweite Teilmenge der akustischen Echokompensationsfilter auszugeben.
 5. Vorrichtung nach Anspruch 4, wobei die Sprachsignalmodifizierungseinheit konfiguriert ist, die Ausgangssignale gleich Null anstelle der für die zweite Teilmenge von akustischen Echokompensationsfiltern (360, 460) bestimmten Ausgangssignale ungleich Null während des Modus mit herabgesetzter Klangqualität auszugeben.
 6. Vorrichtung nach Anspruch 1, wobei das Sprachverbesserungssystems (310) konfiguriert ist, während des Modus mit herabgesetzter Klangqualität eine Teilmenge der akustischen Echokompensationsfilter (360, 460) zu deaktivieren, wobei die Teilmenge der akustischen Echokompensationsfilter der zweiten Teilmenge der Wiedergabekanäle, die Ausgangssignale gleich Null enthalten, entspricht.
 7. Vorrichtung nach Anspruch 1, ferner ein automatisches Spracherkennungs(ASR)-System (315, 415) umfassend, das konfiguriert ist, gemäß eines breiten Hörmodus und eines selektiven Hörmodus zu arbeiten, wobei das ASR-System mit dem Audio-Wiedergabesystem (320, 420) gekoppelt ist, sodass
 - i) das mindestens eines aus dem Sprachverbesserungssystem (310, 410) und dem Audio-Wiedergabesystem angepasst wird, in dem Modus mit voller Klangqualität zu arbeiten, wenn das ASR-System in dem breiten Hörmodus arbeitet, und
 - ii) das mindestens eines aus dem Sprachverbesserungssystem und dem Audio-Wiedergabesystem angepasst wird, in dem Modus mit herabgesetzter Klangqualität zu arbeiten, wenn das ASR-System in dem selektiven Hörmodus arbeitet.
 8. Vorrichtung nach Anspruch 7, wobei eine Teilmenge von mit dem ASR-System (315, 415) gekoppelter Mikrofone (103, 305, 405) konfiguriert ist, während des Modus mit voller Klangqualität aktiviert zu sein, und sämtliche Mikrofone konfiguriert sind, während des Modus mit herabgesetzter Klangqualität aktiviert zu sein.
 9. Vorrichtung nach Anspruch 7, wobei das ASR-System (315, 415) konfiguriert ist, dem mindestens einen aus dem Sprachverbesserungssystem (310, 410) und dem Audio-Wiedergabesystem (320, 420) ein Signal zu senden, welches Folgendes angibt: (i) einen Modus, in den das ASR-System umschaltet; oder (ii) einen Modus, in dem das ASR-System arbeitet.
 10. Vorrichtung nach Anspruch 1, wobei das mindestens eine aus dem Sprachverbesserungssystem (310, 410) und dem Audio-Wiedergabesystem (320, 420) konfiguriert ist, in Reaktion auf einen Umschaltunghinweis zwischen den Modi mit voller und herabgesetzter Klangqualität umzuschalten.
 11. Vorrichtung nach Anspruch 10, wobei der Umschaltunghinweis ein Wort oder mehrere Wörter für Modusumschaltung aus der Spracheingabe oder einen oder mehrere visuelle Hinweise von den möglichen Sprechern beinhaltet.
 12. In einem nicht flüchtigen computerlesbaren Medium codiertes Computerprogrammprodukt zum Bereitstellen verschiedener Stufen von Klangqualität in einem Audio-Entertainment-System, das Produkt umfassend:

Programmcode zum Betreiben mindestens eines aus einem mit Wiedergabekanälen gekoppelten Sprachverbesserungssystem (310, 410)

und einem mit den Wiedergabekanälen gekoppelten Audio-Wiedergabesystem (320, 420), wobei der Programmcode angepasst ist:

ein Signal zu empfangen, das einen Modus des mindestens einem aus einem Sprachverbesserungssystem und einem Audio-Wiedergabesystem angibt;

das mindestens eines aus einem Sprachverbesserungssystem und dem Audio-Wiedergabesystem zu betreiben, sodass in einem Modus mit voller Klangqualität sämtliche Wiedergabekanäle Ausgangssignale ungleich Null enthalten; und

das mindestens eine aus einem Sprachverbesserungssystem und dem Audio-Wiedergabesystem zu betreiben, sodass in einem Modus mit herabgesetzter Klangqualität eine erste Teilmenge der Wiedergabekanäle Ausgangssignale ungleich Null enthält und eine zweite Teilmenge der Wiedergabekanäle Ausgangssignale gleich Null enthält; wobei der Programmcode ferner **dadurch gekennzeichnet ist, dass** er angepasst ist, mindestens eines aus dem Sprachverbesserungssystem und dem Audio-Wiedergabesystem in dem Modus mit herabgesetzter Klangqualität zu betreiben, wenn das Audio-Entertainment-System selektiv nach einem Sprachbefehl horcht.

13. Computerprogrammprodukt nach Anspruch 12, ferner umfassend:

Programmcode zum Deaktivieren eines oder mehrerer Wiedergabekanäle in dem Audio-Wiedergabesystem (320, 420), wenn der Modus dem Modus mit herabgesetzter Klangqualität entspricht.

14. Computerprogrammprodukt nach Anspruch 12, ferner umfassend:

Programmcode zum Umschalten des Audio-Wiedergabesystems (320, 420) in einen Stereo-Modus oder einen Mono-Modus.

15. Computerprogrammprodukt nach Anspruch 12, ferner umfassend:

Programmcode zum Senden von Ausgangssignalen ungleich Null an sämtliche akustische Echokompensationsfilter (360, 460) des Sprachverbesserungssystems (310) in dem Modus mit voller Klangqualität; und Programmcode zum Senden von Ausgangssignalen ungleich Null an eine erste Teilmenge der akustischen Echokompensationsfilter und

Ausgangssignale gleich Null an eine zweite Teilmenge der akustischen Echokompensationsfilter in dem Modus mit herabgesetzter Klangqualität.

Revendications

1. Dispositif pour la fourniture de différents niveaux de qualité de son dans un système de divertissement audio, le dispositif comprenant :

un système d'amélioration vocale (310, 410) avec une unité de modification de signal de référence (350) et une pluralité de filtres d'annulation d'écho acoustique (360, 460), chaque filtre d'annulation d'écho acoustique étant couplé à un canal de reproduction ; et un système de reproduction audio (320, 420) avec des haut-parleurs (325, 425), chaque haut-parleur étant couplé à un canal de reproduction, où au moins un du système d'amélioration vocale et du système de reproduction audio est configuré pour fonctionner selon :

i) un mode de qualité audio intégrale, durant lequel tous les canaux de reproduction contiennent des signaux de sortie différents de zéro, et

ii) un mode de qualité de son réduite, durant lequel un premier sous-ensemble de canaux de reproduction contient es signaux de sortie différents de zéro et un second sous-ensemble de canaux de reproduction contient des signaux de sortie égaux à zéro

où le dispositif est **caractérisé par** au moins un du système d'amélioration vocale et du système de reproduction audio étant configurés pour fonctionner dans le mode de qualité de son réduite lorsque le dispositif est sélectivement à l'écoute d'une commande vocale.

2. Dispositif selon la revendication 1, dans lequel le système de reproduction audio (320, 420) est configuré pour activer tous les canaux de reproduction durant le mode de qualité intégrale et active un sous-ensemble des canaux de reproduction durant le mode de qualité de son réduite.

3. Dispositif selon la revendication 2, dans lequel le mode de qualité de son réduite est un mode stéréo ou un mode mono.

4. Dispositif selon la revendication 1, dans lequel l'unité de modification de signal vocal du système d'amélioration vocale (310, 410) est configurée pour sortir les signaux de sortie différents de zéro vers tous les

- 5 filtres d'annulation d'écho acoustique (360, 460) durant le mode de qualité de son intégrale, et de sortir les signaux de sortie différents de zéro vers un premier sous-ensemble de filtres d'annulation d'écho acoustique et les signaux de sortie égaux à zéro vers un second sous-ensemble de filtres d'annulation d'écho acoustique durant le mode de qualité de son réduite.
5. Dispositif selon la revendication 4, dans lequel l'unité de modification de signal vocal est configurée pour sortir les signaux de sortie égaux à zéro au lieu des signaux de sortie différents de zéro prévus pour le second sous-ensemble de filtres d'annulation d'écho acoustique (360, 460) durant le mode de qualité de son réduite.
6. Dispositif selon la revendication 1, dans lequel le système d'amélioration vocale (310) est configuré pour désactiver un sous-ensemble des filtres d'annulation d'écho acoustique (360, 460) durant le mode de qualité de son réduite, le sous-ensemble des filtres d'annulation d'écho acoustique correspondant au second sous-ensemble de canaux de reproduction qui contient des signaux de sortie égaux à zéro.
7. Dispositif selon la revendication 1, comprenant, en outre, un système de reconnaissance vocale automatique (ASR) (315, 415) qui est configuré pour fonctionner selon un mode d'écoute large et un mode d'écoute sélectif, où le système ASR est couplé au système de reproduction audio (320, 420) de sorte que
- i) l'au moins un du système d'amélioration vocale (310, 410) et du système de reproduction audio est adapté pour fonctionner en mode de qualité de son intégrale lorsque le système ASR fonctionne en mode d'écoute large, et
- ii) l'au moins un du système d'amélioration vocale et du système de reproduction audio est adapté pour fonctionner en mode de qualité de son réduite lorsque le système ASR fonctionne en mode d'écoute sélectif.
8. Dispositif selon la revendication 7, dans lequel un sous-ensemble de microphones (103, 305, 405) couplés au système ASR (315, 415) sont configurés pour être activés durant le mode de qualité de son intégrale et tous les microphones sont configurés pour être activés durant le mode de qualité de son réduite.
9. Dispositif selon la revendication 7, dans lequel le système ASR (315, 415) est configuré pour envoyer à l'au moins un du système d'amélioration vocale (310, 410) et du système de reproduction audio (320, 420) un signal indicatif : (i) d'un mode auquel le système ASR commute ; ou (ii) d'un mode dans lequel le système ASR fonctionne.
10. Dispositif selon la revendication 1, dans lequel l'au moins un du système d'amélioration vocale (310, 410) et du système de reproduction audio (320, 420) est configuré pour commuter entre les modes de qualité de son intégrale et réduite en réponse à un indice de commutation.
11. Dispositif selon la revendication 10, dans lequel l'indice de commutation inclut un ou plusieurs mots de commutation de mode d'une entrée vocale, un ou plusieurs états de dialogue, ou un ou plusieurs indices visuels provenant des possibles locuteurs.
12. Produit de programme informatique encodé dans un moyen lisible par ordinateur non transitoire pour la fourniture de différents niveaux de qualité de son dans un système de divertissement audio, le produit comprenant :
- un code de programme pour faire fonctionner au moins un d'un système d'amélioration vocale (310, 410) couplé à des canaux de reproduction et d'un système de reproduction audio (320, 420) couplé aux canaux de reproduction, le code de programme adapté pour :
- recevoir un signal indicatif d'un mode de l'au moins un d'un système d'amélioration vocale et d'un système de reproduction audio ; faire fonctionner l'au moins un d'un système d'amélioration vocale et du système de reproduction audio de sorte que tous les canaux de reproduction contiennent des signaux de sortie différents de zéro dans un mode de qualité de son intégrale ; et faire fonctionner l'au moins un d'un système d'amélioration vocale et du système de reproduction audio de sorte qu'un premier sous-ensemble des canaux de reproduction contient des signaux de sortie différents de zéro et un second sous-ensemble des canaux de reproduction contient des signaux de sortie égaux à zéro dans un mode de qualité de son réduite ; où le code de programme est en outre **caractérisé en ce qu'il** est adapté pour faire fonctionner au moins un du système d'amélioration vocale et du système de reproduction audio dans le mode de qualité de son réduite lorsque le système de divertissement audio est sélectivement à l'écoute d'une commande vocale.
13. Produit de programme informatique selon la revendication 12, comprenant en outre :

un code de programme pour la désactivation d'un ou plusieurs canaux de reproduction dans le système de reproduction audio (320, 420) lorsque le mode est le mode de qualité de son réduite.

5

14. Produit de programme informatique selon la revendication 12, comprenant en outre :

un code de programme pour la commutation du système de reproduction audio (320, 420) à un mode stéréo ou un mode mono.

10

15. Produit de programme informatique selon la revendication 12, comprenant en outre :

15

un code de programme pour la transmission de signaux de sortie différents de zéro vers tous les filtres d'annulation d'écho acoustique (360, 460) du système d'amélioration vocale (310) dans le mode de qualité de son intégrale ; et
un code de programme pour la transmission de signaux de sortie différents de zéro vers un premier sous-ensemble de filtres d'annulation d'écho acoustique et de signaux de sortie égaux à zéro vers un second sous-ensemble de filtres d'annulation d'écho acoustique dans le mode de qualité de son réduite.

20

25

30

35

40

45

50

55

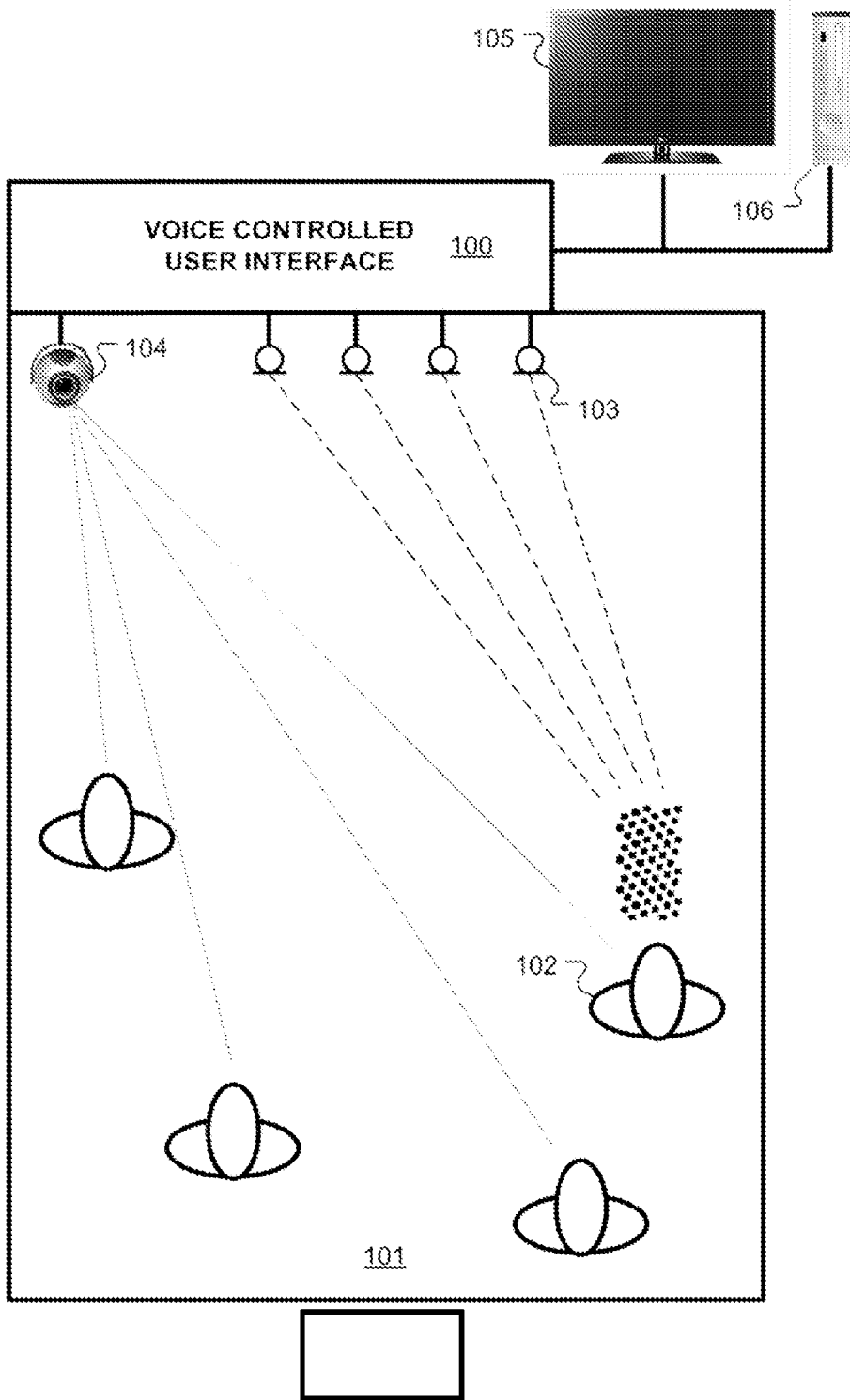


FIG. 1

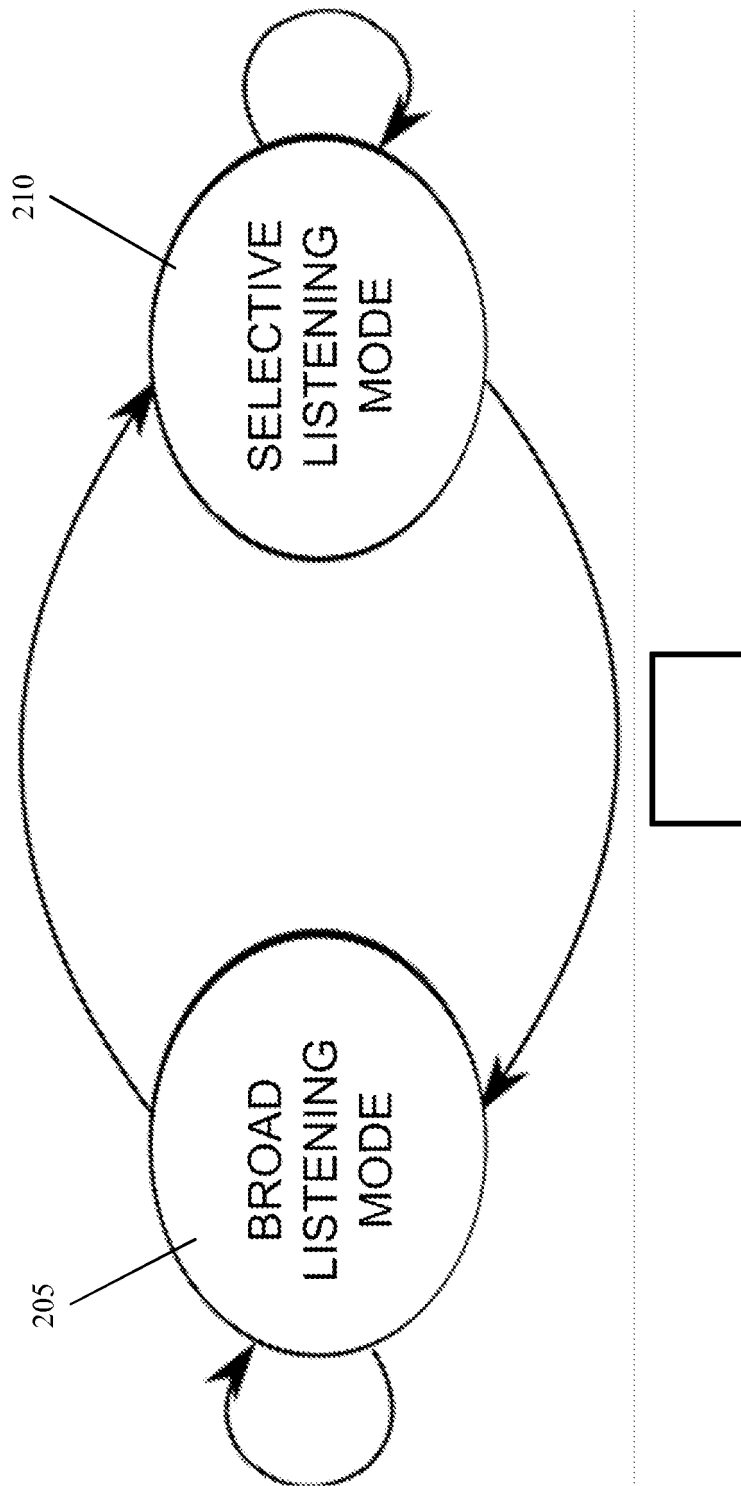


FIG. 2

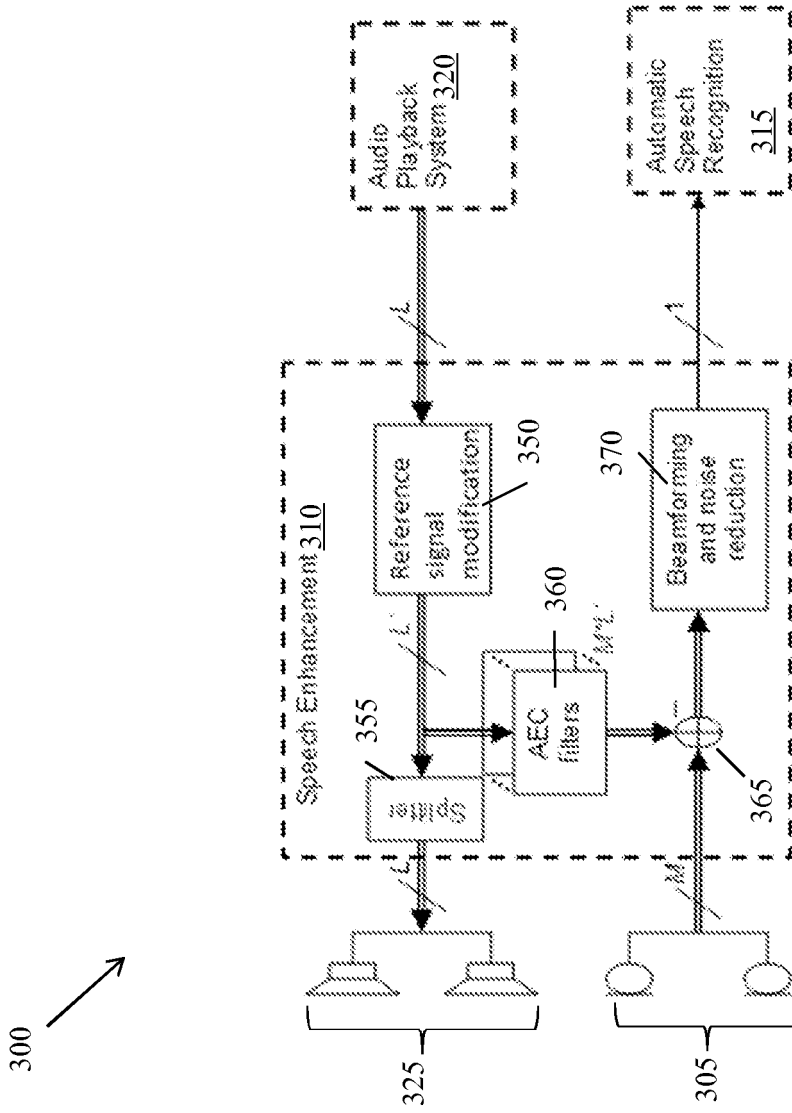


FIG. 3

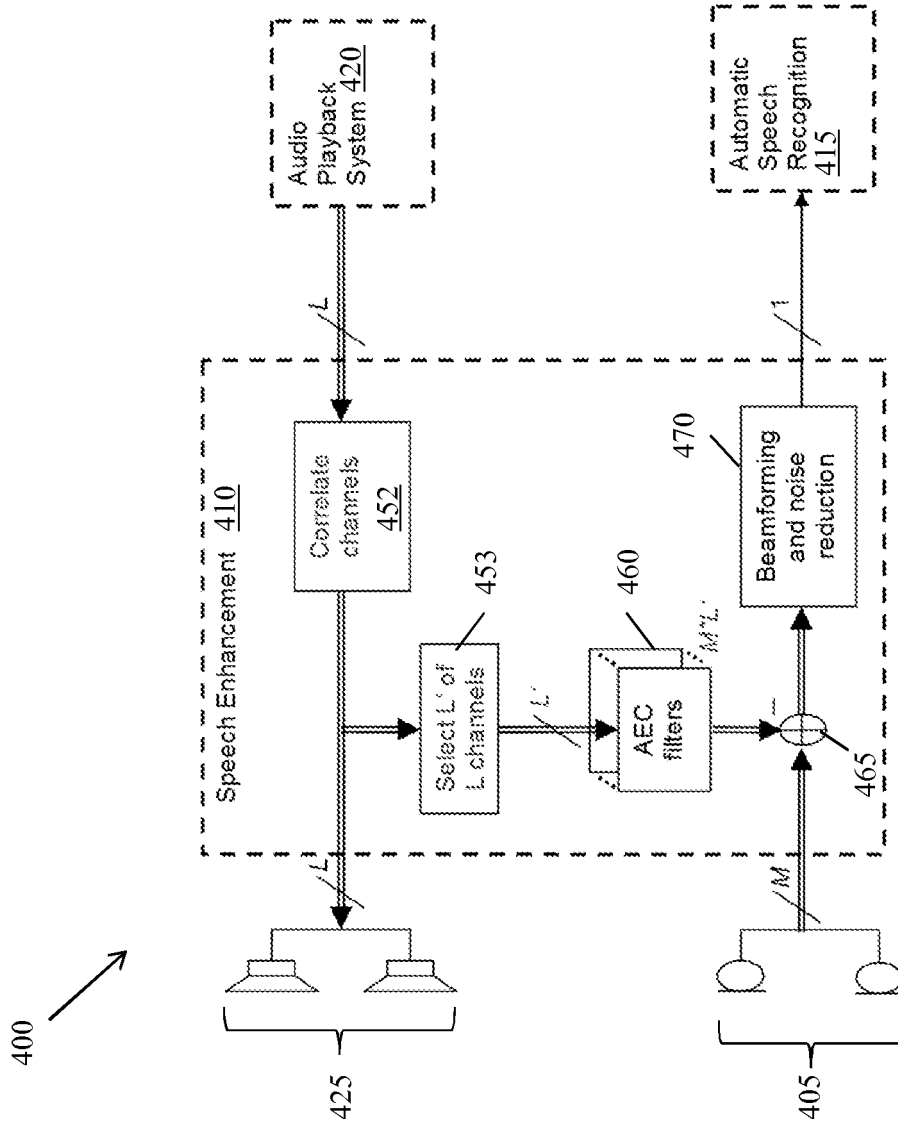


FIG. 4

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 2007280486 A1 [0002]
- US 2009022330 A1 [0003]