



Correlation Based Automatic Volume Control System for Television/ Radio

¹Hussain Mohammed Dipu Kabir, ²Muhammad Enayetur Rahman,
³Arshia Zernab Hassan, ⁴Mohammed Nazim Uddin

^{1,3} Solution Lab, Samsung R&D Institute Bangladesh, Dhaka, Bangladesh

^{2,4} Platform Lab, Samsung R&D Institute Bangladesh, Dhaka, Bangladesh

E-mail address: hmdkabar@ust.hk, enayet.rahman@samsung.com, nazim.uddin@samsung.com

Received 21 Jan.2014, Revised 30 Mar. 2014, Accepted 15 Apr. 2014, Published 1 May. 2014

Abstract: Automatic volume control (AVC) changes volume of TV/Radio based on surroundings noise. A number of hardware and software implementation of AVC is available. Conventional AVC methods concentrates on maintain a constant power-ratio between input of microphone and speaker-output signals. Some new methods are concentrating on preserving intelligibility of speech. They are extracting desired frequency band then comparing their power. This paper proposes Correlation based volume control for radio/Smart Television. Proposed Correlation Method adjusts volume based on linearity of input of microphone and speaker-output signals. A band-pass filter is used to filter-out unwanted frequency bands for comparing sound intelligibility. Conventional system concentrates on maintaining SNR. Proposed method adjusts volume considering delay and wave-shape of input of microphone and speaker-output signals.

Keywords: AVC, Audio, Intelligibility of Speech, Correlation, Adjust Volume

1. INTRODUCTION

Automatic Volume Control (AVC) automatically adjusts the volume of a Television (TV)/ radio according to the surrounding noise or environment with the intelligibility of speech or audio signal from the audio device. AVC in TV/ radio will be used to increase the understanding of the audio signal from the TV/ radio for the user in a noisy environment. Conventional method for the AVC is to keep the signal to noise ratio (SNR) acceptable [1].

This paper presents correlation for automatically adjusting the volume of a TV/ radio. As DC component of the audio signal is not audible, DC component of the signal is discarded filtering. Correlation Method is used for complex/computationally rich systems, as correlation is computationally expensive. Correlation can determine delay. From peak correlation value and index of that value, it is possible to determine linearity between input and output signal. For ideal system input and output signals are proportional, thus linearly dependent. Correlation coefficient between input and output should be closed to 1 for zero-noise condition. Some noise is generated by sound-system (speaker, microphone, wires, conversion loss etc.) and for this reason correlation coefficient can never be one. For any sound-system there will be a maximum value of correlation coefficient. A system constant (K) should be determined by calibrated peak correlation value, user defined volume-high and volume-low level.

Using AVC in a TV/ radio/ speaker is extremely helpful for some reasons. It increases the customer satisfaction and improves the intelligibility of hearing. For example, a user is hearing important news or, watching a live broadcast that has a great importance to him. At that time if by any reason some ambient noises are present there, it would be almost impossible for him to understand that event clearly. So, if there would present such a system that helps him to understand clearly that event by raising volume level, it will surely reduce his sufferings significantly. Another example can be: each user can save his personal profile with his minimum and maximum volume level in TV/ radio. Volume level may vary from user to user. Some user need louder sound level and some may need lower sound level. It will adjust the volume of the TV/ radio according to that particular user profile based on the volume level. So, user doesn't need to bother about the volume controller of the TV/ radio. AVC is also used with ring generators of telephone.

Correlation is used widely in signal processing. J. Benesty used cross-correlation for designing doubletalk detector [2], [3]. M. Collet used correlation for speaker tracking [4]. Canonical correlation is used for multimodal speaker identification [5]. In multimodal identification system speech and lip-texture modalities are fused for recognition. B. Kunka tried to correlate eye-tracking with speech signal for determining interaction between seeing and hearing [6]. Lots of research is going on for reduction of internal and external noises of audio. Internal noises are noises, inserted at the

time of recording. Heejong Yoo tried to suppress noise by filtering, envelop extraction and noise level estimation [7]. PCM signals are used as input if we want to use compressed [8] or encrypted [9]-[11] signal we need to decompress/decrypt it.

2. PRIOR ART

Researchers are designing/analyzing Automatic Volume Control circuit since late 1920's [12]-[16]. As the world is going to be digitized, it is important to design some methods which can be implemented in both hardware and software. Conventional method for the AVC is to keep the Signal to Noise Ratio (SNR) acceptable. The system of Konstantinouet. al is a remote controlled AVC device that periodically measures sound-to-noise ratio to adjust the volume level by keeping the ratio constant at all times. Jubienet. al. built their invention on the idea of an improved adaptive filter which approximates the audio portion of ambient room signal and uses subtraction to generate error signal to ultimately adjust audio signal gain. Automatic volume tracking and control is accumulated in the invention of Hsusan Hui. Here for a predetermined number of audio data, periodic mean-volume is calculated and then adjusted the volume according to max-mean data.

First modern digital AVC was described in 1997 by Helms. F. Felber considered intelligibility of speech in 2011. He developed an Automatic Volume Control method for all devices, containing at least one microphone. He used correlation for adjusting the delay. After delay-adjustment signals are subtracted for obtaining noise of background. This subtracted signal is filtered with band-pass filter and frequency components within selected frequency-band are used to calculate the speech-interference level (SIL). Seefeldt emphasized on simulating the perceived loudness of human auditory system to design an AVC system in 2012. Equivalent Rectangular Band is used to design a frequency filter bank to analyze the audio signal. He further proposed another method using the concept of "Auditory scene analyses" that includes weighting auditory events using skew-ness in the spectra and controlling loudness of the events using the weights.

Kim et. al. proposed Automatic Volume Control for mobile devices in noisy environments and he patented it. Their system first divides audio signal and noise signal into individual frequency bands. Then it calculates the individual band gain for each frequency band and a global band gain of overall frequency band. Energy ratio of original signal and noise is used determining the band gain. These results are used to control the gain and consequently the volume of the audio signal.

3. PROPOSED METHOD FOR VOLUME ADJUSTMENT

This paper presents correlation for volume adjustment. Like conventional methods audio signal is played through speaker and then recorded by microphone at first. Both of played and recorded signals are passed through similar band-pass filter. Filtered signals are compared for volume adjustment. Previous volume is used for avoiding sharp volume change. Proposed system uses minimum and maximum volume levels. As noise annoyance is dependent of personal and social factors [17], minimum and maximum volume levels should be determined by user.

Block diagram of proposed system is shown in Fig. 1. Audio signal is played and then recorded for determining noise level. Audible part of signals is extracted by a band-pass filter. Extracted parts of original audio and recorded signals are compared. Previous volume is also considered before changing volume. 'Adjust volume block' depends on method of volume adjustment.

For complex systems and systems with dedicated audio SoC, more accurate algorithm can be used. Correlation is widely used technique for delay estimation. We may use principle of Pearson product-moment correlation coefficient (PPMCC) [18] for detection of noise and automatic amplitude correction.

PPMCC is a measure of linear dependence between two variables. Equation for PPMCC is as following-

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

Here, r = PPMCC; X_i ; Y_i = Played and recorded Signals.

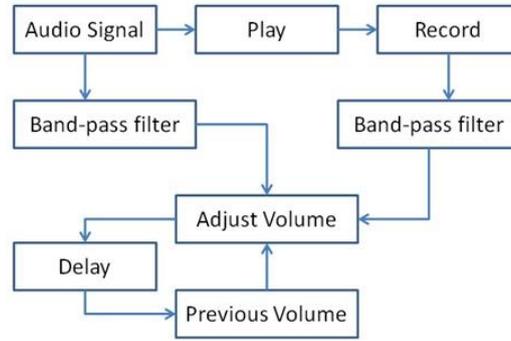


Figure 1 Block Diagram of proposed system

Two variables are played signal (x) and recorded signal (y). Both (x) and (y) are PCM signal. Played signal is multiplied with gain of speaker. Sound of speaker is mixed with other sounds/noises of environment/room. Noise is also introduced by sound system. Finally sound of environment is multiplied by gain of recorder and saved as recorded signal (y). These two signals should be linearly related with a constant delay at zero-noise condition. Their relation is as following-

$$y(t) = \text{Gain} \times x(t - \tau) + \text{noise}(t) \tag{2}$$

Here, τ = Delay, t = Time, Gain = Gain of Speaker \times Gain of Microphone.

For digital audio signals their relation becomes-

$$y(n) = \text{Gain} \times x(n - \tau_n) + \text{noise}(n) \tag{3}$$

Here, τ_n = Delay in number of sample, n = Sample number.

For consideration of delay, maximum value of correlated array is considered as nominator of the PPMCC equation. Duration of recorded and played signal should be at least 4-times larger than delay. Otherwise value of r will be much smaller than 1 for zero noise condition. Value of r will never be 1; because of delay and system-generated noise. Equation for volume correction will be-

$$A_n = \frac{K}{r \cdot A_0} s + A_0(1 - s) \tag{4}$$

Here, A_n = New Amplitude, A_0 = Previous Amplitude, r = Pearson correlation coefficient, K = System parameter, determined by trial and error method, s = selection factor. Selection factor is 0.2 for gradual change of volume.

For any sound-system (K/A_0) is constant. As a simpler process is required for faster calculation of PPMCC, Pearson correlation coefficient is calculated by following equation-

$$r = \frac{\max(\text{corr}(x,y))}{\text{norm}(x)\text{norm}(y)} \tag{5}$$

Here, $\max(a)$ = maximum of array a ; $\text{corr}(x,y)$ = correlation array of x and y ; $\text{norm}(x)$ = norm of x (Euclidean Norm)

Determination of 'r' is computationally expensive, even after simplification. For this reason, it is suitable for complex/ rich system like PC/Smartphone.

Fig. 2 shows adjust volume block for Correlation Method. Fig. 3, showing a sample input signal for understanding that phenomena. Though signal of 10Hz frequency is not audible it is used for clear demonstration. Fig. 4 shows a delayed output at unity gain and zero noise condition. Fig. 5, showing delayed output at noisy condition. Correlation coefficient at zero-noise condition is determined from correlated wave-shape in Fig. 6. Similarly, Correlation coefficient for noisy condition is determined in Fig. 7. For determination of normalized correlated wave, peak of correlated wave is determined at first. From the index of peak value correlated portion of both waves are determined. Finally, wave-shape is normalized by the norms of correlated portions. According to Fig. 6 and Fig. 7 peak of normalized and correlated wave-shape depends on noise. Peak value of normalized correlated wave-shape is 1 at zero noise condition and decreases in increase of noise level.

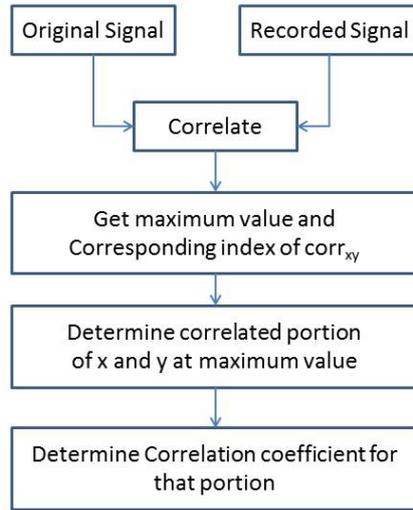


Figure 2 Adjust volume block for correlation method

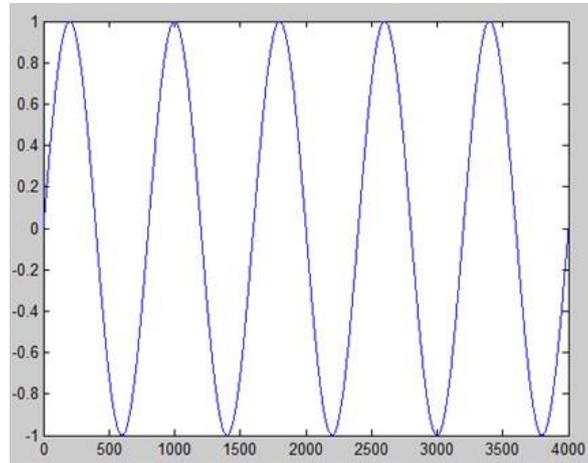


Figure 3 Wave-shape of input signal; containing samples of 0.5 second

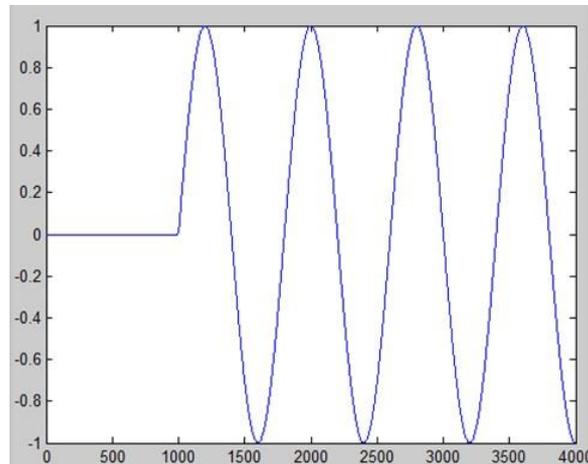


Figure 4 Output signal at zero noise condition; with a constant delay and unity gain

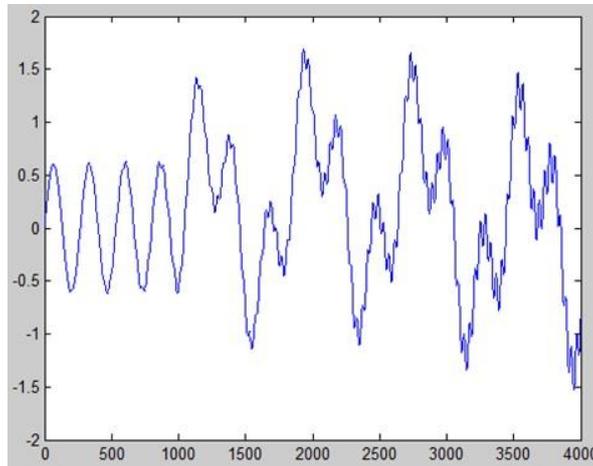


Figure 5 Output signal at noisy condition

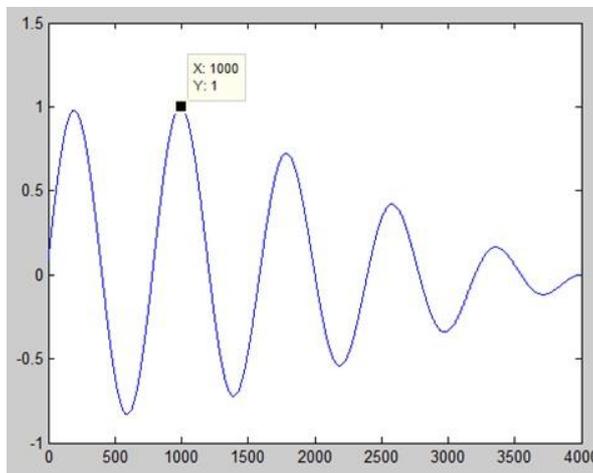


Figure 6 Value of Correlation coefficient from correlation wave-shape at zero-noise condition

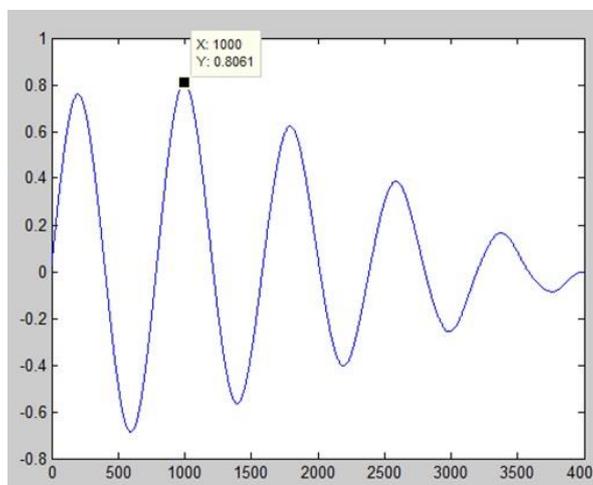


Figure 7 Value of Correlation coefficient from correlation wave-shape at noisy condition

4. EXPERIMENTAL RESULTS

A. Process Flow

Samsung R439 Laptop is used for simulation and demo purpose. When music player and volume updater are different, Windows XP is required. Windows 7 uses an audio mixer. When volume updater updates volume at Windows 7, volume of updater is changed and speaker volume and volume of player remains constant. As Windows XP is not using audio mixer, updater can update speaker volume.

B. Value of Correlation Coefficient

Value of correlation coefficient is observed for different noise level and for different type of signals. When noise is higher than volume of television, average value of correlation coefficient (r) is about 0.08 and for low-noise condition that value (r) becomes 0.4 (on average). For some steps value of correlation coefficient becomes greater than 0.8. Performance is highest for pulse-wave signals.

Fig. 8 and Fig. 9 shows value of correlation coefficients at different steps (iteration) for both high noise and low noise conditions. Blue line is drawn for low noise and red line is drawn for high noise conditions. Fig. 8 is showing curves for pure sine wave or beep signal and Fig. 9 is showing curves for an audio of concert music.

According to curves, correlation coefficient (r) is higher for low noise conditions. Also, sound content is another important factor. Music of concert contains some self-noise and large range of frequency components. When that music is played and recorded, a noticeable amount of noise is inserted due to self-noise of speaker and recording device.

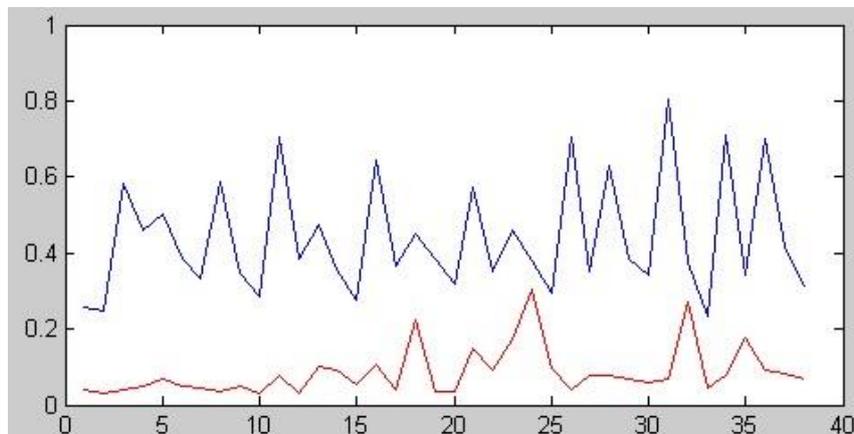


Figure 8 Value of Correlation coefficient for high noise (red curve) and low-noise (blue curve) conditions where signal is

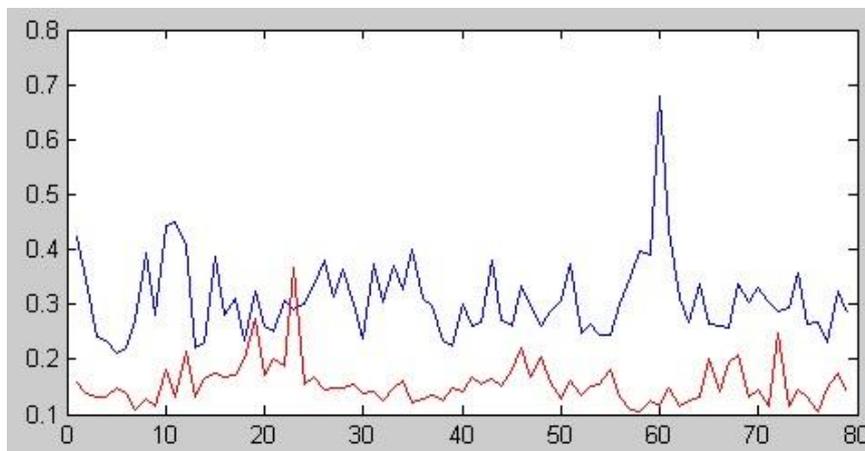


Figure 9 Value of Correlation coefficient for high noise (red curve) and low-noise (blue curve) conditions where signal is audio of concert music

Figures (Fig. 10 to Fig. 14) are showing statistical values of PPMCC. From Fig. 10 and Fig. 11, distribution of PPMCC value is dependent of noise for pulse wave (pure sine wave). That phenomenon is later proved in next two figures and this time for concert music. Though PPMCC value at zero noise condition should be 1(theoretically), PPMCC value is significantly smaller than 1. Reason of smaller PPMCC is noises correspond to speaker and microphone.

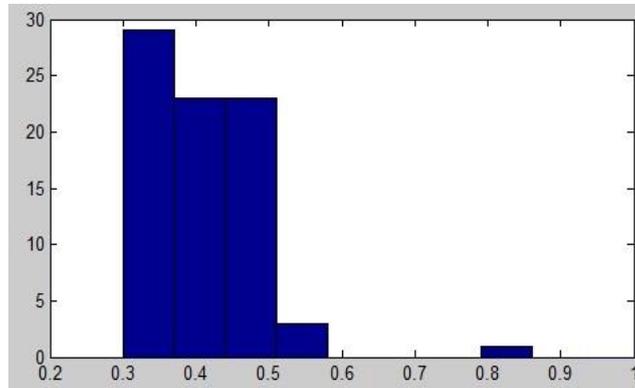


Figure 10 Histogram of correlation coefficient values for low noise and for pulse-wave signals

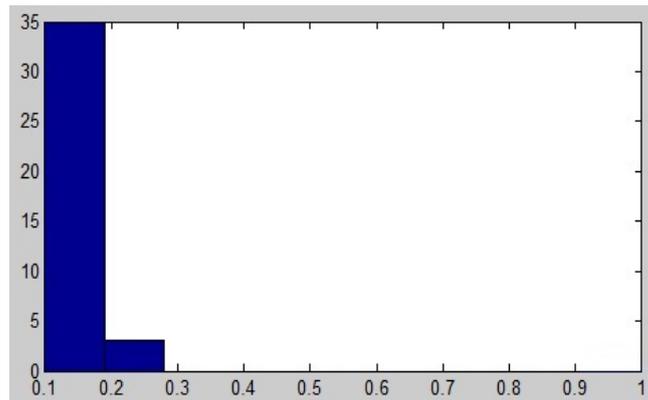


Figure 11 Histogram of correlation coefficient values for high noise and for pulse-wave signals

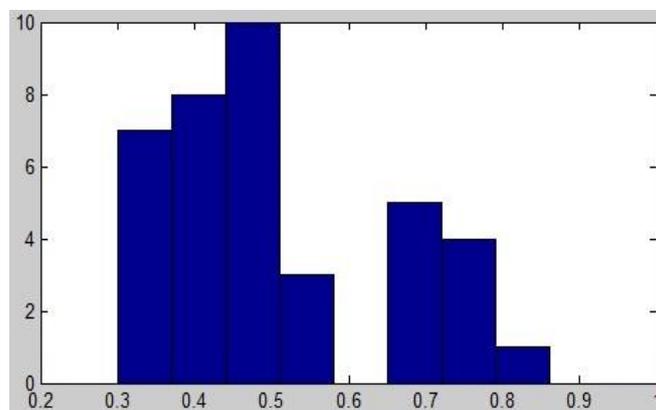


Figure 12 Histogram of correlation coefficient values for low noise and for concert music

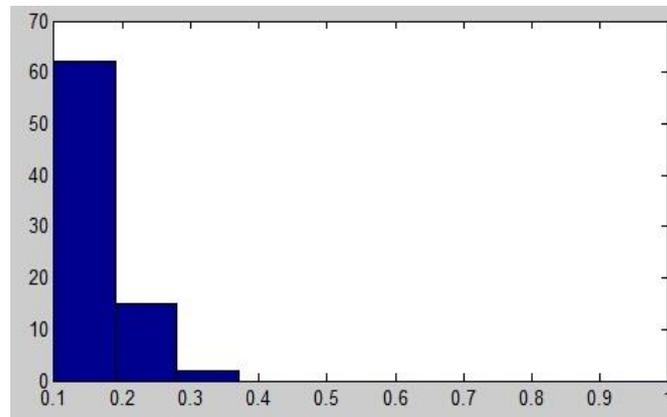


Figure 13 Histogram of correlation coefficient values for high noise and for concert music

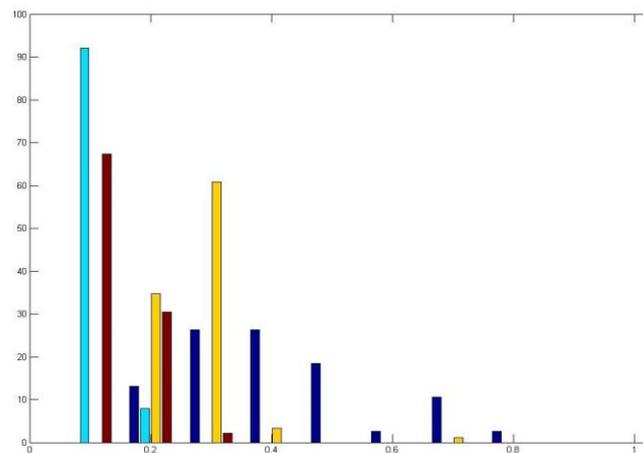


Figure 14 Percentage Bar-chart of correlation coefficient values (blue for low noise, sky-blue for high noise pulse waves; yellow for low noise and red for high noise concert music)

5. DISCUSSION

Watching television or listening to the radio is interfered by various sources noise. Aircraft noise affects people in different ways. It makes conversation difficult, disturbs persons watching television or listening to the radios. People, living within the vicinity of major airport experiences frequent aircraft noise. A lot of research is going on to reduce airport noise or softening its impact. The first is to improve the design of planes so that they make less noise when circling, taxiing, landing, and when taking off. Aircraft noise is extremely loud and researchers succeeded to reduce a few notches on noise meter. People living near railway are also facing similar problems. Railway noise also causes disturbances to local residents. It is possible to reduce railway noise or effect of noise through proper planning and posing permission conditions. In addition noise may come from traffics or viewers of television may start talking suddenly. Viewers of television may sit in a noise insulated room. Noise of that room may increase due to sudden opening of door. Automatic volume control will help user in understanding programs of television in all mentioned noisy situations.

When user-defined volume high level is smaller than volume level required avoiding noise, user may feel further obstruction. While one or more viewers of television/ listeners of radio are talking or making noise, the volume will raise automatically. Listeners will be warned about their noise through raise of volume automatically. As 0.5s-signal streams are compared proposed method correspond to delay of 0.5s for adjusting volume. Traditional Hardware-based AVC's caused minimum delay for changing volume. Felber's proposed AVC corresponds to an approximate delay of 1.4s (according to his time vs. gain curve). Proposed Correlation Method works with intelligibility of speech like Felber's method. It also considers wave-shapes of corresponding signals. Proposed method can be implemented as hardware or as software.

Currently we are working on peak of correlation for each step. That approach is computationally expensive. In future we will try to estimate a constant delay at the beginning steps of simulation. After obtaining delay corresponding correlation coefficient will be calculated. In implementation, constant delay will be calculated at starting period of TV/radio.

6. CONCLUSION

Pearson product-moment correlation coefficient is widely used in determination of linearity between two variables. Here we proposed a correlation based automatic volume control system. Proposed Correlation Method works with intelligibility of speech. When any unwanted signal in a desired frequency range is detected, volume is increased for maintaining sound quality. When amplitude of noise is low, volume decreases and comes to a minimum level and for high-noise condition, volume reaches to a maximum level. Both minimum volume and maximum volume levels are defined by user.

REFERENCES

- [1] Bernhaupt, R., and Michael M. P.. "User interface guidelines for the control of interactive television systems via smart phone applications." *Behaviour & Information Technology* ahead-of-print (2013): 1-16.
- [2] Gänslér, T., and Jacob B.. "A frequency-domain double-talk detector based on a normalized cross-correlation vector." *Signal Processing* 81.8 (2001): 1783-1787.
- [3] Benesty, J., Dennis R. M., and Jun H. C.. "A new class of doubletalk detectors based on cross-correlation." *Speech and Audio Processing, IEEE Transactions on* 8.2 (2000): 168-172.
- [4] Collet, M., Delphine C., and Frdric B.. "A correlation metric for speaker tracking using anchor models." *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, USA. 2005.*
- [5] Sargm, M. E., et al. "Multimodal speaker identification using canonical correlation analysis." *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. Vol. 1. IEEE, 2006.*
- [6] Kunka, B., and Bozena K.. "An new method of audio-visual correlation analysis." *Computer Science and Information Technology, 2009. IMCSIT'09. International Multiconference on. IEEE, 2009.*
- [7] Yoo, Heejong, David V. Anderson, and Paul Hasler. "Continuous-time audio noise suppression and real-time implementation." *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on. Vol. 4. IEEE, 2002.*
- [8] Kabir, H. M. D., et al. "Non-linear Down-sampling and Signal Reconstruction, without Folding." *Computer Modeling and Simulation (EMS), 2010 Fourth UKSim European Symposium on. IEEE, 2010.*
- [9] Kabir, Hussain Mohammed Dipu, et al. "Watermark with Fast Encryption for FPGA Based Secured Realtime Speech Communication." *Consumer Electronics Times.*
- [10] Kabir, H. M. D., and S. B. Alam. "Hardware based realtime, fast and highly secured speech communication using FPGA." *Information Theory and Information Security (ICITIS), 2010 IEEE International Conference on. IEEE, 2010.*
- [11] Alam, S. B., et al. "A secured electronic transaction scheme for mobile banking in Bangladesh incorporating digital watermarking." *Information Theory and Information Security (ICITIS), 2010 IEEE International Conference on. IEEE, 2010.*
- [12] Oliver, B. M. "Automatic volume control as a feedback problem." *Proceedings of the IRE* 36.4 (1948): 466-473.
- [13] Wheeler, Harold A. "Automatic volume control for radio receiving sets." *Radio Engineers, Proceedings of the Institute of* 16.1 (1928): 30-34.
- [14] Egeland, Olav, and Jan Tommy Gravdahl. *Modeling and simulation for automatic control. Vol. 76. Trondheim, Norway: Marine Cybernetics, 2002.*
- [15] Igarashi, Takeo, and John F. Hughes. "Voice as sound: using non-verbal voice input for interactive control." *Proceedings of the 14th annual ACM symposium on User interface software and technology. ACM, 2001.*
- [16] Monk, Andrew, et al. "Why are mobile phones annoying?." *Behaviour & Information Technology* 23.1 (2004): 33-41.
- [17] Guski, Rainer. "Personal and social variables as co-determinants of noise annoyance." *Noise and Health* 1.3 (1999): 45.
- [18] Derrick, Timothy R., Barry T. Bates, and JANET S. Dufek. "Evaluation of time-series data sets using the Pearson



Hussain Mohammed Dipu Kabir was born in Bangladesh in 1988. He received baccalaureate from Department of Electrical and Electronics Engineering (EEE) of Bangladesh University of Engineering & Technology (BUET) in February, 2011. He worked at Solution Lab/ Advanced R&D Department of Samsung Bangladesh R&D Center as Software/Sr. Software Engineer from March 2011 to August 2013. Currently he is a PhD student of The Hong Kong University of Science and Technology (HKUST). He is also serving as a reviewer in IEEE conferences since the time of his undergraduate study. Conferences are PECON, IAPEC, SCORED, ISIEA, ICEDSA, BEIAC, ICCSII etc. His research interests are Embedded System, Audio Processing and Solid-state devices.



Muhammad Enayetur Rahman completed his BSc from Electronics and Communication Engineering (ECE) discipline of Khulna University. Currently he is working at Samsung R&D Institute Bangladesh as Senior Software Engineer.



Arshia Zernab Hassan completed her BSc from Bangladesh University of Engineering and Technology. Currently she is working at Samsung R&D Institute Bangladesh as Software Engineer.



Mohammed Nazim Uddin completed his BSc and MSc in computer science from University of Pune, India. He obtained PhD from Inha University, South Korea in 2012. He was faculty at National University Bangladesh during 2001-2004 and Part Lecturer in Department of Computer Science and Information Engineering at Inha University during 2009-2012. Currently he is working at Samsung R&D Institute Bangladesh as Chief Engineer.