

Auto-calibrating Echo Canceller Software for VoIP Applications on Smartphones

Suehiro Shimauchi, Kazunori Kobayashi, Masahiro Fukui, Sachiko Kurihara, and Hitoshi Ohmuro

Abstract

Automatically calibrating echo canceller software has been developed for voice over Internet protocol (VoIP) applications on smartphones. Because the audio properties of smartphones typically depend on the model, the speech quality of a VoIP application may sometimes degrade, especially during hands-free conversations. We extended the calibration ability of our software in order to handle the variations in smartphone audio properties. As a result, our software exhibited better performance than most conventional software.

Keywords: echo canceller, smartphone, VoIP application

1. Introduction

As smartphones become more widespread, communication is becoming more flexible than that achieved with ordinary fixed or mobile phones or personal computers. When making a voice call using a smartphone, some users may sometimes choose a voice over Internet protocol (VoIP) application instead of the default voice call function embedded in the smartphone. However, to provide good speech quality, the different audio properties of the various types of existing smartphones must be taken into consideration. The audio processing functions for the embedded voice call function can be well tuned to the property of each target smartphone model. In contrast, it is difficult to provide a VoIP application with such well-tuned audio processing functions since the application must be able to run on any model. Therefore, the speech quality of a VoIP application may sometimes degrade, especially during hands-free calls.

We developed auto-calibrating echo canceller software that can be implemented on VoIP applications. The following three functions are included in

the software.

- (1) Acoustic echo cancellation (AEC)
- (2) Noise reduction (NR)
- (3) Automatic gain control (AGC)

These three functions have internal parameters that can be automatically calibrated depending on the property of each smartphone model, which means that VoIP applications with our software can support most smartphone models.

2. Problems with audio processing on VoIP applications for smartphones

When we make a voice call by using a VoIP application on a smartphone, some acoustic problems may disturb our communication with the other party. Typical problems are acoustic echo, noise, and unbalanced sound levels (**Fig. 1**).

Acoustic echo is caused by acoustic coupling between the loudspeaker and microphone, both of which are built into smartphones. The sound reproduced from the built-in loudspeaker is conducted to the built-in microphone. Thus, when the far-end talker's voice is reproduced from the loudspeaker, it

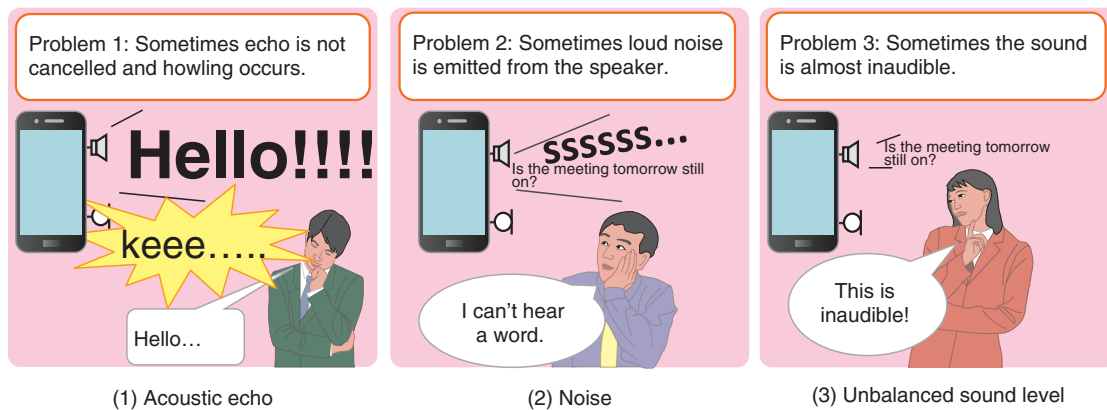


Fig. 1. Typical acoustic problems with hands-free voice calls with smartphones.

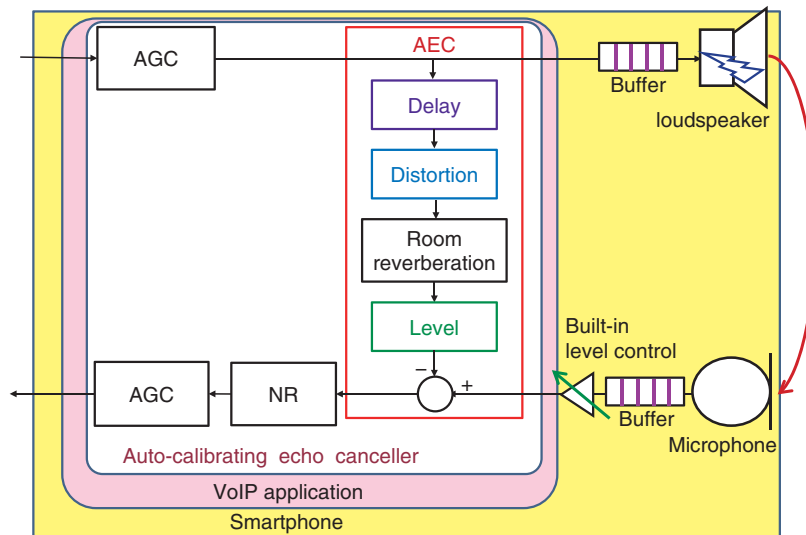


Fig. 2. Functional block diagram of auto-calibrating echo canceller software.

is picked up by the microphone and may be sent back to the talker through the smartphone. This feedback sound is called acoustic echo.

The built-in microphone picks up not only the talker's voice or the above-mentioned acoustic echo, but also background noise. The noise level depends on the environment where the talker is located and the acoustic or electrical properties of the microphone.

Even though a talker may speak into the microphones of various smartphone models at the same volume and from the same distance, the transmitted sound level sometimes differs depending on the model. This is because the microphone sensitivity and amplifier gain of each model are not always

designed to be the same. Also, some smartphones have a built-in level-control device such as a compressor or limiter for processing the microphone signal. The loudness of the loudspeaker also depends on the model.

3. Auto-calibrating echo canceller software

To solve the acoustic echo, noise, and unbalanced sound level problems on any smartphone model, we developed automatically calibrating echo canceller software. A functional block diagram of the software is shown in **Fig. 2**. The software has three main functions, as follows.

3.1 Acoustic echo cancellation

The acoustic echo can be cancelled if the acoustic echo signal is properly predicted by the AEC function. The accuracy of the predicted echo signal depends on how the acoustic echo path is modeled and how the model parameters are adaptively estimated. Most conventional AEC functions adaptively estimate only the room reverberation part of the echo path. This is not sufficient, however, for voice calls carried out using a smartphone. Therefore, we extended the parts of the echo path that can be adaptively estimated by taking into account distortion in the small built-in loudspeaker, variation in the sound buffering delay, and abrupt level changes caused by the built-in level-control function [1]. Because the new AEC function can track a wider range of variations in the echo path, it can predict the echo signal more accurately.

3.2 Noise reduction

The NR function is for transmitting a clear voice signal to the far-end listener by suppressing background noise based on the noise's stationarity. The basic algorithm for the NR function is similar to that used for the supporting technologies for Hikari Living, which provides visual communication via a television set [2].

3.3 Automatic gain control

The AGC function adjusts both the transmitted and received sound levels into an adequate range. Controlling both levels makes it possible to automatically amplify or reduce the sound level by up to 12 dB.

4. Performance evaluation

The performance of the developed software was evaluated on five different smartphone models. In a single-talk case, in which no one speaks into the near-end microphone and the far-end talker's voice is reproduced from the near-end loudspeaker, the

acoustic echo was reduced by more than 40 dB for all tested models. Even in the case of double-talk, in which both talkers spoke simultaneously, the acoustic echo was reduced by more than 20 dB for all tested models. With the conventional software, the acoustic echo was reduced by only 10 dB in the double-talk case for a certain model.

The speech quality of the near-end talker during double-talk was also evaluated using the perceptual evaluation of speech quality (PESQ), which is described in ITU-T (International Telecommunication Union-Telecommunication Standardization Sector) P.862. The PESQ score for our software was 0.45 better than that for the conventional software on average for five models and at most 1.15 better at maximum for a certain model.

5. Conclusion

We developed auto-calibrating echo canceller software for VoIP applications on smartphones. The new software exhibited better performance than the conventional software did due to its extended calibration capability enabling it to handle variations in smartphone audio properties. Although we focused on implementing the new software on VoIP applications, it can also be implemented on the embedded voice call function and exhibit good performance.

References

- [1] M. Fukui, S. Shimauchi, K. Kobayashi, Y. Hioka, and H. Ohmuro, "Acoustic Echo Canceller Software for VoIP Hands-free Application on Smartphone and Tablet Devices," Proc. of the IEEE 32nd International Conference on Consumer Electronics (ICCE 2014), pp. 133–134, Las Vegas, NV, USA, January 2014.
- [2] A. Nakagawa, M. Nakamura, S. Shimauchi, K. Nakahama, T. Matsumoto, R. Tanida, K. Kobayashi, and K. Sugiura, "Supporting Technologies for Hikari Living," NTT Technical Review, Vol. 12, No. 2, February 2014.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201402fa8.html>


Suehiro Shimauchi

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. from Tokyo Institute of Technology in 1991, 1993, and 2007. Since joining NTT in 1993, he has been engaged in research on acoustic signal processing for acoustic echo cancellers. He is a member of the Acoustical Society of Japan (ASJ), the Institute of Electronics, Information and Communication Engineers (IEICE), and IEEE (Institute of Electrical and Electronics Engineers).


Kazunori Kobayashi

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. in electrical and electronic system engineering from Nagaoka University of Technology, Niigata, in 1997, 1999, and 2003. Since joining NTT in 1999, he has been engaged in research on microphone arrays, acoustic echo cancellers, and hands-free systems. He is a member of ASJ and IEICE.


Masahiro Fukui

Deputy Senior Engineer, Terminal Equipment Technology Center, Application Solutions Business Headquarters, NTT Advanced Technology Corporation.

He received the B.E. in information science from Ritsumeikan University, Kyoto, in 2002 and the M.E. in information science from Nara Institute of Science and Technology in 2004. Since joining NTT in 2004, he has been conducting research on acoustic echo cancellers and speech coding. He received the best paper award of ICCE (International Conference on Consumer Electronics) and the technical development award from ASJ in 2014. He is a member of ASJ and IEICE.


Sachiko Kurihara

Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

She joined NTT in 1985. In 1990, she graduated in electronics from the Junior Technical College of the University of Electro-Communications. Since joining NTT, she has conducted research on quality assessment of telephone calls and on speech coding and its quality, and has also been involved in ITU speech coding standardization efforts. She received the Telecommunication Systems Technology Prize awarded by the Telecommunications Advancement Foundation in 1996, the ASJ Technology Research and Development Prize in 1996, and the Director General Prize of Science and Technology Agency for Originality, Ingenuity, and Meritorious service in 1998. In 2009, she received the Encouragement Prize by the Promotion Foundation for Electrical Science and Engineering. She is a member of ASJ.


Hitoshi Ohmuro

Senior Research Engineer, Supervisor, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received the B.E. and M.E. in electrical engineering from Nagoya University, Aichi, in 1988 and 1990. He joined NTT in 1990. He has been engaged in research on highly efficient speech coding and the development of VoIP applications. He is a member of ASJ and IEICE.