DEVELOP … SEPARATION … E FOR IN-VEHICLE INFOTAINMENT SYSTEM (IVIS)

LEE HOW CHEH

This thesis submitted as partial fulfilment of the requirement for the award of the degree of Bachelor of Electrical Engineering (Electronics)

Faculty of Electric & Electronics Engineering

University Malaysia Pahang

19 JUNE 2012

ABSRACTS

In order for in-vehicle infotainment systems (IVISs) to be used using sound detection, it is important for the system to detect a strong and accurate signal. A direct sound source recorded is blended with a mixture of unwanted signal called noise from its surrounding where it does not represent the original sound sample. A clean voice must be obtained in order for the system to recognize the command from user without affecting its quality and to be interpreted precisely. Generally, sampling sound is obtained and matched to the pre-determined sample sound in the system. Unwanted range of signal is eliminated while retaining the expected sound source. In this project, the source taken into consideration is human's voice and white noise. White noise is chosen as it exhibits a random signal with a flat power spectral density which is use to represent the surrounding sound in car as well as sound produced from the engine. The technique used in separating the signal sources is based on Independent Component Analysis (ICA). In this project, method and strategy to separate driver's voice from another sound source has been discussed and the result shown that ICA method can be used to separate sound sources successfully.

ABSTRAK

Bagi membolehkan sistem hiburan dalam kenderaan (IVIS) diguna pakai berpandukan sistem suara, adalah penting untuk sistem berkenaan mengesan isyarat dengan kuat dan tepat. Suara yang dirakam akan mengandungi campuran bunyi yang tidak diingini dan dikenali sebagai "noise" daripada persekitarannya di mana suara tersebut tidak mengambarkan sampel suara yang tulen. Isyarat suara yang bersih mestilah diperolehi bagi membolehkan sistem mengenalpasti arahan daripada pengguna tanpa menjejaskan kualiti dan memastikan arahan tersebut ditafsir dengan tepat. Secara amnya, sampel suara diambil dan disuaikan dengan sampel suara yang sedia ada dalam sistem. Isyarat yang tidak diingini akan disisihkan disamping mengekalkan sumber suara yang diingini. Dalam projek ini, sumber bunyi yang dipertimbangkan adalah suara manusia dan "white noise". "White noise" dipilih kerana bercirikan signal rawak dengan ketumpatan spektrum kuasa rata yang digunakan untuk mewakili bunyi dari persekitaran dalam kereta dan bunyi enjin kereta. Teknik yang akan digunakan untuk mengasingkan sumber suara adalah berdasarkan "Independent Component Analysis (ICA)". Dalam projek ini, kaedah dan strategi untuk mengasingkan suara manusia daripada sumber suara yang lain telah dibincangkan dan keputusan menunjukkan teknik ICA boleh digunakan untuk mengasingkan suara dengan jayanya.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.0 Introduction

In-vehicle Infotainment Systems (IVIS) is a system that has devices installed into vehicle to gives audio or video form of entertainment to the user and deliver navigation which includes GPS system that are widely used recently. Due to safety concern in handling vehicles, it is recommended to build an IVIS system which will not interrupt the driver's attention and focus while driving. In short a hand-free interface which works on sound recognition will be a better way of operating the system while driving. The system will receive instruction verbally and activated based on the instruction. The use of sound signal as interface involves the theory of sound source separation before it can be fully implemented into the IVIS. In this project, Sound Source Separation (SSS) system will be studied in detail for further use in the development of the system.

Human has the ability to differentiate and interpret different sources of sound at the same time but for computer or devices to mimic this ability of human is a tough job. A continuous study has been made from the past on sound source separation where mixture of sounds is separated into different set of signals in which one of it will carry the original sound source. These separated set of signals will ease the work of transcription and recognition by computer or devices. For the IVIS system to understand the instruction given, a clear, accurate and acceptable sound signal is required. The signal received must match to the original signal in the database of the system and able to separate the mixture of speaker's sound from surrounding noise. Unfortunately, direct sound signal received will be a mixture of many sounds and noises. Noises referred here include the white noise from air ventilation, noise from the engine of the car, radio and et cetera. A study on how to separate the mixture of sound signal into individual component is to be made in this project.

1.1 Project Objectives

The objective of this project is:

i. To identify characteristic which determine the differences between noise and original sound required.
ii. To propose a method for sound source separation.
iii. To separate mixture of sound from noise or interrupting signal in order to obtain clean sound signal.

1.2 Project Scope

There are several scopes that need to be proposed for this project and these include:

i. The mixture sound used in this project consists of unpolluted sound signal by a male speaker and white noise only.

ii. MATLAB will be used as software medium to interface sample signal and to perform its operation in separating mixture of sample sound signals. This is due to the fact that MATLAB application is a wide and common tool used for research and study purposes.

iii. The parameter used to eliminate and separate out noise from sound sample will be based on ICA technique.

1.2 Thesis Outline

This thesis consists of five chapters. Each chapter will be elaborated in details.

Chapter 1 gives an overview of the project which includes the objectives, scope of project and the thesis outline.

Chapter 2 includes the literature review where different methods of sound separation and software will be discussed in detail.

Chapter 3 discusses the methods of ICA separation where each process involve in ICA will be discussed.

Chapter 4 describes the result and discussion of the method used. The signal to noise ratio also will be discussed.

Chapter 5 concludes the outcome of the project. For further development and studies purposes, recommendation on this project is also included in this part.

CHAPTER 2

LITERATURE REVIEW

2.0 Sound Source Separation Methods

Sound source separation (SSS) is used in separating mixture of sound sources into individual signal in which retaining its original required sound signal. The main use of this SSS algorithm has plays a vital role in providing a better outcome in audio, image processing and biomedical applications and has become one of the most active research areas in signal processing. SSS always comes with the problem of limited amount of information shown in the sources and also in determining the nature of the mixtures. Besides, the number of sources and mixture channels determine its complexity. There are three types of case involving it. If the number of sources is less than the number of observation channel, it is called over-determined case. If the number of sources is equal the number of observation channel, it is called determined case and lastly if the sources more than observation channel, it is categorized as underdetermined case.

In order to understand more on SSS, first it is wise to understand the characteristics of audio sources which include speech and white noise. Speech is the most important sound produced by humans using the vocal folds [1]. It is based upon the syntactic combination of lexicals and names that are drawn from very large vocabularies. Speech signal is a continuous and varying signal and will have different amplitude and frequency. Sinusoidal partials are multiples of a single frequency called the fundamental frequency, or pitch. The pitch varies over time, but stays within a range of about40 Hz centered an average of 140 Hz for male and 200 Hz for female speakers [2].

On the other hand, noise is an unwanted signal that exists in a signal or data. It can be categorize into acoustic noise and non-acoustic noise. Example of acoustic noise is noise present in audio signal while example of non-acoustic noise is electronic noise, thermal noise and visual noise. Noise in radio transmission will produces hissing sound while noise in data transmission will result in inaccurate data transmission or reduces the strength of the signal. High level of noise will resulted in interference, undelivered message or distortion of the signal sent. There are several colors of noise which include white noise, pink noise, brown noise and grey noise. In this project, the noise that will be study is white noise. White noise is a most common and easily obtained signal which shows constant amplitude and frequency in other word it has a flat power spectral density.

Blind Source Separation (BSS) is known by its name as not much information about the sources or mixing process is known. Its' approaches are very popular in the statistical signal processing and machine learning areas. Generally, several observations of the mixture are available in these kinds of methods, but their performance is closely related to the mixing environment [2].

BSS has always been closely related to Independent Component Analysis (ICA). Independent Component Analysis (ICA) was introduced in the early 1990's [6]. ICA assumes that the source signals are statistically independent and non-Gaussian [3]. These assumptions are usually sufficient to carry out separation in the linear complete case, when there are as many mixture channels as sources available. When the problem is underdetermined, i.e. there are more sources than mixture channels, the difficulty is even higher and stronger assumptions are taken, generally related to the sparsity of the sources under some signal transformation [4]. ICA is a method of separating out independent sources from linearly mixed data, and belongs to the class of general linear models. ICA provides a better decomposition than other well-known models such as Principal Component Analysis (PCA). ICA seeks a linear transformation (an un-mixing matrix) to coordinates in which the data are maximally statistically independent. Statistical independence is measured by the mutual information between the unmixed variables and ICA algorithms seek unmixed matrices which minimize the mutual information.

According to Derry FitzGerald, the mathematical model underlying the ICA can be stated by assuming that there are *N* independent sources, s_i, which transmit signals[5]. The signals are measured by *M* sensors. The signals measured by the sensors, x_i, can be mapped to the sources using an unknown function f_i, resulting in:

$$x_i = f_i(s_1, \ldots\ldots, s_N)$$

In this case, the contributions of each of the *N* sources are assumed to be added together linearly to create each x_i. The equation can be written using the matrix notation as below

$$\begin{aligned} & x = As \\ \text{with}\quad & x^T = [\, x_1 \ldots x_M] \\ & s^T = [s_1 \ldots . s_N] \end{aligned}$$

and A is an M x N invertible matrix. Since the number of sensor has to be equal the number of sources, A will be of N x X size. ICA then attempts to estimate the matrix A or equivalently to find and un-mixing matrix W such that

$$y = Wx = WAs$$

to gives an estimate of the original source signals where $y=[y_1....y_N]$ and W is of size $N \times N$.

Due to amplitude and ordering limitation of ICA, dimensional reduction cannot be implemented on this technique. ICA can only be used on sound sources which has the characteristic of non-Gaussian properties. As for comparison, PCA can be used for dimensional reduction but at the expense of not characterizing non-Gaussian sources properly [5]. ICA cannot separate sources with more than one Gaussian source. It will be decorrelated but the original source directions will remain unknown [8]. In addition, the estimated sources studied were of an unknown scaling and perturbation. The variance of the original sources cannot be determined and the order they originally appeared cannot be determined too.

To perform ICA estimation, FASTICA algorithm also can be used. It is a highly efficient computational method which uses a fixed-point iteration scheme and gives a faster processing [9]. The convergence is cubic which is in contract to ordinary ICA algorithm that based on stochastic gradient methods where its convergence is linear only. However, in this FASTICA algorithm, research still carrying on because there is an explicit normalization step which can affect the convergence speed due to extra rotational.

Apart from Independent Component Analysis, Principle Component Analysis (PCA) is also used in SSS. PCA is also known as Karhunen-Loeve Transform and was formulated by Karl Pearson in 1901. T his analysis aimed to transform set of data with related variables into set of independent values variable known as principle components by using the concept of orthogonal transformation. This analysis is commonly used in exploratory data analysis and predictive models. PCA works by considering that the number of principle components must be less or equal to the number of original variable in which there are a few methods to carry out PCA and these include eigenvalue decomposition of data covariance matrix and singular value decomposition of a data matrix. The expected outcome from this PCA is measured by component scores and loadings.

In the function of PCA, the first principle component will contains the largest amount of total variance as possible and each successive principle component contains as much of the remaining variance as possible. Due to this fact, PCA can be used as one of the method of dimensional reduction, through the discarding of components that contribute minimal variance to overall data [5]

Singular Value Decomposition is the most used method. It decomposes Y, and n x m matrix into

$$\mathbf{Y} = USV^{T}$$

Where U is an n x n orthogonal matrix, V is a n x m orthogonal matrix and S is an n x m diagonal matrix of singular values. By getting rid of singular vectors that cause minimal variance to the data, dimensional reduction can be obtained. In this case, if the number of principal components is known in the earlier stage, the computational load of carrying out PCA can also be reduced by calculating only the required number of eigenvalues and eigenvectors [5].

The third method used in SSS is Computational Auditory Scene Analysis (CASA). CASA is a technique where it adopted the ability of human to hear voices from surrounding which able to detect and differentiate mixture of sound present at the same time. This is more appropriate to referred it as the cocktail party problems in which human are able to identify different sound source in a crowded place. This ability is difficult to be gained by the computer system through signal processing. In contrast to pure mathematical techniques, CASA methods are aimed at designing machines capable of hearing the way humans do. This is done by providing a system which able to convert live and real sound samples into representation that represent what is listened by human. Since CASA is based on human's audio hearing model, this system works with not more than two audio detectors to detect surrounding acoustic sound.

However, CASA has a few limitation as it only re-synthesis via filter mask and only on periodic targets. The features used by these models are very specific in which limit the percentage of successful separation under very specific situations. When an intrusion is removed from the auditory scene, it leaves a "gap" in the spectrum which can often be heard in the resynthesized waveform [7]. According to Cooke and Brown, the search strategy unable to group components which are widely separated in time such as a sequence of speech sounds from a single speaker. Therefore their applicability is not as wide as the one of BSS approaches.

2.1 Software Review

In this project, MATLAB is used to separate the mixture of sound source sample. Generally, MATLAB is a high-level language and interactive environment that enables performance of computationally intensive tasks faster than with traditional programming languages such as C, C++, and FORTRAN (see MATLAB, 2009). Essential feature of an education automatic control system program in control engineering can be developed. The MATLAB system has five main parts which are MATLAB language, MATLAB working environment, handle graphics, MATLAB mathematical function library and MATLAB Application Program Interface (API).

MATLAB allows solving of many technical computing problems and specifically problems deal with matrix and vector formulations. It also has toolbox which allows user to learn and apply specialized technology. It is a collection of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems (see MATLAB 2009). Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.

Another software that can be use is ICALAB for Signal Processing. This is a package for MATLAB. It consists of many efficient algorithms for independent component analysis for example higher order statistics, blind source separation, second order statistics, linear prediction, and blind signal extraction. Although this package is easy to use, it is not chosen in this project as all algorithms is already provided internally where the parameters of ICA cannot be tuned accordingly.

CHAPTER 3

METHODOLOGY

3.0 Introduction

In this chapter, the method that is used to separate the combined signals will be discussed in detail. Independent Component Analysis (ICA) method which is one of the available methods as per discussed in Chapter 2 is chosen for the implementation on this project as it is based on its simplicity and effectiveness in carrying out the separating process.

This project starts with the collection of sound samples. The first signal is the clean human sound without any interference. The second sound is the white noise signal sound. Human sounds are obtained from open, sources in the internet while white noise is recorded by using sound recorder.

The second step of this project is to mix the two sound sources to obtained two different combined sounds. This is followed by development of Sound Source Separation algorithm based on Independent Component Analysis (ICA) using MATLAB application.

When the mixture of sound is successfully separated, the result will consists of two sound signals which are original sound from human and white noise signal. If output sound is not properly separated, the process of running MATLAB application using ICA technique is repeated all over again by modifying and varying the parameter set in ICA algorithm. The flow chart in Figure 3.1 shows how the project is done from step to step flow.

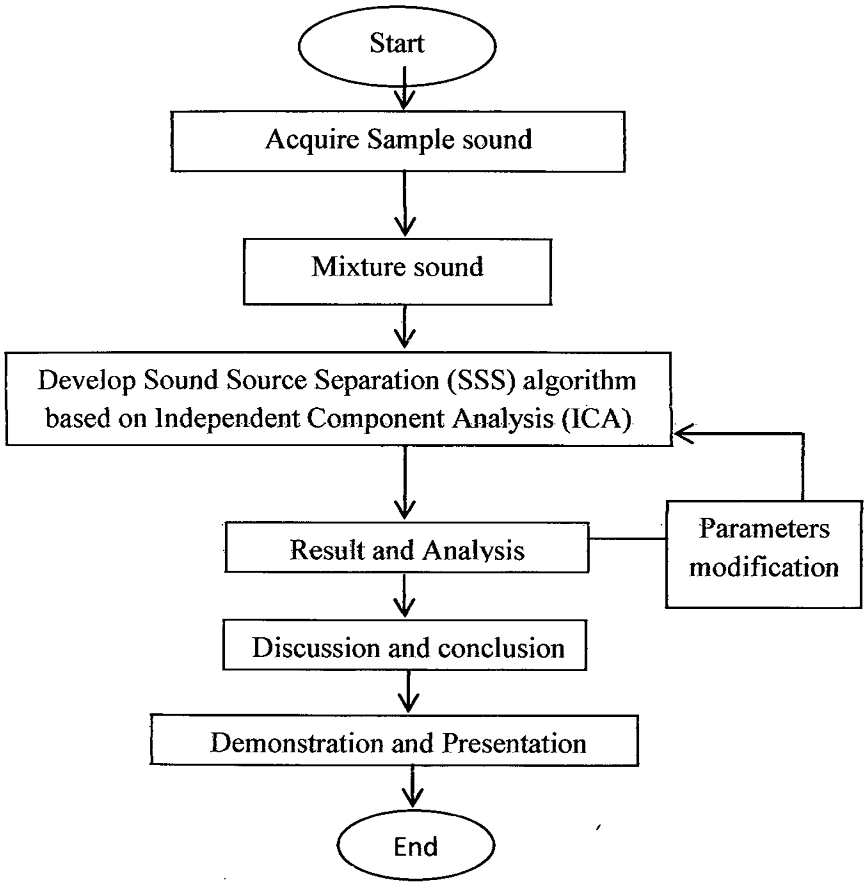

Figure 3.1: Project flowchart

3.1 Sample Sound Acquisition

There are many sample sounds which can be used in this research purposes. These varies from sound of human talking in the car, sound of the engine, radio, wind blow, breathing, air conditioner, mixture of a few sounds and many more. In order to simplify and limit the scope of study and ease of analysis, the sound signal in this project will be concentrated on white noise which is a random signal with a flat power spectral density as well as human voice.

White noise is chosen as it gives a very close representation of surrounding sound especially in a fully closed car and also the engine sound which has a flat power spectral density too. This white noise is recorded with SONY IC Recorder ICD-SX56. The human sample sound proposed are taken from the archive through the web services. This is because to obtained samples sounds which are of pure mixture of the twos investigated sound are impossible by using microphones. It must be perform in studio. As comparison, a variety of separated individual sample sound can be obtained easily from the archive in internet. All sample sounds used are of mono channel with frequency of 11.025kHz.

3.2 Mixture sound

The independent sample sounds are joined together using AUDACITY software. AUDACITY is free, open source, cross-platform software for recording and editing sounds. This software enables editing of an original sound such as amplification, noise removal, pitch change, tempo and speed editing, combines wave signals, frequency change and etcetera. The first mixed signals are of white