when the user is speaking and substantially different at least in phase when the user is not. In other words, the output signals from the feedback mics have approximately the same speech data components, and thus exhibit high coherence (in phase relationship) when the user is speaking. Moreover, the distance from ambient noise to the two mics are usually not equal or asymmetric, as indicated by paths P1 and P2 in Fig. 1, meaning that the coherence will be rotated in phase. (It still may be coherent, meaning the magnitude may still be very close to 1, but the angle will not be 0, which indicates an 'in-phase' situation.

[0028] Thus, the exemplary embodiment uses the complex coherence of the left and right feedback error mic signals fbL and fbR within a threshold test to determine if user voice activity is likely to be present or not. If the threshold criteria is satisfied, then a mute command signal is provided to mute module 136, and if not, an unmute command signs is provided. More specifically, the threshold test takes the form of

[0029] If mean (real(Cxy(freq_range))) > CT then Unmute; else Mute,

where mean () denotes an average function (over frequency); real () denotes the real part of a complex argument; Cxy (freq_range) denotes the complex coherence of signals x and y over a frequency range denoted freq_range; subscript x denotes the fbL error mic signal (left fb mic signal, FBA) and subscript y denotes the fbR error mic signal (right fb mic signal, FBB); CT denotes the coherence threshold. The real part of the complex coherence Cxy is used as this is a measure of how “in phase” the two signals are. Note: abs(Cxy) is 1 if only one average is taken, but this can still be useful as the important information is still in the phase. The freq_range over which the mean is computed can vary. However, because the body-conducted acoustics of the human voice box are primarily low frequency, some embodiments use the range of 70-700Hz or 100-500Hz. In response to an indication from the Cxy detector circuitry or logic that the user is not talking (that is user speech is absent), the exemplary embodiment mutes or attenuates the output of one or more microphones, such as the voice mic, and/or the feedforward microphones, using mute module 136.

[0030] Also, in some embodiments, the coherence threshold CT is in the range of 0.7 to 0.9, inclusive, with slight variations in performance. Setting it too close to 1 makes the detector more prone to false negatives (speech not detected when it exists) if there is significant background noise, and setting it too low yields more false positives (that is, unmuting when there is no speech. In some embodiments, the coherence threshold may be

adjusted dynamically based on system or environmental parameters, such as signal (speech) to noise ratio. That is, if the speech is much stronger than noise, then some embodiments shift the threshold closer to 1, for example 0.9, and if it is very noisy, these embodiments decrease the threshold level to avoid muting user speech. An exemplary threshold between 0.7 and 0.8, such as 0.75, is a good compromise.

[0031] This left-right symmetry based approach is effective, but may be vulnerable to on-axis ambient noise which can occur, for example, if a second speaker is on axis with the user (or generally equidistant from the left and right earpieces). (It also suffers from poor signal-to-noise ratio (SNR.) On-axis noise will exhibit high coherence at the feedback error microphones and thus be falsely recognized as speech. Interestingly, one way to confirm this vulnerability is to place a headset using this UVAD approach (for example with left and right feedforward or left and right feedback mics, or more generally two mics configured to be approximately equidistant from the vocal region) on a conventional Head and Torso Simulator (HATS), such as B&K Type 4128-C simulator manufactured by Brüel & Kjær Sound & Vibration Measurement A/S of Nærum, Denmark. The headsets can then be seen to automatically mute in response to noise sources approximately equidistance from the mics, for example, directly in front, behind, and above the HATS.

[0032] To reject on-axis ambient noise sources, such as non-user speech, and enable better performance in noisy environments, some embodiments use a second method of user voice activity detection, which has the added advantage of also being useful for a single-earpiece systems, or more generally to any situation where two mics having asymmetric coupling to a user and ambient sound. This method leverages the insight that the magnitude of energy occurring at the feedback mics (more generally, the mic with better acoustic coupling to the user vocal region) is substantially greater than that occurring at the feedforward mics when the user is speaking than when not speaking, due to asymmetry in the acoustic paths of voice signals to the feedback and feedforward mics. The feedback mics are positioned in the front cavity of the earpieces in or near the user ear canal. With this positioning, the feedback mics receive user voice energy via bone conduction paths 104A and 104B with very low attenuation and receive ambient noise via air conduction with strong attenuation due to the seal that is typically present in ANR systems. Therefore, the feedback mic to feedforward mic output ratios (fb/ff energy ratio) when user voice is present is much