

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
25 November 2004 (25.11.2004)

PCT

(10) International Publication Number
WO 2004/102372 A2

- (51) International Patent Classification⁷: **G06F 3/033**
- (21) International Application Number:
PCT/US2004/008587
- (22) International Filing Date: 19 March 2004 (19.03.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
10/435,231 9 May 2003 (09.05.2003) US
- (71) Applicant: **INTEL CORPORATION (a Delaware Corporation)** [US/US]; 2200 Mission College Boulevard, Santa Clara, CA 95052 (US).
- (72) Inventors: **RAYKAR, Vikas**; 555 East EL Camino Real, #719, Sunnyvale, CA 94087 (US). **KOZINTSEV, Igor**; 4913 Anna Drive, San Jose, CA 95124 (US). **LIENHART, Rainer**; 2727 Forbes Avenue, Santa Clara CA 95051 (US).
- (74) Agents: **MALLIE, Michael, J.** et al.; Blakely, Sokoloff, Taylor & Zafman, 12400 Wilshire Boulevard, 7th Floor, Los Angeles, CA 90025 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 2004/102372 A2

(54) Title: THREE-DIMENSIONAL POSITION CALIBRATION
DISTRIBUTED COMPUTING PLATFORM

Error creating image for scaled gra
C:\spidi\tmp\fmt-source-dwg.g4

(57) Abstract: A first computing device transmitting a wireless signal to a second and third computing devices, the signal requesting an actuator of the second computing device generate an acoustic signal to be received by a sensor of the third computing device, wherein the actuator and sensor are unsynchronized. The first computing device computes, based on an time estimate for the acoustic signal to travel from the actuator of the second computing device to the sensor of the third computing device, a physical location of the actuator of the second computing device and the sensor of the third computing device.

THREE-DIMENSIONAL POSITION CALIBRATION OF AUDIO SENSORS AND ACTUATORS ON A DISTRIBUTED COMPUTING PLATFORM

TECHNICAL FIELD

[0001] Embodiments described herein relate to position calibration of audio sensors and actuators in a distributed computing platform.

BACKGROUND

[0002] Many emerging applications like multi-stream audio/video rendering, hands free voice communication, object localization, and speech enhancement, use multiple sensors and actuators (like multiple microphones/cameras and loudspeakers/displays, respectively). However, much of the current work has focused on setting up all the sensors and actuators on a single platform. Such a setup would require a lot of dedicated hardware. For example, to set up a microphone array on a single general purpose computer, would typically require expensive multichannel sound cards and a central processing unit (CPU) with larger computation power to process all the multiple streams.

10

[0003] Computing devices such as laptops, personal digital assistants (PDAs), tablets, cellular phones, and camcorders have become pervasive. These devices are equipped with audio-visual sensors (such as microphones and cameras) and actuators (such as loudspeakers and displays). The audio/video sensors on different devices can be used to form a distributed network of sensors. Such an ad-hoc network can be used to capture different audio-visual scenes (events such as business meetings, weddings, or public

15

events) in a distributed fashion and then use all the multiple audio-visual streams for an emerging applications. For example, one could imagine using the distributed microphone array formed by laptops of participants during a meeting in place of expensive stand alone speakerphones. Such a network of sensors can also be used to detect, identify, locate and track stationary or moving sources and objects.

[0004] To implement a distributed audio-visual I/O platform, includes placing the sensors, actuators and platforms into a space coordinate system, which includes determining the three-dimensional positions of the sensors and actuators.

BRIEF DESCRIPTION OF DRAWINGS

[0005] **Figure 1** illustrates a schematic representation of a distributed computing platform in accordance with one embodiment.

[0006] **Figure 2** is a flow diagram describing the process of generating position calibration for audio sensors and actuators in accordance with one embodiment.

[0007] **Figure 3** illustrates a computation scheme to generate position coordinates.

DETAILED DESCRIPTION

[0008] Embodiments of a three-dimensional position calibration of audio sensors and actuators in a distributed computing platform are disclosed. In the following description, numerous specific details are set forth. However, it is understood that embodiments may be practiced without these specific details. In other instances, well-known circuits, structures and techniques have not been shown in detail in order not to obscure the understanding of this description.

[0009] Reference throughout this specification to “one embodiment” or “an embodiment” indicate that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the phrases “in one embodiment” or “in an embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0010] **Figure 1** illustrates a schematic representation of a distributed computing platform consisting of a set of General-Purpose Computers (GPC) 102 – 106 (sometimes referred to as computing devices). GPC 102 is configured to be the master, and performs the location estimation. The GPCs (102 – 106) shown in **Figure 1** may include a personal computer (PC), laptop, PDA, tablet PC, or other computing devices. In one embodiment, each GPC is equipped with audio sensors 108 (e.g., microphones), actuators 110 (e.g., loudspeakers), and wireless communication capabilities 112. As is explained in more detail below, the sensors and actuators of the multiple GPCs are used to estimate their respective physical locations.

[0011] For example, in one embodiment, given a set of M acoustic sensors and S acoustic actuators in unknown locations, one embodiment estimates their respective three dimensional coordinates. The acoustic actuators are excited using a predetermined calibration signal such as a maximum length sequence or chirp signal, and the time of arrival (TOA) is estimated for each pair of the acoustic actuators and sensors. In one embodiment, the TOA for a given pair of microphone and speakers is defined as the time for the acoustic signal to travel from the speaker to the microphone. Measuring the TOA

and knowing the speed of sound in the acoustical medium, the distance between each acoustical signal source and the acoustical sensors can be calculated, thereby determining the three dimensional positions of the actuators and the sensors.

[0012] **Figure 2** is a flow diagram describing, in greater detail, the process of generating the three-dimensional position calibration of audio sensors and actuators in a distributed computing platform, according to one embodiment. The process described in the flow diagram of **Figure 2** periodically references the GPCs of the distributed computer platform illustrated in **Figure 1**.

[0013] In block 202, a first GPC 102, which may be considered the master GPC of the distributed platform, transmits a wireless signal to a surrounding set of GPCs in the distributed platform (the actual number of GPCs included in the distributed platform may vary based on implementation). The signal from the first GPC 102 includes a request that a specific actuator of one of the GPCs (e.g., second GPC 103) be excited to generate an acoustic signal to be received by the sensors of the surrounding GPCs (e.g., GPC 102, 104 - 106). In one embodiment, the initial wireless signal from the master GPC 102 identifies the specific actuator 110 to be excited.

[0014] In response to the signal from the master GPC 102, in block 204 the second GPC 103 excites the actuator 110 to generate an acoustic signal. In one embodiment, the acoustic signal may be a maximum length sequence or chirp signal, or another predetermined signal. In block 206, the second GPC 103 also transmits a first global time stamp to the other GPCs 104 - 106. In one embodiment, the global time stamp identifies when the second GPC 103 initiated the actuation of the actuator 110 for the second GPC

103. In block 208, the sensors of the GPCs 102, 104 - 106, receive the acoustic signal generated by the second GPC 103.

[0015] In block 210, the time for the acoustic signal to travel from the actuator 110 of the second GPC 103 to the respective sensors (hereinafter referred to as Time of Arrival (TOA)), is estimated. In one embodiment, the TOA for a given pair of a microphone and speaker is defined as the time taken by the acoustic signal to travel from the speaker to the microphone.

[0016] In one embodiment, the GPCs that receive the acoustic signal via their sensors, proceed to estimate the respective TOAs. In one embodiment, there exists a common clock in the distributed platform so that GPCs 102 - 106 are able to determine the time of arrival of audio samples captured by the respective sensors. As a result, the TOA can be estimated based on the difference between the first global time stamp issued by the second GPC 103 and the time of when the acoustic signal is received by a sensor.

[0017] Considering, however, that sensors are distributed on different platforms, the audio stream among the different GPCs are typically not synchronized in time (e.g., analog-to-digital and digital-to-analog converters of actuators and sensors of the different GPCs are unsynchronized). As a result, the estimated TOA does not necessarily correspond to the actual TOA. In particular, the TOA of the acoustic signal may include an emission start time, which is defined as the time after which the sound is actually emitted from the speaker (e.g., actuator 110) once the command has been issued from the respective GPC (e.g., GPC 103). The actual emission start time is typically never zero and can actually vary in time depending on the sound card and processor load of the respective GPC.

[0018] Therefore, to account for the variations in the emission start time, multiple alternatives may be used. For example, in one embodiment, if multiple audio input channels are available on the GPC exciting an actuator, then one of the output channels can be connected directly to one of the input channels forming a loop-back. Source emission start time can then be estimated for a given speaker, and can be globally transmitted to the other GPCs 102, 104 – 106 to more accurately determine the respective TOAs. Furthermore, in one embodiment, in the case of using the loop-back, the estimated emission start time will be included in the global time stamp transmitted by the respective GPC.

[0019] Once the TOAs for the acoustic signal have been estimated by the receiving GPCs 104-106, which may include accounting for the unknown emission start time as described above, the TOAs are transmitted to the master GPC 102. In an alternative embodiment, the TOAs can be computed by the master GPC 102, in which case each sensor of GPCs 104 – 106 generate a second global timestamp of when the acoustic signals arrived, respectively. In the alternative embodiment, the master GPC 102 uses the first global time stamp (identifying when the second GPC 103 initiated the actuation of the actuator 110) and the second global time stamps to estimate the TOAs for the respective pairs of actuators and sensors. In such as case, the master GPC 102 may also estimate the emission start time of the acoustic signal to estimate the TOAs.

[0020] In decision block 212, if additional actuators remain in the distributed platform, the processes of blocks 202 – 210 are repeated to have each of the actuators in the platform generate an acoustic signal to determine the TOAs with respective receiving sensors. In

an alternative embodiment, multiple separate actuators may be actuated in parallel, wherein the actuator signals are multiplexed by each actuator using a unique signal (e.g., different parameters for chirp or MLS signals). In the case of actuating the multiple separate actuators in parallel, the master GPC 102 identifies to each actuator a unique signal parameters to be used when exciting the actuator.

[0021] Once all of the TOAs for the different pairs of actuators and sensors have been computed and transmitted to the master GPC 102, in block 214 the master GPC 102 computes the coordinates of the sensors and the actuators. More specifically, as illustrated in the position computation scheme of **Figure 3**, in one embodiment the master GPC 102, utilizes a nonlinear least squares (NLS) computation 302 to determine the coordinates 304 of the actuators and/or sensors. In one embodiment, the NLS computation 302 considers the TOAs 306, the number of microphones 308 and the number of speakers 310 in the platform, along with an initial estimation 312 at the coordinates of the actuators and speakers. The actual computation used by the master GPC 102 to compute the coordinates of the actuators and sensors based on the TOAs may vary based on implementation. For example, in an alternative embodiment to compute the positions of sensors and actuators with unknown emission start times, the NLS procedure is used to jointly estimate the positions and the emission times. Emission times add extra S (number of actuators) variables to the computation procedure.

[0022] To provide the initial estimation as used by the NLS, several alternatives are available. For example, if an approximate idea of the microphone and speaker positions is available, then the initialization may be done manually. In another embodiment, the use of one or more cameras may provide a rough estimate to be used as the initial estimation.

[0023] An additional embodiment to generate an initial estimation includes assuming that microphones and speakers on a given computing platform are approximately at the same position, and given all estimates of the pairwise distances between the separate GPCs, a multidimensional scaling approach may be used to determine the coordinates from, in one embodiment, the Euclidean distance matrix. The approach involves converting the symmetric pairwise distance matrix to a matrix of scalar products with respect to some origin and then performing a singular value decomposition to obtain the matrix of coordinates. The matrix coordinates in turn, may be used as the initial guess or estimate of the coordinates for the respective GPCs, and the microphones and speakers located on them.

[0024] The techniques described above can be stored in the memory of one of the computing devices or GPCs as a set of instructions to be executed. In addition, the instructions to perform the processes described above could alternatively be stored on other forms of computer and/or machine-readable media, including magnetic and optical disks. Further, the instructions can be downloaded into a computing device over a data network in a form of compiled and linked version.

[0025] Alternatively, the logic to perform the techniques as discussed above, could be implemented in additional computer and/or machine readable media, such as discrete hardware components as large-scale integrated circuits (LSI's), application-specific integrated circuits (ASIC's), firmware such as electrically erasable programmable read-only memory (EEPROM's); and electrical, optical, acoustical and other forms of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); etc.

[0026] These embodiments have been described with reference to specific exemplary embodiments thereof. It will, however, be evident to persons having the benefit of this disclosure that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the embodiments described herein. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

CLAIMS

What is claimed is:

1. A method comprising:
 - an actuator of a first computing device generating an acoustic signal;
 - a sensor of a second computing device receiving the acoustic signal;
 - generating an estimate of a time for the acoustic signal to travel from the actuator to the sensor, wherein the sensor and actuator are unsynchronized; and
 - a third computing device computing, based on the time estimate, a physical location of at least one of a set including the sensor of the second computing device and the actuator of the first computing devices.

2. The method of claim 1, wherein the method further includes:
 - a sensor of a fourth computing device receiving the acoustic signal; and
 - generating a second estimate of a time for the acoustic signal to travel from the actuator to the sensor of the fourth computing device, wherein the actuator of the first computing device and the sensor of the fourth computing device are unsynchronized; and
 - the third computing device computing, based on the second estimate of time, a physical location of at least one of a set including the sensor of the fourth computing device and the actuator of the first computing device.

3. The method of claim 2, wherein the method further includes:
 - an actuator of the second computing device generating a second acoustic signal;
 - a sensor of a first computing device receiving the acoustic signal;

generating a third estimate of a time for the second acoustic signal to travel from the actuator of the second computing device to the sensor of the first computing device; and

the third computing device computing, based on the third estimate of time, a physical location of at least one of a set including the sensor of the first computing device and the actuator of the second computing device.

4. The method of claim 3, wherein the actuator is a speaker.

5. The method of claim 4, wherein the sensor is a microphone.

6. The method of claim 5, wherein the acoustic signals generated by the first and second computing devices are selected from a group comprising of maximum length sequence signal and a chirp signal.

7. The method of claim 1, further including:

Estimating an emission start time of when the acoustic signal was emitted from the actuator.

8. The method of claim 7, wherein the estimating of the emission start time includes the first computing device using a loopback device to estimate the emission start time; and using the emission start time to determine the estimate of a time for the acoustic signal to travel from the actuator of the first computing device to the sensor second computing device.

9. The method of claim 1, wherein the method further includes prior to the third computing device computing the physical locations, computing an initial estimation of the physical location of the sensor of the second computer and actuator of the first computer via a video modality.

10. The method of claim 1, wherein the method further includes prior to the third computing device computing the physical locations of the sensor of the second computer, computing an initial estimation of the physical location of the sensor of the second computer and actuator of the first computer via multidimensional scaling.

11. A method comprising:

a first computing device transmitting a wireless signal to a second and third computing devices, the signal requesting an actuator of the second computing device generate an acoustic signal to be received by a sensor of the third computing device, wherein the actuator and sensor are unsynchronized; and

the first computing device computing, based on a time estimate for the acoustic signal to travel from the actuator of the second computing device to the sensor of the third computing device, a physical location of at least one of a set including the actuator of the second computing device and the sensor of the third computing device.

12. The method of claim 11, wherein the method further includes:

the first computing device transmitting the wireless signal to the second and a fourth computing devices, the signal requesting the actuator of the second computing device generate the acoustic signal to be received by a sensor of the fourth computing device, wherein the actuator and the sensor of the fourth computing device are unsynchronized; and

the first computing device computing, based on an time estimate for the acoustic signal to travel from the actuator of the second computing device to the sensor of the fourth computing device, a physical location of at least one of a set including the actuator of the second computing device and the sensor of the fourth computing device.

13. The method of claim 12, wherein the method further includes:

the first computing device transmitting a second wireless signal to the second and third computing devices, the signal requesting an actuator of the third computing device generate an acoustic signal to be received by a sensor of the second computing device, wherein the actuator of the third computing device and a sensor of the second computing device are unsynchronized; and

the first computing device computing, based on an time estimate for the acoustic signal to travel from the actuator of the third computing device to the sensor of the second computing device, a physical location of at least one of a set including the actuator of the third computing device and the sensor of the second computing device.

14. The method of claim 13, wherein the actuator is a speaker.

15. The method of claim 14, wherein the sensor is a microphone.

16. The method of claim 15, wherein the acoustic signals to be generated by the second and third computing devices are selected from a group comprising of maximum length sequence signal, and a chirp signal.

17. The method of claim 11, further including: estimating an emission start time of when the acoustic signal was emitted from the actuator of the second computing device.

18. The method of claim 17, wherein estimating the emission start time includes the second computing device using a loopback device to estimate the emission start time; using emission start time to determine the estimate of a time for the acoustic signal to travel from the actuator of the second computing device to the sensor third computing device.

19. The method of claim 11, wherein the method further includes:
prior to the first computer determining a physical location of at least one of the group including the actuator of the second computing device and the sensor of the third computer, the first computer computing an initial estimation of the physical location of the actuator of the second computer and the sensor of the third computing device via video modality.

20. The method of claim 11, wherein the method further includes:
prior to the first computer determining a physical location of at least one of the group including the actuator of the second computing device and the sensor of the third computer, the first computer computing an initial estimation of the physical location of the actuator of the second computer and the sensor of the third computing device via video modality.

21. A machine readable medium having stored thereon a set of instructions, which when executed, cause the machine to perform a method comprising of:

an actuator of a first computing device generating an acoustic signal;
a sensor of a second computing device receiving the acoustic signal;
generating an estimate of a time for the acoustic signal to travel from the actuator to the sensor, wherein the sensor and actuator are unsynchronized; and
a third computing device computing, based on the time estimate, a physical location of at least one of a set including the sensor of the second computing device and the actuator of the first computing devices.

22. The machine readable medium of claim 21, wherein the method further includes:

a sensor of a fourth computing device receiving the acoustic signal; and
generating a second estimate of a time for the acoustic signal to travel from the actuator to the sensor of the fourth computing device, wherein the actuator of the first computing device and the sensor of the fourth computing device are unsynchronized; and
the third computing device computing, based on the second estimate of time, a physical location of at least one of a set including the sensor of the fourth computing device and the actuator of the first computing device.

23. The machine readable medium of claim 22, wherein the method further includes:

an actuator of the second computing device generating a second acoustic signal;
a sensor of a first computing device receiving the acoustic signal;
generating a third estimate of a time for the second acoustic signal to travel from the actuator of the second computing device to the sensor of the first computing device;
and

the third computing device computing, based on the third estimate of time, a physical location of at least one of a set including the sensor of the first computing device and the actuator of the second computing device.

24. The machine readable medium of claim 23, wherein the actuator is a speaker.

25. The machine readable medium of claim 24, wherein the sensor is a microphone.

26. The machine readable medium of claim 25, wherein the acoustic signals generated by the first and second computing devices are selected from a group comprising of maximum length sequence signal and a chirp signal.

27. The machine readable medium of claim 21, further including:
estimating an emission start time of when the acoustic signal was emitted from the actuator.

28. The machine readable medium of claim 27, wherein the estimating of the emission start time includes the first computing device using a loopback device to estimate the emission start time; and

using the emission start time to determine the estimate of a time for the acoustic signal to travel from the actuator of the first computing device to the sensor second computing device.

29. The machine readable medium of claim 21, wherein the method further includes:

prior to the third computing device computing the physical locations, computing an initial estimation of the physical location of the sensor of the second computer and actuator of the first computer via a video modality.

30. The machine readable medium of claim 21, wherein the method further includes:

prior to the third computing device computing the physical locations, computing an initial estimation of the physical location of the sensor of the second computer and actuator of the first computer via multidimensional scaling.

31. A machine readable medium having stored thereon a set of instructions, which when executed, cause the machine to perform a method comprising of:

a first computing device transmitting a wireless signal to a second and third computing devices, the signal requesting an actuator of the second computing device generate an acoustic signal to be received by a sensor of the third computing device, wherein the actuator and sensor are unsynchronized; and

the first computing device computing, based on a time estimate for the acoustic signal to travel from the actuator of the second computing device to the sensor of the third computing device, a physical location of at least one of a set including the actuator of the second computing device and the sensor of the third computing device.

32. The machine readable medium of claim 31, wherein the method further includes:

the first computing device transmitting the wireless signal to the second and a fourth computing devices, the signal requesting the actuator of the second computing device generate the acoustic signal to be received by a sensor of the fourth computing

device, wherein the actuator and the sensor of the fourth computing device are unsynchronized; and

the first computing device computing, based on an time estimate for the acoustic signal to travel from the actuator of the second computing device to the sensor of the fourth computing device, a physical location of at least one of a set including the actuator of the second computing device and the sensor of the fourth computing device.

33. The machine readable medium of claim 32, wherein the method further includes:

the first computing device transmitting a second wireless signal to the second and third computing devices, the signal requesting an actuator of the third computing device generate an acoustic signal to be received by a sensor of the second computing device, wherein the actuator of the third computing device and a sensor of the second computing device are unsynchronized; and

the first computing device computing, based on an time estimate for the acoustic signal to travel from the actuator of the third computing device to the sensor of the second computing device, a physical location of at least one of a set including the actuator of the third computing device and the sensor of the second computing device.

34. The machine readable medium of claim 33, wherein the actuator is a speaker.

35. The machine readable medium of claim 34, wherein the sensor is a microphone.

36. The machine readable medium of claim 35, wherein the acoustic signals to be generated by the second and third computing devices are selected from a group comprising of maximum length sequence signal, and a chirp signal.

37. The machine readable medium of claim 31, further including:
estimating an emission start time of when the acoustic signal was emitted from the actuator of the second computing device.
38. The machine readable medium of claim 37, wherein estimating the emission start time includes the second computing device using a loopback device to compute the emission start time;
using emission start time to determine the estimate of a time for the acoustic signal to travel from the actuator of the second computing device to the sensor third computing device.
39. The machine readable medium of claim 31, wherein the method further includes:
prior to the first computer determining a physical location of at least one of the group including the actuator of the second computing device and the sensor of the third computer, the first computer computing an initial estimation of the physical location of the actuator of the second computer and the sensor of the third computing device via video modality.
40. The machine readable medium of claim 31, wherein the method further includes:
prior to the first computer determining a physical location of at least one of the group including the actuator of the second computing device and the sensor of the third computer, the first computer computing an initial estimation of the physical location of the actuator of the second computer and the sensor of the third computing device via video modality.

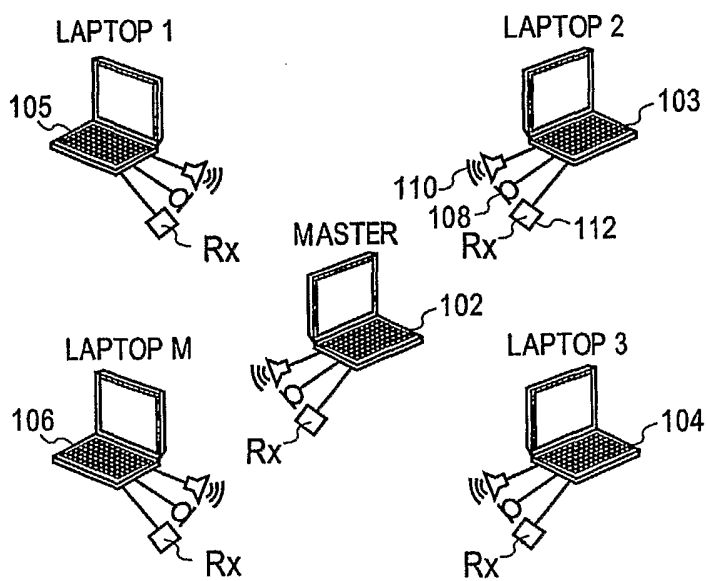


FIG. 1

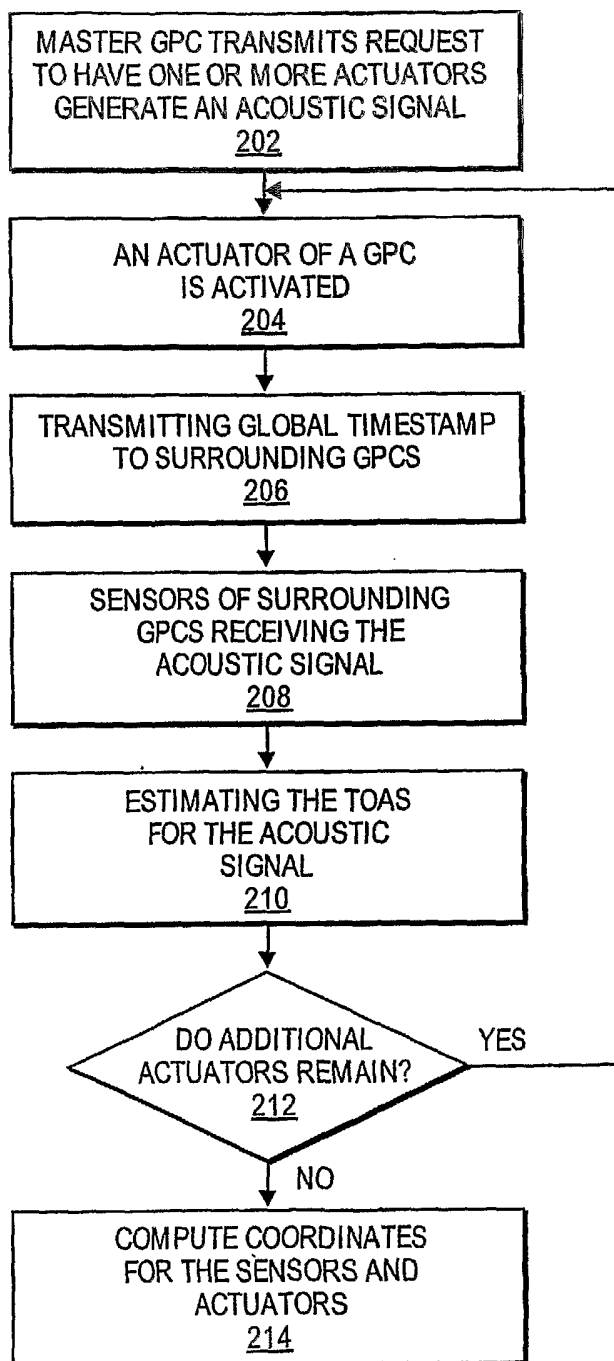


FIG. 2

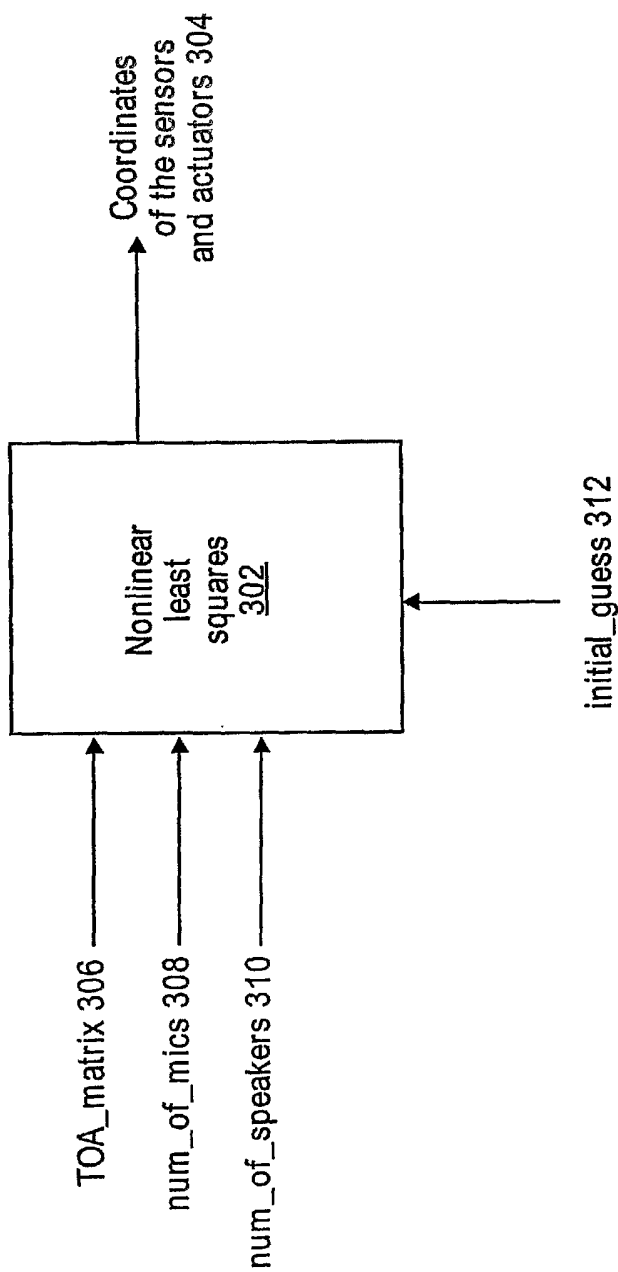


FIG. 3