Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

Performance Analysis of Audio and Video Synchronization using Spreaded Code Delay Measurement Technique

A.Thenmozhi
PG scholar, Department of ECE, Anna University, Chennai
Email: thenmozhi94.a@gmail.com
Dr.P.Kannan
Professor & HOD, Department of ECE, Anna University, Chennai
Email: deepkannan@yahoo.co.in

ABSTRACT

The audio and video synchronization plays an important role in speech recognition and multimedia communication. The audio-video sync is a quite significant problem in live video conferencing. It is due to use of various hardware components which introduces variable delay and software environments. The objective of the synchronization is used to preserve the temporal alignment between the audio and video signals. This paper proposes the audio-video synchronization using spreading codes delay measurement technique. The performance of the proposed method made on home database and achieves 99% synchronization efficiency. The audio-visual signature technique provides a significant reduction in audio-video sync problems and the performance analysis of audio and video synchronization in an effective way. This paper also implements an audio- video synchronizer and analyses its performance in an efficient manner by synchronization efficiency, audio-video time drift and audio-video delay parameters. The simulation result is carried out using mat lab simulation tools and simulink. It is automatically estimating and correcting the timing relationship between the audio and video signals and maintaining the Quality of Service.

Keywords-**Audio spreading codes, Hamming distance correlation, Spectrograph, Synchronization, Video spreading codes.**

Date of Submission: April 17, 2018 Date of Acceptance: June 23, 2018

1. INTRODUCTION

The Audio and Video synchronization is defined as the relative temporal distinction between the sound (audio) and image (video) during the transmission and reception. It is also known as audio-video sync, A/V sync and Audio/Video sync. Lip synchronization (lip sync or lip synch) refers to the voice is synchronized with lip movements. Human can able to detect the distinction between the audio and corresponding video presentation less than 100ms in lip sync problem. The lip sync becomes a significant problem in the digital television industry, filming, music, video games and multimedia application. It is corrected and maintained by audio-video synchronizers. In multimedia technology, the audio and video synchronization plays an important role in synchronizing audio and video streams. With the advancement of interactive multimedia application, distinct multimedia services like content on demand services, visual collaboration, video telephony, distance education and E-learning are in huge demand. In multimedia system applications, audio-visual streams are saved, transmitted, received and broadcasted. During an interaction time, the timing relations between audio-video streams have to be conserved in order to provide the finest perceptual quality.

2. PROPOSED METHODOLOGY

The proposed framework is automatically measuring and maintaining the perfect synchronization between audio and video using audio-visual spreading codes. Fig.1. shows the proposed framework for audio and video synchronization based on audio-visual spreading codes.

During transmission, the audio and video signals are processed individually. The audio spreading code is extracted from the spectrograph of the input audio which is broken up into chunks. The spectrogram is the visual way of representing the spectrum of sounds and it can be used to display the spoken word phonetically. It is also called spectral waterfalls, voice grams or voiceprints. The video spreading code is computed by the absolute difference the consecutive video frames where the input video is broken up into video frames and finally to attain a coarse absolute difference image. The audio-visual signature or A/V sync signature based on content and don't change excessively. It is an authentication mechanism and formed by taking hash of the original audio-video streams. The robust hash filters the little changes in the signal processing and reduces the audio-visual spreading code sizes. It is based on the difference between the successive audio and video frames. Within the communication network, the audio-video streams encounter different signal processing namely audio

Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

compression, video compression, format conversion, audio down sampling, video down sampling etc.., and their relative temporal alignment between audio and video signals may be altered.

At the detection, the processed audio and video codes are extracted from the processed audio and video streams. During synchronization, the processed audio and video

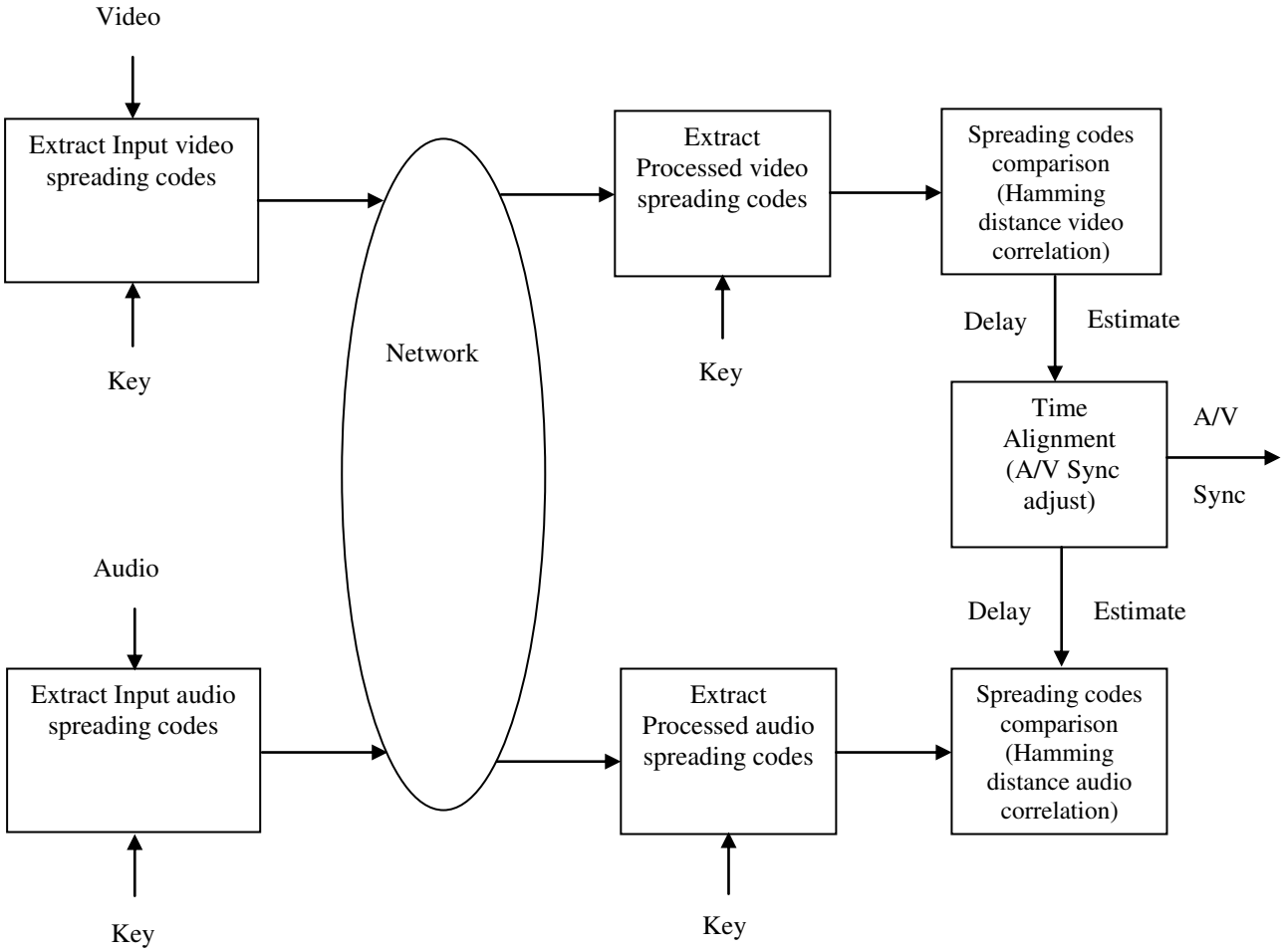

Fig.1. Audio-Video sync using audio-visual spreading codes.

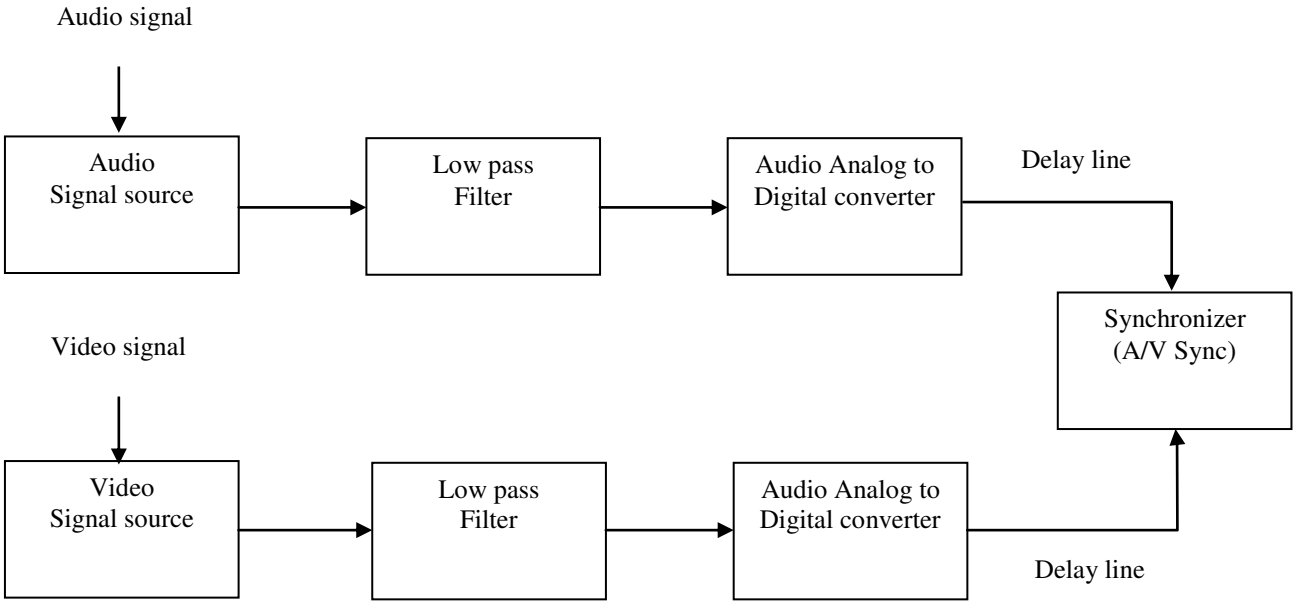

Fig.2 Audio – Video Synchronizer

Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

spreading codes are compared with the corresponding input audio and video signatures using Hamming distance correlation. The output of the Hamming distance is used to estimate the temporal misalignment between the audio-visual streams. Finally the measured delays are used to correct the relative misalignment between audio-visual streams.

The test content used for the performance assessment of the system consisted of A/V clips of a variety of content types such as scripted dramas; animation program, music concert, news programs, sports and lives music. The input audio and video is taken from the recorded dataset for the audio and video synchronization. The input video is divided into frames. A low-pass filter (LPF) is a filter that passes low frequency signals & attenuates high frequency signals by the cutoff frequency. It prevents the high pitches and removes the short-term fluctuations in the audio – video signals. It also produces the smoother form of a signal.

An analog-to-digital converter (ADC) converts an input analog voltage or current to a digital magnitude of the voltage or current. It converts a continuous time and continuous amplitude analog signal to a discrete time and a discrete amplitude signal. The delay line produces a specific delay in the audio and video signal transmission path. The Synchronizer is a variable audio delay used to correct and maintain the audio and video synchronization or timing.

3. EXPERIMENTAL RESULTS

A/V CONTENT

The test content used for the performance assessment of the system consisted of 5 seconds A/V clips of a variety of content types such as scripted dramas, talk programs, sports and live music. Fig.3, shows the input audio and video is taken from the recorded dataset for the audio and video synchronization. The frame number is given as the input for synchronization purpose.

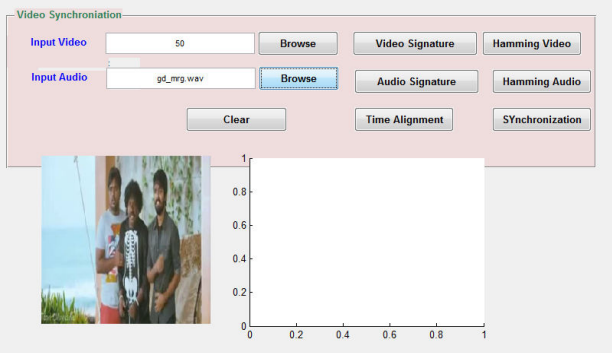

Fig.3. Input audio and video.

VIDEO FRAME

The input video is divided into frames for generating the video spreading codes. The input video is divided into 30 frames/seconds. There are totally 74 frame are available in the input video frame. Fig.4, shows the frame conversion of the input video.

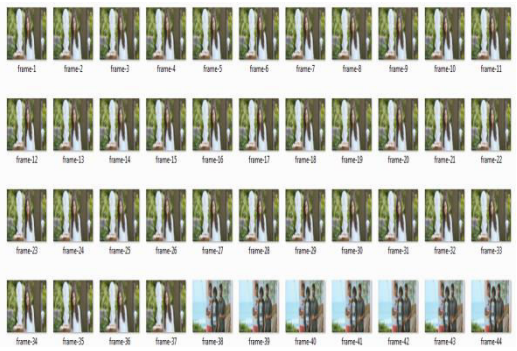

Fig.4. The frame conversion.

AUDIO SPREADING CODES GENERATION

The audio spreading code is primarily based on the coarse representation of the spectrograph onto random vectors. Fig.5, shows the audio spreading codes generation using spectrograph.

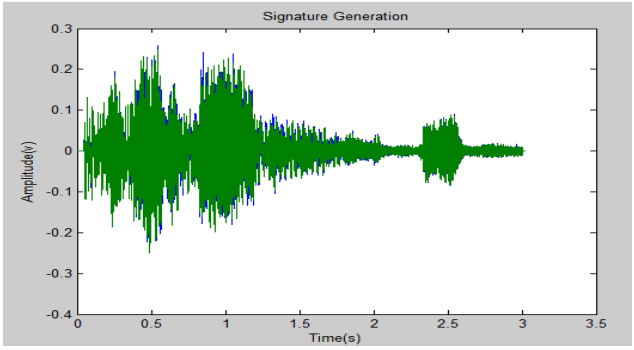

Fig.5. The audio spreading codes extraction using spectrogram.

VIDEO SPREADING CODES GENERATION

The video spreading code is based on coarse illustration of the distinction image between two consecutive. Fig.6, shows the video spreading codes generation.

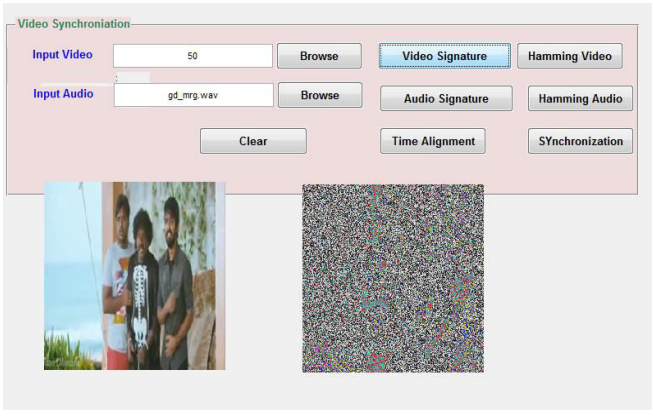

Fig.6. The video spreading codes generation.

HAMMING VIDEO AND HAMMING AUDIO

The Hamming distance correlation is used to calculate the temporal misalignment between audio-visual streams and the quality of the audio-video synchronization can be

Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

measured. Fig.7, shows the hamming video and Fig.8, shows the hamming audio.

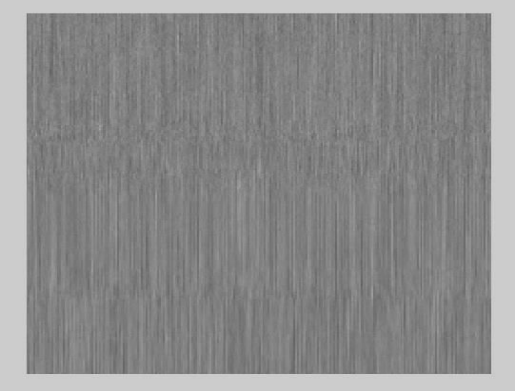

Fig.7. The hamming video.

Fig.8. The hamming audio.

Table.1 The hamming distance for video and audio.

INPUT IMAGE	ENCODE IMAGE	HAMMING VIDEO	HAMMING AUDIO
1010	1010	0	1
1010	1010	0	1
1010	1010	0	1
1010	1010	1	1
1110	1110	0	1
1010	1010	0	1
1010	1010	0	1
1110	1110	0	1
1110	1101	2	1
1010	1010	0	1
1110	1110	0	1
1110	1010	1	1
1001	1001	0	1
1001	1001	0	1
1010	1100	2	1
1001	1101	1	1
1110	1110	0	1
1110	1110	0	1
1011	1001	1	1
1000	1000	0	1
1000	1000	0	1
1011	1010	1	1
1011	1011	0	1

From the Table 4.1, it is inferred that the hamming distance for the video and audio. Hamming code is a set of error-correction code s that can be used to detect and correct bit errors. It is used to find the misalignment between the audio and video streams.

TIME ALIGNMENT

The estimated relative misalignment is used to achieve the same alignment between the audio and video streams that was present before processing. It aligns the audio and video frame in an appropriate manner. Fig.9, shows the relative time alignment between the audio and video stream. It decodes the corresponding video frame that is given as input in Fig. 3.3 with proper time alignment between the input and processed video frames. Fig.10, shows the decoded input video frame.

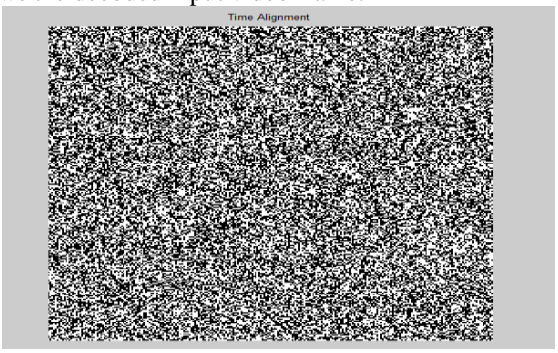

Fig.9. The audio-video stream time alignment.

Fig.10. The decoded video frame.

AUDIO – VIDEO SYNCHRONIZATION

Fig.11, shows the audio and video synchronization using signature. The A/V sync using spreading codes provides perfect synchronization between the corresponding audio and video streams. Finally, the reliability measures along with the estimated delays can be used to detect or correct the relative misalignment between the A/V streams. It can detect and maintain the audio - video sync accuracy.

Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

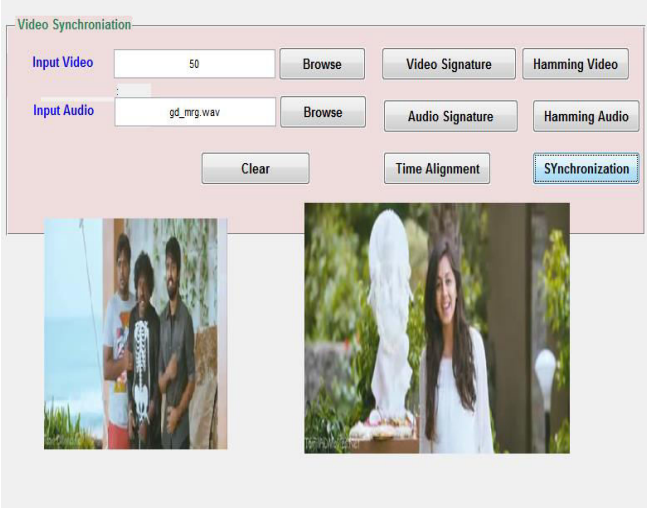

Fig.11. The audio – video synchronization.

AV SIGNALS

The test content used for the performance assessment of the system consisted of 5 seconds A/V clips. Fig. 12, shows the input audio is taken from the recorded dataset for the audio and video synchronization and every 10 msec for audio.

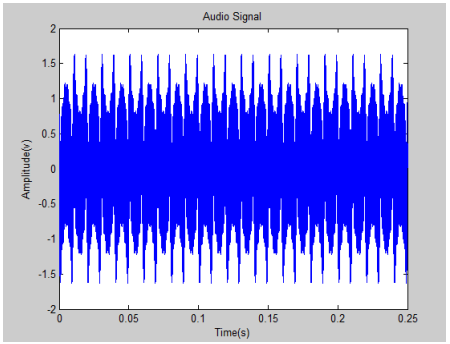

Fig.12 Input audio signal.

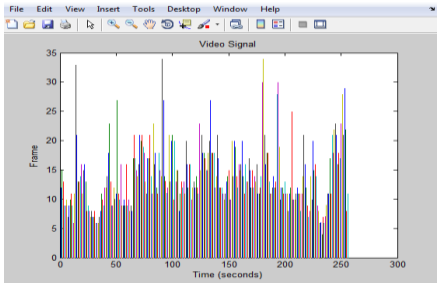

Fig.13 Input video signal.

Fig. 13, shows the input video is taken from the recorded dataset for the audio and video synchronization. Every video frame plays 3 msec. The input video is divided into 50 frames/seconds. There are totally 74 frame are available in the input video.

NOISE REMOVAL

Fig.14 shows the audio low pass filtered output. The filter allows the frequencies below the cut off frequency but the high frequencies in the input signal are attenuated.

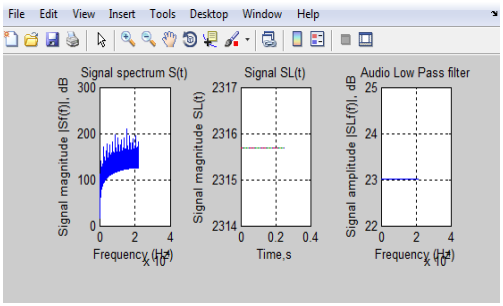

Fig.14 Audio Low passes filter output.

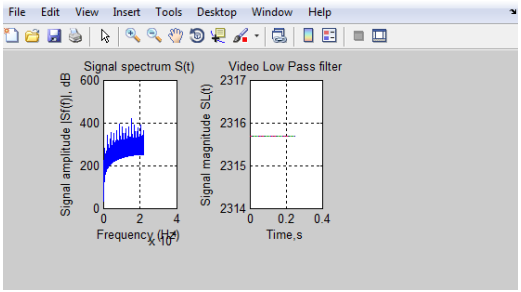

Fig.15 Video Low passes filter output.

Fig.15 shows the video low pass filtered output. The filter allows the frequencies below the cut off frequency but the high frequencies in the input signal are attenuated.

ANALOG TO DIGITAL CONVERSION

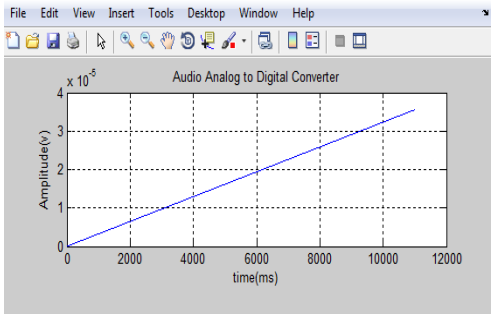

Fig. 16 Audio Analog to Digital converter.

Fig.16 shows the audio analog to digital converter output. ADC converts the analog audio signal into digital representing the amplitude of the voltage.

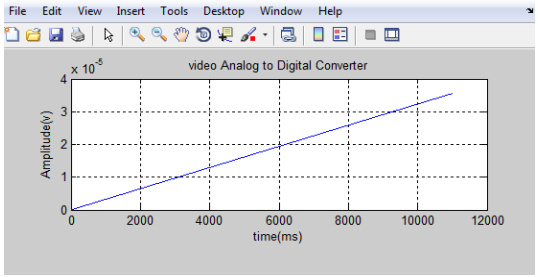

Fig.17 Video Analog to Digital converter.

Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

Fig.17 shows the video analog to digital converter output. ADC converts the analog video signal into digital representing the amplitude of the voltage.

A/V SYNCHRONIZATION

The objective of the synchronization is to line up both the audio and video signals that are processed individually.

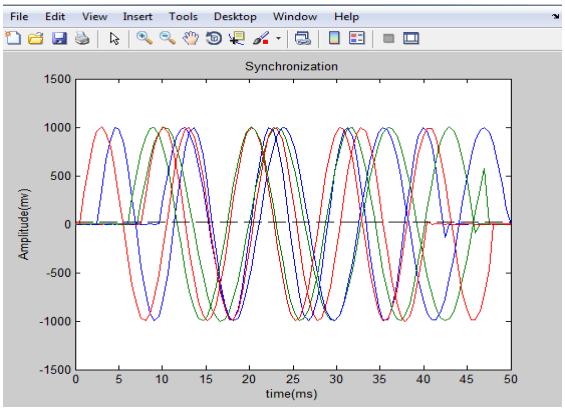

Fig. 18 A/V signal Synchronization.

Fig.18 shows the A/V signal synchronization. It aligns both audio and video signals. The synchronization is guaranteeing that the audio and video streams matched after processing.

4. PERFORMANCE ANALYSIS OF AUDIO-VIDEO SYNCHRONIZATION

SYNCHRONIZATION EFFICIENCY

The Synchronization efficiency is the process of establishing consistency among A/V content from a source to a processed or target A/V content storage in percentage. If the synchronization efficiency is high, then the audio and video are perfectly synchronized. Otherwise the audio and video synchronization will be poor. It is expressed as

$$\eta = P_{out}/P_{in} \qquad (1)$$

Where,
η = the Synchronization Efficiency.
P_{out} = the synchronized A/V stream.
P_{in} = the unsynchronized A/V stream.

AUDIO–VIDEO TIME DRIFT

The Time drift is defined as the amount of time the audio departs from perfect synchronization with the video where a positive number indicates the audio leads the video while the negative number indicates the audio lags the video. The audio – video time drift can be represented as

$$t_{A/V} = t_r - t_p \qquad (2)$$

Where,
$t_{A/V}$ = A/V Time drift.
t_r = Source time.
t_p = deviation time.

AUDIO TO VIDEO DELAY

The Audio to Video Delay is referred as the relative time alignment delay between the audio and video streams. The amount of the visual data is much bigger than audio data and the delays which are generated to the audio and video streams are typically unequal. The solution to audio to video delay is to add fixed delays to match the video delay. Finally, the estimated delays are used to correct the relative misalignment between the audio and video streams. The A/V delay is given as

$$D = t \pm t_0 \qquad (3)$$

Where,
D = The Audio – Video delay.
t = the audio/video time.
t_0 = the extra audio/video time.

Table.2.Performance analysis for audio and video synchronization.

PARAMETERS	A/V CONTENT
Synchronization Efficiency	99 %
Audio and video sync time drift	16 ms
Audio to video delay	16 ms

From the Table.2, it is inferred that the audio and video synchronization parameter. The synchronization efficiency is very high. The audio – video sync time drift and audio to video delay are very less.

5. CONCLUSION

Thus the audio and video synchronization using spreading codes technique was implemented and their performances were analyzed sufficiently and appropriately. The proposed system would automatically estimate and preserve the perfect synchronization between the audio and video streams and it would maintain the perceptual quality of audio and video. This method provides high-quality accuracy and low computational complexity. The experimental test results were shown the guarantee and quite simple process applicable for the real world multimedia application and offline applications. This method is suitable for content distribution network, communication network and traditional broadcast networks. In future work, the proposed framework will be developed with modified structures to provide vast improvement in real time application. Improvement of future work may also include improvement of the signature matching and thus increase the synchronization rate. Also we want to automate the detection of time drift to achieve a completely unsupervised synchronization process.

Volume: 10 Issue: 01 Pages: 3728-3734 (2018) ISSN: 0975-0290

6. ACKNOWLEDGMENT

At first, I thank Lord Almighty to give knowledge to complete the survey. I would like to thank my professors, colleagues, family and friends who encouraged and helped us in preparing this paper.

REFERENCES

[1] Alka Jindal, Sucharu Aggarwal, "Comprehensive overview of various lip synchronization techniques" *IEEE International transaction on Biometrics and Security technologies, 2008.*

[2] Anitha Sheela.k, Balakrishna Gudla, Srinivasa Rao Chalamala, Yegnanarayana.B, "Improved lip contour extraction for visual speech recognition" *IEEE International transaction on Consumer Electronics,pp.459-462, 2015.*

[3] N. J. Bryan, G. J. Mysore and P. Smaragdis, "Clustering and synchronizing multicamera video via landmark cross-correlation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2012, pp. 2389–2392.*

[4] Claus Bauer, Kent Terry, Regunathan Radhakrishnan, "Audio and video signature for synchronization" *IEEE International conference on Multimedia and Exposition Community (ICME), pp.1549-1552, 2008.*

[5] N. Dave, N. M. Patel. "Phoneme and Viseme based Approach for Lip Synchronization.", *International Journal of Signal Processing, Image Processing and Pattern Recognition, pp. 385-394, 2014.*

[6] Dragan Sekulovski, Hans Weda, Mauro Barbieri and Prarthana Shrestha, "Synchronization of Multiple Camera Videos Using Audio-Visual Features," in *IEEE Transactions On Multimedia, Vol. 12, No. 1, January 2010.*

[7] Fumei Liu, Wenliang, Zeliang Zhang, "Review of the visual feature extraction research" *IEEE 5th International Conference on software Engineering and Service Science, pp.449-452, 2014.*

[8] Josef Chalaupka, Nguyen Thein Chuong, "Visual feature extraction for isolated word visual only speech recognition of Vietnamese" *IEEE 36th International conference on Telecommunication and signal processing (TSP), pp.459-463, 2013.*

[9] K. Kumar, V. Libal, E. Marcheret, J. Navratil, G.Potamianos and G. Ramaswamy, "Audio-Visual speech synchronization detection using a bimodal linear prediction model". in *Computer Vision and Pattern Recognition Workshops, 2009, p. 54.*

[10] Laszlo Boszormenyi, Mario Guggenberger, Mathias Lux, "Audio Align-synchronization of A/V streams based on audio data" *IEEE International journal on Multimedia, pp.382-383, 2012.*

[11] Y. Liu, Y. Sato, "Recovering audio-to-video synchronization by audiovisual correlation analysis". in *Pattern Recognition, 2008, p. 2.*

[12] C. Lu and M. Mandal, "An efficient technique for motion-based view-variant video sequences synchronization," in *IEEE International Conference on Multimedia and Expo, July 2011, pp. 1–6.*

[13] Luca Lombardi, Waqqas ur Rehman Butt, "A survey of automatic lip reading approaches" *IEEE 8th International Conference Digital Information Management (ICDIM), pp.299-302, 2013.*

[14] Namrata Dave, "A lip localization based visual feature extraction methods" *An International journal on Electrical and computer Engineering, vol.4, no.4, December 2015.*

[15] P. Shrstha, M. Barbieri, and H. Weda, "Synchronization of multi-camera video recordings based on audio," in *Proceedings of the 15th international conference on Multimedia 2007, pp.545–548.*

Author Details

A. Thenmozhi (S.Anbazhagan) completed B.E (Electronics and Communication Engineering) in 2016 from Anna University, Chennai. She has published 2 papers in National and International Conference proceedings. Her area of interest includes Electronic System Design, Signal Processing, Image Processing and Digital Communication.

Dr. P. Kannan (Pauliah Nadar Kannan) received the B.E. degree from Manonmaniam Sundarnar University, Tirunelveli, India, in 2000, the M.E. degree from the Anna University, Chennai, India, in 2007, and the Ph.D degree from the Anna University Chennai, Tamil Nadu, India, in 2015. He has been Professor with the Department of Electronics and Communication Engineering, PET Engineering College Vallioor, Tirunelveli District, Tamil Nadu, India. His current research interests include computer vision, biometrics, and Very Large Scale Integration Architectures.