November 07, 2017

Noise and howling mitigation in audio/video conferencing

Christian Schuldt

Robin Lundberg

Noise and howling mitigation in audio/video conferencing

ABSTRACT

Loudspeaker-to-microphone feedback causes echoes in audio/video conferences. A microphone typically has a background-noise level estimator, which provides noise-level estimates for the purpose of setting parameters within acoustic echo cancelers (AEC), noise reducers, comfort noise generators (CNG), etc. Output from a noisy loudspeaker may be picked up by the microphone, adding to background noise already present at the microphone. Loudspeaker noise, when picked up by the microphone, causes a misestimation of background noise, which in turn leads to incorrect parameter setting. This can lead to divergence of adaptive AEC circuitry, audible echoes, howling, etc.

Techniques of this disclosure improve estimates of the levels of loudspeaker and background noises, as captured at the microphone. Thereby, AEC, CNG, and noise-reduction parameters are set and adapted in a manner robust to ambient noise levels. Howling is controlled by adjusting loop gains, based on loudspeaker and background noise estimates.

KEYWORDS

- Acoustic echo cancellation
- Residual echo suppressor
- Howling
- Noise estimator
- Loudspeaker-microphone feedback

BACKGROUND

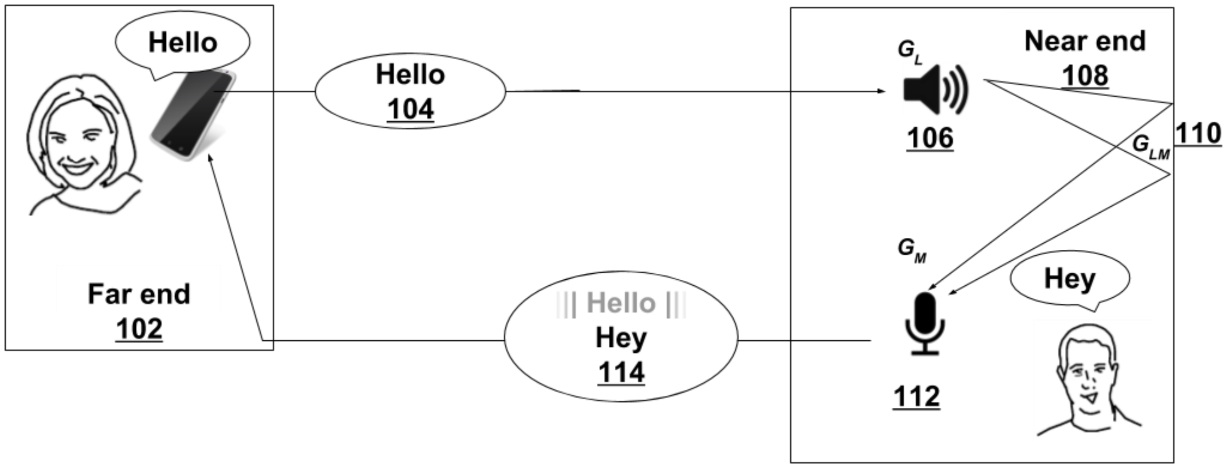

Fig. 1: The formation of echoes

Fig. 1 illustrates the formation of acoustic echoes in electronically mediated voice communication, e.g., a telephone call. At a far end (102), a participant in the call says "Hello." The speech is transmitted (104) to a near end (108), where it is output by a speaker (106). Aside from reaching human ears at the near end, the output from the speaker also reflects off walls and objects (110) at the near end. Near end microphone (112) picks up the near-end participant's speech signal ("Hey") along with the reflections. The microphone signal (114), comprising both near-end speech as well as reflections of signal received from the far-end, is sent to the far end.

DESCRIPTION

Fig. 2 illustrates echo cancellation techniques per the present disclosure. Analysis of signals emitted by speakers and detected by microphones is performed only if participant users provide permission for such analysis. The techniques are not implemented if users do not provide such permission.

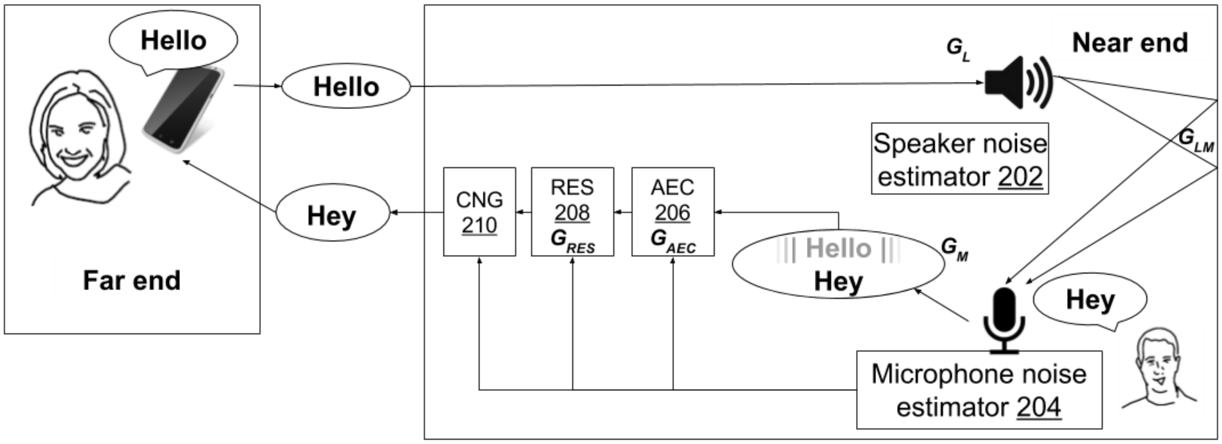

Fig. 2: Cancellation of echoes

An acoustic echo canceler (AEC, 206) in the near-end microphone to far-end speaker path, cancels echoes using linear filtering. Acoustic echo cancelers estimate the acoustic path, e.g., transfer function, between the speaker and the microphone. Once the acoustic path is known, an echo is canceled by synthesizing a negative echo at the delay of the echo.

Following the AEC is a residual echo suppressor (RES, 208), which suppresses, using linear and non-linear techniques, any echoes not removed by the AES. The RES is also referred to as a non-linear processor (NLP). By removing echoes of small magnitude, the RES also removes noise of similar magnitude, such that the far end subjectively senses unnatural silence during periods when the near end is silent. To alleviate this dead-connection like aural sensation, comfort noise is added by the comfort noise generator (CNG, 210).

Each component along the signal path of Fig. 2 has a gain associated with it. For example, the speaker at the near end has a gain denoted G_L, the microphone at the near end has a gain denoted G_M, and the multi-path between loudspeaker and microphone on the near-end has a gain denoted G_{LM}. The acoustic echo canceler has a gain G_{AEC}, and the residual echo suppressor has a gain G_{RES}. For stable operation, e.g., to prevent howling, the product of all these gains is less than 0 dB.

The parameters controlling the operations of the AEC, the RES and the CNG depend on noise estimates at the microphone and the speaker. For example, the background noise level in the room, e.g., based on an estimate of the noise level at the microphone, is used to automatically adjust the step-size for the adaptive filters of the AEC, and for setting parameters within noise-reduction componentry. An incorrect setting for the adaptive step-size, e.g., due to an inaccurate background noise estimate, can cause the adaptive filter to diverge or to freeze, thereby negating echo-cancellation function of the AEC.

As another example, noise generated by the CNG is generally set equal to or lower than the background noise in the room. Too high a CNG noise output risks the onset of howling. In order that parameters are set accurately for the AEC, RES and CNG, a noise estimator (202) is provided at the loudspeaker and a noise estimator (204) is provided at the microphone. The RES and the AEC also use input from the microphone noise estimator for their respective signal processing.

The noise estimator at the microphone nominally estimates the level of background noise, e.g., thermal noise, and noise originating from air-conditioning, fans, etc. in the room. It is this level of background noise that is used to set the AEC, RES and CNG parameters. However, sometimes noise is generated at the loudspeaker, and this noise, known as a noise echo, is picked up by the microphone. The presence of the noise echo at the microphone leads to an increase in total microphone noise and to inaccurate estimates of the level of background noise at the microphone. Thus, estimating the noise level, typically in frequency sub-bands, but also in the full-band signal, on both loudspeaker and microphone, is an important audio signal processing component in such configurations.

A sudden surge in loudspeaker noise is picked up by the microphone. Due to a pre-existing level of background noise, the percentage increase in total (background plus noise-echo) noise at the microphone is small compared to the size of the noise surge at the loudspeaker. The microphone noise estimator thus attributes the relatively small increase in its measured noise to a change in background noise, thereby misestimating background noise. The microphone noise estimator reaches steady-state before the loudspeaker noise estimator.

During the time when the microphone noise estimator has reached steady-state but not the loudspeaker noise estimator, the estimate of the noise echo (and hence, the estimate of background noise) on the microphone is inaccurate. As explained earlier, an inaccurate estimate of the level of background noise at the microphone can lead to divergence of the acoustic echo canceler adaptive filters, re-emergence of echoes, howling, etc.

Per techniques of this disclosure, a sudden surge in loudspeaker noise, as reported by the loudspeaker noise estimator, causes the microphone noise estimator to pause its estimation process. When the loudspeaker noise stabilizes, and the loudspeaker noise level is reliably estimated, the microphone noise estimator resumes operation. The reliable estimate of loudspeaker noise level is used by the microphone noise estimator to arrive at an estimate of background noise level.

The loop-gain in a phone is the total gain in all frequencies applied to a signal as it is received by the conference unit, played out on the loudspeaker, received as echo on the microphones of the same unit, and then transmitted to the far-end. To prevent howling, the combined loop-gain of the devices at the far- and near-ends needs to be below 0 dB. Typically, nothing is known about the device at the far-end, but a safe practice is to ensure that the loop-gain on the devices under near-end user control is below 0 dB.

Mathematically, the product $G_L G_M G_{LM} G_{AEC} G_{RES}$ is to be maintained at a value less than unity, in order that howling not occur. When howling occurs, a tendency of the participants on the call is to reduce the microphone gain (G_M) or the loudspeaker volume (G_L). However, this also attenuates useful signal. Gain reduction at the AEC is not always possible, for example, after a device reboot or during periods of echo-path non-stationarity.

The loudspeaker-to-microphone gain G_{LM} is a function of room-and-occupant geometry, and is not tunable. Per techniques of this disclosure, howling is prevented by automatically modifying, e.g., reducing, the gain G_{RES} of the residual echo suppressor such that the product of all gains along the signal path falls below unity (0 dB). In particular, the acoustic feedback between the loudspeaker and microphone(s) is measured as part of echo-cancellation processes and is hence known. The loudspeaker noise level is estimated during runtime, as explained previously. The echo return loss enhancement (ERLE) is estimated during runtime. By knowledge of the aforementioned quantities, the noise-echo level at the microphone is known. If the noise echo from the loudspeaker is a significant part of the total microphone noise, or possibly above a predetermined threshold, the RES gain is reduced to keep the total loop gain below 0 dB.

Changing the gain of the RES can inadvertently affect the output level of the comfort noise generator, since the CNG conventionally introduces only as much as noise as is removed by the RES. Per techniques of this disclosure, the level of comfort noise is based upon the background noise level, as determined by the microphone noise estimator, itself based on a stable, reliable estimate of loudspeaker noise level. In particular, if the noise-echo is significant, the amount of comfort noise is reduced, or allowed to increase only slowly. Furthermore, when noise echo is large, the AEC adapts to the noise echo, obviating the need for RES to further reduce noise echo to prevent howling.

In this manner, techniques of this disclosure estimate background noise level at a microphone while being robust to changes in loudspeaker noise level. By estimating background noise level, parameters for adaptive loops within acoustic echo-cancelers are set appropriately, e.g., so that the AEC doesn’t diverge, or allow re-emergence of echoes. Per techniques of this disclosure, howling is controlled automatically by reducing the gain of residual echo suppressor. Concomitantly, the comfort noise generator generates noise at a level commensurate with the noise-echo and background noise levels. In this manner, comfort noise is generated at not so high a level that howling occurs, and at not so low a level that dead-connection sensation is felt. The described techniques have low computational complexity.

In situations in which certain implementations discussed herein may collect or use personal information about users (e.g., user data, information about a user’s social network, user's location and time at the location, user's biometric information, user's activities and demographic information), users are provided with one or more opportunities to control whether information is collected, whether the personal information is stored, whether the personal information is used, and how the information is collected about the user, stored and used. That is, the systems and methods discussed herein collect, store and/or use user personal information specifically upon receiving explicit authorization from the relevant users to do so. For example, a user is provided with control over whether programs or features collect user information about that particular user or other users relevant to the program or feature. Each user for which personal information is to be collected is presented with one or more options to allow control over the information collection relevant to that user, to provide permission or authorization as to whether the information is collected and as to which portions of the information are to be collected. For example, users can be provided with one or more such control options over a communication network. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. As one example, a user's identity may be treated so that no personally identifiable information can be determined. As another example, a user's geographic location may be generalized to a larger region so that the user's particular location cannot be determined.

CONCLUSION

Techniques are described to enable robust operation of audio-video conferencing equipment, e.g., acoustic echo cancelers, residual echo suppressors, comfort noise generators, noise reducers, etc. The techniques work by providing accurate estimates of background noise and loudspeaker noise as captured at the microphone. By thus providing accurate estimates, howling and acoustic echoes are minimized. The techniques have low computational complexity.