International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:11, No:7, 2017

Subjective Evaluation of Spectral and Time Domain Cascading Algorithm for Speech Enhancement for Mobile Communication

Harish Chander, Balwinder Singh, Ravinder Khanna

***Abstract*—**In this paper, we present the comparative subjective analysis of Improved Minima Controlled Recursive Averaging (IMCRA) Algorithm, the Kalman filter and the cascading of IMCRA and Kalman filter algorithms. Performance of speech enhancement algorithms can be predicted in two different ways. One is the objective method of evaluation in which the speech quality parameters are predicted computationally. The second is a subjective listening test in which the processed speech signal is subjected to the listeners who judge the quality of speech on certain parameters. The comparative objective evaluation of these algorithms was analyzed in terms of Global SNR, Segmental SNR and Perceptual Evaluation of Speech Quality (PESQ) by the authors and it was reported that with cascaded algorithms there is a substantial increase in objective parameters. Since subjective evaluation is the real test to judge the quality of speech enhancement algorithms, the authenticity of superiority of cascaded algorithms over individual IMCRA and Kalman algorithms is tested through subjective analysis in this paper. The results of subjective listening tests have confirmed that the cascaded algorithms perform better under all types of noise conditions.

***Keywords*—**Speech enhancement, spectral domain, time domain, PESQ, subjective analysis, objective analysis.

I. Introduction

MOBILE communication has become the need of the modern living. Although the mobile phone is also used for multimedia and data communication but primarily it is used for speech communication only. Because of the ease that a mobile phone provides to the user, it is being used in all the places irrespective of the surrounding noise level. The surrounding noise mostly affects the speech communication. Although the researchers in the past have developed several speech enhancement algorithms to de-noise the corrupted speech, but none of these algorithms perform uniformly well in the different noise environments and under different SNR conditions [1]. Most of these algorithms are based on processing the signal either in spectral domain or in the time domain that have their own advantages and disadvantages. There is a need to develop an algorithm that could perform uniformly well in all types of noise environments and under different SNR conditions.

A speech enhancement algorithm based on cascading of IMCRA [2] in the spectral domain and Kalman filter [3] in time domain was proposed by [4]. The comparative performance of these algorithms was adjudged in terms of three widely used objective parameters; Global SNR, Segmental SNR and PESQ. The results of objective evaluation have proven that cascaded algorithms perform better than the IMCRA and Kalman filter algorithms alone. The objective measures alone do not certify the quality of speech and thus the performance of speech enhancement algorithms cannot be certified by objective measures alone [5]. To ascertain the authenticity of the results obtained in objective measures, the enhanced speech is put to subjective listening tests. In fact subjective listening evaluation proves to be more authenticated to decide the quality of speech than the objective evaluation [6]. To ascertain the performance of cascaded algorithms the comparative analysis of subjective evaluation of IMCRA, Kalman filtering and cascaded algorithms is presented in this paper.

II. Subjective Evaluation

A. Subjective Evaluation Procedure

Subjective evaluation of speech quality is a method in which speech is made to listen by the subjects who then rate the quality of the speech. Three subjective measures are generally used to measure the quality of speech. These measures are Signal Distortion (SIG), Background Noise (BAK) and Overall Quality (OVRL) which are described as [7]:

- SIG – The listener observes only the signal and rates it on the five point scale as given in Table I.
- BAK – The listener observes only the BAK and rates it on the five point scale as given in Table I.
- OVRL – The listener observes the speech signal as a whole and rates it on the five point scale as given in Table I.

The pure speech corpus is taken from the TIMIT database [8]. The noise corpus used is recorded in real time using a Nokia mobile phone under typical noise environments [4]. The pure speech signal is mixed at different SNR values with standard real time noises and is processed through speech enhancement algorithms under test. The processed speech is then put into listening tests to the subjects.

Each trial of the listening test consists of three subsamples of the processed signal. Each subsample is at least of 4

Harish Chander is pursuing his PhD from I K Gujral Punjab Technical University, Jalandhar, India (corresponding author, phone: +918860020897; e-mail: harishrajni@rediffmail.com).

Balwinder Singh is with Centre for Development of Advanced Computing, Mohali, India (e-mail: balwinder@cdac.in).

Ravinder Khanna is with the Electronics & Communication Engineering Department, Maharishi Markandeshwar University, Sadopur, India (e-mail: ravikh2006@gmail.com).

seconds duration followed by a silent voting period of at least 1 second duration. After listening each subsample the listener has to rate only one of the parameters, SIG, BAK and OVRL, of the speech signal on a five point rating scale as mentioned in Table I. The order of rating for the first half of the trials is kept as 'BAK, SIG, OVRL', and for the second half of the trials, it is kept as 'SIG, BAK and OVRL'. The change in order balances the effects of rating scale order within the experiments.

TABLE I
DESCRIPTION OF SCALES FOR SIG, BAK AND OVRL

Scale	SIG	BAK	OVRL
5	No distortion	Not noticeable	Excellent
4	Slight distortion	Slightly noticeable	Good
3	Somewhat distortion	Noticeable but not intrusive	Fair
2	Fairly high distortion	Somewhat intrusive	Poor
1	Very high distortion	Very intrusive	Bad

B. Test Setup for Subjective Evaluations

As per P.835 standards [7], a total of 32 listeners are deputed to perform the test. To limit the length of the test only two sentences, one from male speaker and the other from female speaker, obtained from the TIMIT database as given in Table II, are tested in this paper. The pure speech signal is mixed at SNR levels of -5dB and 5dB with five different types of noises as described in Table III. Degraded speech is processed through MATLAB using IMCRA, Kalman, cascaded IMCRA-Kalman and cascaded Kalman-IMCRA algorithms. Samples of the processed speech signals are prepared as discussed in Section II *A* and are stored in PCs connected with headphone as per P.835 standards. Listeners are seated visibly separated from each other. The tests are conducted at Multimedia Laboratory of 'Centre for Development of Advanced Computing' C-DAC, Mohali, Punjab, India. The following guidelines are followed while selecting the listeners for the test:

1. Listeners have normal hearing ears.
2. Listeners are about 18 to 50 years of age.
3. Listeners can speak and understand the English language.
4. No listener has participated in the similar test in the last three months.
5. There are an equal number of male and female listeners.

TABLE II
LIST OF PURE SPEECH SENTENCES USED

S. No.	Speaker	Sentence
sp01	Male	The birch canoe slid on the smooth planks
sp02	Female	The friendly gang left the drug store

TABLE III
TYPES OF NOISE SIGNALS USED

Scale	Type of Noise
ns01	Multitalker babble noise
ns02	Railway platform train arrival
ns03	Car inside with windows closed
ns04	Exhaust fan noise
ns05	Street noise in running auto rickshaw

Before the start of the actual test, listeners are issued with the instructions they have to follow during the test. They are also given practice sessions for familiarization. The complete test is divided into blocks and in between tests, listeners are given a short break.

III. EVALUATED ALGORITHMS

A. IMCRA Algorithm

IMCRA [2] is a spectral domain algorithm in which noise is estimated by averaging past spectral power values using a smoothing parameter which is controlled by the minimum values of the smoothing parameter that is further adjusted by the speech presence probability in sub bands. Since efficiency of any speech enhancement algorithm depends upon accurate estimation of noise signal and speech presence interval detection in the degraded speech, the detection of the speech presence in IMCRA algorithm is carried out in two iterations. In the first iteration speech presence periods are estimated roughly and in the second iteration stronger speech components are eliminated, thus ensuring minimum tracking during speech presence [9], [10]. Speech presence probability is controlled more in the speech, absence period and very less during speech presence periods.

B. Kalman Filtering

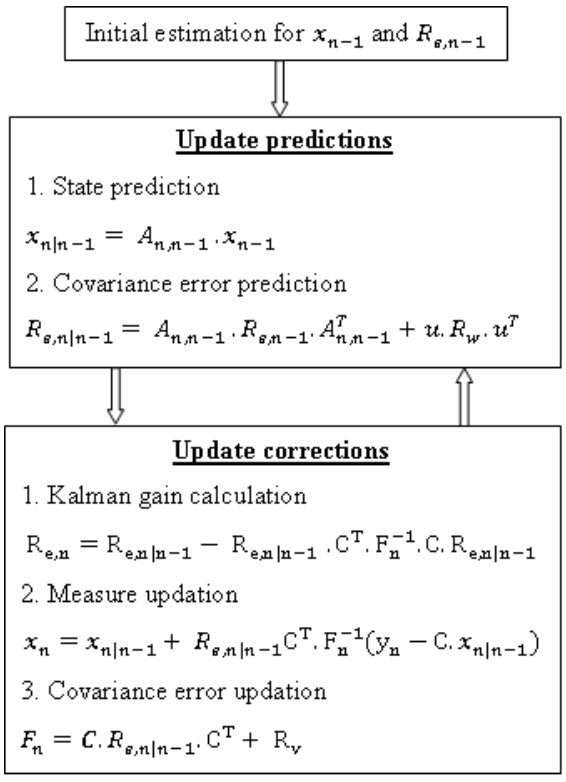

Fig. 1 Complete operation of Kalman filter

The Kalman filter is one of the finest time domain filters that provide optimum recursive solutions using least square method [3]. Kalman filter works on the principle of prediction

and correction with feedback control. The complete algorithm is described with the help of the block diagram in Fig. 1.

C. Cascading Algorithms of IMCRA and Kalman Filtering

Time domain and frequency domain speech enhancement algorithms alone do not provide complete enhancement of the noisy speech which can improve the speech quality as well as speech intelligibility under all types of noisy environments and SNR conditions [1]. To avail the benefits of spectral and time domain both, a cascaded algorithm using IMCRA in the spectral domain and Kalman filter in time domain was proposed in [4]. In the cascaded algorithm the degraded speech is processed first through an IMCRA/Kalman algorithm followed by Kalman/IMCRA algorithm. Fig. 2 shows the block diagram representation of the cascaded algorithms. The performance of the cascaded algorithms, individual IMCRA and Kalman algorithms was tested by [4] through objective parameters, Global SNR, Segmental SNR and PESQ and it was reported that with cascaded algorithms there is a substantial improvement in these parameters.

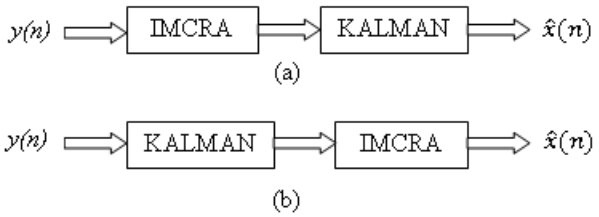

Fig. 2 Block diagram representation of spectral and time domain cascading algorithms, (a) IMCRA-KALMA, (b) KALMAN-IMCRA

IV. Result Analysis

Figs. 3-5 portray Mean Opinion Score (MOS) of three subjective parameters SIG, BAK and OVRL recorded by all the listeners under SNR conditions of 0dB, 5dB and 10dB respectively. Within each figure, (a), (b), (c), (d) and (e) portray MOS under noises ns01, ns02, ns03, ns04 and ns05, as described in Table III, respectively. Each value for respective parameter in the graph is the mathematical average of all the 32 listeners and for speech signals sp01 and sp02 as described in Table II. From the figures, it can easily be analyzed that in most of the cases, MOS for cascaded algorithms is higher than the MOS of the individual Kalman or IMCRA algorithms.

V. Conclusion

From the comparative subjective analysis of individual Kalman and IMCRA algorithms and the cascaded algorithms, it is concluded that the cascaded algorithms of Kalman and IMCRA substantially improve the enhanced speech. The subjective results presented here authenticate the findings of the authors described in [4].

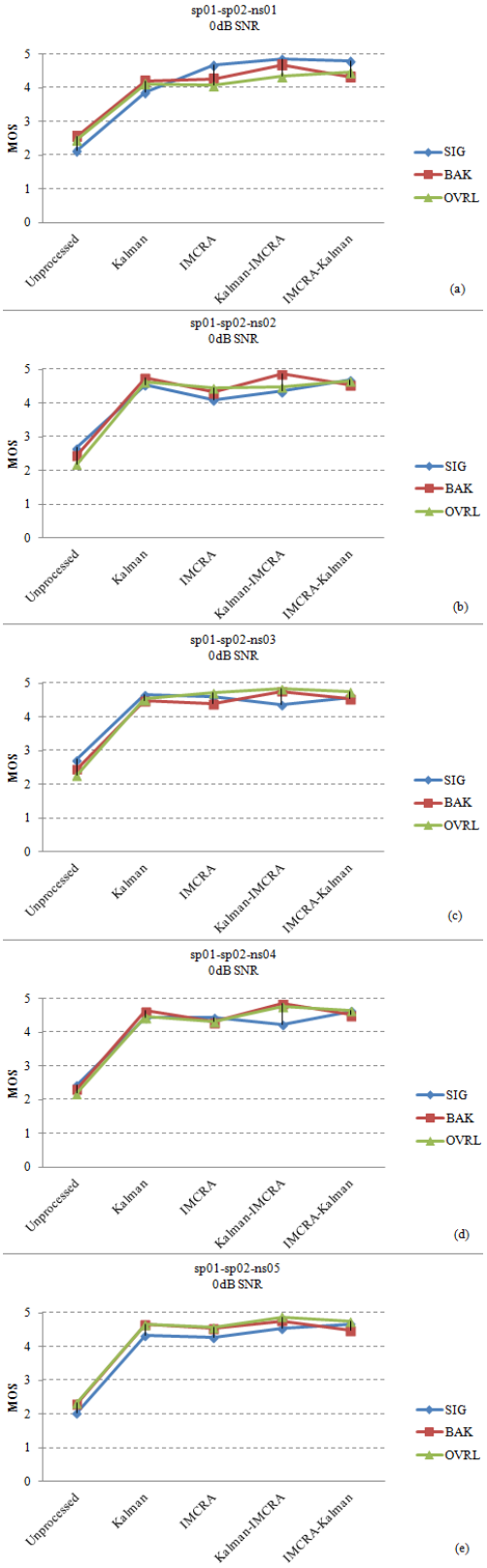

Fig. 3 MOS of subjective listening tests under 0dB SNR condition, average of 32 listeners and two speech signals sp01 and sp02, corrupted with noise, (a) ns01, (b) ns02, (c) ns03, (d) ns04, (e) ns05

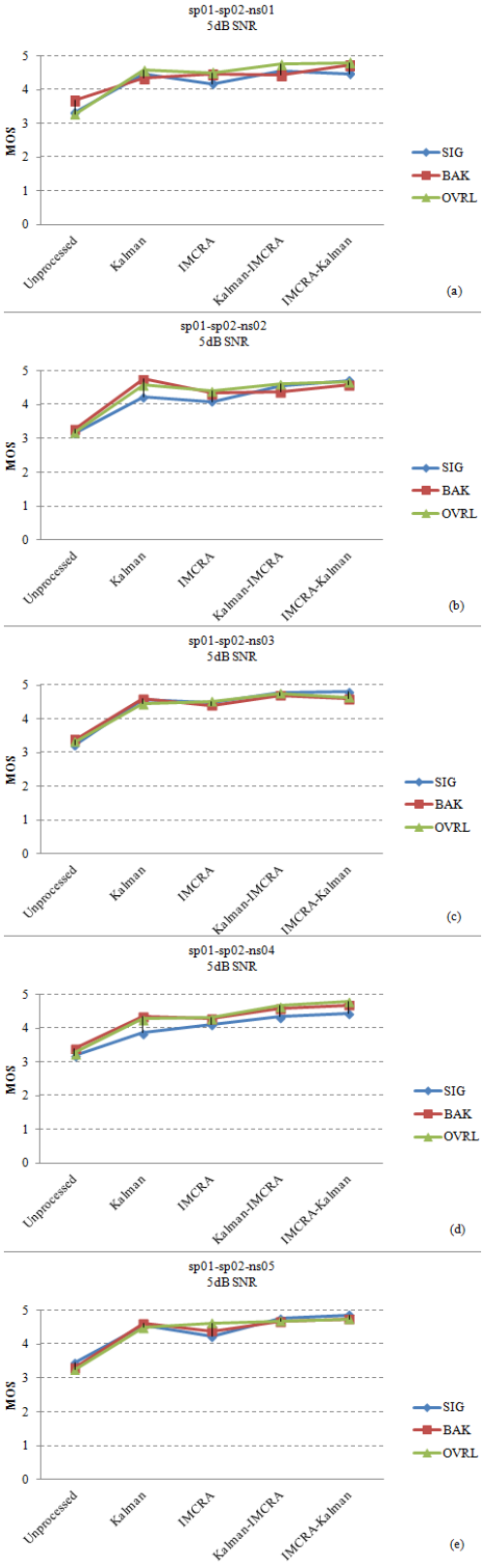

Fig. 4 MOS of subjective listening tests under 5dB SNR condition, average of 32 listeners and two speech signals sp01 and sp02, corrupted with noise, (a) ns01, (b) ns02, (c) ns03, (d) ns04, (e) ns05

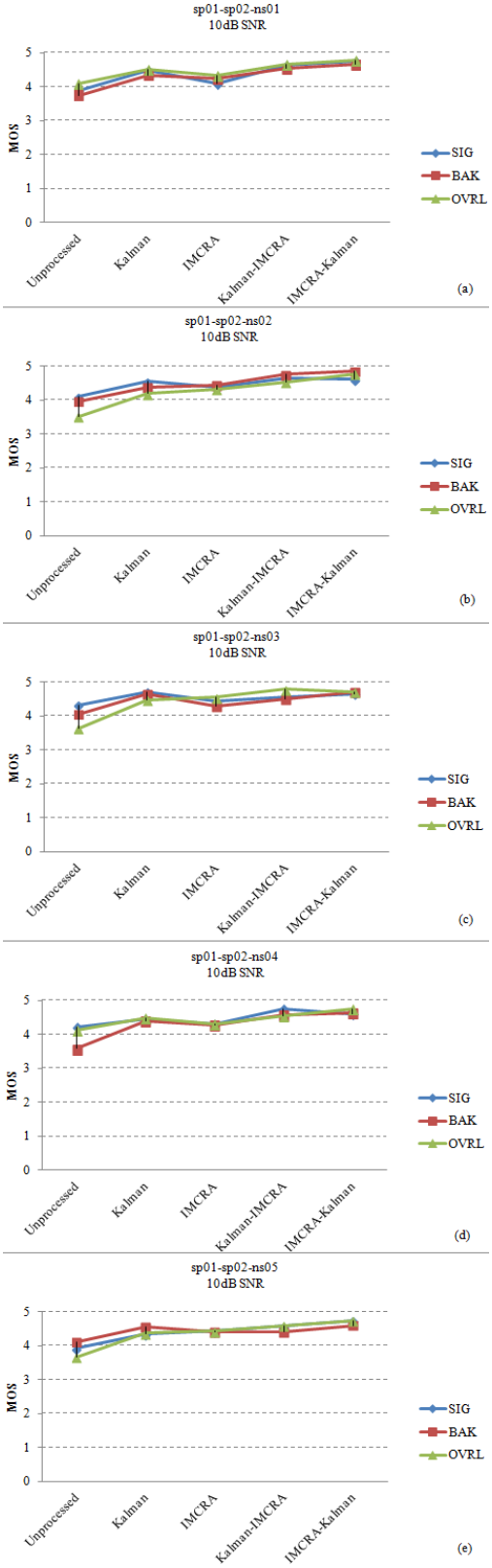

Fig. 5 MOS of subjective listening tests under 10dB SNR condition, average of 32 listeners and two speech signals sp01 and sp02, corrupted with noise, (a) ns01, (b) ns02, (c) ns03, (d) ns04, (e) ns05

REFERENCES

[1] S. Quackenbush, T. Berwell and M. Clements, *Objective Measures of Speech Quality. Englewood Cliffs*. NJ Prentice-Hall, 1988.
[2] Yi Hu, Philipos C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *NIH Speech Commun*., 49 (7) pp. 588–601, July 2007.
[3] Philipos C. Loizou, Gibak Kim, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, No.1, pp. 47-56, 2011.
[4] Harish Chander, Balwinder Singh, Ravinder Khanna, "Effective speech enhancement for mobile communication using cascading of frequency and time domain techniques," *SERSC, International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 10, No. 5, pp. 45-56, May 2017.
[5] Israel Cohen, "Noise spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, issue 5, pp. 466-475, Sep. 2003.
[6] Kybic, B. J., *Kalman Filtering and Speech Enhancement*, Polytechnic Thesis, Ecole Polytechnique De Lausanne, 1998.
[7] ITU-T P.835, *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*, ITU-T Recommendation P.835, 2003.
[8] John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett and Nancy L. Dahlgren, *Getting started With the DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database*, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, 1993.
[9] I. Cohen and B. Berdigo, "Spectral enhancement by tracking speech presence probability in subbands," *Proc. IEEE workshop on hands free speech Communication*, HSC2001 Kypto, Japan, pp. 95-98, Apr. 2001.
[10] I. Cohen and B. Berdigo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, No. 11, pp. 2403-2418, Nov. 2001.

Harish Chander, born in 1968, has graduated in Electronics and Communication Engineering from Institution of Engineers, India, in 1994. He has completed his Masters in Digital Systems from MNNIT, Allahabad, India, in 2002. He has served the Indian Air Force as electronics & telecommunication engineer for 20 years. For the last 11 years he has been serving various academic organizations in different positions and presently he is working as Controller of Examinations with Institution of Electronics & Telecommunication Engineers, India. He is pursuing his PhD from I K Gujral Punjab Technical University, Jalandhar, India. His areas of research are signal processing, speech processing and wireless communication.

Dr. Balwinder Singh has graduated in Electronics and Communication Engineering from National Institute of Technology, Jalandhar, India, in 2002. He has completed his Masters in Microelectronics from Punjab University, Chandigarh, India, in 2004 and PhD from Guru Nanak Dev University, Amritsar, India, in 2014. He is presently working as Senior Engineer & Coordinator ACS Division, Centre for Development of Advanced Computing (C-DAC), Mohali, India. His Research Interest are Low power VLSI Design and Testing, Digital IP cores and analog modules, Sensor and MEMS design and modeling, FPGA based embedded systems and Image Processing for Embedded systems.

Dr. Ravinder Khanna has graduated in Electrical Engineering from Indian Institute of Technology (IIT), Delhi in 1970 and has completed his Masters and Ph.D in Electronics and Communication Engineering from the same Institute in 1981 and 1990 respectively. He has worked as an Electronics Engineer in Indian Defense Forces for 24 Years where he was involved in teaching, research and project management of some of the high tech weapon systems. Since 1996 he has switched to academics. He has worked in many premiere technical institutes in India and abroad. Currently he is Professor and Dean Research with Maharishi Markandeshwar University, Sadopur, Haryana, India. He is active in the general areas of Computer Networks, Image Processing and Natural Language Processing.