(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2013 (10.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/006700 A2

(51) International Patent Classification: *G10L 21/02* (2013.01)

(21) International Application Number: PCT/US2012/045556

(22) International Filing Date: 5 July 2012 (05.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

1111474.1	5 July 2011 (05.07.2011)	GB
13/212,688	18 August 2011 (18.08.2011)	US

(71) Applicant (*for all designated States except US*): **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **STROMMER, Stefan**; c/o Skype, International Patents, 70 Sir John Rogerson's Quay, Dublin 2 (IE). **SORENSEN, Karsten Vandborg**; c/o Skype, International Patents, 70 Sir John Rogerson's Quay, Dublin 2 (IE).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: PROCESSING AUDIO SIGNALS

(57) Abstract: A computer-implemented system and method are described for improving the QoE of real-time video sessions between mobile users. For example, a method according to one embodiment of the invention comprises: configuring one or more servers on the perimeter of a service provider network; receiving a request from a first mobile device to establish a real-time communication session with a second mobile device; providing the first and second mobile devices with networking information for connecting to the servers; and establishing the realtime communication session through the server.

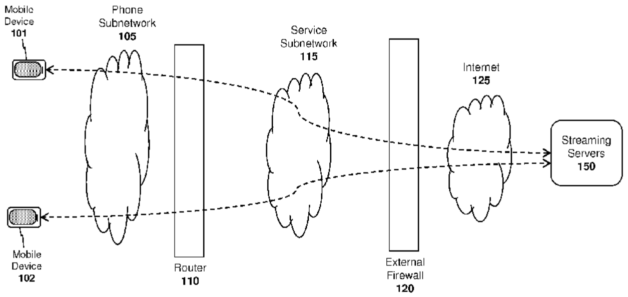

Fig. 1a
(prior art)

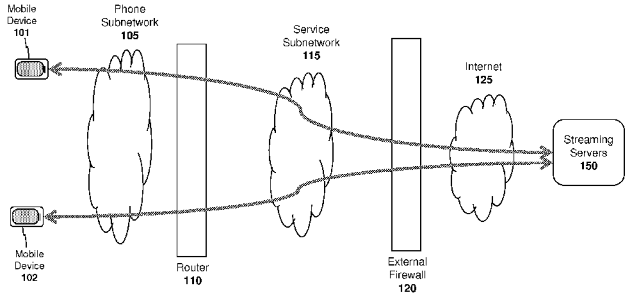

Fig. 1b
(prior art)

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

PROCESSING AUDIO SIGNALS

Field of the Invention

This invention relates to processing audio signals during a communication session.

Background

Communication systems allow users to communicate with each other over a network. The network may be, for example, the internet or the Public Switched Telephone Network (PSTN). Audio signals can be transmitted between nodes of the network, to thereby allow users to transmit and receive audio data (such as speech data) to each other in a communication session over the communication system.

A user device may have audio input means such as a microphone that can be used to receive audio signals, such as speech from a user. The user may enter into a communication session with another user, such as a private call (with just two users in the call) or a conference call (with more than two users in the call). The user's speech is received at the microphone, processed and is then transmitted over a network to the other user(s) in the call.

As well as the audio signals from the user, the microphone may also receive other audio signals, such as background noise, which may disturb the audio signals received from the user.

The user device may also have audio output means such as speakers for outputting audio signals to the user that are received over the network from the user(s) during the call. However, the speakers may also be used to output audio signals from other applications which are executed at the user device. For example, the user device may be a TV which executes an application such as a communication client for communicating over the network. When the user device is engaging in a call, a microphone connected to the user device is intended to receive speech or other audio signals provided by the user intended for transmission to the other user(s) in the call. However, the microphone may pick up unwanted audio signals which are output from the speakers of the user device. The unwanted audio signals output from the user device may contribute to disturbance to the audio signal received at the microphone from the user for transmission in the call.

In order to improve the quality of the signal, such as for use in the call, it is desirable to suppress unwanted audio signals (the background noise and the unwanted audio signals output from the user device) that are received at the audio input means of the user device.

The use of stereo microphones and microphone arrays in which a plurality of microphones operate as a single device are becoming more common. These enable use of extracted spatial information in addition to what can be achieved in a single microphone. When using such devices one approach to suppress unwanted audio signals is to apply a beamformer. Beamforming is the process of trying to focus the signals received by the microphone array by applying signal processing to enhance sounds coming from one or more desired directions. For simplicity we will describe the case with only a single desired direction in the following, but the same method will apply when there are more directions of interest. The beamforming is achieved by first estimating the angle from which wanted signals are received at the microphone, so-called Direction of Arrival ("DOA") information. Adaptive beamformers use the DOA information to filter the signals from the microphones in an array to form a beam that has a high gain in the direction from which wanted signals are received at the microphone array and a low gain in any other direction.

While the beamformer will attempt to suppress the unwanted audio signals coming from unwanted directions, the number of microphones as well as the shape and the size of the microphone array will limit the effect of the beamformer, and as a result the unwanted audio signals suppressed, but remain audible.

For subsequent single channel processing, the output of the beamformer is commonly supplied to single channel noise reduction stage as an input signal. Various methods of implementing single channel noise reduction have previously been proposed. A large majority of the single channel noise reduction methods in use are variants of spectral subtraction methods.

The spectral subtraction method attempts to separate noise from a speech plus noise signal. Spectral subtraction involves computing the power spectrum of a speech-plus-noise signal and obtaining an estimate of the noise spectrum. The power spectrum of the speech-plus-noise signal is compared with the estimated noise spectrum. The noise reduction can for example be implemented by subtracting the magnitude of the noise spectrum from the magnitude of the speech plus noise spectrum. If the speech-plus-noise signal has a high Signal-plus-Noise to Noise Ratio (SNNR) only very little noise reduction is applied. However if the speech-plus-noise signal has a low SNNR the noise reduction will significantly reduce the noise energy.

A problem with spectral subtraction is that it often distorts the speech and results in temporally and spectrally fluctuating gain changes leading to the appearance of a type of residual noise often referred to as musical tones, which may affect the transmitted speech quality in the call. Varying degrees of this problem also occur in the other known methods of implementing single channel noise reduction.

Summary of Invention

According to a first aspect of the invention there is provided a method of processing audio signals during a communication session between a user device and a remote node, the method comprising: receiving a plurality of audio signals at audio input means at the user device including at least one primary audio signal and unwanted signals; receiving direction of arrival information of the audio signals at a noise suppression means; providing to the noise suppression means known direction of arrival information representative of at least some of said unwanted signals; and processing the audio signals at the noise suppression means to treat as noise, portions of the signal identified as unwanted dependent on a comparison between the direction of arrival information of the audio signals and the known direction of arrival information.

Preferably, the audio input means comprises a beamformer arranged to: estimate at least one principal direction from which the at least one primary audio signal is received at the audio input means; and process the plurality of audio signals to generate a single channel audio output signal by forming a beam in the at least one principal direction and substantially suppressing audio signals from any direction other than the principal direction.

Preferably, the single channel audio output signal comprises a sequence of frames, the noise suppression means processing each of said frames in sequence.

Preferably, direction of arrival information for a principal signal component of a current frame being processed is received at the noise suppression means, the method further comprising: comparing the direction of arrival of information for the principal signal component of the current frame and the known direction of arrival information.

The known direction of arrival information includes at least one direction from which far-end signals are received at the audio input means. Alternatively, or additionally, the known direction of arrival information includes at least one classified direction, the at least one classified direction being a direction from which at least one unwanted audio signal arrives at the audio input means and is identified based on the signal characteristics of the at least one unwanted audio signal. Alternatively, or additionally, the known direction of arrival information includes at least one principal direction from which the at least one primary audio signal is received at the audio input means. Alternatively, or additionally, the known direction of arrival information further includes the beam pattern of the beamformer.

In one embodiment, the method further comprises: determining whether the principal signal component of the current frame is an unwanted signal based on said comparison; and applying maximum attenuation to the current frame being processed if it is determined that the principal signal component of the current frame is an unwanted signal. The principal signal component of the current frame may be determined to be an unwanted signal if: the principal signal component is received at the audio input means from the at least one direction from which far-end signals are received at the audio input means; or the principal signal component is received at the audio input means from the at least one classified direction; or the principal signal component is not received at the audio input means from the at least one principal direction.

The method may further comprise: receiving the plurality of audio signals and information on the at least one principal direction at signal processing means;

processing the plurality of audio signals at the signal processing means using said information on the at least one principal direction to provide additional information to the noise suppression means; and applying a level of attenuation to the current frame being processed at the noise suppression means in dependence on said additional information and said comparison.

Alternatively, the method may further comprise: receiving the single channel audio output signal and information on the at least one principal direction at signal processing means; processing the single channel audio output signal at the signal processing means using said information on the at least one principal direction to provide additional information to the noise suppression means; and applying a level of attenuation to the current frame being processed at the noise suppression means in dependence on said additional information and said comparison.

The additional information may include: an indication on the desirability of the principal signal component of the current frame, or a power level of the principal signal component of the current frame relative to an average power level of the at least one primary audio signal, or a signal classification of the principal signal component of the current frame, or at least one direction from which the principal signal component of the current frame is received at the audio input means.

Preferably, the at least one principal direction is determined by: determining a time delay that maximises the cross-correlation between the audio signals being received at the audio input means; and detecting speech characteristics in the audio signals received at the audio input means with said time delay of maximum cross-correlation.

Preferably, audio data received at the user device from the remote node in the communication session is output from audio output means of the user device.

The unwanted signals may be generated by a source at the user device, said source comprising at least one of: audio output means of the user device; a source of activity at the user device wherein said activity includes clicking activity comprising button clicking activity, keyboard clicking activity, and mouse clicking activity. Alternatively, the unwanted signals are generated by a source external to the user device.

Preferably, the at least one primary audio signal is a speech signal received at the audio input means.

According to a second aspect of the invention there is provided user device for processing audio signals during a communication session between a user device and a remote node, the user terminal comprising: audio input means for receiving a plurality of audio signals including a at least one primary audio signal and unwanted signals; and noise suppression means for receiving direction of arrival information of the audio signals and known direction of arrival information representative of at least some of said unwanted signals, the noise suppression means configured to process the audio signals by treating as noise, portions of the signal identified as unwanted dependent on a comparison between the direction of arrival information of the audio signals and the known direction of arrival information.

According to a third aspect of the invention there is provided a computer program product comprising computer readable instructions for execution by computer processing means at a user device for processing audio signals during a communication session between the user device and a remote node, the instructions comprising instructions for carrying out the method according to the first aspect of the invention.

In the following described embodiments, direction of arrival information is used to refine the decision of how much suppression to apply in subsequent single channel noise reduction methods. As most single channel noise reduction methods have a maximum suppression factor that is applied to the input signal to ensure a natural sounding but attenuated background noise, the direction of arrival information will be used to ensure that the maximum suppression factor is applied when the sound is arriving from any other angle than what the beamformer focuses on. For example, in the case of a TV playing out, maybe with a lowered volume, through the same speakers as are used for playing out the far end speech, a problem is that the output will be picked up by the microphone. With described embodiments of the present invention, it would be detected that the audio is arriving from the angle of the speakers and a maximum noise reduction would be applied in addition to the attempted suppression by the beamformer. As a result, the undesired signal would be less audible and therefore less disturbing to the far end speaker, and due to the reduced energy it would lower the average bit rate used for transmitting the signal to the far end.

Detailed Description

For a better understanding of the present invention and to show how the same may be put into effect, reference will now be made, by way of example, to the following drawings in which:

Figure 1 shows a communication system according to a preferred embodiment;

Figure 2 shows a schematic view of a user terminal according to a preferred embodiment;

Figure 3 shows an example environment of the user terminal;

Figure 4 shows a schematic diagram of audio input means at the user terminal according to one embodiment;

Figure 5 shows a diagram representing how DOA information is estimated in one embodiment.

In the following embodiments of the invention, a technique is described in which, instead of fully relying on the beamformer to attenuate sounds that are not coming from the direction of focus, using the DOA information in the subsequent single channel noise reduction method ensures maximum single channel noise suppression of sounds from any other direction than the ones the beamformer is focussed on. This is a significant advantage when the undesired signal can be distinguished from the desired nearend speech signal by using spatial information. Examples of such sources are loudspeakers playing music, fans blowing, and doors closing.

By using signal classification the direction of other sources can also be found. Examples of such sources could be, e.g. cooling fans / air conditioning systems, music playing in the background, and keyboard taps.

Two approaches can be taken: Firstly, undesired sources that are arriving from certain directions can be identified and the angles excluded from the angles where a noise suppression gain higher than the one used for maximum suppression is allowed. It would for example be possible to ensure that segments of audio from a certain undesired direction are scaled down as if the signal contained only noise. In practice the noise estimate can be set equal to the input signal for such a segment and consequently the noise reduction method would then apply maximum attenuation.

Secondly, noise reduction can be made less sensitive to speech in any other direction than the ones where we expect nearend speech to arrive from. That is, when calculating the gains to apply to the noisy signal as a function of the signal-plus-noise to noise ratio, the gain as a function of signal-plus-noise to noise ratio would also depend on how desired we consider the angle of the incoming speech to be. For desired directions the gain as a function of a given signal-plus-noise to noise ratio would be higher than for a less desired direction. The second method would ensure that we do not adjust based on moving noise sources which do not arrive from the same direction as the primary speaker(s), and which also have not been detected to be a source of noise.

Embodiments of the invention are particularly relevant in monophonic sound reproduction (often referred to as mono) applications with a single channel. Noise reduction in stereo applications (where there is two or more independent audio channels) is not typically carried out by independent single channel noise reduction methods, but rather by a method which ensures that the stereo image is not distorted by the noise reduction method.

Reference is first made to Figure 1, which illustrates a communication system 100 of a preferred embodiment. A first user of the communication system (User A 102) operates a user device 104. The user device 104 may be, for example a mobile phone, a television, a personal digital assistant ("PDA"), a personal computer ("PC") (including, for example, Windows™, Mac OS™ and Linux™ PCs), a gaming device or other embedded device able to communicate over the communication system 100.

The user device 104 comprises a central processing unit (CPU) 108 which may be configured to execute an application such as a communication client for communicating over the communication system 100. The application allows the user device 104 to engage in calls and other communication sessions (e.g. instant messaging communication sessions) over the communication system 100. The user device 104 can communicate over the communication system 100 via a network 106, which may be, for example, the Internet or the Public Switched Telephone Network (PSTN). The user device 104 can transmit data to, and receive data from, the network 106 over the link 110.

Figure 1 also shows a remote node with which the user device 104 can communicate over the communication system 100. In the example shown in Figure 1, the remote node is a second user device 114 which is usable by a second user 112 and which comprises a CPU 116 which can execute an application (e.g. a communication client) in order to communicate over the communication network 106 in the same way that the user device 104 communicates over the communications network 106 in the communication system 100. The user device 114 may be, for example a mobile phone, a television, a personal digital assistant ("PDA"), a personal computer ("PC") (including, for example, Windows™, Mac OS™ and Linux™ PCs), a gaming device or other embedded device able to communicate over the communication system 100. The user device 114 can transmit data to, and receive data from, the network 106 over the link 118. Therefore User A 102 and User B 112 can communicate with each other over the communications network 106.

Figure 2 illustrates a schematic view of the user terminal 104 on which is executed the client. The user terminal 104 comprises a CPU 108, to which is connected a display 204 such as a screen, input devices such as keyboard 214 and a pointing device such as mouse 212. The display 204 may comprise a touch screen for inputting data to the CPU 108. An output audio device 206 (e.g. a speaker) is connected to the CPU 108. An input audio device such as microphone 208 is connected to the CPU 108 via noise suppression means 227. Although the noise suppression means 227 is represented in Figure 2 as a stand alone hardware device, the noise suppression means 227 could be implemented in software. For example the noise suppression means 227 could be included in the client.

The CPU 108 is connected to a network interface 226 such as a modem for communication with the network 106.

Reference is now made to Figure 3, which illustrates an example environment 300 of the user terminal 104.

Desired audio signals are identified when the audio signals are processed having been received at the microphone 208. During processing, desired audio signals are identified based on the detection of speech like qualities and a principal direction of a main speaker is determined. This is shown in Figure 3 where the main speaker (user 102) is shown as a source 302 of desired audio signals that arrives at the microphone 208 from a principal direction d1. Whilst a single main speaker is shown in Figure 3 for simplicity, it will be appreciated that any number of sources of wanted audio signals may be present in the environment 300.

Sources of unwanted noise signals may be present in the environment 300. Figure 3 shows a noise source 304 of an unwanted noise signal in the environment 300 that may arrive at the microphone 208 from a direction d3. Sources of unwanted noise signals include for example cooling fans, air-conditioning systems, and a device playing music.

Unwanted noise signals may also arrive at the microphone 208 from a noise source at the user terminal 104 for example clicking of the mouse 212, tapping of the keyboard 214, and audio signals output from the speaker 206. Figure 3 shows the user terminal 104 connected to microphone 208 and speaker 206. In Figure 3, the speaker 206 is a source of an unwanted audio signal that may arrive at the microphone 208 from a direction d2.

Whilst the microphone 208 and speaker 206 have been shown as external devices connected to the user terminal it will be appreciated that microphone 208 and speaker 206 may be integrated into the user terminal 104.

Reference is now made to Figure 4 which illustrates a more detailed view of microphone 208 and the noise suppression means 227 according to one embodiment.

Microphone 208 includes a microphone array 402 comprising a plurality of microphones, and a beamformer 404. The output of each microphone in the microphone array 402 is coupled to the beamformer 404. Persons skilled in the art will appreciate that to implement beamforming multiple inputs are needed. The microphone array 402 is shown in Figure 4 as having three microphones, it will be understood that this number of microphones is merely an example and is not limiting in any way.

The beamformer 404 includes a processing block 409 which receives the audio signals from the microphone array 402. Processing block 409 includes a voice activity detector (VAD) 411 and a DOA estimation block 413 (the operation of which will be described later).The processing block 409 ascertains the nature of the audio signals received by the microphone array 402, and based on detection of speech like qualities detected by the VAD 11 and DOA information estimated in block 413,one or more principal direction(s) of main speaker(s) is determined. The beamformer 404 uses the DOA information to process the audio signals by forming a beam that has a high gain in the direction from the one or more principal direction(s) from which wanted signals are received at the microphone array and a low gain in any other direction. Whilst it has been described above that the processing block 409 can determine any number of principal directions, the number of principal directions determined affects the properties of the beamformer e.g. less attenuation of the signals received at the microphone array from the other (unwanted) directions than if only a single principal direction is determined. The output of the beamformer 404 is provided on line 406 in the form of a single channel to be processed to the noise reduction stage 227 and then to an automatic gain control means (not shown in Figure 4).

Preferably, the noise suppression is applied to the output of the beamformer before the level of gain is applied by the automatic gain control means. This is because the noise suppression could theoretically slightly reduce the speech level (unintentionally) and the automatic gain control means would increase the speech level after the noise suppression and compensate for the slight reduction in speech level caused by the noise suppression.

DOA information estimated in the beamformer 404 is supplied to the noise reduction stage 227 and to signal processing circuitry 420.

The DOA information estimated in the beamformer 404 may also be supplied to the automatic gain control means. The automatic gain control means applies a level of gain to the output of the noise reduction stage 227. The level of gain applied to the channel output from the noise reduction stage 227 depends on the DOA information that is received at the automatic gain control means. The operation of the automatic gain control means is described in British Patent Application No. 1108885.3 and will not be discussed in further detail herein.

The noise reduction stage 227 applies noise reduction to the single channel signal. The noise reduction can be carried out in a number of different ways including by way of example only, spectral subtraction (for example, as described in the paper “Suppression of acoustic noise in speech using spectral subtraction” by Boll, S in Acoustics, Speech and Signal Processing, IEEE Transactions on, Apr 1979, Volume 27, Issue 2, pages 113 - 120).

This technique (as well as other known techniques) suppress components of the signal identified as noise so as to increase the signal-to-noise ratio, where the signal is the intended useful signal, such as speech in this case.

As described in more detail later, the direction of arrival information is used in the noise reduction stage to improve noise reduction and therefore enhance the quality of the signal.

The operation of DOA estimation block 413 will now be described in more detail with reference to Figure 5.

In the DOA estimation block 413, the DOA information is estimated by estimating the time delay e.g. using correlation methods, between received audio signals at a plurality of microphones, and estimating the source of the audio signal using the *a priori* knowledge about the location of the plurality of microphones.

Figure 5 shows microphones 403 and 405 receiving audio signals from an audio source 516. The direction of arrival of the audio signals at microphones 403 and 405 separated by a distance, d can be estimated using equation (1):

$$\theta = arcsin\left(\frac{\tau_D v}{d}\right) \qquad (1)$$

where *v* is the speed of sound, and τ_D is the difference between the times the audio signals from the source 516 arrive at the microphones 403 and 405 - that is, the time delay. The time delay is obtained as the time lag that maximises the cross-correlation between the signals at the outputs of the microphones 403 and 405. The angle θ may then be found which corresponds to this time delay.

It will be appreciated that calculating a cross-correlation of signals is a common technique in the art of signal processing and will not be describe in more detail herein.

The operation of the noise reduction stage 227 will now be described in further detail below. In all embodiments of the invention the noise reduction stage 227 uses DOA information known at the user terminal and represented by DOA block 227 and receives an audio signal to be processed. The noise reduction stage 227 processes the audio signals on a per-frame basis. A frame can, for example, be between 5 and 20 milliseconds in length, and according to one noise suppression technique are divided into spectral bins, for example, between 64 and 256 bins per frame.

The processing performed in the noise reduction stage 227 comprises applying a level of noise suppression to each frame of the audio signal input to the noise reduction stage 227. The level of noise suppression applied by the noise reduction stage 227 to each frame of the audio signal depends on a comparison between the extracted DOA information of the current frame being processed, and the built up knowledge of DOA information for various audio sources known at the user terminal. The extracted DOA information is passed on alongside the frame, such that it is used as an input parameter to the noise reduction stage 227 in addition to the frame itself.

The level of noise suppression applied by the noise reduction stage 227 to the input audio signal may be affected by the DOA information in a number of ways.

Audio signals that arrive at the microphone 208 from directions which have been identified as from a wanted source may be identified based on the detection of speech like characteristics and identified as being from a principal direction of a main speaker.

The DOA information 427 known at the user terminal may include the beam pattern 408 of the beamformer. The noise reduction stage 227 processes the audio input signal on a per-frame basis. During processing of a frame, the noise reduction stage 227 reads the DOA information of a frame to find the angle from which a main component of the audio signal in the frame was received at the microphone 208. The DOA information of the frame is compared with the DOA information 427 known at the user terminal. This comparison determines whether a main component of the audio signal in the frame being processed was received at the microphone 208 from the direction of a wanted source.

Alternatively or additionally, the DOA information 427 known at the user terminal may include the angle ∅ at which farend signals are received at the microphone 208 from speakers (such as 206) at the user terminal (supplied to the noise reduction stage 227 line 407).

Alternatively or additionally, the DOA information 427 known at the user terminal may be derived from a function 425 which classifies audio from different directions to locate a certain direction which is very noisy, possibly as a result of a fixed noise source.

When the DOA information 427 represents the principal wanted direction, and it is determined by comparison that a main component of the frame being processed is received at the microphone 208 from that principal direction. The noise reduction stage 227 determines a level of noise suppression using conventional methods described above.

In a first approach, if it is determined that a main component of the frame being processed is received at the microphone 208 from a direction other than a principal direction, the bins associated with the frame are all treated as though they are noise (even if a normal noise reduction technique would identify a good signal-plus-noise to noise ratio and thus not significantly suppress the noise). This may be done by setting the noise estimate equal to the input signal for such a frame and consequently the noise reduction stage would then apply maximum attenuation to the frame. In this way, frames arriving from directions other than the wanted direction can be suppressed as noise and the quality of the signal improved.

As mentioned above, the noise reduction stage 227 may receive DOA information from a function 425 which identifies unwanted audio signals arriving at the microphone 208 from noise source(s) in different directions. These unwanted audio signals are identified from their characteristics, for example audio signals from key taps on a keyboard or a fan have different characteristics to human speech. The angle at which the unwanted audio signals arrive at the microphone 208 may be excluded where a noise suppression gain higher than the one used for maximum suppression is allowed. Therefore when a main component of an audio signal in a frame being processed is received at the microphone 208 from an excluded direction the noise reduction stage 227 applies maximum attenuation to the frame.

A verification means 423 may be further included. For example, once one or more principal directions have been detected (based on the beam pattern 408 for example in the case of a beamformer), the client informs the user 102 of the detected principal direction via the client user interface and asks the user 102 if the detected principal direction is correct. This verification is optional as indicated by the dashed line in Figure 4.

If the user 102 confirms that the detected principal direction is correct, then the detected principal direction is sent to the noise reduction stage 227 and the noise reduction stage 227 operates as described above. The communication client may store the detected principal direction in memory 210, once the user 102 logs in to the client and has confirmed that a detected principal direction is correct, following subsequent log-ins to the client if a detected principal direction matches a confirmed correct principal direction in memory the detected principal direction is taken to be correct. This prevents the user 102 having to confirm a principal direction every time he logs into the client.

If the user indicates that the detected principal direction is incorrect, then the detected principal direction is not sent as DOA information to the noise reduction stage 227. In this case, the correlation based method (described above with reference to Figure 5) will continue to detect the principal direction and will only send the detected one or more principal directions once the user 102 confirms that the detected principal direction is correct.

In the first approach, the mode of operation is such that maximum attenuation can be applied to a frame being processed based on DOA information of the frame.

In a second approach, the noise reduction stage 227 does not operate in such a strict mode of operation.

In the second approach, when calculating the gains to apply to the audio signal in the frame as a function of the signal-plus-noise to noise ratio, the gain as a function of signal-plus-noise to noise ratio depends on additional information. This additional information can be calculated in a signal processing block (not shown in Figure 4).

In a first implementation the signal processing block may be implemented in the microphone 208. The signal processing block receives as an input the far-end audio signals from the microphone array 402 (before the audio signals have been applied to the beamformer 404), and also receives the information on the principal direction(s) obtained from the correlation method. In this implementation, the signal processing block outputs the additional information to the noise reduction stage 227.

In a second implementation the signal processing block may be implemented in the noise reduction stage 227 itself. The signal processing block receives as an input the single channel output signal from the beamformer 404, and also receives the information on the principal direction(s) obtained from the correlation method. In this implementation the noise reduction stage 227 may receive information indicating that the speakers 206 are active and can ensure that the principal signal component in the frame being processed is handled as noise only, provided that it is different from the angle of desired speech.

In both implementations the additional information calculated in the signal processing block is used by the noise reduction stage 227 to calculate the gain to apply to the audio signal in the frame being processed as a function of the signal-plus-noise to noise ratio.

The additional information may include for example the likelihood that desired speech will arrive from a particular direction/angle.

In this scenario the signal processing block provides, as an output, a value that indicates how likely the frame currently being processed by the noise reduction stage 277, contains a desired component that the noise reduction stage should preserve. The signal processing block quantifies the desirability of angles from which incoming speech is received at the microphone 208. For example if audio signals are received at the microphone 208 during echo, the angle at which these audio signals are received at the microphone 208 is likely to be an undesired angle since it is not desirable to preserve any far-end signals received from speakers (such as 206) at the user terminal.

In this scenario, the noise suppression gain as a function of signal-plus-noise to noise ratio applied to the frame by the noise reduction stage 227 is dependent on this quantified measure of desirability. For desired directions the gain as a function of a given signal-plus-noise to noise ratio would be higher than for a less desired direction i.e. less attenuation is applied by the noise reduction stage 227 for more desired directions.

The additional information may alternatively include the power of the principal signal component of the current frame relative to the average power of the audio signals received from the desired direction(s). In this scenario, the noise suppression gain as a function of signal-plus-noise to noise ratio applied to the frame by the noise reduction stage 227 is dependent on this quantified power ratio. The closer the power of the principal signal component is relative to the average power from the principal directions, the higher the gain as a function of a given signal-plus-noise to noise ratio applied by the noise reduction stage 227, i.e. less attenuation is applied.

The additional information may alternatively be a signal classifier output providing a signal classification of the principal signal component of the current frame. In this scenario, the noise reduction stage 227 may apply varying levels of attenuation to a frame wherein the main component of the frame is received at the microphone array 402 from a particular direction in dependence on the signal classifier output. Therefore if an angle is determined to be a non-desired direction, the noise reduction stage 227 may reduce noise from the non-desired direction more than speech from the same non-desired direction. This is possible and indeed practical if desired speech is expected to arrive from the non-desired direction. However, it has the major drawback that the noise will be modulated, i.e. the noise will be higher when the desired speaker is active, and the noise will be lower when an undesired speaker is active. Instead, it is preferable to slightly reduce the level of speech in signals from this direction. If not handling it exactly as noise by making sure to apply the same amount of attenuation, then by handling it as somewhere in between desired speech and noise. This can be achieved by using a slightly different attenuation function for non-desired directions.

The additional information may alternatively be the angle itself from which the principal signal component of the current frame is received at the audio input means. i.e. $\emptyset$ supplied to the noise reduction stage 227 on line 407. This enables the noise reduction stage to apply more attenuation as the audio source moves away from the principal direction(s).

In this second approach, more granularity is provided as the noise reduction stage 227 is able to operate in between the two extremes of handling a frame as noise only and as traditionally done in single-channel noise reduction methods. Therefore the noise reduction stage 227 can be made slightly more aggressive for audio signals arriving from undesired directions without handling it fully as if it was nothing but noise. That is, aggressive in the in the sense that we for example will apply some attenuation to the speech signal.

Whilst the embodiments described above have referred to a microphone 208 receiving audio signals from a single user 102, it will be understood that the microphone may receive audio signals from a plurality of users, for example in a conference call. In this scenario multiple sources of wanted audio signals arrive at the microphone 208.

While this invention has been particularly shown and described with reference to preferred embodiments, it will be understood to those skilled in the art that various changes in form and detail may be made without departing from the scope of the invention as defined by the appendant claims.

CLAIMS

1. A method of processing audio signals during a communication session between a user device and a remote node, the method comprising:

receiving a plurality of audio signals at audio input means at the user device including at least one primary audio signal and unwanted signals;

receiving direction of arrival information of the audio signals at a noise suppression means;

providing to the noise suppression means known direction of arrival information representative of at least some of said unwanted signals; and

processing the audio signals at the noise suppression means to treat as noise, portions of the signal identified as unwanted dependent on a comparison between the direction of arrival information of the audio signals and the known direction of arrival information.

2. The method according to claim 1, wherein the audio input means comprises a beamformer arranged to:

estimate at least one principal direction from which the at least one primary audio signal is received at the audio input means; and

process the plurality of audio signals to generate a single channel audio output signal by forming a beam in the at least one principal direction and substantially suppressing audio signals from any direction other than the principal direction, wherein the single channel audio output signal comprises a sequence of frames, the noise suppression means processing each of said frames in sequence.

3. The method according to any preceding claim, wherein direction of arrival of information for a principal signal component of a current frame being processed is received at the noise suppression means, the method further comprising:

comparing the direction of arrival of information for the principal signal component of the current frame and the known direction of arrival information, wherein the known direction of arrival information includes at least one of: (i) at least one direction from which far-end signals are received at the audio input means; (ii) at least one classified direction, the at least one classified direction being a direction from which at least one unwanted audio signal arrives at the audio input means and is identified based on the signal characteristics of the at least one unwanted audio signal; (iii) at least one principal direction from which the at least one primary audio signal is received at the audio input means; and (iv) the beam pattern of the beamformer.

4. The method according to claim 3, further comprising:

determining whether the principal signal component of the current frame is an unwanted signal based on said comparison;

applying maximum attenuation to the current frame being processed if it is determined that the principal signal component of the current frame is an unwanted signal; and determining that the principal signal component of the current frame is an unwanted signal if:

the principal signal component is received at the audio input means from the at least one direction from which far-end signals are received at the audio input means; or

the principal signal component is received at the audio input means from the at least one classified direction; or

the principal signal component is not received at the audio input means from the at least one principal direction.

5. The method according to claim 3, further comprising:

receiving the plurality of audio signals and information on the at least one principal direction at signal processing means;

processing the plurality of audio signals at the signal processing means using said information on the at least one principal direction to provide additional information to the noise suppression means; and

applying a level of attenuation to the current frame being processed at the noise suppression means in dependence on said additional information and said comparison, wherein the additional information includes one of: (i) an indication on the desirability of the principal signal component of the current frame, (ii) a power level of the principal signal component of the current frame relative to an average power level of the at least one primary audio signal; (iii) a signal classification of the principal signal component of the current frame; and (iv) at least one direction from which the principal signal component of the current frame is received at the audio input means.

6. The method according to any of claim 4 to 8, further comprising:

receiving the single channel audio output signal and information on the at least one principal direction at signal processing means;

processing the single channel audio output signal at the signal processing means using said information on the at least one principal direction to provide additional information to the noise suppression means; and

applying a level of attenuation to the current frame being processed at the noise suppression means in dependence on said additional information and said comparison, wherein the additional information includes one of: (i) an indication on the desirability of the principal signal component of the current frame, (ii) a power level of the principal signal component of the current frame relative to an average power level of the at least one primary audio signal; (iii) a signal classification of the principal signal component of the current frame; and (iv) at least one direction from which the principal signal component of the current frame is received at the audio input means.

7. The method according to any of claims 2 to 6, wherein the at least one principal direction is determined by:

determining a time delay that maximises the cross-correlation between the audio signals being received at the audio input means; and

detecting speech characteristics in the audio signals received at the audio input means with said time delay of maximum cross-correlation.

8. The method according any preceding claim, wherein the unwanted signals are generated by a source external to the user device or a source at the user device, said source comprising at least one of: audio output means of the user device; a source of activity at the user device wherein said activity includes clicking activity comprising button clicking activity, keyboard clicking activity, and mouse clicking activity.

9. A user device for processing audio signals during a communication session between the user device and a remote node, the user device comprising:

audio input means for receiving a plurality of audio signals including a at least one primary audio signal and unwanted signals; and

noise suppression means for receiving direction of arrival information of the audio signals and known direction of arrival information representative of at least some of said unwanted signals, the noise suppression means configured to process the audio signals by treating as noise, portions of the signal identified as unwanted dependent on a comparison between the direction of arrival information of the audio signals and the known direction of arrival information.

10. A computer program product comprising computer readable instructions for execution by computer processing means at a user device for processing audio signals during a communication session between the user device and a remote node, the instructions comprising instructions for carrying out the method according to any of claims 1 to 8.

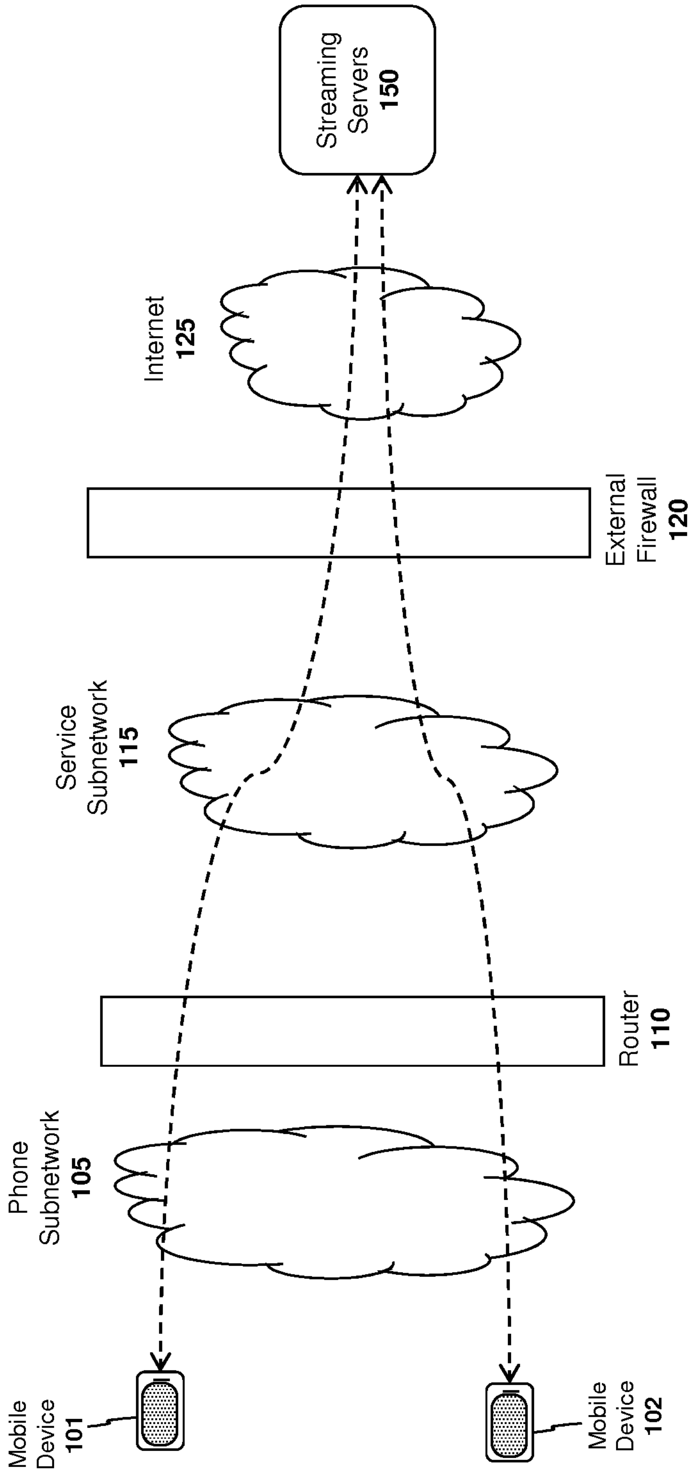

Fig. 1a
(prior art)

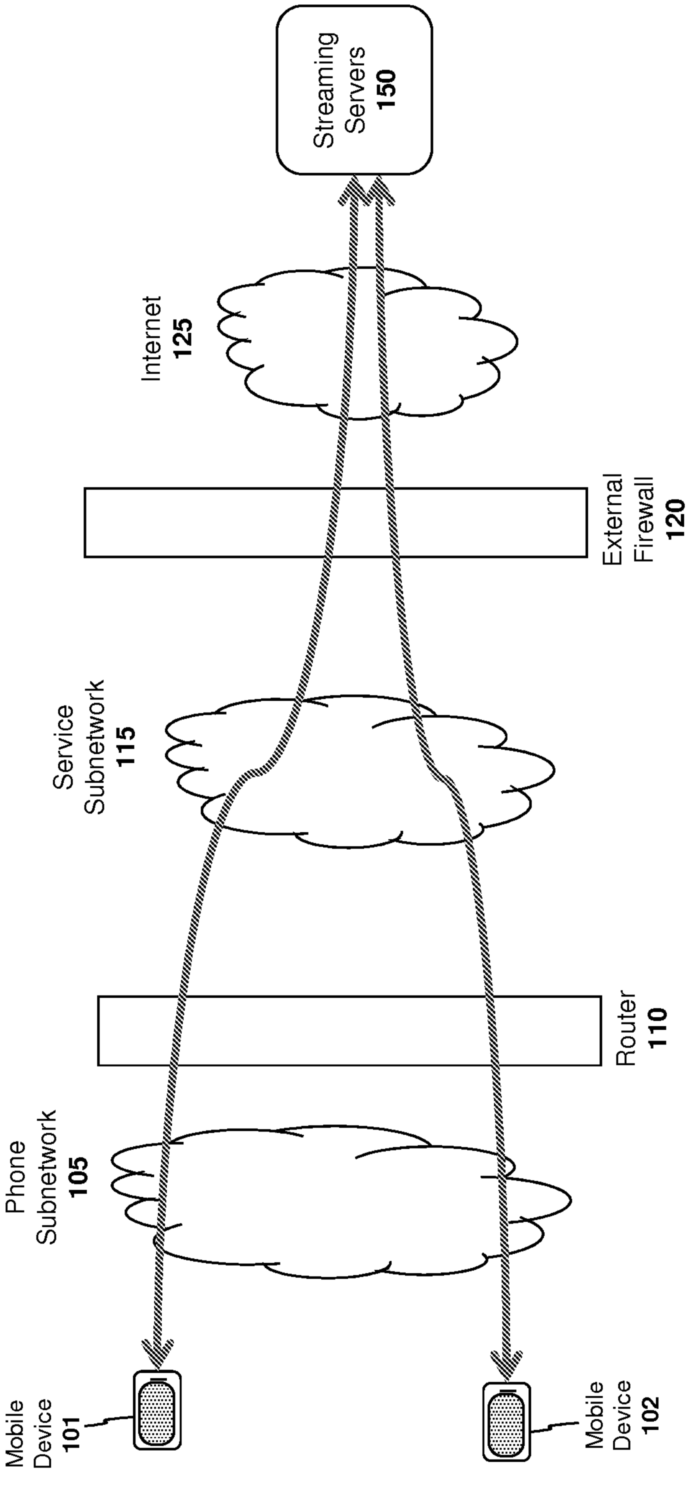

Fig. 1b
(prior art)

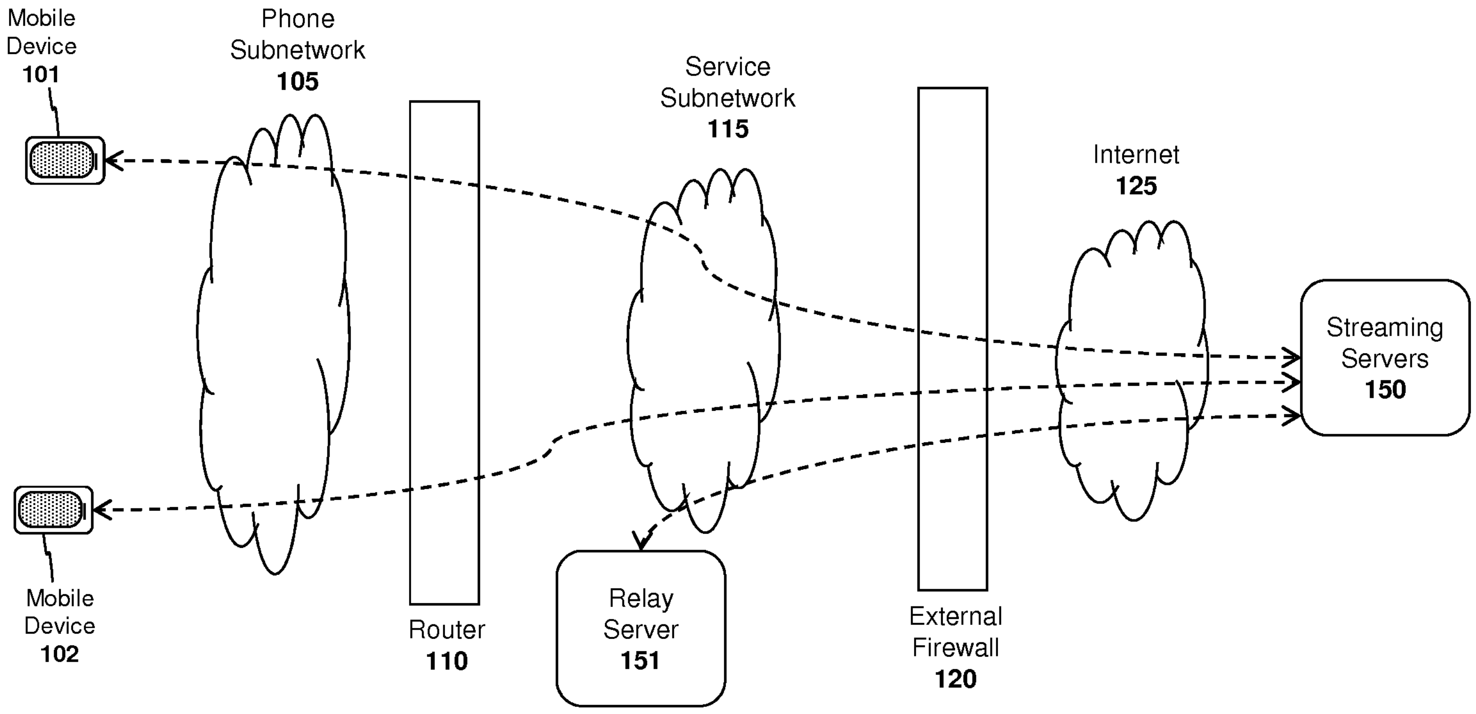
Mobile
Device
101
Phone
Subnetwork
105
Service
Subnetwork
115
Internet
125
Streaming
Servers
150
Mobile
Device
102
Router
110
Relay
Server
151
External
Firewall
120

Fig. 2a

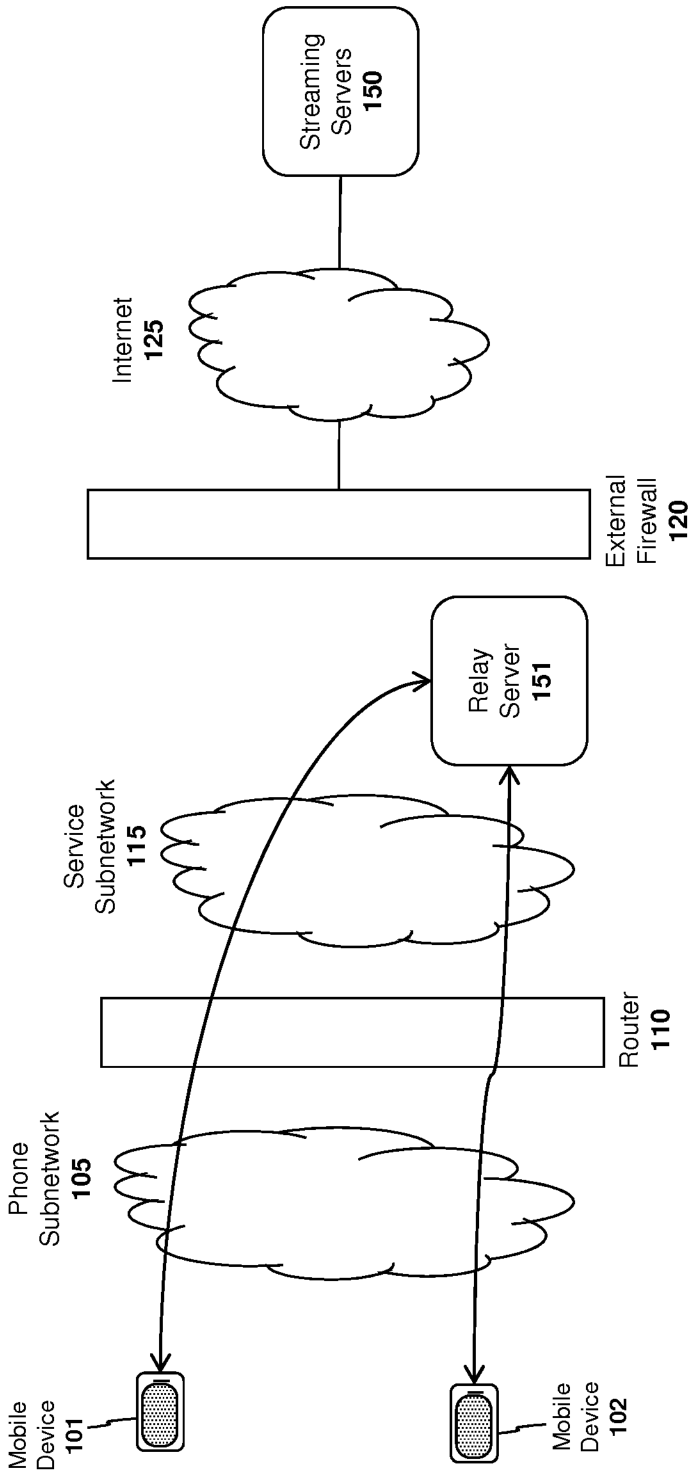
Streaming Servers 150
Internet 125
External Firewall 120
Relay Server 151
Service Subnetwork 115
Router 110
Phone Subnetwork 105
Mobile Device 101
Mobile Device 102

Fig. 2b

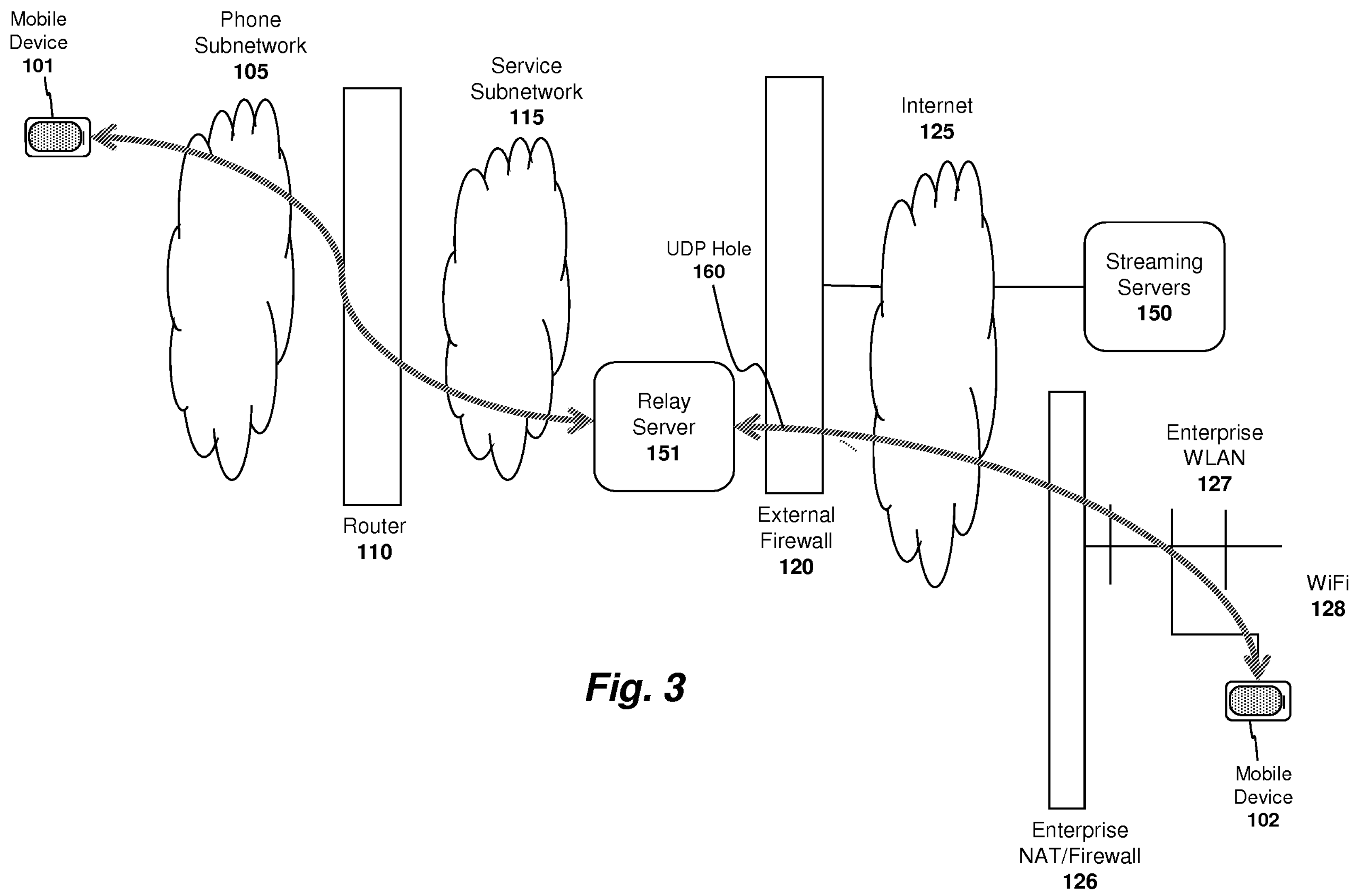
Mobile
Device
101
Phone
Subnetwork
105
Router
110
Service
Subnetwork
115
Relay
Server
151
UDP Hole
160
External
Firewall
120
Internet
125
Streaming
Servers
150
Enterprise
NAT/Firewall
126
Enterprise
WLAN
127
WiFi
128
Mobile
Device
102

Fig. 3

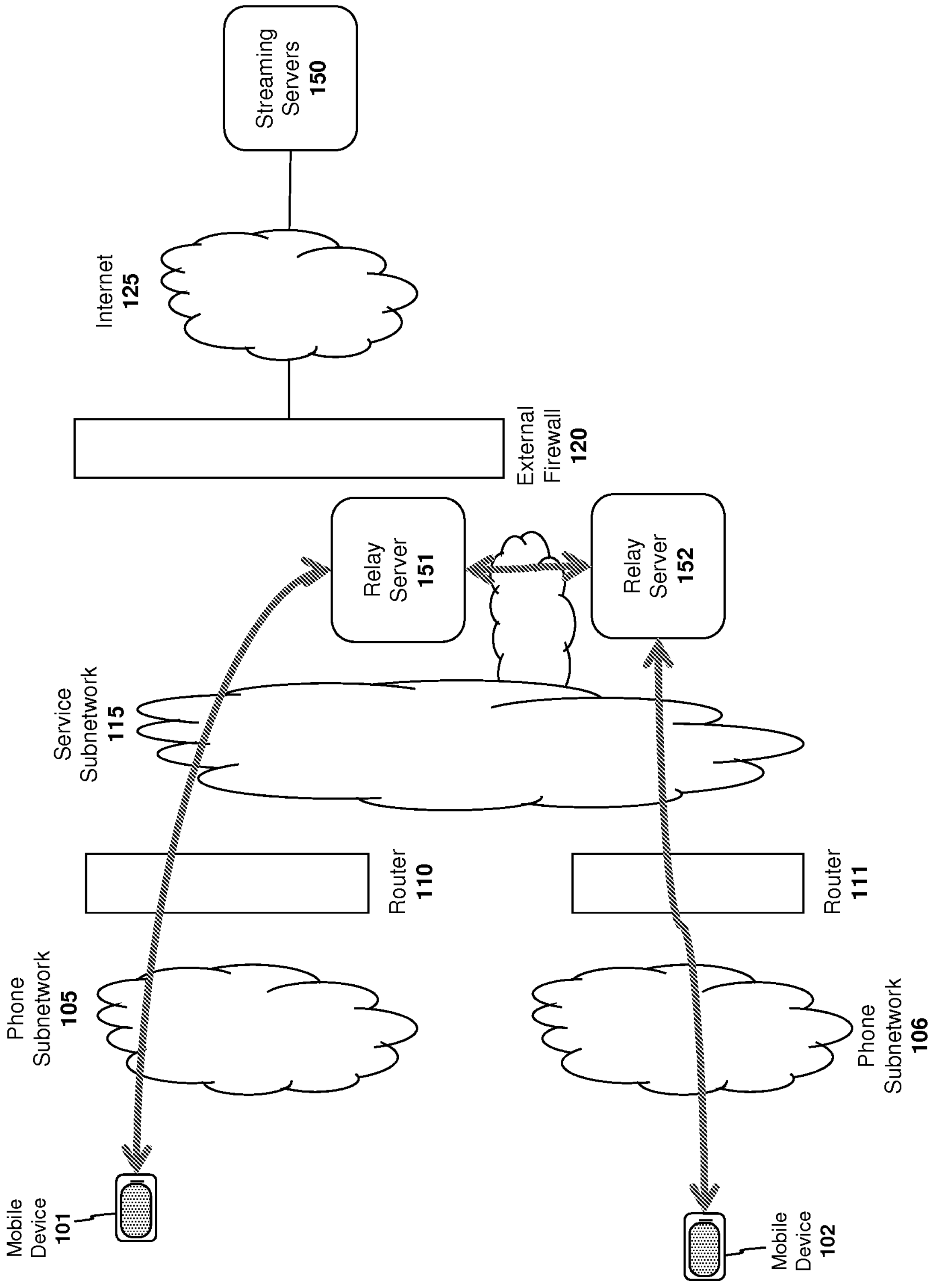
Mobile Device 101
Phone Subnetwork 105
Router 110
Service Subnetwork 115
Relay Server 151
External Firewall 120
Internet 125
Streaming Servers 150
Relay Server 152
Router 111
Phone Subnetwork 106
Mobile Device 102

Fig. 4a

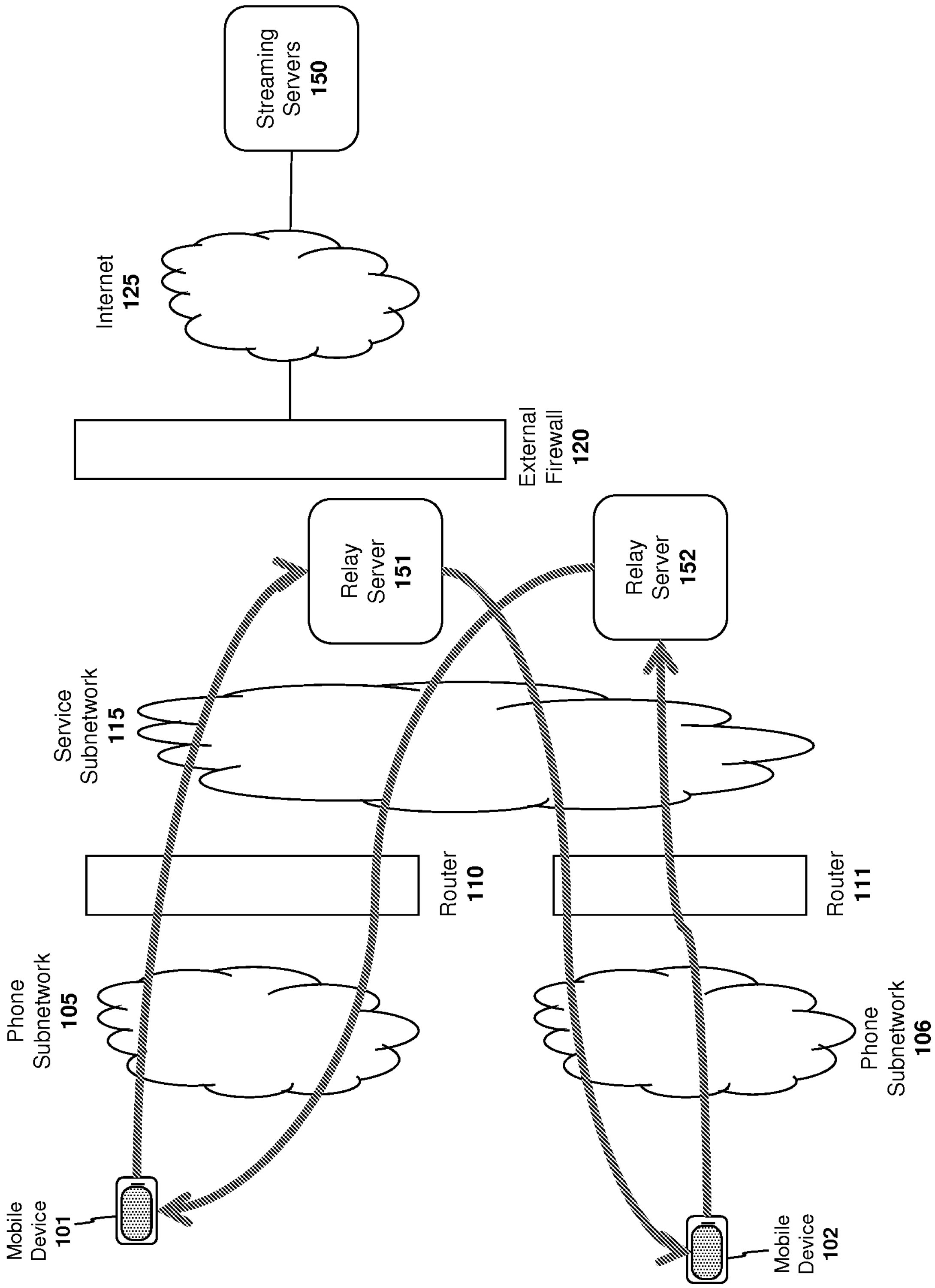
Streaming Servers 150
Internet 125
External Firewall 120
Relay Server 151
Relay Server 152
Service Subnetwork 115
Router 110
Router 111
Phone Subnetwork 105
Phone Subnetwork 106
Mobile Device 101
Mobile Device 102

Fig. 4b