Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 100 (2015) 1469 – 1474
Procedia Engineering
www.elsevier.com/locate/procedia

25th DAAAM International Symposium on Intelligent Manufacturing and Automation, DAAAM 2014

Control Approaches for Audio Signal Quality Improvement in the Developed Conference System Based on the Personal User Devices

Khassanov Alisher, Krupenkin Alexander, Borgul Alexandr*

St. Petersburg National Research University of Information Technologies, Mechanics & Optics, 197101, 49 Kronverkskiy av., Saint Petersburg, Russian Federation

Abstract

The paper describes developed conference system based on the interaction between personal user mobile devices and administrator's PC. This system can be used instead of expensive, sophisticated equipment for conferences with the appropriate quality of audio signal. Analysis of different approaches and technics and final algorithms are presented. Phase modulation, gain reduction, spatial filtering, room modeling methods, and particularly AGC, AEQ, notch howling suppression, pitch shifting were taken as the basic elements of signal processing procedures. Practical realization of the final system is described as well. Main part of the tuning goes in automatic mode but there is a fine-tuning possibility for the users in the manual mode.

Peer-review under responsibility of DAAAM International Vienna

Keywords: conference system; signal processing; filtering; modern control theory

1. Introduction

The problem of the acoustic feedback control attracted a lot of attention last 50 years and nowadays researches in this field continue as well. This paper introduces the problem considered in case of audio signal processing in the developed software for usage of smartphones as microphones in conference systems. That software is a system with the server-client architecture, where server part installed on PC and client's part is a unity of mobile applications. This software can be used by administrators and organizers of the meetings and other public events. The main goal

* Corresponding author. Tel.: +7-951-664-47-88.
E-mail address: borgulalexandt@gmail.com, 148586@niuitmo.ru

Peer-review under responsibility of DAAAM International Vienna
doi:10.1016/j.proeng.2015.01.518

of the sound processing subsystem in conference system is to provide sufficiently gained speech signal without speech intelligibility loss and noise or such unpleasant effects. But in such case the microphone and loudspeaker are connected via the conference hall and there are existence of positive acoustic feedback in closed loop system. Acoustic feedback determines the existence of echo and if total gain of the sound system is higher than the value which called maximum stable gain (MSG) the system becomes unstable and howling appears. Nowadays four groups of the methods of automatic sound processing are developed for the increasing of MSG with the comfortable speaker voice's listening. Exhaustively definition of MSG and detailed description of those methods can be found in [3]. Here we just briefly mention each of them for arguing the further work. They are: phase modulation; gain reduction, spatial filtering and room modelling methods.

Methods of the third group (spatial filtering) require setup of the additional acoustic equipment (microphones and loudspeakers). That fact in our special case doesn't match the problem statement. It doesn't allow to process signal for amplifying and playing by acoustic system using only PC without any additional equipment. Let's consider other methods.

Methods with the room modelling are often use real-time tuned filters based on the systems with the finite impulse response [4-11]. This problem appears in case of calculation processes necessity for the filter tuning and filtration. We should develop the filter with the length of 4-12 thousands of delay elements even on the 8 kHz acoustic signal discretization frequency in the room with the impulse responses 0.5-1.5 second long. That filter is not available for realization on the PC due to calculation difficulties of these group algorithms. For example, authors had 17-second long processing procedure per 32-millisecond frame. Calculation was executed on the Intel Core i7 processor. It was also performed for the 0.5 second with the parallel processing on the embedded GPU of the processor for matrix multiplying. It's better in case of using the same equipment for processing. Though it's not enough for this methods realization in the task of acoustic feedback control for the conference system with sound processing on the PC.

Methods of gain reducing usually use manual tuning of the sound processing system experience and propose three automatic approaches of tuning such as automatic gain control (AGC), automatic equalization (AEQ) and automatic howling filtration by the notch filters in the frequency field (Notch Howling Suppression). All of them are good for realization on the PC. The most interesting one is a filtration with the notch filters. Synthesis of these filters should be done during the system working process with the parameters that allow increasing of MSG as follows. Signal goes through the sound processing system then analyzes in the frequency field and evaluates with the some of criterions. If the narrow band can be characterized as the undesirable then it's so called howling of the acoustic system. The next step is the locating of the place in the frequency domain and notch filter synthesizing. That filter should decrease the intensity of the signal on the howling frequency.

Server part of the developed software may be installed on the PC and available on the Windows, MAC OS and Linux Ubuntu platforms. Mobile applications are available on android, iOS, Windows Phone. Developed system is a solution for the problem of public events organizing when it requires presence of expensive, hard-to-use equipment with a lot of difficult settings, with a short range of distances, high level of noises, or renting of special conference halls. People in the room can use their personal mobile devices as the both microphones, systems for voting and scheduling with the described system.

Though there are a lot of advantages in the developed system there are some problems as well. The problem is decreasing of the transferred audio signal's quality due the presence of delays and saturations in the system. The methods used and adapted for the problem of the sound's quality preservation in the presented environment with acoustic feedback are shown further in the paper. The structure of the paper presented as follows. The problem statement' description is presented in the second part. Analysis and description of applied methods for the signal processing presented in the third part. Final structure and the practical realization presented in the fourth part of the paper.

2. Problem statement

Let's formulate the problem in terms of acoustic environment. Path of the audio signal may be described as follows. Speech of the speaker receives by the microphone of the smartphone, and then follows through wireless network to the signal processing and amplifying system (server part on the PC). Then amplified signal plays by loudspeaker.

Larsen effect (also known as audio feedback) is a special kind of positive feedback, which occurs when a sound loop exists between an audio input and an audio output (loudspeaker). In general, the signal received by the microphone is amplified and passed out of the loudspeaker. The sound from the loudspeaker can be received by the microphone again, amplified further, and then passed out through the loudspeaker again. The system become unstable in case the amplification coefficient is more than the maximum stable gain (MSG) of the signal attenuation in the environment between loudspeaker and microphone [1, 2]. It characterizes by the high speed of the noise generation on the maximal volume of the loudspeakers.

3. Methods of automatic audio signal processing. Analysis and application

We used multiple criterions detection where is the howling frequencies were detected by every of the criterions. We used criterions such as PAPR, PHPR, PNPR, IMSD well described in [12] and [15]. PAPR, PHPR and PNPR detect howling when the signal's part volume on the particular frequency is higher than the threshold. Meanwhile IMSD detects howling when the signal volume on the one of the particular frequencies is not changed over several frames. This frequency becomes a candidate to be howling. So, using some of the criterions together and unifying the results of comparing with the logical AND allows to decrease the number of detected errors. Otherwise some of the criterions can be incorrectly applied during the recognition of the speech or other acoustic signal as the howling, which is not. Quality of criterions and their combinations comparing described in [11] and [12].

Let's give more detailed description of the PHPR (Peak-to-Harmonic Power Ratio):

$$\mathrm{PHPR}(\breve{w}_i,\mathrm{t},\mathrm{m})[\mathrm{dB}] = 10\log_{10}\frac{|Y(\breve{w}_i,t)|^2}{|Y(m\breve{w}_i,t)|^2}, \tag{1}$$

where w_i - analysed frequency, t – time, m (=2,3,4) – distance multiplier for the harmonic component, Y – component of the analysed signal in the frequency field at the frequency w_i. It should be taken into account that this criterion is available for applying only in range of frequencies below the:

$$\mathrm{f}_s/\max(m), \tag{2}$$

where f_s – discretization frequency of the acoustic signal. It can be applied because the absence of the harmonic components above the Nyquist frequency. It doesn't need to use this criterion for the other frequencies. So, the howling detection at the frequencies above that one can be performed only based on the criterions.

ISMD criterion described in [14] and works only after processing of some number of the frames. This number should be enough to calculate components depends on given length of detection criterion memory (m) that given as the relation between Q_M and P.

$$\mathrm{IMSD}(\breve{w}_i,\mathrm{t},\mathrm{m}) = \frac{1}{\mathrm{Q}_M-1}\sum_{m=1}^{\mathrm{Q}_M-1}\left[\frac{1}{\mathrm{Q}_M}\sum_{j=0}^{\mathrm{Q}_M-1}\frac{1}{\mathrm{Q}_M-j}\left(20\log_{10}|Y(\breve{w}_i,t-jP)| - 20\log_{10}|Y(\breve{w}_i,t-\mathrm{Q}_M P)|\right) - 1mj=0\mathrm{m}-11\mathrm{m}-j(20\log10|Y(wi,t-jP)|-20\log10|Y(wi,t-\mathrm{m}P)|\right), \tag{3}$$

The example of the criterions applying illustrated in fig. 1.

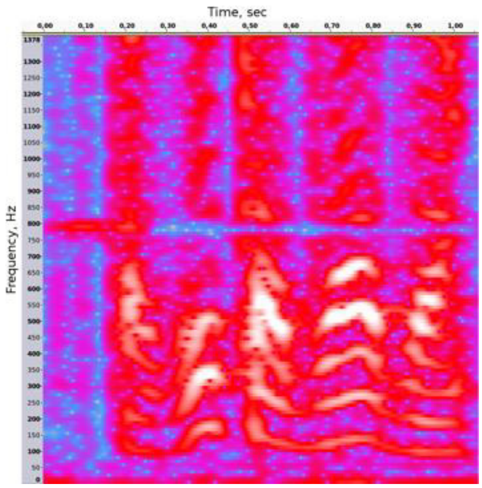

Fig. 1. Howling detected at the frequency of 790 kHz in the first frame and then filtrated with the narrow band notch filter.

This criterion shouldn't be considered for the howling detection first because in case of implementing through the logical AND with the other criterions in the beginning of the session it wouldn't allow to suppress the howling. It's experimentally noticed that the speaker using smartphone, as the microphone doesn't talk for the few milliseconds. At this moment absence of ISMD or work of other criterions on the voice doesn't make any disturbances in the quality of audio signal. Authors also noticed during the experimental work that howling frequency can be shifted in case of significant delays (more than 50 millisecond) and microphone placing changes related to the loudspeakers. Synthesis of the new filter may take a few milliseconds when the howling may appear again. This example illustrated in fig. 2. The next algorithm may be used for this problem solution. Basic level of the new filter should be chosen the same as the closest filters in case of howling frequency detection nearby the frequency with another applied filter.

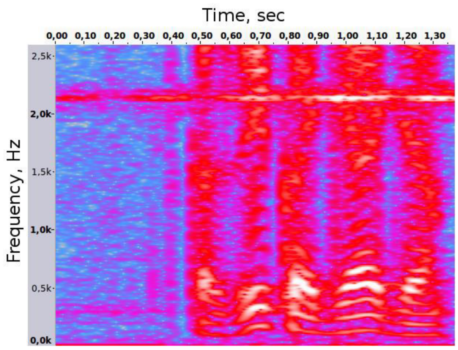

Fig. 2. Unpleasant effect – howling at the frequency about 2.2 kHz.

Phase-Modulation Methods are really attractive as well by their realization and calculation simplicity. The same may be applied to the loudness decreasing methods. Authors applied method of signal's tone shifting in the frequency domain. It cannot be noticed by bare ears but allows significant increasing speech loudness (MSG). Shifting algorithm supposes to shift the pitch of the played signal up or down on the small constant with the period equals the length of the impulse response of the room with the installed conference system. The influence of the pitch shifting on the speech components is shown in the highlighted region in the fig. 3.

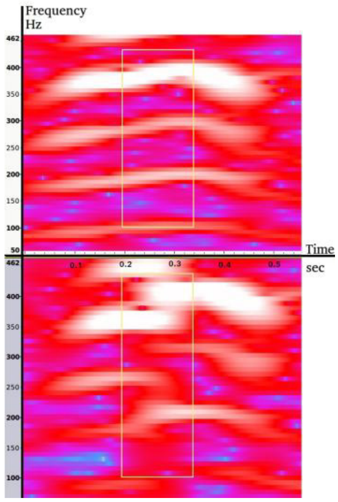

Fig. 3. Changes in the speech components with the frequency shifting.

4. Final structure and practical realization

Wireless network (Wi-Fi) was used for the microphone and PC connection, and for the packet transmitting protocol (UDP). Smartphones are used as the microphones and the length of the packet is different depends on the model. That's why the significant dynamic delay (5-200 millisecond) of the playing signal appears and it's not comfortable to process the signals with the different lengths. The compensation of the both jitter and packet length difference was applied with the help of buffering for the proper system work in this conditions. Data stores in the program buffer. The constant range sends to the sound processing system with the period and then goes to the amplifying and playing system. And the rest of the data with variable length allows jitter's compensation and frame length equalization. Though it makes additional algorithmic delays.

The final component of sound processing system is the broadband equalizer. It allows characteristics increasing of the speech such as brightness, intelligibility and existence for every room. That function is manual and available for the moderators as the 6-banded equalizer. Approximation is proposed for the narrow and broad components equalization. It's done to exclude unpleasant howling frequencies and save the form of the curve proposed by the moderator for the sound quality increasing. Partially it was done based on the previous research in this area [16].

The final structure of the proposed conference system for the realization on the PC looks like:

- Noise gate and suppressor
- Automatic dynamic compressor
- Howling detection
- Pitch shifter
- User equalizer
- Final equalizer

Noise gate and suppressor allow system to supress the residual echo after speaker's speech is done and to work in noisy environment. Dynamic compressor is needed to align dynamic range of speaker's speech.

Conclusion

The paper presents the developed conference system based on the PC and user's mobile devices. The problems of this system developing are described. The solution of the audio signal quality improvement is developed. Methods and approaches of the quality of audio signal increasing are presented. There is developed algorithm of MSG increasing, different noise and notch filters, filters of unpleasant effects, pitch shifting methods, and automatic and manual equalizers in the sound processing scheme. Final solution is a bit complex itself but allows using the equipment with the small computational power. The final scheme of the system that reduces main unpleasant effects in the audio signal is presented. It uses the combination of advantages of the entire described separate signal processing algorithms with the fastest available speed. None of the known algorithms allow achieving desirable quality of audio signal. But our approach allows uniting the parts of the current algorithms with our inputs and receiving the best available quality.

We consider using of the methods from the paper, which increase MSG but don't have an echo cancellation effect as the future researches. There are a lot of studies on this topic but no good practical results. And that's why it's actual to continue developing AFC methods and not only increase MSG, but quality of the audio signal as well. We are planning to propose one common criterion for the speech intelligibility and echo cancellation, and use it as a common parameter in the common cases.

Acknowledgements

This work was partially financially supported by Government of Russian Federation, Grant 074-U01.

This work was supported by the Ministry of Education and Science of Russian Federation (Project 14.Z50.31.0031).

References

[1] M. R. Schroeder, "Improvement of acoustic-feedback stability by frequency shifting," J. Acoust. Soc. Amer., vol. 36, no. 9, pp. 1718–1724, Sep. 1964.

[2] R. W. Guelke and A. D. Broadhurst, "Reverberation time control by direct feedback," Acustica, vol. 24, pp. 33–41, 1971

[3] Toon van Waterschoot and Marc Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges", Proc. IEEE, vol. 99, no. 2, Feb. 2011, pp. 288-327.

[4] A. Goertz - An adaptive subtraction filter for feedback cancellation in public address sound systems, in Proc. 15th Int. Congr. Acoust., Trondheim, Norway, Jun. 1995, pp. 69–72.

[5] J. Hellgren and U. Forssell - Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification,[IEEE Trans. Speech Audio Process., vol. 9, no. 7, pp. 906–913, Nov. 2001.

[6] J. Hellgren - Analysis of feedback cancellation in hearing aids with Filtered-X LMS and the direct method of closed-loop identification,[IEEE Trans. Speech Audio Process., vol. 10, no. 2, pp. 119–131, Feb. 2002.

[7] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen - Acoustic feedback suppression for long acoustic paths using a nonstationary source model,[in Proc. 13th Eur. Signal Process. Conf., Antalya, Turkey, Sep. 2005.

[8] G. Rombouts, T. van Waterschoot, and M. Moonen -Robust and efficient implementation of the PEM-AFROW algorithm for acoustic feedback cancellation, J. Audio Eng. Soc., vol. 55, no. 11, pp. 955–966, Nov. 2007.

[9] T. van Waterschoot, G. Rombouts, and M. Moonen - On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address systems, in Proc. 4th IEEE Benelux Signal Process. Symp., Hilvarenbeek, The Netherlands, Apr. 2004, pp. 167–170.

[10] Geert Rombouts, Toon van Waterschoot, Kris Struyve, and Marc Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model", IEEE Trans. Signal Process., vol. 54, no. 9, Sept. 2006, pp. 3426-3434.

[11] A. Spriet, S. Doclo, M. Moonen, J. Wouters - Handbook of Speech processing, chapter 48., Springer-Verlag Berlin Heidelberg, 2008.

[12] S. Braun - Evaluation of various Algorithms to detect acoustic Feedback, Betreuung: DI Hannes Pomberger, Graz, 2012.

[13] M. H. Er, T. H. Ooi, L. S. Li, and C. J. Liew, "A DSP-based acoustic feedback canceller for public address systems," in Proc. Int. Conf. Signal Process. (ICSP '93), Beijing, China, Oct. 1993, pp. 1251–1254.

[14] N. Osmanovic, V. E. Clarke, and E. Velandia, "An in-flight low latency acoustic feedback cancellation algorithm," in Preprints AES 123rd Convention, New York, NY, USA, Oct. 2007, AES Preprint 7266.

[15] Toon van Waterschoot and Marc Moonen, "Comparative evaluation of howling detection criteria in notch-filter-based howling suppression", J. Audio Eng. Soc., vol. 58, no. 11, Nov. 2010, pp. 923-940.

[16] Bobtsov, A.A., Nikolaev, N.A., Slita, O.V., Borgul, A.S., Aranovskiy, S.V. The new algorithm of sinusoidal signal frequency estimation, (2013) IFAC Proceedings Volumes (IFAC-PapersOnline), 11 (PART), pp. 182-186.