other devices. For example, in many noisy environments, users wearing headsets with their mobile phones are typically required to manually switch their phones to a mute mode to prevent background noise from being transmitted by the phone to other phones and devices. Although this is effective in preventing the unwanted noise transmission, it often leads to gaps in communications as users who muted their phones forget to unmute them when they start speaking again. Moreover, in conference calls with many users manually muting and unmuting, the potential for communications gaps is even greater.

In attempting to solve the problem, some designers have provided circuitry that monitors the voice mic output and automatically mutes or unmutes based on comparison of the mic output level to one or more thresholds. However, the monitoring circuitry suffers from at least two problems that has inhibited its use in mobile phones. First, it consumes considerable power and thus shortens battery life. Second, the circuitry creates its own communications gaps not only because its slow reaction to speaking stops and starts, but also because it confuses external noises, for example voices of nearby people, with the user voice.

Accordingly, the present inventors have recognized a need for better ways of reducing transmission of ambient noise through ANC and other types of headsets.

Summary

To address one or more of these and/or other needs or problems, the present disclosure includes, among other things, one or more exemplary systems, kits, methods, devices, assemblies, and/or components for automatically detecting the presence or absence of user speech and operating associated muting or other voice-or speech-dependent functionality. Some embodiments include a voice mic, at least two control mics, and processing circuitry, with the processing circuitry configured to mute the voice mic based on a relationship of the control mic output signals.

In an aspect, there is provided a communication assembly comprising:

an incoming audio signal node configured for connection to a communication device that provides an incoming audio signal Rx;

signal processing circuitry configured to output a voice activity detection signal indicative of a determined absence or presence of user speech, the signal processing circuitry configured to determine absence or presence of user speech based on a first signal ϕ, and a second signal θ, , with the ϕ and θ signals derived from two or more transducers responsive to at least the ambient noise a, user speech s, and an incoming audio signal u, with the ϕ and θ signals being determined from a transform matrix

$$\begin{bmatrix}\theta \\ \phi\end{bmatrix} = \boldsymbol{T}\begin{bmatrix}S \\ a \\ u\end{bmatrix}$$

where T denotes the matrix of transfer functions, and is defined as

$$\boldsymbol{T}\begin{bmatrix}T_{s\theta} & T_{a\theta} & T_{u\theta} \\ T_{s\phi} & T_{a\phi} & T_{u\phi}\end{bmatrix}$$

wherein $T_{s\theta}$ represents a transfer function from user speech s to signal θ; $T_{u\theta}$ represents a transfer function from incoming audio u to signal θ;

$T_{a\theta}$ represents a transfer function from ambient noise a to signal θ;

$T_{s\Phi}$ represents a transfer function from user speech s to signal Φ; and

$T_{a\phi}$ represents a transfer function from ambient noise a to signal ϕ; and

wherein the third signal is based at least in part on an externally generated audio signal Rx or at least one audio notification signal, the signal processing circuitry providing an indication that the user is speaking responsive to a ratio of a prediction error for θ to a variance of Φ exceeding a detection threshold.

In an embodiment, there is provided a method of operating a communication assembly having a voice mic output signal and a set of two or more transducers, the method comprising: determining whether a user of the communication assembly is speaking based on the output of the two or more transducers; and in response to a determination that the use is not speaking, changing the operating state of one or more speech-dependent modules associated with the communication assembly to a resource conservation state.

In an embodiment there is provided a communication assembly comprising:

first and second transducers configured to be worn on a user head, with the first transducer configured for substantially greater acoustic coupling to a user vocal region than the second transducer; and

signal processing circuitry configured to output a voice activity detection signal indicative of a determined absence or presence of user speech in a predetermined frequency range when the first and second transducers are mounted to a user head; and

wherein the signal processing circuitry is further configured to indicate absence of user speech when the first and second transducers are positioned on a standard head and torso simulator (HATS) that is outputting a speech signal in the predetermined frequency range.

In a first embodiment, the two control mics, for example a left earpiece mounted mic and a right earpiece mounted mic, are configured for approximately symmetric acoustic coupling to a user vocal region. The processing circuitry determines whether a user is speaking based on measured temporal symmetry -- for example phase relationship-- of output signals from the two mics, with high symmetry indicative of user speech (arriving at both mics approximately simultaneously) and low symmetry indicative of ambient noise or non-user speech. The two control mics in

measurements). *H*(*t*) is a regression coefficient that explains the contribution of the rx/music signal to the feedback mic measurement.

Some embodiments use the following process model:

$$\begin{bmatrix} G(t+1) \\ H(t-1) \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} G(t) \\ H(t) \end{bmatrix} + \begin{bmatrix} (1-\alpha_1^2)^{1/2} & 0 \\ 0 & (1-\alpha_1^2)^{1/2} \end{bmatrix} \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix}$$

where $q_1(t)$ and $q_2(t)$ are independent Gaussian random variables with zero means and unit variances $\alpha_1$and α_2 are tunable parameters that govern how quickly *G*(*t*) and *H*(*t*) can change over time.

The state-space framework is useful because there are efficient algorithms for state estimation Recursive Least Square (RLS), Least Mean Square (LMS), and Kalman fileter, for example). Some embodiments estimate the sates, *G*(*t*) and *H*(*t*), in several frequency bins by using a Kalman filter in each bin.

Some embodiments incorporate a more generalized approach to user voice activity detection which avoids the need for explicitly characterizing the electroacoustical parameters of the system. The generalized approach uses two signals θ (theta) and ϕ (phi) which are each a function of at least two of user speech s, ambient noise a, and/or incoming audio u, according to the matrix of transfer functions T. (Incoming audio u can be a mix of externally generated audio signals such as Rx/music from a communication device and internally generated audio signals, such as system event prompts, notifications, or alarms) The matrix T of transfer functions is determined by how speech, ambient and incoming audio u appear on the two or more transducers, and how the transducers are combined with a reference to the Rx/music to generate phi and theta. Mathematically, this can be expressed as

$$\begin{bmatrix} \theta \\ \phi \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} s \\ a \\ u \end{bmatrix}$$

where *T* denotes the matrix of transfer functions, and is defined as

$$T = \begin{bmatrix} T_{s\theta} & T_{\alpha\theta} & T_{u\theta} \\ T_{s\phi} & T_{\alpha\phi} & T_{u\phi} \end{bmatrix}$$

wherein $T_{s\theta}$ represents a transfer function from user speech s to signal θ; $T_{u\theta}$ represents a transfer function from incoming audio u to signal θ;$T_{a\theta}$ represents a transfer function from ambient noise a to signal θ; $T_{s\Phi}$ represents a transfer function from user speech s to signal Φ; and $T_{a\phi}$ represents a transfer function

WHAT IS CLAIMED IS:

1. A communication assembly comprising:

an incoming audio signal node configured for connection to a communication device that provides an incoming audio signal Rx;

signal processing circuitry configured to output a voice activity detection signal indicative of a determined absence or presence of user speech, the signal processing circuitry configured to determine absence or presence of user speech based on a first signal ϕ, and a second signal θ, with the ϕ and θ signals derived from two or more transducers responsive to at least the ambient noise a, user speech s, and an incoming audio signal u, with the ϕ and θ signals being determined from a transform matrix

$$\begin{bmatrix}\theta \\ \phi\end{bmatrix} = \boldsymbol{T}\begin{bmatrix}S \\ a \\ u\end{bmatrix}$$

where T denotes the matrix of transfer functions, and is defined as

$$\boldsymbol{T}\begin{bmatrix}T_{s\theta} & T_{a\theta} & T_{u\theta} \\ T_{s\phi} & T_{a\phi} & T_{u\phi}\end{bmatrix}$$

wherein $T_{s\theta}$ represents a transfer function from user speech s to signal θ; $T_{u\theta}$ represents a transfer function from incoming audio u to signal θ;$T_{a\theta}$ represents a transfer function from ambient noise a to signal θ; $T_{s\Phi}$ represents a transfer function from user speech s to signal Φ; and $T_{a\phi}$ represents a transfer function from ambient noise a to signal ϕ; and $T_{u\phi}$ represents a transfer function from incoming audio u to signal Φ; and wherein the incoming audio signal u is based at least in part on an externally generated audio signal Rx or at least one audio notification signal, the signal processing circuitry providing an indication that the user is speaking responsive to a ratio of a prediction error for θ to a variance of Φexceeding a detection threshold.

2. The communication assembly of claim 1, wherein the two or more transducers include first and second mics, with the first mic positioned with substantially greater acoustic coupling to a user vocal region than the second mic.

3. The communication assembly of claim 2, wherein the first mic is positioned to receive more acoustic energy from the user vocal region via a bone conduction path than the second mic.

4. The communication assembly of claim 1, further comprising:

an automatic muting module configured to attenuate a voice mic output signal in response to the voice activity detection signal indicating a determined absence of user speech and to de-attenuate the voice mic output signal in response to the voice activity detection signal indicating presence of user speech.

5. The communication assembly of claim 1, wherein the automatic muting module is configured to mute or attenuate the voice mic output signal a predetermined amount of time after receipt of the voice activity detection signal.

6. The communication assembly of claim 1, further comprising:

a keyword recognition module configured to recognize one or more keywords within user speech signals derived from a voice mic output signal and responsive to the voice activity detection signal indicating a determined absence of user speech to enter a shutdown or sleep state to conserve power or memory.

7. The communication assembly of claim 1, further comprising: a side tone module responsive to user speech to provide a side tone signal to a loud speaker within a headset earpiece and responsive to the voice activity detection signal indicating a determined absence of user speech to enter a shutdown or sleep state to conserve power or memory.

8. The communication assembly of claim 1, further comprising: a side tone module responsive to user speech to provide a side tone signal to a loud speaker within a headset earpiece and responsive to the voice activity detection signal indicating a determined absence of user speech to enter a shutdown or sleep state to conserve power or memory.

9. The communication assembly of claim 1, further comprising: a noise reduction module responsive to the voice activity detection signal indicating a determined presence of user speech to reduce noise in a first way and responsive to the voice activity detection signal indicating a determined absence of user speech to reduce noise in a second way.

10. The communication assembly of claim 1, further comprising: at least one speech-dependent module configured for operation in a manner dependent on speech content of a voice mic output signal and responsive to the voice activity detection signal indicating a determined absence of user speech to enter a power and/or memory conservation mode.

11. The communication assembly of claim 1,

wherein the assembly includes a first node configured for connection to a loud speaker associated with a headset earpiece; a second node configured for connection to a feedforward ambient mic associated with the headset earpiece, and a third node configured for connection to a feedback error mic associated with headset earpiece;

wherein signal processing circuitry is coupled to the first, second and third nodes, and configured to provide an anti-noise signal to the loud speaker in response to signals originating from the feedforward ambient mic and the feedback error mic; and

wherein the ϕ and θ signals are derived from signals originating from the feedforward ambient mic and the feedback error mic.

12. The communication assembly of claim 11, wherein the signal processing circuitry is further configured to indicate absence of the user speech when the first and second transducers are placed on a standard head and torso simulator (HATS) that is outputting a speech signal in the range.

13. The communication assembly of claim 11, wherein the headset earpiece in an in-the-ear earpiece having an ear-canal portion containing the feedback error mic.