VOCAL REMOVED FROM MULTIOBJECT AUDIO

Sneha Jadiye$^1$, Aarti Bakshi$^2$, Manjiri Gogate$^3$
$^1$Electronics and Telecommunications, SLRTCE
$^2$Information Technology, SPIT
$^3$Electronics, SLRTCE

Abstract—This project presents a new approach for the implementation of Karaoke Machine using Least Mean Square filter by outcome the filter coefficients that narrate to creating the smallest mean squares of the fault signal. It is a stochastic slope descent method in that the filter is only modified based on the error at the current time. The basic idea behind Least Mean Square filter is to method the optimum filter weights by updating the filter weights in a manner to unite to the best filter weight. The process starts by presumptuous a minor weights (zero in most cases), and at each step, by finding the slope of the mean square error, the bulks are informed. The basic use of Least Mean Square filter is to remove undesirable noise from the Music signal by using Out Of Phase Stereo model.

Keywords—Karaoke Machine, MATLAB, Out of Phase Stereo (OOPS), Simulink, LMS

I. INTRODUCTION

Traditional audio services provide users with audio signals produced by mixing several audio objects such as vocal and musical instruments. However, users can control only the overall volume of the selected music because the producer predetermined the music. Since users demands for an advanced audio service is continuously increasing, an interactive audio service (IAS) and MUSIC 2.0 had been introduced. In the IAS, individual audio objects and the preset information are transmitted to users instead of the already mixed music. The IAS usually has two modes. One is a preset and the other is an interactive mode. In the preset mode, users reproduce the music using the transmitted audio objects and the preset information. Therefore, they can only select audio objects and control their volume. On the other hand, in the interactive mode, users can generate their own music by using the transmitted audio objects. The interactive mode can support the Karaoke and the personal studio services. The Karaoke service supports the background music for singing users, and the personal studio one, remixing of their voices with the background music. Although the present IAS can satisfy these demands, the service is still not suitable for mobile environments. Needed bit rate is directly proportional to the number of audio objects and the whole coded information has to be transmitted to the users. In other words, the present IAS is appropriate to use only when a broadband network or broadcasting system, and a personal device with a mass storage are guaranteed. In order for the IAS to be more useful and effective, the required bit rate has to be lowered as much as possible.

Karaoke from Japanese kara, "empty" or "void"(and kesutora, orchestra") is a form of entertainment in which users sing along with recorded music using a microphone and a PA system. The music is typically a well-known song in which the voice of the original singer is removed or educed in volume. Lyrics are usually displayed on a video screen, along with a moving symbol or changing color, to guide the singer.

In some countries, karaoke with video lyrics display capabilities is called KTV. Karaoke machines and numerous applications in studios, theatres, in DVD and VCD players, music or voice recording as well as in games and reality shows and singing contests. In karaoke, a popular entertainment, the quality of the accompanying sound track directly affects the aesthetic experience of the user singer.
Reproduction of an accompaniment requires lots of time and labor, yet the music is probably not close enough to the original mix. Thus, removing the vocal or lead instrument track from the original recording remains an attractive topic. For voice suppression, there are methods like applying Equalizers to the original recording attenuating the vocal frequency bands of the spectrum. This can be done by the means of multiple filters of varying pass band frequencies. But this cannot remove the vocal completely and the spectral composition of the original material is altered a lot. Also, Filtering changes/distorts the pitch of the original voice. On the other hand bass component gets suppressed and distorted.

Among many adaptive algorithms that exist in the open literature, the class of approaches which are derived from the minimization of the mean squared error between the output of the adaptive filter and some desired signal, seems to be the most popular. Probably the simplest algorithm belonging to this class is the Least Mean Squared (LMS) algorithm which has the advantage of low complexity and simplicity of implementation. One of the main concerns in all practical situations is to develop algorithms which provide fast convergence of the adaptive filter coefficients and in the same time good filtering performance. There are four main classes of applications where the adaptive filters were applied with success, namely: system identification, inverse modeling, prediction and interference cancelling.

II. PROPOSED MODEL

The model is shown in following figure. The model is designed in MATLAB Simulink. The music signal whose voice speech has to be eliminated is given from a file. A part of the speech signal is eliminated using an Out of Phase Stereo algorithm which eliminate the voiced signal to some extent. The remainder of the music signal is applied to a LMS filter, one for each phase and the output is given concatenated and saved. The project operation is initiated by executing MATLAB Version R2013a. The stored multimedia file is imported in MATLAB and decode the music for obtaining MP3 format irrespective of any format in which it is saved (e.g. MP2, wma, wav). The program will obtain the audio samples and pass it to a column separator. By using out of phase stereo method the individual mono signal will be separated from the stereo signal. Signal inversion takes place in the first column and the addition of the inverted with the original signal takes place. The signal is then passed on to an LMS (Least Mean Square) filter which will remove the additional speech signal. And finally an audio signal without speech is obtained as the output.

Figure 1. Proposed Karaoke Machine

III. SIMULATION RESULTS

The Model is the real time processing engine of the OOPS algorithm. It provides enhanced capabilities by the implementation of a normalized LMS filter. This filter has an adaptive response and it changes its response for each new processing frame to match the vocals and remove the
vocals. This allows us to achieve a much more lifelike non-vocal output as compared to the standard OOPS algorithm.

We have to see what are the workspace variables that we can extract from the model for the purpose of post processing and visualization. For that, extract some portion of the song, one sample is the first 20 seconds recording of the input music from one channel while output for the same duration.

To verify the algorithm, it is necessary that the data sets be chosen randomly and for the ease of processing we will take a set of 5000 data points each using 4 random seeding points. Here I take some English and Hindi songs as an input. In following figures, the red color indicates the input signal and blue indicates the output signal. The system takes simultaneous signals.

The OOPS+ LMS removes the pitch and format. From following figure, there are some blue lines even when there are no red lines these are caused by superposition of two signals. This can be seen between both the results of OOPS and results of LMS+OOPS. This figure is zoomed view of the pitch and formant area only it shows the frequency domain response of the system for up to 1000 Hz.

![Figure 2. OOPS result for wake me up song](image-url)
Figure 3. Combined output result for wake me up song

Figure 4. OOPS output result for give me some sunshine song
A novel approach to removing the vocal signals from the music signal is proposed. Unlike the used karaoke algorithms which completely removes the low frequency signal assuming it’s the vocal signal, we have proposed an algorithm which targets the vocal signal by detecting vocal frequency and eliminating it using the LMS filter. The LMS filter coefficients are modified due to the quasi periodic nature of the speech signal and this has been observed in the results. Modifying the filter coefficients we have developed an adaptive filter and have effectively removed the vocal signal from the music which have been shown in the results. The proposed model can be easily extended to further suppress the vocal signal and can be used to develop better karaoke machine which cannot eliminate desired music elements.

REFERENCES

2. Lawrence R. Rabiner, Member, Ronald W. Schafer, Member, 'Recursive and Non recursive Realizations of Digital Filters Designed by Frequency Sampling Techniques', IEEE Transactions on Audeo and Electroacoustics vol 19, no. 3, 1971.