ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 4, April 2015

Comparative Study of Filter Performance for Separation of Singing Voice from Music Accompaniment

Harshada P. Burute[1], Madhuri Patil[2], Kirtimalini Chaudhari[3], Dr. Pradeep B. Mane[4]

Department of Electronics, All India Shri Shivaji Memorial Society, Institute of Information Technology, Pune, Maharashtra, India[1,4]

Department of Electronics, All India Shri Shivaji Memorial Society, College of Engineering, Pune, Maharashtra, India[2,3]

Abstract: An audio signal is a representation of sound. Audio signals have frequency range 20 to 20 kHz. Audio signals may be synthesized directly. A mixture refers to the physical combination of two or more substances on which the identities and are mixed in the form to separate out. An audio signal classification system should be able to categorize different audio input formats (speech, background noise, and music). Audio signal classification system analyzes the input audio signal and describes the signal at the output. These are used to characterize both music and speech signals. The categorization can be done on the basis of pitch, music content, music tempo and rhythm. From the comparative results it is observed that the wiener filter is better for noise reduction than others. We refer SEGSNR parameter for study because of its improved filter performance. Separating singing voice from music is very useful in many applications.

Keywords: music genre, mixture, audio signal classification, pitch, music tempo, music rhythm.

I. INTRODUCTION

Speech is an acoustic signal produced from a speech production system. An audio signal is a representation of sound. Audio signals have frequencies in the audio frequency range of roughly 20 to 20,000 Hz. It is well known that the human auditory system has a remarkable capability in separating sounds from different sources [1]. In an influential book [2], Bregman proposed that the auditory system employs a process called auditory scene analysis (ASA) for different sound sources. The work by Mellinger represents the first computational auditory scene analysis (CASA) system attempt to musical sound separation [1].

Singing pitch estimation and singing voice separation are challenging due to the presence of music accompaniments that are often non stationary and harmonic [4]. Speech separation is a very challenging task in signal processing. An Audio signal classification system detecting the audio type of a signal (speech, background noise and musical genres).

Singing is used to produces musically relevant sounds by the human voice, and it is employed in most cultures for entertainment or self-expression. The singing voice becomes immediately the main focus of attention when we listen to musical pieces with a vocal part. Recently, along with the development of multimedia technology, a variety of speech communication services using speech commands have become popular [6]. Most songs, especially popular songs, are mixtures of singing voice and music together. Music recording are either monaural (single channel) or stereo (two channel).

Ozerov [3] introduce a general formalism for source model adaptation which is expressed in the framework of

Bayesian models. Particular cases of the proposed approach are then investigated experimentally on the problem of separating voice from music in popular songs. The obtained results show that an adaptation scheme can improve consistently and significantly the separation performance in comparison with nonadapted models.

Comb Filter [12], Kalman Filter [13], and Wiener Filter [16], [17] are mostly used for Sound Separation in many research papers. Audio signal separation has been a topic of research for many years.

A singing voice separation system has its applications in areas such as automatic lyrics recognition and alignment, singer identification, musical information retrieval, karaoke, musical genre classification, melody extraction, audio signal classification[1],[6],[7], etc.

The organization of the paper is as follows. In Section II literature survey of sound separation and adaptive filters. Section III shows that comparative results of filter performance. In Section IV, we summarize conclusion.

II. LITERATURE SURVEY

Li and Wang [1], proposed a computational auditory scene analysis (CASA) system to separate singing voice from music accompaniment for monaural recordings. System consist of singing voice detection stage, pitch detection

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 4, April 2015

stage used hidden Markov model (HMM) and separation stage. Singing voice separation from monaural recordings where only one channel is available. Kim identified the large majority of sounds generated during singing is voiced (about 90%), while speech has a larger amount of unvoiced sounds Wang [1] described. Ozerov *et al.* [3] focused on the difference of spectral distribution (timbre) of singing voice and instruments, and modelled them by Gaussian mixture model (GMM). In their method, the GMM was trained in advance in a supervised way, and tuned adaptively for each input. Some studied utilized the pitch information of singing voice. Tandem algorithm [4] that estimates the singing pitch and separates the singing voice jointly and iteratively. Algorithm detects multiple pitch contours and separates the singer by estimating the ideal binary mask (IBM). System having trend estimation algorithm first estimates the pitch ranges of the singing voice. Many source separation algorithm have been developed including computational auditory scene analysis [1], independent component analysis[16], blind source separation[14], hidden Markov models, support vector machines, sinusoidal modelling, and non-negative matrix factorization (NMF) [5], [9].Tachibana *et al.* [6] focused on the fluctuation of a singing voice and considered to detect it by exploiting two differently resolved spectrograms.

In real-world audio signals several sound sources are usually mixed. The process in which individual sources are estimated from the mixture signal is called sound source separation. Adaptive filter is important in the signal processing. Adaptive filter is used to reject unwanted signal & take pure signal. An adaptive filter has an adaptation algorithm that is meant to monitor the environment & vary the filter transfer function accordingly. Based in actual signal received attempts to find optimum filter design [10].

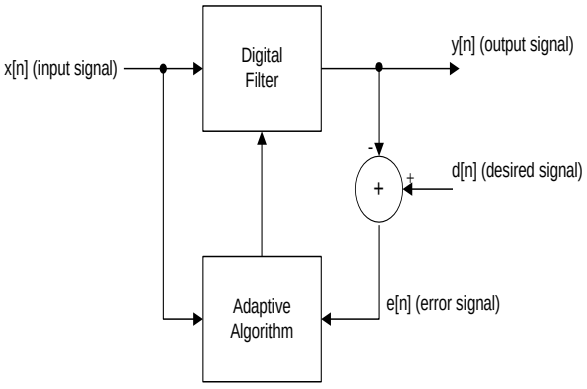

Fig.1. Basic Adaptive filter model [10]

In figure of basic adaptive filter model has only one input signal x[n] and one output signal y[n]. For adaptive filter, d[n] and e[n] signals are required. When the filter is operating in an unknown environment these required quantities need to be found from the accumulated data. The basic operation now involves 2 processes as –

1) **Filtering process-** which produces an output signal in response to given input signal.
2) **Adaptation process**- which aims to adjust the filter parameters to the environment. Because of complexity of optimization algorithms most adaptive filters are digital filter that perform digital signal processing. When processing analog signal the adaptive filter is then preceded by ADC and DAC converter [8], [10].

The removal of unwanted signals through the use of optimization (minimization) theory is becoming popular, basically in the area of adaptive filtering. Adaptive filters have a self-adjusting ability [10]. It can eliminate unwanted signals from useful information. These filters minimize the mean square of the error signal.

Sound source separation mostly researchers used Comb Filter [12], Kalman Filter [13] and Wiener Filter [16], [17]. In this paper we discuss about above three filters.

Gainza*et al.* [12] developed a method for separating harmonic sound sources using FIR **comb filters**. In this method a pre-processing task is performed by a multipitch estimator to detect the pitches [11] that the signal is composed of. Then, a method based on the Short Time Fourier Transform (STFT) is utilized to iteratively extract the harmonics belonging to a given source by using FIR comb filters.

Goh*et al.* [13] proposed a (single) speech model which can satisfactorily describe both voiced and unvoiced speech, as well as silence. Since it originates from autoregressive modeling, the long-term characteristics of noise are naturally taken care of. Coupling the proposed speech model with the popular additive white-Gaussian-noise model, they are able to treat the enhancement problem quite realistically on a theoretical basis. Main objective is to obtain an optimal estimate of the clean speech in the minimum-mean-square-error (MMSE) sense, using this model. To obtain this, first reformulate the model equations so as to facilitate a subsequent application of the well-established Kalman filter for computing the desired estimate. Performance assessment based on spectrogram plots,objective measures and informal subjective listening tests allindicate that this method gives consistently good results. As far assignal-to-noise ratio is concerned, the improvements over existingmethods can be as high as 4dB.

Chen *et al.* [18] described about the problem of noise reduction has attracted a considerable amount of research attention over the past several years. Among the numerous techniques that were developed, the optimal Wiener filter can be considered as one of the most fundamental noise reduction approaches, which has been delineated in different forms and adopted in various applications. Although it is not a secret that the **Wiener filter** [9], [15] may cause some detrimental effects to the speech signal (appreciable or even significant degradation in quality or intelligibility), few efforts have been reported to show the inherent relationship between noise reduction and speech distortion. By defining a speech-distortion index to

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 4, April 2015

measure the degree to which the speech signal is deformed and two noise-reduction factors to quantify the amount of noise being attenuated, this paper studies the behavior quantitative performance of the Wiener filter in the context of noise reduction. Results show that in the single-channel case the a *posteriori* signal-to-noise ratio (SNR) [17] (defined after the Wiener filter) is greater than or equal to the a *priori* SNR (defined before the Wiener filter), indicating that the Wiener filter is always able to achieve noise reduction.

III. COMPARATIVE STUDY

Main performance parameters are SNR (Signal to Noise Ratio), Segmental SNR, SDR (Signal to Distortion Ratio), source-to-interferences ratio (SIR), and sources-to artifacts Ratio (SAR). We refer SEGSNR [dB] with noise level for comparative study of filter performance because of its better noise reduction parameter than others. When noise level increases SEGSNR (Segmental SNR) also increases which improves the filter performance for sound separation. TABLE I shows that all comparative parameters values [dB].

TABLE. I. COMPARATIVE PARAMETERS

NO.	Para-meters	Ref. No.	Filter Used	Values
1	**SNR** (signal-to-noise ratio) [dB]	[13]	Kalman Filter	-5, 0,5,10.
		[18]	Wiener Filter	20, 15, 10.
2	**SEGSNR** (segment-al SNR) [dB]	[13]	Kalman Filter	-11.72, -6.73, -1.71, 3.29.
		[15]	Spectro-temporal filter	6.59, 0, 10.99, 12.86.
3	**SAR** (signal to artifacts ratio) [dB]	[16]	Kalman Filter	11.15, 10.05, 10.08.
4	**SDR** (signal-to-distort-ion ratio) [dB]	[16]	Kalman Filter	3.05, 2.23, 1.58.
5	**SIR** (signal-to-interfe-rence ratio) [dB]	[16]	Kalman Filter	10.35, 6.47, 4.2.

Differences between the Kalman and Wiener theories are listed below [19].

- The Kalman theory allows consideration of nonstationary processes, including a finite initial time; the Wiener theory does not.
- The Wiener theory does not draw great distinction between colored and white measurement noise. The Kalman theory in the first instance demands white measurement noise, but extension of the theory to the colored noise case is possible by modelling colored noise as the output of a linear system driven by white noise [13].
- The Kalman theory is essentially concerned with finite-dimensional systems. The Wiener theory permits infinite-dimensional systems, although the task of spectral factorization becomes much more difficult, and is still central to application of sound separation.

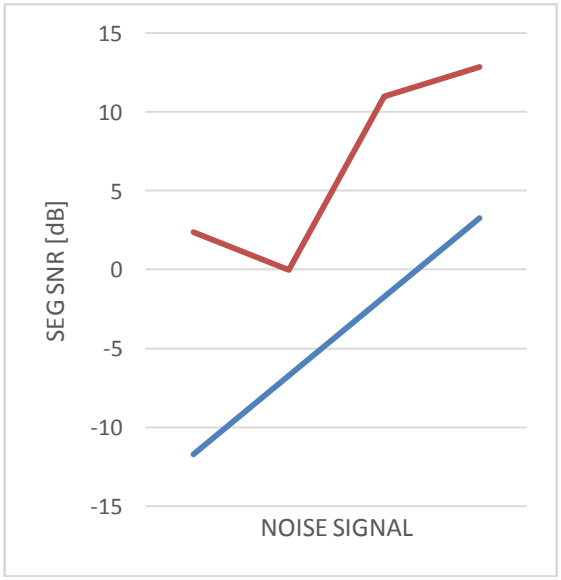

Fig. 2. Comparative result of Segmental SNR in Noisy signal [13], [15].

Fig.2 shows that, the Spectro-temporal filter (Wiener Filter) is most suitable for separation of singing voice from music accompaniment because of its improved filter performance (SEGSNR).

IV. CONCLUSION

From the comparative results, it is observed that the Wiener Filter is better for noise reduction than Comb Filter and Kalman Filter with respect to noise level and SEGSNR (Segmental SNR). The Wiener Filter is most suitable for separation of singing voice from music accompaniment because of its improved filter performance (SEGSNR).

ACKNOWLEDGMENT

Firstly, author would like to thank Dr. P. B. Mane and **Prof. Kirtimalini Chaudhary** for his valuable guidance, advice & support. Secondly, author would also like to thank Dr. D. K. Shedge, Head of Electronics Engineering Department for providing his valuable support. Author would lastly thank all the staff of PG section, friends and parents for their guidance & support.

REFERENCES

[1] Yipeng Liand DeLiang, "Separation of Singing Voice From Music Accompaniment for Monaural Recordings", IEEE

TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 4, MAY 2007, pp. 1475 – 1487.

[2] A.S.Bregman, Auditory scene analysis. Cambridge, MA: MIT press, 1990, pp.1-45,455-528.

[3] Alexey Ozerov, Pierrick Philippe, Frederic Bimbot, and RemiGribonval, "Adaptation of Bayesian models for single channel source separation and its application to voice/music separation in popular songs", IEEE Transactions on audio, speech, and language processing, vol. 15, no.5, July 2007, pp. 1564-1578.

[4] Chao-Ling Hsu, DeLiang Wang, Jyh-Shing Roger Jang, and Ke Hu, "A Tandem Algorithm for Singing Pitch Extraction and voice Separation from Music accompaniment", IEEE Transactions on audio, speech, and language processing, vol.20, no.5, July 2012, pp.1482-1491.

[5] Hideyuki Tachibana, Nobutaka Ono, and Shigeki Sagayama, "Singing voice enhancement in monaural music signals based on two-stage harmonic/percussive sound separation on multiple resolution spectrograms" , IEEE Transactions on audio, speech, and language processing, vol.22, no.1, January 2014, pp.228-237.

[6] Bilei Zhu, Wei Li, Ruijiang Li, and XiangyangXue, "Multi-stage non negative matrix factorization for monaural singing voice separation", IEEE Transactions on audio, speech, and language processing, vol.21, no.10, October 2013, pp.2096-2107.

[7] S. Umesh and Rohit Sinha, "A study of filter bank smoothing in MFCC features", IEEE Transactions on audio, speech, and language processing, vol.15, no.8, November 2007, pp.2418-2430.

[8] Abdullah Celik, MilutinStanacevic and GertCauwenberghs, "Mixed signal real time adaptive blind source separation".

[9] Zafar Rafii, Francois G. Germain, Dennis L. Sun, and Gautham J. Mysore, "Combining modelling of singing voice and background music for automatic separation of musical mixtures", ISMIR, 2013.

[10] Ifeachor and Jervis, "digital signal processing: a practical approach", second edition, Pearson educations, pp.645-680.

[11] Michael Stark, Michael Wohlmayr, and Franz Pernkopf, "Source-filter-based single-channel speech separation using pitch information", IEEE Transactions on audio, speech, and language processing, vol.19, no.2, February 2011, pp.242-255.

[12] Mikel Gainza, Robert Lawlor, and Eugene Coyle, " Harmonic Sound Source Separation using FIR Comb Filters" , Audio Research Group at ARROW@DIT, October 2004.

[13] ZentonGoh, Kah-Chye Tan and B. T. G. Tan, "Kalman-Filtering Speech Enhancement Method Based on a Voiced-Unvoiced Speech Model", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 7, NO. 5, SEPTEMBER 1999, pp. 510 – 524.

[14] Benxu Liu, VaninirappuputhenpurayilGopalanReju, Andy W. H. Khongand VinodVeeraReddy,"A GMM Post-Filter for Residual Crosstalk Suppression in Blind Source Separation", IEEE SIGNAL PROCESSING LETTERS, VOL. 21, NO. 8, AUGUST 2014, pp. 942 – 946.

[15] Yu Gwang Jin, Jong Won Shinand Nam Soo Kim, "Spectro - Temporal Filtering for Multichannel Speech Enhancement in Short-Time Fourier Transform Domain", IEEE SIGNAL PROCESSING LETTERS, VOL. 21, NO. 3, MARCH 2014, pp. 352 – 355.

[16] AlirezaMasnadi-Shiraziand Bhaskar D. Rao, "An ICA-SCT-PHD Filter Approach for Tracking and Separation of Unknown Time-Varying Number of Sources", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 4, APRIL 2013, pp. 828- 841.

[17] Ibrahim Almajai and Ben Milner, "Visually Derived Wiener Filters for Speech Enhancement", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 6, AUGUST 2011, pp. 1642 – 1651.

[18] Jingdong Chen, Jacob Benesty, Yiteng (Arden) Huangand Simon Doclo,"New Insights Into the Noise Reduction Wiener Filter", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 4, JULY 2006, pp. 1218 – 1234.

[19] B. D. O. Anderson and J. B. Moore, Optimal Filtering. Englewood Cliffs, NJ: Prentice-Hall, 1979.