



US009270807B2

(12) **United States Patent**
Shivappa et al.

(10) **Patent No.:** **US 9,270,807 B2**
(45) **Date of Patent:** **Feb. 23, 2016**

(54) **AUDIO LOCALIZATION USING AUDIO SIGNAL ENCODING AND RECOGNITION**

(56) **References Cited**

U.S. PATENT DOCUMENTS

(75) Inventors: **Shankar Thagadur Shivappa**, Beaverton, OR (US); **Tony F. Rodriguez**, Portland, OR (US)

6,356,905 B1 3/2002 Gershman et al.
7,739,705 B2 * 6/2010 Lee et al. 725/10
8,175,297 B1 * 5/2012 Ho et al. 381/122

(Continued)

(73) Assignee: **Digimarc Corporation**, Beaverton, OR (US)

FOREIGN PATENT DOCUMENTS

JP 2008242810 10/2008
JP 2009198209 9/2009

(Continued)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 877 days.

OTHER PUBLICATIONS

(21) Appl. No.: **13/033,372**

International Search Report and Written Opinion in PCT/US12/026361 dated Jun. 28, 2012.

(Continued)

(22) Filed: **Feb. 23, 2011**

Primary Examiner — Davetta W Goins

Assistant Examiner — Daniel Sellers

(65) **Prior Publication Data**

US 2012/0214544 A1 Aug. 23, 2012

(74) *Attorney, Agent, or Firm* — Digimarc Corporation

(57) **ABSTRACT**

(51) **Int. Cl.**
H04R 29/00 (2006.01)
G06F 17/00 (2006.01)
H04M 1/725 (2006.01)
G10L 19/018 (2013.01)
G06Q 30/02 (2012.01)
H04R 3/00 (2006.01)

A positioning network comprises an array of signal sources that transmit signals with unique characteristics that are detectable in signals captured through a sensor on a mobile device, such as a microphone of a mobile phone handset. Through signal processing of the captured signal, the positioning system distinguishes these characteristics to identify distinct sources and their corresponding coordinates. A position calculator takes these coordinates together with other attributes derived from the received signals from distinct sources, such as time of arrival or signal strength, to calculate coordinates of the mobile device. A layered protocol is used to introduce distinguishing characteristics in the source signals. This approach enables the use of low cost components to integrate a positioning network on equipment used for other functions, such as audio playback equipment at shopping malls and other venues where location based services are desired.

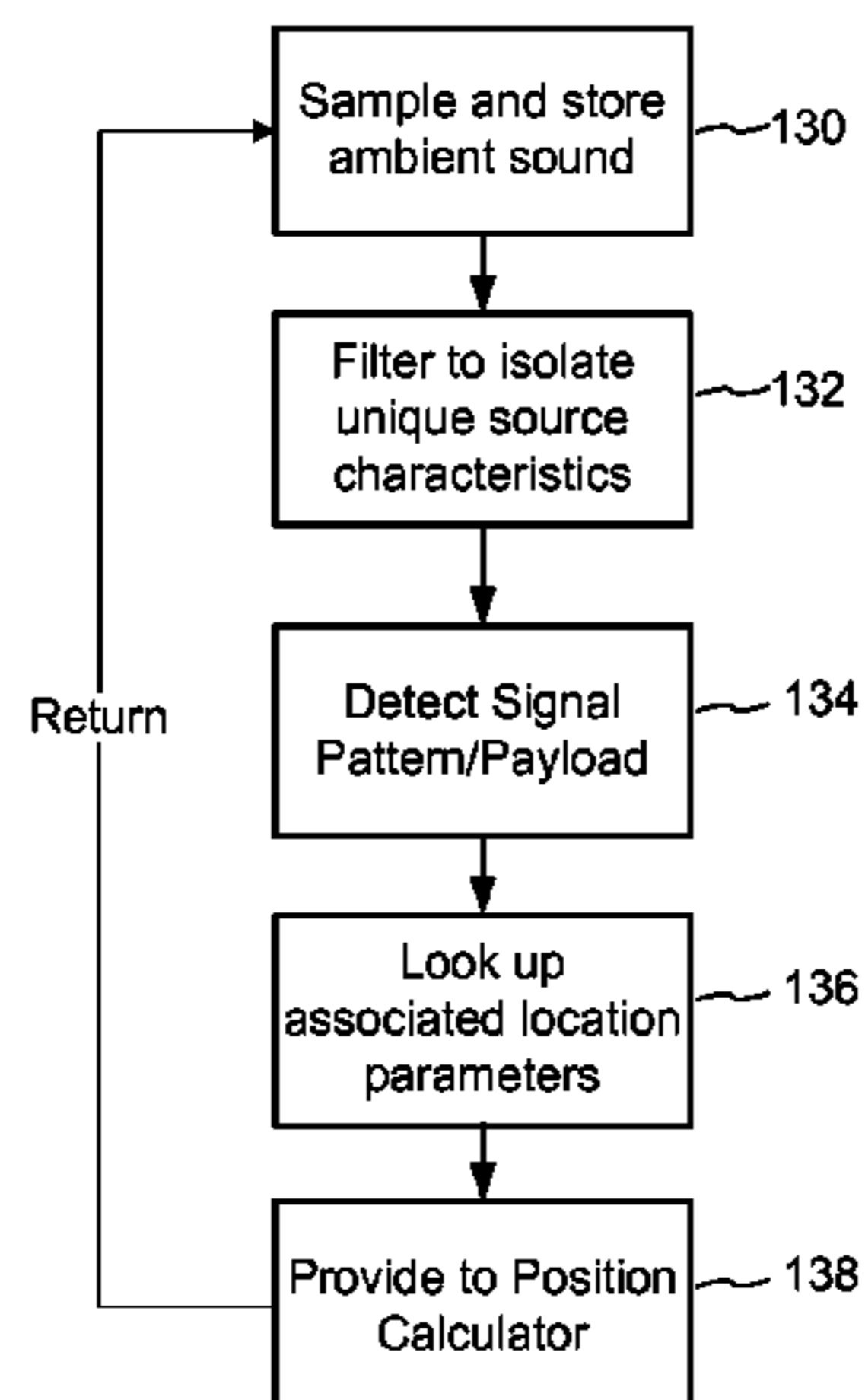
(52) **U.S. Cl.**
CPC **H04M 1/72572** (2013.01); **G06Q 30/0261** (2013.01); **G10L 19/018** (2013.01); **H04R 29/007** (2013.01); **G06Q 30/02** (2013.01); **H04R 3/005** (2013.01)

(58) **Field of Classification Search**
CPC .. G10L 19/018; G06Q 30/02; G06Q 30/0252; G06Q 30/0259; G06Q 30/0261; G06Q 30/0267; H04R 29/007; H04R 3/005

USPC 381/58; 700/94

See application file for complete search history.

20 Claims, 4 Drawing Sheets



(56)

References Cited

U.S. PATENT DOCUMENTS

8,219,384 B2 * 7/2012 Lloyd et al. 704/8
 8,508,357 B2 * 8/2013 Srinivasan et al. 340/534
 8,606,293 B2 * 12/2013 Kim et al. 455/456.1
 8,660,581 B2 2/2014 Davis et al.
 9,064,398 B2 6/2015 Davis et al.
 2003/0215110 A1 11/2003 Rhoads et al.
 2004/0111320 A1 6/2004 Schlieffers et al.
 2004/0128514 A1 * 7/2004 Rhoads 713/176
 2008/0104227 A1 5/2008 Birnie et al.
 2008/0242275 A1 10/2008 Yuba et al.
 2009/0262975 A1 10/2009 Rhoads et al.
 2010/0114344 A1 * 5/2010 Hannaby 700/94
 2010/0134278 A1 6/2010 Srinivasan et al.
 2010/0199296 A1 8/2010 Lee et al.
 2010/0280826 A1 11/2010 Bakish
 2010/0322035 A1 12/2010 Rhoads et al.
 2011/0019999 A1 1/2011 George et al.
 2011/0028160 A1 2/2011 Roeding et al.
 2011/0029359 A1 2/2011 Roeding et al.
 2011/0029360 A1 2/2011 Gollapalli
 2011/0029362 A1 2/2011 Roeding et al.
 2011/0029364 A1 2/2011 Roeding et al.
 2011/0029370 A1 2/2011 Roeding et al.
 2011/0054890 A1 * 3/2011 Ketola et al. 704/231
 2011/0150240 A1 6/2011 Akiyama et al.
 2011/0161076 A1 6/2011 Davis et al.

2012/0020189 A1 * 1/2012 Agevik et al. 367/127
 2012/0150578 A1 * 6/2012 Mangat et al. 705/7.13
 2012/0165046 A1 6/2012 Rhoads et al.

FOREIGN PATENT DOCUMENTS

WO WO2010016589 2/2010
 WO WO2010070526 6/2010

OTHER PUBLICATIONS

Rishabh, "Indoor Localization Using Controlled Ambient Sounds", 2012 International Conference on Indoor Positioning and Indoor Navigation, Nov. 13-15, 2012, 10 pages.
 Mandal, "Beep: 3D Indoor Positioning Using Audible Sound", IEEE Consumer Communications and Networking Conference (CCNC'05), Las Vegas, Jan. 2005, 7 pages.
 F. Blackmon and J. Pollock, "Blue Rose Perimeter Defense and Security System," Technologies for Homeland Security and Homeland Defense V, Proceedings of SPIE, vol. 6201, pp. 620123, 2006.
 Indoor Navigation System for Handheld Devices, Manh Hung V. Le, Dimitris Saragas, and Nathan Webb, Oct. 22, 2009.
 U.S. Appl. No. 14/748,126, filed Jun. 23, 2015.
 Notice of Reasons for Rejection dated Nov. 26, 2015, in JP2013-555568.
 Supplementary Partial European Search Report dated Oct. 22, 2015 in European Patent application EP12757448.

* cited by examiner

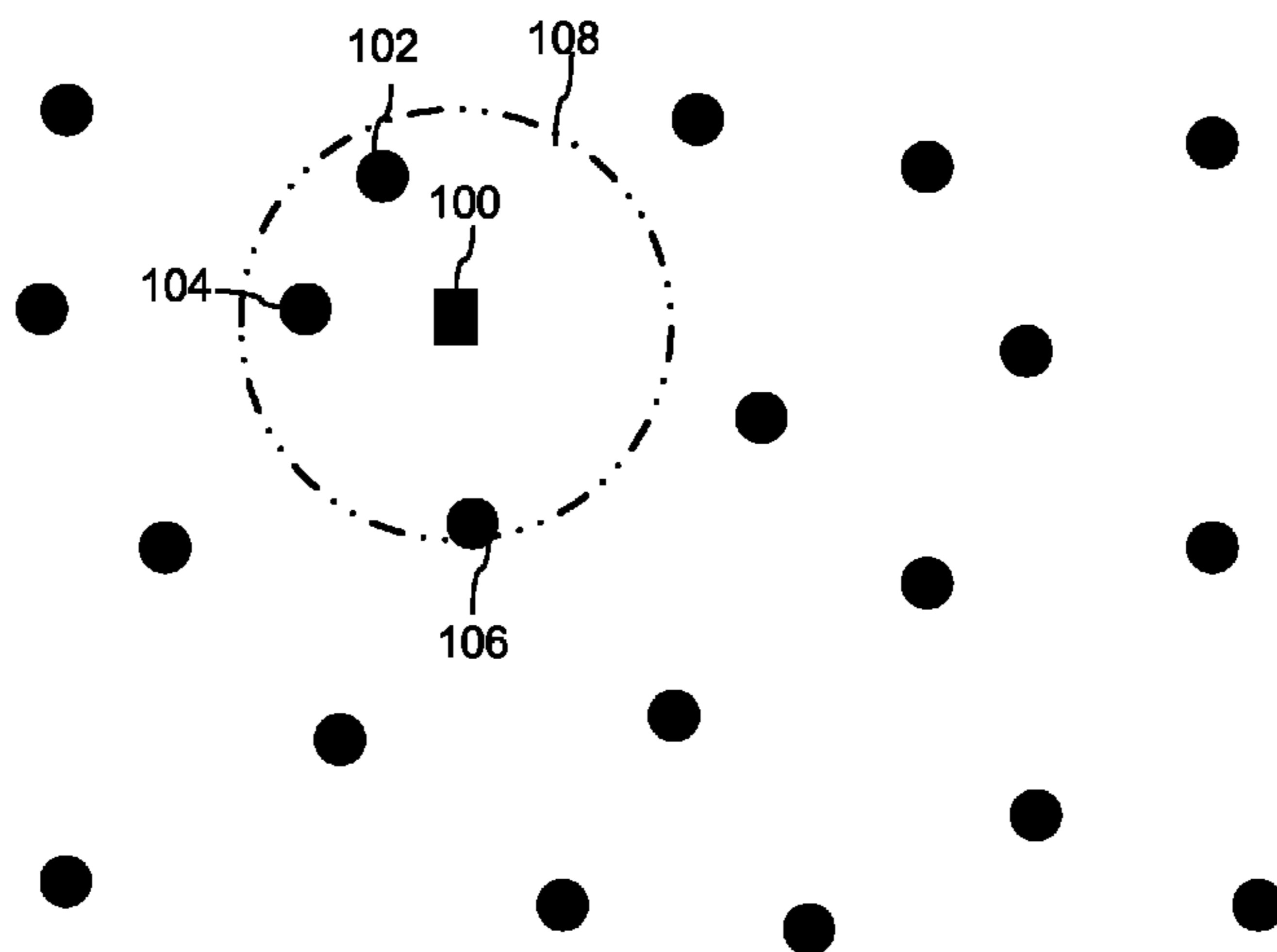


Fig. 1

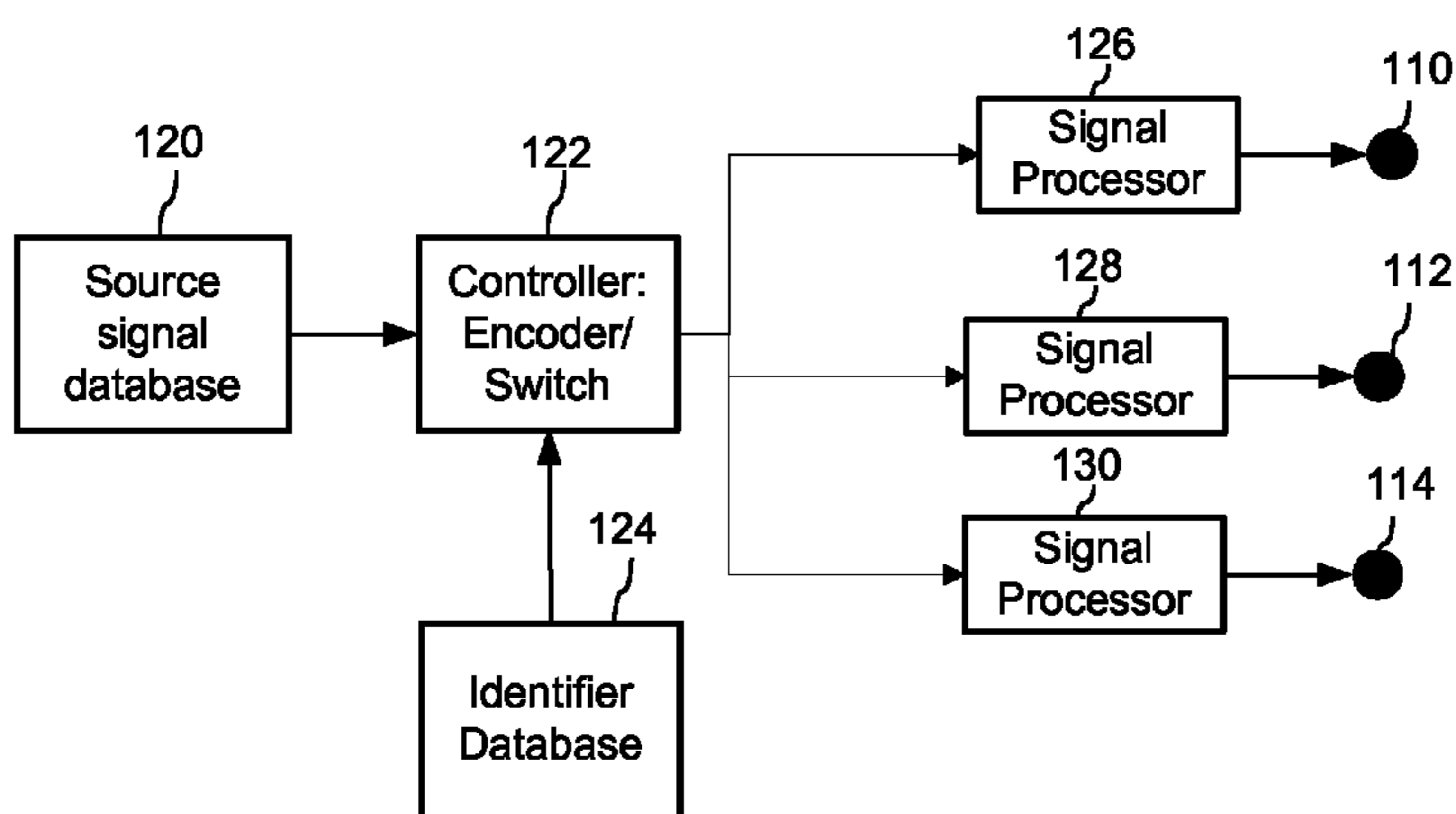


Fig. 2

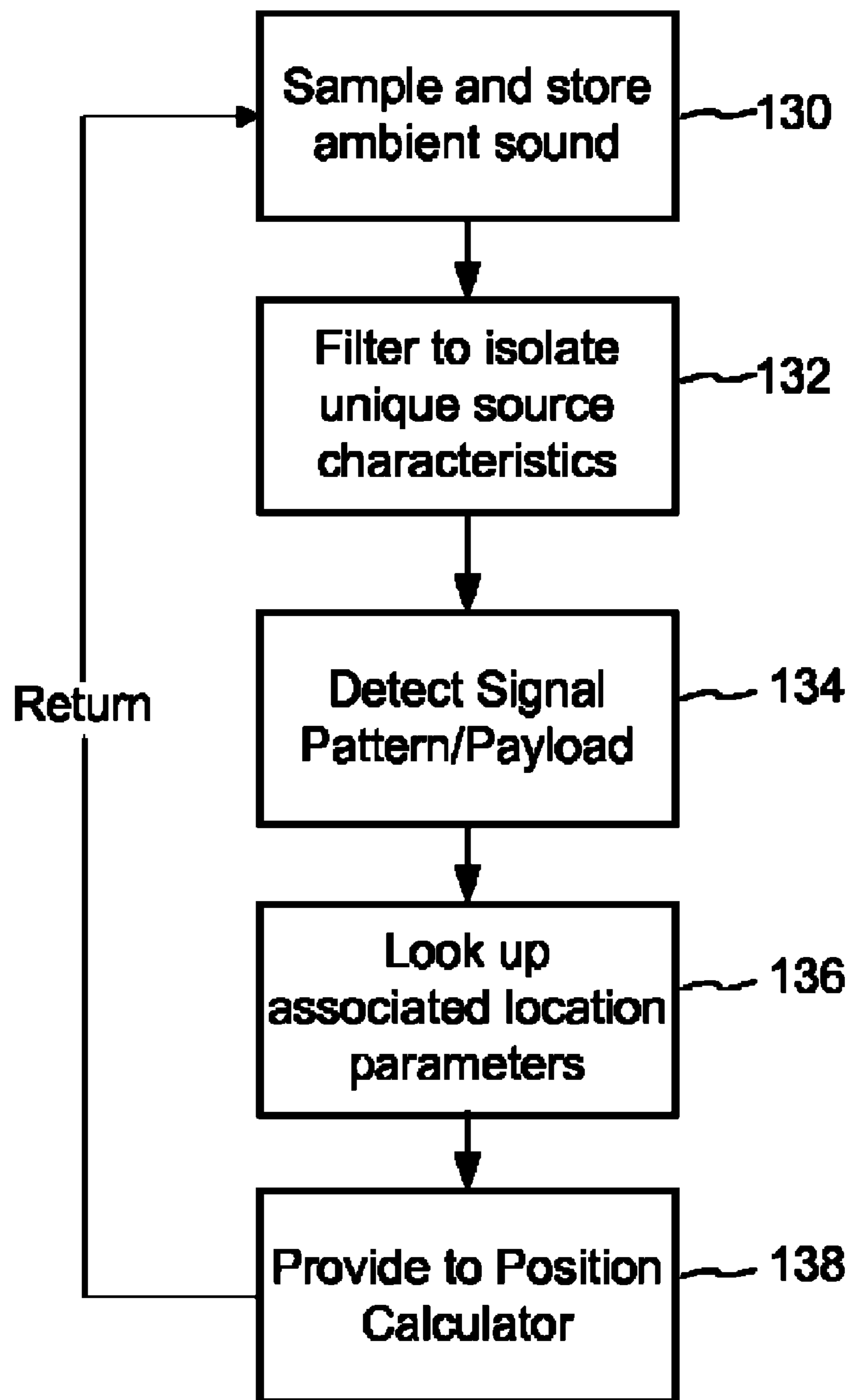


Fig. 3

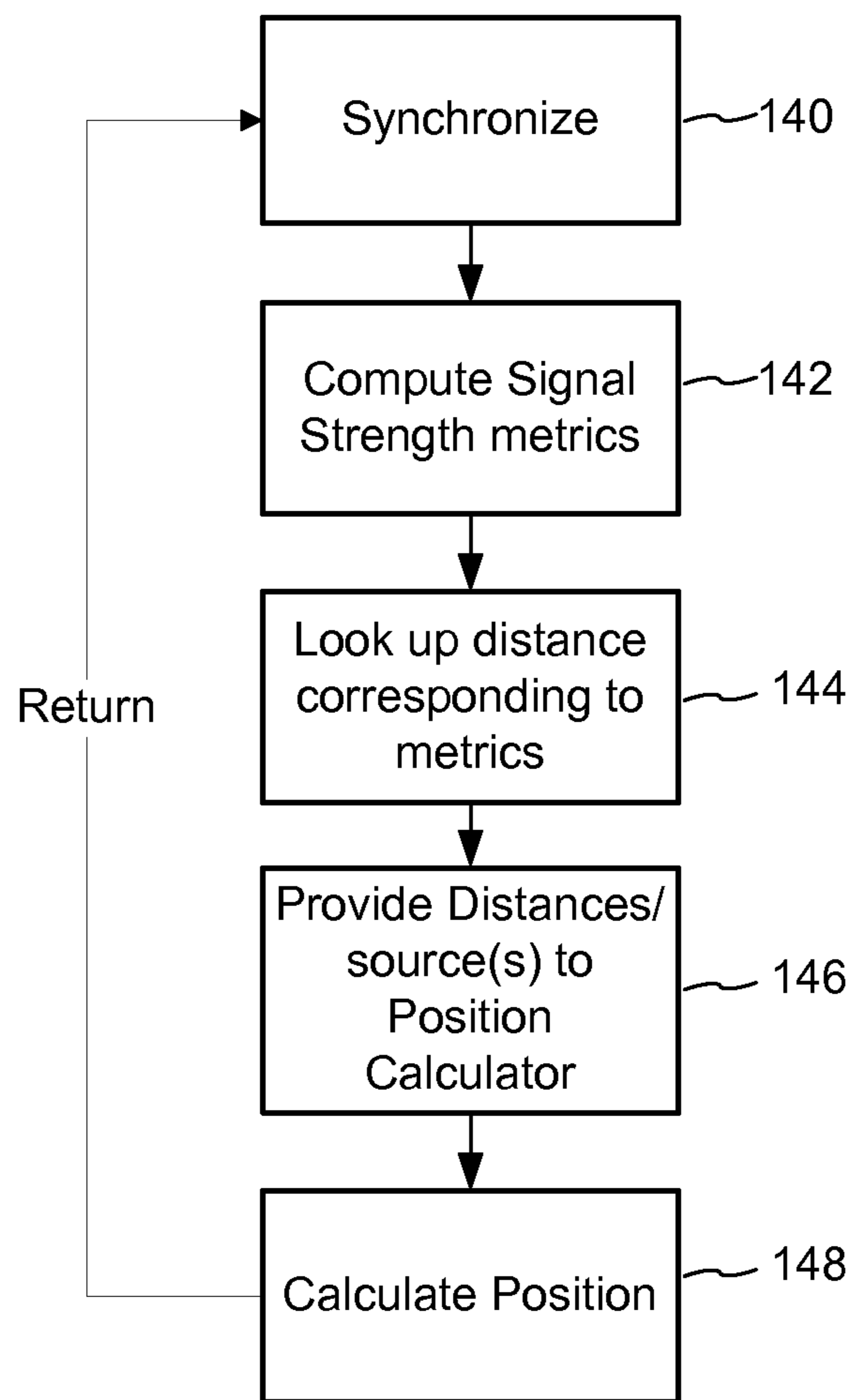


Fig. 4

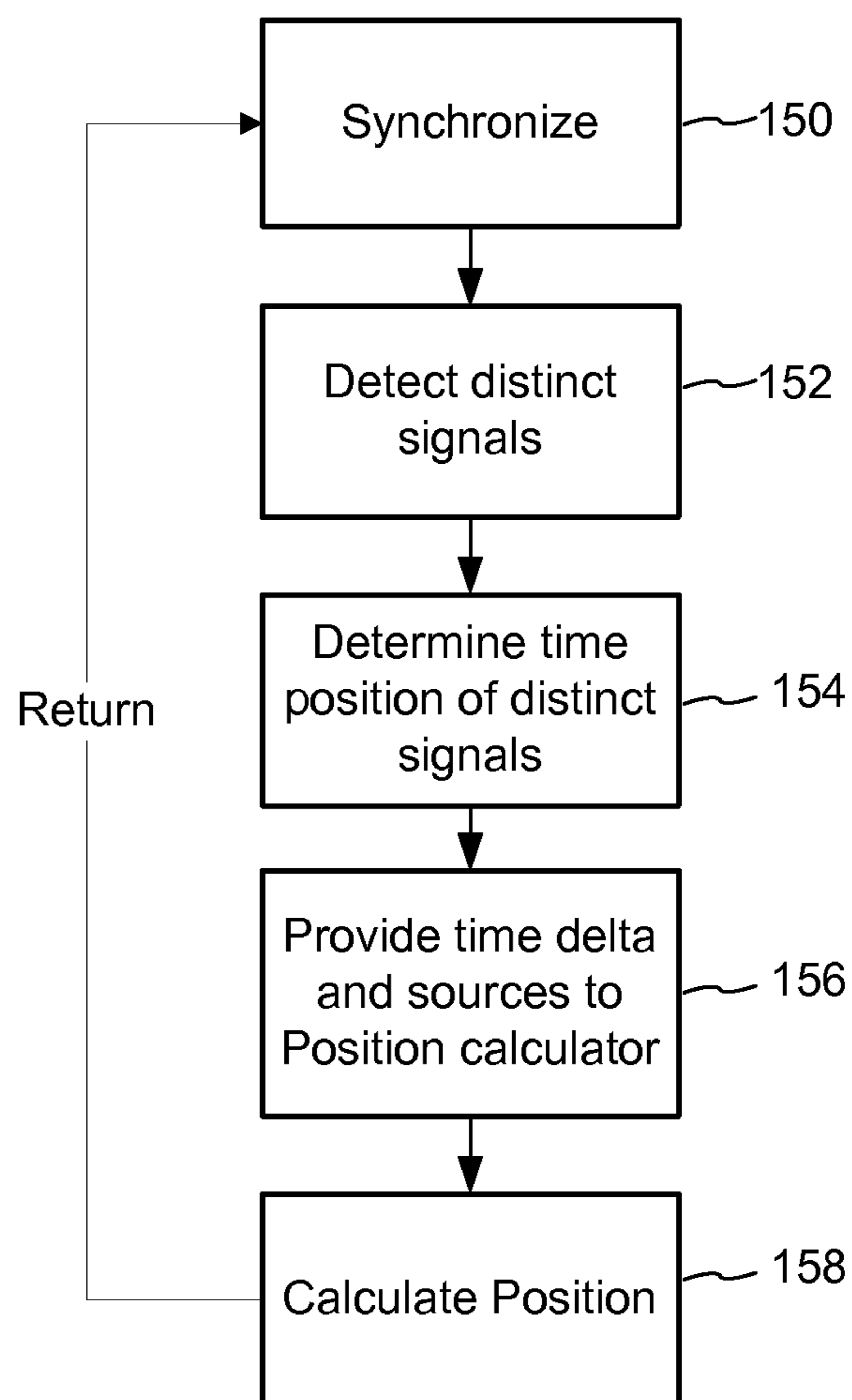


Fig. 5

AUDIO LOCALIZATION USING AUDIO SIGNAL ENCODING AND RECOGNITION

TECHNICAL FIELD

The invention relates to audio positioning systems, and more specifically, relates to audio signal processing for positioning systems.

BACKGROUND AND SUMMARY

Audio source localization uses one or more fixed sensors (microphones) to localize a moving sound source. The sound source of interest usually is a human voice or some other natural source of sound.

Reversing this scenario, sound signals transmitted from known locations can be used to determine the position of a moving sensor (e.g., a mobile device with a microphone) through the analysis of the received sounds from these sources. At any point of time, the relative positioning/orientation of the sources and sensors can be calculated using a combination of information known about the sources and derived from the signals captured in the sensor or a sensor array.

While traditional Global Positioning System (GPS) technologies are finding broad adoption in a variety of consumer devices, such technologies are not always effective or practical in some applications. Audio signal-based positioning can provide an alternative to traditional GPS because audio sources (e.g., loudspeakers) and sensors (e.g., microphones on mobile devices) are ubiquitous and relatively inexpensive, particularly in application domains where traditional GPS is ineffective or not cost effective. Applications of this technology include indoor navigation, in store browsing, games and augmented reality.

Audio based positioning holds promise for indoor navigation because sound systems are commonly used for background sound and public address announcements, and thus, provide a low cost infrastructure in which a positioning network can be implemented. Audio based positioning also presents an alternative to traditional satellite based GPS, which is not reliable indoors. Indoor navigation enabled on a mobile handset enables the user to locate items in a store or other venue. It also enables navigation guidance to the user via the mobile handset via directions and interactive maps presented on the handset.

Audio based positioning also enables in-store browsing based on user location on mobile handsets. This provides benefits for the customer, who can learn about products at particular locations, and for the store owner, who can gather market intelligence to better serve customers and more effectively configure product offerings to maximize sales.

Audio based positioning enables location based game features. Again, since microphones are common on mobile phones and these devices are increasingly used as game platforms, the combination of audio based positioning with game applications provides a cost effective way to enable location based features for games where other location services are unreliable.

Augmented reality applications use sensors on mobile devices to determine the position and orientation of the devices. Using this information, the devices can then "augment" the user's view of surrounding area with synthetically generated graphics that are constructed using a spatial coordinate system of the neighboring area constructed from the devices location, orientation and possible other sensed context information. For example, computer generated graphics

are superimposed on a representation of the surrounding area (e.g., based on video captured through the device's camera, or through an interactive 2D or 3D map constructed from a map database and location/orientation of the device).

Though audio positioning systems hold promise as an alternative to traditional satellite based GPS, many challenges remain in developing practical implementations. To be a viable low cost alternative, audio positioning technology should integrate easily with typical consumer audio equipment that is already in use in environments where location based services are desired. This constraint makes systems that require the integration of complex components less attractive.

Another challenge is signal interference and degradation that makes it difficult to derive location from audio signals captured in a mobile device. Signal interference can come from a variety of sources, such as echoes/reverberation from walls and other objects in the vicinity. Data signals for positioning can also encounter interference from other audio sources, ambient noise, and noise introduced in the signal generation, playback and capture equipment.

Positioning systems rely on the accuracy and reliability of the data obtained through analysis of the signals captured from sources. For sources at fixed locations, the location of each source can be treated as a known parameter stored in a table in which identification of the signal source indexes the source location. This approach, of course, requires accurate identification of the source. Positioning systems that calculate position based on time of arrival or time of flight require synchronization or calibration relative to a master clock. Signal detection must be sufficiently quick for real time calculation and yet accurate enough to provide position within desired error constraints.

Positioning systems that use signal strength as a measure of distance from a source require reliable schemes to determine the signal strength and derive a distance from the strength within error tolerances of the application.

These design challenges can be surmounted by engineering special purpose equipment to meet desired error tolerances. Yet such special purpose equipment is not always practical or cost effective for wide spread deployment. When designing a positioning system for existing audio playback equipment and mobile telephone receivers, the signal generation and capture processes need to be designed for ease of integration and to overcome the errors introduced in these environments. These constraints place limits on the complexity of equipment that is used to introduce positioning signals. A typical configuration is comprised of conventional loudspeakers driven by conventional audio components in a space where location based services add value and other forms of GPS do not work well, such as indoor shopping facilities and other public venues.

The audio playback and microphone capture in typical mobile devices constrain the nature of the source signal. In particular, the source signal must be detectable from an ambient signal captured by such microphones. As a practical matter, these source signals must be in the human audible frequency range to be reliably captured because the frequency response of the microphones on these devices is tuned for this range, and in particular, for human speech. This gives rise to another constraint in that the source audio signals have to be tolerable to the listeners in the vicinity. Thus, while there is some flexibility in the design of the audio signal sources, they must be tolerable to listeners and they must not interfere with other purposes of the audio playback equipment, such as to provide background music, information messages to shoppers, and other public address functions.

Digital watermarking presents a viable option for conveying source signals for a positioning system because it enables integration of a data channel within the audio programming played in conventional public address systems. Digital watermarks embed data within the typical audio content of the system without perceptibly degrading the audio quality relative to its primary function of providing audio programming such as music entertainment and speech. In addition, audio digital watermarking schemes using robust encoding techniques can be accurately detected from ambient audio, even in the presence of room echoes and noise sources.

Robustness is achieved using a combination of techniques. These techniques include modulating robust features of the audio with a data signal (below desired quality level from a listener perspective) so that the data survives signal degradation. The data signal is more robustly encoded without degrading audio quality by taking human auditory system into account to adapt the data signal to the host content. Robust data signal coding techniques like spread spectrum encoding and error correction improve data reliability. Optimizing the detector through knowledge of the host signal and data carrier enable weak data signal detection, even from degraded audio signals.

Using these advances in robust watermarking, robust detection of audio watermarks is achievable from ambient audio captured through the microphone in a mobile device, such as a cell phone or tablet PC. As a useful construct to design audio watermarking for this application, one can devise the watermarking scheme to enhance robustness at two levels within the signal communication protocol: the signal feature modulation level and the data signal encoding level. The signal feature modulation level is the level that specifies the features of the host audio signal that are modified to convey an auxiliary data signal. The data signal encoding level specifies how data symbols are encoded into a data signal. Thus, a watermarking process can be thought of as having two layers of signal generation in a communication protocol: data signal formation to convey a variable sequence of message symbols, and feature modulation to insert the data signal into the host audio signal. These protocol levels are not necessarily independent. Some schemes take advantage of feature analysis of the host signal to determine the feature modification that corresponds to a desired data symbol to be encoded in a sequence of message symbols. Another consideration is the use of synchronization and calibration signals. A portion of the data signal is allocated to the task of initial detection and synchronization.

When designing the feature modulation level of the watermarking scheme for a positioning application in mobile devices, one should select a feature modulation that is robust to degradation expected in ambient capture. Robust audio features that are modulated with an auxiliary data signal to hide the data in a host audio program in these environments include features that can be accumulated over a detection window, such as energy at frequency locations (e.g., in schemes that modulate frequency tones adapted using audio masking models to mask audibility of the modulation). The insertion of echoes can also be used to modulate robust features that can be accumulated over time, like autocorrelation. This accumulation enables energy from weak signals to be added constructively to produce a composite signal from data can be more reliably decoded.

When designing the data signal coding level for a positioning application, one should consider techniques that can be used to overcome signal errors introduced in the context of ambient capture. Spread spectrum data signal coding (e.g., direct sequence and channel hopping), and soft decision error

correction improve robustness and reliability of audio watermarks using these modulation techniques. Direct sequence spread spectrum coding spreads a message symbol over a carrier signal (typically a pseudorandom carrier) by modulating the carrier with a message symbol (e.g., multiplying a binary antipodal carrier by 1 or -1 to represent a binary 1 or 0 symbol). Alternatively, a symbol alphabet can be constructed using a set of fixed, orthogonal carriers. Within the data signal coding level, additional sub-levels of signal coding can be applied, such as repetition coding of portions of the message, and error correction coding, such as convolution coding and block codes. One aspect of data signal coding that is directly related to feature modulation is the mapping of the data signal to features that represent candidate feature modulation locations within the feature space. Of course, if the feature itself is a quantity calculated from a group of samples, such as time segment of an audio clip, the feature modulation location corresponds to the group of samples and the feature of that group.

One approach is to format a message into an encoded data signal packet comprising a set of encoded symbols, and then multiplex packets onto corresponding groups of feature modulation locations. The multiplexing scheme can vary the mapping over time, or repeat the same mapping with each repetition of the same packet.

The designer of the data encoding scheme will recognize that there is interplay among the data encoding and mapping schemes. For example, elements (e.g., chips) of the modulated carrier in a direct sequence spread spectrum method are mapped to features in a fixed pattern or a variable scattering. Similarly, one way to implement hopping is to scatter or vary the mapping of encoded data symbols to feature modulation locations over the feature space, which may be specified in terms of discrete time or frequencies.

Robust watermark readers exploit these robustness enhancements to recover the data reliably from ambient audio capture through a mobile device's microphone. The modulation of robust features minimizes the impact of signal interference on signal degradation. The reader first filters the captured audio signal to isolate the modulated features. It accumulates estimates of the modifications made to robust features at known feature modulation locations. In particular, it performs initial detection and synchronization to identify a synchronization component of the embedded data signal. This component is typically redundantly encoded over a detection window so that the embedded signal to noise ratio is increased through accumulation. Estimates are weighted based on correspondence with expected watermark data (e.g., a correlation metric or count of detected symbols matching expected symbols). Using the inverse of the mapping function, estimates of the encoded data signal representing synchronization and variable message payload are distinguished and instances of encoded data corresponding to the same encoded message symbols from various embedding locations are aggregated. For example, if a spreading sequence is used, the estimates of the chips are aggregated through demodulation with the carrier. Periodically, buffers storing the accumulated estimates of encoded data provide an encoded data sequence for error correction decoding. If valid message payload sequences are detected using error detection, the message payload is output as a successful detection.

While these and other robust watermarking approaches enhance the robustness and reliability in ambient capture applications, the constraints necessary to compute positioning information present challenges. The positioning system preferably should be able to compute the positioning information quickly and accurately to provide relevant location

and/or device orientation feedback to the user as he or she moves. Thus, there is a trade-off between robustness, which tends toward longer detection windows, and real time response, which tends toward a shorter detection window. In addition, some location based techniques based on relative time of arrival rely on accurate synchronization of source signal transmissions and the ability to determine the difference in arrival of signals from different sources.

Alternative approaches that rely on strength of signal metrics can also leverage watermarking techniques. For example, the strength of the watermark signal can be an indicator of distance from a source. There are several potential ways to design watermark signals such that strength measurements of these signals after ambient capture in a mobile device can be translated into distance of the mobile device from a source. In this case, the watermarks from different sources need to be differentiated so that the watermark signal from each can be analyzed.

The above approaches take advantage of the ability to differentiate among different sources. One proposed configuration to accomplish this is to insert a unique watermark signal into each source. This unique signal is assigned to the source and source location in a database. By identifying the unique signal, a positioning system can determine its source location by finding it in the database. This approach potentially increases the implementation cost by requiring additional circuitry or signal processing to make the signal unique from each source. For audio systems that comprise several speakers distributed throughout a building, the cost of making each signal unique yet and reliably identifiable can be prohibitive for many applications. Thus, there is a need for low cost means to make a source or a group of neighboring sources unique for the purpose of determining where a mobile device is within a network of sources.

Digital watermarks can be used to differentiate streams of audio that all sound generally the same. However, some digital watermark signaling may have the disadvantage that the host audio is a source of interference to the digital watermark signal embedded in it. Some forms of digital watermarking use an informed embedding in which the detector does not treat the host as interfering noise. These approaches raise other challenges, particularly in the area of signal robustness. This may lead the signal designer to alternative signaling techniques that are robust techniques for conveying source identification through the audio being played through the audio playback system.

One alternative is to use a form of pattern recognition or content fingerprinting in which unique source locations are associated with unique audio program material. This program material can be music or other un-obtrusive background sounds. To differentiate sources, the sounds played through distinct sources are selected or altered to have distinguishing characteristics that can be detected by extracting the unique characteristics from the received signal and matching them with a database of pre-registered patterns stored along with the location of the source (or a neighborhood area formed by a set of neighboring sources that transmit identical sounds). One approach is to generate unique versions of the same background sounds by creating versions from a master sound that have unique frequency or phase characteristics. These unique characteristics are extracted and detected by matching them with the unique characteristics of a finite library of known source signals.

The approaches of inserting a digital watermark or generating unique versions of similarly sounding audio share some fundamental principles in that the task is to design a signaling means in which sources sound the same, yet the detector can

differentiate them and look up locations parameters associated with the unique signal payload or content feature pattern. Hybrid approaches are also an option. One approach is to design synthetic signals that convey a digital payload like a watermark, yet are themselves the background sound that is played into the ambient environment of a building or venue where the audio based positioning system is implemented. For example, the data encoding layer of a watermark system can be used to generate data signal that is then shaped or adapted into a pleasing background sound, such as the sound of a water feature, ocean waves or an innocuous background noise. Stated another way, the data signal itself is selected or altered into a form that has some pleasing qualities to the listener, or even simulates music. Unique data signals can be generated from structured audio (e.g., MIDI representations) as distinct collections of tones or melodies that sound similar, yet distinguish the sources.

One particular example of a system for producing “innocuous” background sound is a sound masking system. This type of system adds natural or artificial sound into an environment to cover up unwanted sound using auditory masking. One supplier of these types of systems is Cambridge Sound Management, LLC, of Cambridge, Mass. In addition to providing sound masking, these systems include auxiliary inputs for paging or music distribution. The system comprises control modules that control zones, each having zone having several speakers (e.g., the module independently controls the volume, time of day masking, equalization and auto-ramping for each zone). Each control modules is configurable and controllable via browser based software running on a computer that is connected to the module through a computer network or direct connection.

Another hardware configuration for generating background audio is a network of wireless speakers driven by a network controller. These systems reduce the need for wired connections between audio playback systems and speakers. Yet there is still a need for a cost effective means to integrate a signaling technology that enables the receiver to differentiate sources that otherwise would transmit the same signals.

In this disclosure, we describe methods and systems for implementing positioning systems for mobile devices. There is a particular emphasis on using existing signal generation and capture infrastructure, such as existing audio or RF signal generation in environments where traditional GPS is not practical or effective.

One aspect of the invention is a method of determining position of a mobile device. In this method, the mobile device receives audio signals from two or more different audio sources via its microphone. The audio signals are integrated into the normal operation of an audio playback system that provides background sound and public address functionality. As such, the audio signals sound substantially similar to a human listener, yet have different characteristics to distinguish among the different audio sources. The audio signals are distinguished from each other based on distinguishing characteristics determined from the audio signals. Based on identifying particular audio sources, the location of the particular audio sources is determined (e.g., by finding the coordinates of the source corresponding to the identifying characteristics). The position of the mobile device is determined based on the locations of the particular audio sources.

Particular sources can be identified by introducing layers of unique signal characteristics, such as patterns of signal alterations, encoded digital data signals, etc. In particular, a first layer identifies a group of neighboring sources in a network, and a second layer identifies a particular source. Once the sources are accurately distinguished, the receiver then

looks up the corresponding source coordinates, which then feed into a position calculator. Position of the mobile device is then refined based on coordinates of the source signals and other attributes derived from the source signals.

Additional aspects of the invention include methods for generating the source signals and associated positioning systems.

These techniques enable a variety of positioning methods and systems. One such system determines location based on source device location and relative time of arrival of signals from the sources. Another determines location based on relative strength of signal from the sources. For example, a source with the strongest signal provides an estimate of position of the mobile device. Additional accuracy of the location can be calculated by deriving an estimate of distance from source based on signal strength metrics.

The above-summarized methods are implemented in whole or in part as instructions (e.g., software or firmware for execution on one or more programmable processors), circuits, or a combination of circuits and instructions executed on programmable processors.

Further features will become apparent with reference to the following detailed description and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram illustrating a mobile device in the midst of a network of signal sources.

FIG. 2 is a diagram illustrating a system for generating unique audio source signals for use in a position system.

FIG. 3 is a flow diagram of a process for analyzing an ambient audio signal to detect and identify an audio source signal.

FIG. 4 is a flow diagram of a process for determining distance from an audio source signal by analyzing strength of signal metrics.

FIG. 5 is a flow diagram of a process for determining the time difference of arrival of audio signals from distinct audio sources.

DETAILED DESCRIPTION

Sensor and Source Configurations

Before getting to the details of a particular localization approach, we start with a discussion of sensor and source configurations and an overview of location information that can be derived from each. In the case of audio localization, the sensors are microphones and the sources are audio transmitters (e.g., loudspeakers). Each can be present in many different configurations, and we review the main categories here. We are particularly interested in applications where the sensor is a common component of a consumer device that is popular among consumers, such as a mobile phone or tablet computer. As such, our examples of configurations use these devices. Later, we provide particular examples of the methods applicable to each of the configurations.

Configurations can be organized according to the three following categories: 1) the number of sources, 2) the number of microphones on the mobile device; and 3) the number of mobile devices collaborating with each other.

To illustrate, we use a general example of a network of signal sources. FIG. 1 is a diagram illustrating a mobile device **100** in the midst of a network of signal sources (represented as dots, e.g., **102**, **104** and **106**). At a given position within the network of audio sources in FIG. 1, there is a subset

of the network comprising one or more sources within the range of the mobile device. This range is depicted as a dashed circle **108**.

One Loudspeaker:

A positioning system can be configured to detect or measure the proximity of the sensor to one source (e.g., such as the closest source). Even within a network of signal sources as shown in FIG. 1, the system can be reduced to a single source, e.g., **102**, within the range of the mobile device **100**. At a minimum, the mobile device knows that it is within the neighborhood of source **102**. With additional information, such as the strength of signal or direction of the source, more position information can be computed and provided to the user of the mobile device.

Two or Preferably More than Two Loudspeakers:

Two or more speakers enable triangulation to estimate the relative position of the sensor. Referring to FIG. 1, sources **102**, **104** and **106** are in the range of the mobile device **100**. The relative arrival time of the audio signal from these sources to the mobile device provide sufficient data to determine location. For example, each pair of source to mobile device **100** within the range **108** provides input to a set of equations that can be solved to calculate a location. The relative arrival time to the mobile device from two different sources provides a location approximation of the mobile device along a hyperboloid. Adding another pair enables calculation of the mobile device as the intersection of the hyperboloids calculated for the two pairs. As the number of pairs of sources within range of the mobile device increase, the system can include them in the data used to calculate a solution. Also, the particular sources used are preferably vetted before data obtained from them is included according to signal metrics, such as signal strength of a detected embedded signal from the source.

This approach is sometimes referred to as multilateration or hyperbolic positioning. In this case, we locate a receiver by measuring the time difference of arrival (TDOA) of a signal from different transmitters. Phase difference of two transmitters can be used as well. With multiple transmitters, the TDOA approach is solved by creating a system of equations to find the 3D coordinates (e.g., x, y and z) of the receiver based on the known coordinates of each transmitter and the TDOA for each pair of transmitters to the receiver. This system of equations can then be solved using singular value decomposition (SVD) or Gaussian elimination. A least squares minimization can be used to calculate a solution to the receiver's position.

Additional assumptions simplify the calculation, such as assuming that the mobile device is on the ground (e.g., simplifying a 3D to a 2D problem), and using a map of the network site to limit the solution space of positions of a mobile device to particular discrete positions along paths where users are expected to travel. In the latter, rather than attempting to solve a system of equations with a SVD method, the system can step through a finite set of known positions in the neighborhood to determine which one fits the data best.

The accuracy of the calculations may dictate that the location is accurate within some error band (e.g., the intersection of two or more error bands along the two or more hyperboloids for corresponding two or more pairs of sources relative to the mobile device).

Another approach using two or more sources is to approximate distance from the source using strength of signal metrics that provide a corresponding distance within an error band from each source to the mobile device. For example, a watermark detection metric, such as correlation strength or degree

of signal correspondence between detected and expected signals is used to approximate the distance of the source from the mobile device. The strength of signal is a function of the inverse square of the distance from the source. The strength of signals at higher frequencies decreases more quickly than lower frequencies. Strength of signal metrics that determine the relative strength of low to high frequency signals can be used to estimate distance from source. Accuracy may be improved by tuning the metrics for a particular source location and possible receiver locations that represent the potential position solution space for the positioning system. For instance, for a given installation, the relationship between a strength of signal metric and the distance from a particular sound source is measured and then stored in a look up table to calibrate the metric to acoustic properties at that installation.

One Microphone or Closely Spaced Microphones:

This is the state of typical mobile devices, and as such, they are not suited to perform direction of arrival estimation as in the case of microphone arrays.

Microphone Array with Two or More Microphones:

Using a microphone array to provide direction of arrival of a sound is practical in devices such as tablet PCs that have the required physical dimensions to accommodate the microphone array. With such an array, the localization method can identify the direction of the sound source relative to the orientation of the receiving device and enable better triangulation schemes. This direction information simplifies the calculation of the receiver's position to finding the point along a line through the source and receiver where the receiver is located. When the receiver can determine direction and orientation relative to two or more sources, the positioning system computes position as the intersection of these lines between the receiver and each source. With the orientation provided by a microphone array, one can enable mapping applications (e.g., display a map showing items in an orientation based on the direction of where the user is headed).

In order to determine the direction of a distinct source among two or more sources, the system first identifies the unique sources. The signal properties of each unique source signal than are used to filter the source signal to isolate the signal from a particular source. For example, a matched filter is used to isolate the received signal from a particular source. Then, the system uses microphone array processing to determine the direction of that isolated signal. This microphone array processing detects relative phase delay between the isolated signals from the different microphones in the array to provide direction of arrival relative to the orientation of the array.

In one embodiment, the source signal is unique as a result of direct sequence spread spectrum watermark that is added to the host audio signal. A correlation detector detects the carrier signal and then isolates the watermark signal. The phase delays between pairs of carrier signals detected from each microphone are then used to determine direction of arrival.

Single Mobile Device:

This is a scenario in which a single mobile device captures distinct audio from one or more sources and derives localization from data that it derives from this captured audio about the source(s) such as source identity, location, direction, signal strength and relative characteristics of signals captured from different sources.

Multiple Mobile Devices:

In this scenario, localization of the sources may be enhanced by enabling the devices to collaborate with each other when they are in the vicinity of each other. This collaboration uses a wireless communication protocol for exchange of information among devices using known means

of inter-device communication between neighboring devices (e.g., Bluetooth, Wi-Fi standard, etc.).

Having reviewed various configurations, we now turn to a description of audio signal positioning systems. One scheme, from which many variants can be derived, is to configure a space with loudspeakers that continuously play some identifiable sound. The microphone(s) on the mobile device capture this audio signal, identify the source, and determine the relative proximity/positioning of the source.

Within this type of configuration, there are three main aspects to consider: 1. The means to identify the sound source; 2. The means to perform ambient detection of signals from the source (e.g., ambient refers to capture of ambient sounds through a microphone); and 3. The means to determine sound source proximity and position estimation.

1. Identifiable Sound Source

Existing sound source localization schemes focus on locating the dominant sound sources in the environment. In contrast, we need the ability to locate specific (maybe non-dominant) sound sources, even in the presence of other sources of sound in the neighborhood. One way to achieve this is to look for the presence of an encoded data signal (e.g., such as a non-audible digital watermark; or data signal constructed to be tolerable as background sound). Another way is to use a content fingerprinting technique to recognize a specific sound source as being present in the neighborhood of the mobile device.

2. Ambient Detection of the Source

We need to ensure that the embedded signals used to convey information within the audio signal (e.g., digital watermark or synthesized sound conveying data within the audio source signal) can be recovered reliably from ambient captured audio, especially in noisy environments such as in a shopping mall. One way to increase robustness of a digital watermark, among others, is to sense the ambient "noise" level and adjust the watermark strength embedded in the transmitted signals in real-time so that detection is reliable.

3. Sound Source Proximity/Position Estimation

After the source is identified, the proximity information is estimated. If microphone arrays are available on the mobile device, the relative direction of the source is determined from the microphone array. One approach described further below is to use strength of signal metrics such as metric that measures watermark signal degradation of a combination of robust and fragile digital watermarks. This metric is then provided to a look up table to translate it into an estimate of the distance from the source to the microphone. For example in one implementation, watermarks are embedded at different robustness levels whose detection is dependent on distance from the source. As distance from the source decreases, the ability to recover watermarks at successively lower signal strength or robustness increases. The weakest watermark to be detected provides an indicator of distance from the source because the point at which the next weakest watermark is no longer detected corresponds to a distance from the source.

As another example, detection metrics of the embedded signal can be used to measure the strength of the signal from a particular source. In one implementation, an embedded digital watermark is encoded by modulating frequency tones at selected higher frequencies (e.g., higher frequencies still within the audible range of the microphone on a mobile device). The strength of these tones is attenuated as distance from the source grows. Thus, a detection metric such as the ratio of the high frequency tones to the low frequency tones of the embedded signal provides a detection metric that corresponds to a distance from the source.

11

In some applications, proximity from multiple sources might need to be estimated simultaneously, to allow for triangulation-based position estimation.

Below, we provide details of some alternative system implementations, including:

1. Different approaches to introduce a digital watermark into an audio stream;
2. Sensing ambient audio level and adjusting the watermark strength based on the psycho-acoustic modeling of the ambient audio level for real-time masking computation; and
3. A proximity estimation enabled watermarking scheme.

The ability to identify the source uniquely allows localization of a receiving device in the presence of background noise and other sources that might interfere with the source signals. Initially, the localization method seeks to determine whether the mobile device being located is close to any relevant source.

We have devised a variety of methods for determining the closest source. These methods include a watermarking approach for arbitrary host content, a content fingerprinting approach using a defined set of audio source signals, and synthetic audio approach where audio is constructed to convey particular information.

FIG. 2 is a block diagram illustrating a configurable system for generating unique audio signals within a network of audio sources. The task of this system is to generate unique signals from audio sources (e.g., loudspeakers **110**, **112**, **114**) that are identified through analysis of ambient audio captured at a receiving device. Continuing the theme from FIG. 1, these loudspeakers are representative of the source nodes in a positioning network. Each one has an associated location that is registered with the system in an initialization stage at a venue where the positioning system is implemented. In some implementations, the source signals are adapted for the particular room or venue acoustics to minimize interference of echoes and other distortion. Further, as noted, the solution space for discrete positions of a mobile device within a particular venue can be mapped and stored in conjunction with the identifiers for the network nodes. This information is then fed to the position calculation system based on identification of the nodes from the received signals captured in a mobile device.

The strength of signal metrics for a received strength of signal system (RSS) are tuned based on taking signal measurements at discrete locations within the venue and storing the relationship between the value of one or more signal metrics for a particular source signal at the network node along with the corresponding distance from a source, which is identified through the source identifier(s) of the source signal(s) at that network location.

The system of FIG. 2 is preferably designed to integrate easily in typical audio equipment used to play background music or other programming or background sounds through a network of speakers at a venue. This audio equipment includes pre-amplifiers, audio playback devices (e.g., CD player or player of digital audio stream from a storage device), a receiver-amplifier and ultimately, the output speaker. As noted in the summary, these devices are preferably controllable via control modules that control the audio playback in zones and are each configurable and controllable through software executing on a remote computer connected to the controllers via a network connection.

Audio processing to make unique audio source signals can be inserted at various points in the audio signal generation and transmission path. FIG. 2 shows several different options. First, the audio signal originates from a database **120**. In a mode where the unique signal is generated by selecting a

12

unique signal with corresponding unique fingerprint, or is generated as a synthetic audio signal conveying an identifier, the system has a controller that selects the unique audio signal for a particular source and sends that signal down a path to the loudspeaker for output. The role of an identifier database **124** in this case is to store an association between the unique signal fingerprints or payload of the synthetic signal with the corresponding source (e.g., loudspeaker) location. To simplify configuration of the system, the database can store a pointer to location parameters that are set when the loudspeaker locations are set. These parameters may also include other parameters that adapt the position calculation to a particular network location or source signal (such as a discrete set of position locations, strength of signal characteristics, unique source signal characteristics to aid in pre-filtering or detection, etc.).

In the case where a digital watermark signal stream is embedded to identify the location, the controller **122** includes a digital watermark embedder that receives the audio stream, analyzes it, and encodes the digital watermark signal according to an embedding protocol. This protocol specifies embedding locations within the feature space where one or more data signal layers are encoded. It also specifies format parameters, like data payload structure, redundancy, synchronization scheme, etc. In this type of implementation, the identifier database stores the association between the encoded source identifier and location of the source.

In a watermarking approach, each loudspeaker plays a uniquely watermarked sound. The controller **122** switches the uniquely watermarked audio signals onto the transmission paths of the corresponding speakers (e.g., **110**, **112**, **114**).

Alternatively, if it is not practical to implement unique embedding for each loudspeaker, a set of loudspeakers within a neighborhood play the same watermarked signal, but they have additional signatures that enable the receiver to distinguish the source. For instance, using the example of FIG. 2, the controller sends the same audio signal to the transmission path of a subset of loudspeakers in a particular area of the building. Then, a signal processor (e.g., **126**, **128**, **130**) within the transmission path of each particular source introduces a unique signature into the audio signal. This signature is stored in addition to the source identifier in the database **124** to index the particular location of the loudspeaker that receives the signature altered audio signal at the end of the transmission path.

Since the signal processors (e.g., **126**, **128**, **130**) are needed for several locations in the network of audio sources, they are preferably inexpensive circuits that can be added in-line with the analog transmission path to each loudspeaker. For example, a tapped delay line circuit is connected in-line to introduce a unique set of echoes that is detectable at the receiver to distinguish the audio signals within the subset of sources of the network sharing the same identifier. One approach to construct a tapped delay line circuit is to use a bucket brigade device. This is a form of analog shift register constructed from an NMOS or PMOS integrated circuit.

The speakers in this area are assigned a neighborhood location. If no further position data can be derived at the receiver than the identity of the source, this neighborhood location can at least provide a position accurate to within an area defined as the proximity to the location of the speaker subset. If the signature is detectable from a dominant source, this detection from the dominant source provides a position accurate to within the proximity of the dominant source. Finally, when two more signatures are detected in the cap-

tured audio, then additional position calculations are enabled as explained previously based on TDOA, direction of arrival, triangulation, etc.

A multi-layered watermarking scheme enables a hierarchical scheme of identifying sources within a network. In such a scheme, a first encoded data signal identifies a first larger area of the source network (e.g., a circle encompassing a subset of network nodes that share the same top level identifier). Additional information extracted from the received signal provide additional metrics that narrow the location to a smaller set of sources, a particular source, a particular distance from the source, and finally a particular location within some error tolerance bubble. The simplest of this type of scheme is a two layered approach in which there two watermark layers from each source: a common watermark embedded in the signals output at by a set of speakers in a network (e.g., a set of speakers in a particular area that defines a local neighborhood for mobile devices in this area) and a lower level watermark that is easy to introduce and has a smaller payload, just enough to distinguish between the set of speakers. Techniques for this type of watermarking include: a direct sequence spread spectrum (DSSS) watermark, an echo based watermark, an amplitude or frequency modulation based watermark, and combinations of these methods, which are not mutually exclusive. As described further below, DSSS is used in one embodiment to formulate an encoded data signal, which then is used to modulate features of the signal, such as time and/or frequency domain samples according to a perceptual masking model. An echo based technique is also used to modulate autocorrelation (e.g., echo modulation detected at particular delays). A set of masked frequency tones is also used to encode a data signal onto host audio.

In one particular implementation, we designed a two layer watermark scheme as follows. For a first layer of watermark, a watermark encoder generates a DSSS data signal. The encoder then maps the encoded data chips to corresponding consecutive time blocks of audio to spread the signal over time. For the time portion corresponding to a particular chip, the data signal is adapted to the audio signal for that portion using an audio masking model. The perceptual adaptation generates a particular adjustment for the audio signal in the time block to encode the corresponding chip. This can include frequency domain analysis to adapt the data signal to the audio based on frequency domain masking model. The chip signal may be conveyed in one band or spread over some frequency bands (e.g., spreading of the signal may be both in time and frequency). This first layer conveys an identifier of a portion of the network comprises a set of neighboring network nodes.

For a second layer, a signal processor introduces a distinct echo pattern into the audio signal to identify a particular source within the neighboring network nodes identified by the first layer.

The first layer reliability is enhanced by spreading the signal over time and averaging detection over a period of time encompassing several segments of the entire chipping sequence. This period can be around 1 to 5 seconds.

The second layer reliability is enhanced by using a distinct combination of echoes to represent a particular source within a subset of sources. A symbol alphabet is constructed from a combination of echoes within a maximum delay of 50 milliseconds. This maximum delay minimizes the perception of the echoes by humans, particularly given the ambient noise present in the applications where the positioning system is to be used. Each combination of echoes forms an echo pattern

corresponding to a symbol. The source identifier in the second layer is formed from a set of one or more symbols selected from the alphabet.

Robustness is further enhanced by using a combination of strong echoes that are spaced apart (e.g., 5 milliseconds apart) and selected to minimize conflict with room echoes and other “non-data” echoes or noise sources. For example, the echo patterns used to distinguish sources from room effects have a time (combination of delays) and frequency configuration that is distinguishable from room echoes. The frequency configuration can be selected by selecting pre-determined echoes within pre-determined frequency bands (e.g., selected from a range of high, mid, low bands within a signal coding range selected to not be audible by humans, but still within audible capture range of a typical cell phone microphone).

Robustness and reliability is further enhanced by signal detector design. Detector design includes pre-filtering the signal to remove unwanted portions of the signal and noise. It also includes accumulating energy over time to improve signal to noise ratio. For example, a detector uses a series of correlators that measure the autocorrelation in the neighborhood of the predetermined discrete delays in the symbol alphabet. The energy accumulated over time at the pre-determined delays is evaluated to identify whether an echo pattern corresponding to a data symbol or symbols is present.

Preferably, the signal processor that introduces the second layer is an inexpensive circuit that is connected in line in the electrical path of the audio signal from the sound system amplifier to the loudspeaker. One implementation of such a circuit is the bucket brigade circuit described in this document. These circuits can be made to be configurable by selective turning on or adjusting the gain of the delay signals that are introduced into the audio signal passing through the device.

An alternative way to implement the second layer is to introduce a set of frequency tones. These tones can be adjusted in amplitude according to audio masking models. One form of signal processor for inserting these tones is to add oscillator circuits at selected frequencies (e.g., three or four selected tones from a set of 10 predetermined tones). A composite signal is constructed by selecting a combination of oscillator outputs preferably high enough in the human auditory range to be less audible, yet low enough to be robust against ambient noise and other noise sources introduced through microphone capture. Also the selected tones must be reliably detected by the microphone, and thus, must not be distorted significantly in the microphone capture process.

Complementary detectors for this form of frequency modulation use filter banks around the pre-determined frequency tones. Energy at these frequencies is accumulated over time and then analyzed to identify a combination of tones corresponding to a predetermined identifier or data symbol.

Yet another way to differentiate a source or group of sources is to introduce a temporal perturbation or jitter. In this approach, time scale changes are applied to corresponding portions of an audio signal in a pattern associated with a source or group of sources to distinguish that source or group from other sources. This pattern of time scale changes can be detected by, for example, synchronizing with a chip sequence. For example, a search for a correlation peak of the chip sequence at different time scales indicates that time scale shift relative to a known time scale at which the chip sequence was encoded.

In a content fingerprint approach, the receiver uses content fingerprinting to identify the source. For a particular implementation, there is a well defined set of possible clips that will

be used for a localization scheme, and each is registered in a content fingerprint database. Sound segments captured in the receiver are processed to derive fingerprints (e.g., a robust hash or vector of features) that are then matched against the registered fingerprints in the database. The matching fingerprint in the database indicates the source.

In an implementation using synthesized audio, each loudspeaker plays specially designed audio clip that sounds pleasant to the ear but carries the hidden payload—maybe by slight adjustment of the frequencies on a MIDI sequence or shaping a watermark signal to sound like ocean waves or fountain sounds.

The closest source can be identified based on its unique identifier, using any of the identifications schemes above. It may also be determined using strength of signal analyses. One particular analysis using watermarks is to encode watermarks at successively different strengths and then determine the closest source as the one in which the weakest of these watermarks is detected.

When two or more sources can be detected in the audio captured at the mobile device, forms of triangulation based positioning can be performed using estimates of direction or distance of the mobile devices relative to the sources.

Ambient Capture

Previously, we outlined techniques for uniquely identifying the source by generating source signals that can be identified in the receiver. This application requires design of signaling techniques that do not degrade the quality of the background sound and yet are reliably detected from ambient sound captured through a mobile device's microphone.

FIG. 3 is a flow diagram of a process for analyzing an ambient audio signal to detect and identify an audio source signal. This process is preferably implemented within the mobile device. However, aspects of the process can be distributed to another device by packaging data for a processing task and sending to another computer or array of computers for processing and return of a result (e.g., to a cloud computing service). In block 130, control of the audio stream captured in the microphone is obtained. The audio stream is digitized and buffered.

In block 132, the buffered audio samples are filtered to isolate modulated feature locations (in the case of a digital watermark or synthetic data signal) or to isolate features of a content fingerprint.

Next, in block 134, a digital watermark decoder analyzes the filtered content to decode one or more watermark signals. As explained previously, encoded data is modulated onto features by modifying the features. This modulation is demodulated from features to produce estimates of the encoded data signal. These estimates are accumulated over a detection window to improve signal detection. The inverse of the data encoding provides a payload, comprising an identifier. For example, one embodiment mentioned above uses a spread spectrum carrier and convolution codes to encode a first watermark layer. In one implementation, the first layer conveys a 32 bit payload and a 24 bit CRC computed from the 32 bit payload. The combined 56 bits are encoded with a one-third rate convolution encoder to generate 168 encoded bits. Each of these bits modulates a 100 chip carrier signal in a DSSS protocol. The 100 chip sequence are mapped sequentially in time, with each chip mapping to 2-3 audio samples at 16 KHz sample rate.

The detector demodulates the carrier signal which provides a weighted bit estimate. A soft error correction decoder uses a Viterbi decoder for convolution decoding of a payload of data symbols. The demodulation is implemented as a sliding correlator that extracts chip estimates. These chip estimates

are weighted by a correlation metric and input to the Viterbi decoder, which in turn, produces a 56 bit decoded output. If the CRC succeeds, the first layer identifier is deemed detected. If not, the sliding correlator shifts and repeats the process. This first robust watermark layer provides a source identifier, identifying at least the network neighborhood in which the receiving device is located.

A second layer detector then operates portions of audio from which the first layer was successfully detected and decodes a second layer identifier, if present. This detector applies an echo or frequency tone detector, for example, using the approach described previously. The autocorrelation detector, for instance, takes a low pass filtered version of the audio, and then executes a shift, multiply and add to compute autocorrelation for pre-determined delays.

For content fingerprints, the features are hashed into a feature vector that is matched with pre-registered feature vectors in a database. For an application of this type, the library of unique content fingerprints is relatively small and can be stored locally. If necessary, however, the fingerprint matching can be done remotely, with the remote service executed on a server returning the source identifier of the matching source signal.

The source identifier obtained from processing block 134 is used to look up the associated location parameters for the source. If two or more source identifiers are detected, a further analysis is done on detection metrics to estimate which is the dominant source. The source identifier with the stronger detection metrics is identified as the closest source.

FIG. 4 is a flow diagram of a process for determining distance from an audio source signal by analyzing strength of signal metrics. This process is designed to follow initial detection of a source signal, such as the process of FIG. 3. In block 140, the detection of a robust signal layer provides a frame of reference within the buffered audio in the device to make more granular assessments of weak watermark data. For example, the block boundaries of the chip sequences for which the first layer payload is successfully detected provide synchronization for further operations. In block 142, signal metrics are computed. One metric is a correlation metric in which the detected watermark's encoded data signal is re-generated after error correction and then compared with the input to the soft decision decoder. This comparison provides a measure of correlation strength between the expected signal and the extracted signal prior to error correction. This approach allows the payload to provide a source identifier, and the strength metric to provide an estimate of distance from the source. The correlation strength metric may be further refined by measuring the encoded source signal energy at particular frequencies, and providing a series of signal strength metrics at these frequencies. For instance, frequency components of the first layer or a separate second layer are distinctly measured. One signal strength metric based on these measurements is to compute a ratio of encoded data signal strength at low frequency feature locations to higher frequency feature locations. This particular metric can be derived from a special purpose watermark signal layer that is designed to estimate distance from source. Alternatively, the modulation of frequency tones can provide the source identifier, and the strength ratios computed between high and low frequency components of distinct watermarks provide the strength metric. In both cases, as distance increases from the source, the strength metric decreases.

In block 144, the detection metrics are used to look up distance estimates. In block 146, the source identifiers and associated detection metrics are supplied to a position calculator. The position calculator looks up location of the sources

from the source IDs and then enters location and distance parameters and solves for an estimate of position of the mobile device location. To simplify the calculation, the solution set is reduced to a set of discrete locations in the network. The position is determined by finding the solution that intersects the position of these discrete locations.

FIG. 5 is a flow diagram of a process for determining the time difference of arrival of audio signals from distinct audio sources. In one implementation, the detector measures the difference in arrival time of distinct source signals that are encoded using the DSSS data signal approach described previously. For this implementation, we select a chip sequence length based on the spacing of nodes in the positioning network. In particular, we choose a length of chip sequence at least equal to the largest delay between source signal arrivals that we expect. If the maximum speaker distance is 50 feet, then the maximum difference in distance from source 1 to source 2 is around 50 feet. At a sample rate of 16 kHz, the chip sequence should be at least 800 samples.

In block 150, the detector executes a search for the encoded data signals. For the DSSS data encoding protocol, the detector executes a slide, correlate, and trial decode process to detect a valid watermark payload. In block 152, it then seeks to differentiate source signals from different sources. This differentiation is provided by the unique payloads and/or unique signal characteristics of the source signals.

In block 154, the detector measures the time difference between one or more pairs of distinct signal sources. The identifier and time differences for a pair of distinct source signals received at the device is then provided to a position calculator in block 156.

In block 158, a position calculator uses the data to estimate the mobile device position. It uses the TDOA approach outlined previously.

We have described alternative approaches for integrating audio positioning signals into an audio sound system to calculate position of a mobile device from analysis of the source signal or signals captured through the microphone of the device. These approaches can be used in various configurations and combinations to provide position and navigation at the mobile device. There are a variety of enhancements that can be used without interfering with the primary function of the audio playback equipment to provide background and public address programming.

An enhancement is to adapt watermark strength based on sensing the ambient sound level. As ambient sound level increases, the watermark signal is increased accordingly to stay within the higher masking threshold afforded by the ambient sound.

Another enhancement is to provide the host signal sets to the receiver, which is then used to do non-blind watermark detection. In such detection, the knowledge of the host signal is used to increase recoverability of the encoded data. For example, it can be used to remove host signal interference in cases where the host signal interferes with the watermark signal. As another example, it can be used to ascertain content dependent parameters of the watermark encoding, such as the gain applied to the watermark signal based on the host signal characteristics.

Another enhancement is to model the room acoustics for a particular neighborhood of speakers in the location network, and then use this model to enable reversal of room acoustic effects for audio captured by receivers in that neighborhood.

The range of the loudspeakers is limited, so triangulation may not always be necessary to deduce location of the mobile device. One can infer proximity information from just one loud-speaker.

A combination of fragile and robust watermarks can be used—at farther distances, fragile watermarks will not be recovered, which provides an indicator of distance from a source. Source signals are encoded with a primary identifier in a first layer, and then additional secondary layers, each at robustness level (e.g., amplitude or frequency band) that becomes undetectable as distance from the source increases.

Additionally, multiple phones in the same neighborhood can communicate with each other (e.g., using Wi-Fi protocols or Bluetooth protocols) and exchange information based on relative positioning.

Various aspects of the above techniques are applicable to different types of source signals that are detectable on mobile devices, such as mobile telephones. For example, mobile phones are equipped with other types of sensors that can detect source signals corresponding to network locations, such as RFID or NFC signals.

CONCLUDING REMARKS

Having described and illustrated the principles of the technology with reference to specific implementations, it will be recognized that the technology can be implemented in many other, different, forms. To provide a comprehensive disclosure without unduly lengthening the specification, applicants incorporate by reference the patents and patent applications referenced above.

The methods, processes, and systems described above may be implemented in hardware, software or a combination of hardware and software. For example, the signal processing operations for distinguishing among sources and calculating position may be implemented as instructions stored in a memory and executed in a programmable computer (including both software and firmware instructions), implemented as digital logic circuitry in a special purpose digital circuit, or combination of instructions executed in one or more processors and digital logic circuit modules. The methods and processes described above may be implemented in programs executed from a system's memory (a computer readable medium, such as an electronic, optical or magnetic storage device). The methods, instructions and circuitry operate on electronic signals, or signals in other electromagnetic forms. These signals further represent physical signals like image signals captured in image sensors, audio captured in audio sensors, as well as other physical signal types captured in sensors for that type. These electromagnetic signal representations are transformed to different states as detailed above to detect signal attributes, perform pattern recognition and matching, encode and decode digital data signals, calculate relative attributes of source signals from different sources, etc.

The above methods, instructions, and hardware operate on reference and suspect signal components. As signals can be represented as a sum of signal components formed by projecting the signal onto basis functions, the above methods generally apply to a variety of signal types. The Fourier transform, for example, represents a signal as a sum of the signal's projections onto a set of basis functions.

The particular combinations of elements and features in the above-detailed embodiments are exemplary only; the interchanging and substitution of these teachings with other teachings in this and the incorporated-by-reference patents/applications are also contemplated.

We claim:

1. A method of determining position of a mobile device comprising:

receiving, at one location, audio signals from two or more different audio sources that overlap in time, in a microphone of the mobile device, wherein the audio signals sound substantially similar to a human listener, yet have different characteristics to distinguish among the different audio sources;

distinguishing the audio signals from each other based on two or more layers of distinguishing characteristics determined from the audio signals, wherein a first layer provides information to identify a group of audio sources, and a second layer provides information to identify a particular audio source within the group;

based on identifying particular audio sources, determining location of the particular audio sources;

determining position of the mobile device based on the locations of the particular audio sources.

2. The method of claim 1 comprising determining position of the mobile device based on the locations of the audio sources and relative attributes of the received audio signals.

3. The method of claim 2 wherein the relative attributes comprise time of arrival of the received audio signals.

4. The method of claim 2 wherein the relative attributes comprise strength of signal metrics derived from analyzing strength of audio signals from different sources.

5. The method of claim 1 wherein the mobile device comprises a mobile telephone.

6. The method of claim 1 wherein distinguishing the audio signals comprises detecting a digital watermark encoded into host audio content.

7. The method of claim 1 wherein distinguishing the audio signals comprises differentiating source by performing a content fingerprint recognition.

8. The method of claim 1 wherein the distinguishing comprises detecting a pattern of frequency tones.

9. The method of claim 1 wherein the distinguishing comprises detecting a pattern of alterations introduced into the audio signals prior to output of the audio signals from respective source devices, wherein the alterations are separately detectable, yet the output audio signals are perceived to be the same signal by human listeners.

10. The method of claim 9 wherein the pattern of alterations is inserted by a signal processing circuit in a path from an audio playback system to a speaker.

11. The method of claim 9 wherein the pattern comprises a temporal jitter.

12. A method of determining position of a mobile device comprising:

receiving audio signals from two or more different audio sources in a microphone of the mobile device, wherein the audio signals sound substantially similar to a human listener, yet have different characteristics to distinguish among the different audio sources;

distinguishing the audio signals from each other based on two or more layers of distinguishing characteristics determined from the audio signals, wherein a first layer provides information to identify a group of audio sources, and a second layer provides information to identify a particular audio source within the group;

based on identifying particular audio sources, determining location of the particular audio sources; and

determining position of the mobile device based on the locations of the particular audio sources; wherein the distinguishing comprises detecting an echo pattern associated with a group of sources or particular audio source.

13. A position system comprising:

a microphone for receiving, at one location, two or more time-overlapping audio source signals in an audible range and converting to an electronic signal, wherein the audio signals sound substantially similar to a human listener, yet have different characteristics to distinguish among the different audio sources; and

one or more processors for accessing the electronic signal corresponding to received audio signals and distinguishing said two or more time-overlapping audio signals from each other based on two or more layers of distinguishing characteristics determined from the audio signals, wherein a first layer provides information to identify a group of audio sources, and a second layer provides information to identify a particular audio source within the group, and for determining location of the particular audio sources based on identifying the particular audio sources and determining position of the mobile device based on the locations of the particular audio sources.

14. An audio signal generation system comprising:

a controller for controlling an audio signal output by an audio playback device, the controller establishing a first layer of characteristics in the audio signal for identifying a group of loudspeakers connected to the audio playback device; and

a signal processor connected between the audio playback device and a first loudspeaker to introduce a second layer of signal characteristics into the audio signal to distinguish the audio signal from the first loudspeaker to which the signal processor is connected; and

a database storing an association between layers of unique characteristics of the audio signals and position of the loudspeakers, the database being responsive to queries to provide position of a loudspeaker corresponding to unique characteristics derived from audio signals from the loudspeakers.

15. The system of claim 14 wherein the signal processor comprises frequency oscillators for introducing a pattern of frequency tones associated with a particular loudspeaker to which the signal processor is connected.

16. An audio signal generation system comprising:

a controller for controlling an audio signal output by an audio playback device, the audio signal comprising a first layer of characteristics for identifying a group of loudspeakers connected to the audio playback device; and

a signal processor connected between the audio playback device and a first loudspeaker to introduce a second layer of signal characteristics into the audio signal to distinguish the audio signal from the first loudspeaker to which the signal processor is connected; and

a database storing an association between layers of unique characteristics of the audio signals and position of the loudspeakers, the database being responsive to queries to provide position of a loudspeaker corresponding to unique characteristics derived from audio signals from the loudspeakers; wherein the signal processor comprises a delay line circuit for introducing a pattern of echoes associated with a particular loudspeaker to which the delay line circuit is connected.

17. A method of determining position of a mobile device comprising:

receiving, at one location, source signals from two or more different sources that overlap in time, in a sensor of the mobile device;

21

distinguishing the source signals from each other based on two or more layers of distinguishing characteristics determined from the source signals, wherein a first layer provides information to identify a group of sources, and a second layer provides information to identify a particular source within the group;
 based on identifying particular sources, determining location of the particular sources; and
 determining position of the mobile device based on the locations of the particular sources and relative attributes of the received source signals.

18. A method of determining position of a mobile device comprising:

receiving, at one location, audio signals from two or more different audio sources that overlap in time, in a microphone of the mobile device, wherein the audio signals sound substantially similar to a human listener, yet have different characteristics to distinguish among the different audio sources;

22

distinguishing the audio signals from each other based on distinguishing characteristics determined from the audio signals, wherein the distinguishing characteristics provides information to identify a particular audio source;

based on identifying particular audio sources, determining location of the particular audio sources; and

determining position of the mobile device based on the locations of the particular audio sources and a relative attribute of the received audio signals.

19. The method of claim **18** wherein the relative attribute comprises time of arrival of distinct audio signals.

20. The method of claim **18** wherein the relative attribute comprises strength of signal from distinct audio signal sources.

* * * * *