COMPARATIVE STUDY OF HYBRID SPECTRAL SUBTRACTION SPEECH ENHANCEMENT ALGORITHMS

FATIN NABIHAH BINTI PIDRUS

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Information Security)

Faculty of Computing
Universiti Teknologi Malaysia

JUNE 2018

"My dearest mum, my everlasting father's soul, family, and friends"
This is for you

ACKNOWLEDGEMENT

Initially, I would like to express my special thanks of gratitude to my dearest family for their endless support and motivation to push me forward. This project would not have been possible without them.

I also warmly appreciate for my supervisor Dr Mazura Mat Din who helped me in doing a lot of research and I came to know about so many new things during my master project. Special thanks to my closest friends and classmates for their endless motivation, unstoppable urge to push me forward, generously share their experiences without any hesitation and support me during my weakest time.

Additionally, I would like to thank the authority of Universiti Teknologi Malaysia (UTM) for giving me the chance to pursue my dream. Finally, I made it.

ABSTRACT

Speech is one of the fundamental means of communication. Speech is always accompanied by the background noise thus speech enhancement has been a long standing problem in signal processing. Speech enhancement algorithms are important components in many systems where speech plays a vital part, including telephony, hearing aids, voice over IP, and automatic speech recognizers. Speech enhancement is generally concerned with the problem of enhancing the quality of speech signals. The aim in speech enhancement is to improve the intelligibility and quality of an audio recording. Speech enhancement often involves with poor quality and intelligibility recordings as these audio were under circumstances that did not optimized signal-to-noise ratio. Usually the quality of the audio recordings obtained is damaged by many factors such as poor equalization, additive noise, distortion, or excessive vibration. Many studies involving noise reduction of recorded speech have been founded over decades. Researchers have suggested using hybrid of several algorithms for speech enhancement purpose as it works best, however the comparative studies among these hybrid algorithms are not much found in the field. Speech enhancement can help solves many communication problems and the major achievement is when there are many situations where speech enhancement is used in audio forensics to help reveal the events and able to clarify the dialogue in forensic investigations. In this research three existing hybrid speech enhancement algorithms are studied and compared to see which one gives the best result in speech quality performance. The only way to tell is to evaluate the enhanced speech by using Perceptual Evaluation of Speech Quality (PESQ).

ABSTRAK

Ucapan adalah salah satu cara asas dalam berkomunikasi. Ucapan sentiasa diiringi dengan bunyi latar belakang dan telah menyebabkan peningkatan pertuturan menjadi masalah utama dalam pemprosesan isyarat. Algoritma penambahbaikan ucapan adalah komponen penting dalam banyak sistem dimana ucapan memainkan peranan penting, termasuk telefoni, alat bantuan pendengaran, suara melalui IP, dan penceramah ucapan automatik. Peningkatan ucapan biasanya berkaitan dengan masalah meningkatkan kualiti isyarat pertuturan. Tujuan penambahbaikan pertuturan adalah untuk meningkatkan kecerdasan dan kualiti rakaman audio. Peningkatan ucapan sering melibatkan rakaman kualiti dan kecekapan yang tidak baik kerana audio ini berada di bawah keadaan yang tidak dioptimumkan isyarat kepada nisbah hingar. Biasanya kualiti rakaman audio yang diperoleh rosak oleh banyak faktor seperti penyamaan miskin, bunyi tambahan, gangguan, atau getaran berlebihan. Banyak kajian yang melibatkan pengurangan bunyi ucapan yang direkodkan telah diasaskan selama beberapa dekad. Penyelidik telah mencadangkan menggunakan hibrid beberapa algoritma untuk tujuan peningkatan pertuturan kerana ia berfungsi dengan lebih baik, namun kajian perbandingan antara algoritma hibrid ini tidak banyak ditemui di lapangan. Peningkatan ucapan boleh membantu menyelesaikan banyak masalah komunikasi dan pencapaian utama adalah apabila terdapat banyak situasi di mana peningkatan pertuturan digunakan dalam forensik audio untuk membantu mendedahkan peristiwa dan dapat menjelaskan dialog dalam penyiasatan forensik. Dalam kajian ini, tiga algoritma peningkatan hibrid yang sedia ada dikaji dan dibandingkan dengan melihat mana yang memberikan hasil terbaik dalam prestasi kualiti pertuturan. Satu-satunya cara untuk memberitahu ialah untuk menilai ucapan yang dipertingkatkan dengan menggunakan penilaian persepsi ucapan.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREAVIATIONS

CD	Compact Disc
dB	Decibel
DVD	Digital Versatile Disc
Hz	Hertz
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IJAREEIE	International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering
IRS	Intermediate Reference System
ITU-T	International Telecommunication Union
MBSS	Multiband Spectral Subtraction
MMSE	Minimum Mean Square Error
NOISEUZ	Noisy Speech Corpus
PC	Personal Computer
PESQ	Perceptual Evaluation of Speech Quality File Format
SP	Speech pauses
SPL	Sound Pressure level
SSF	Spectral Subtraction Filter
STFT	Short-time Fourier transform
SVD	Singular Value Decomposition
TDT	Tucker Davis Technologies
VoIP	Voice over internet protocol
WAV	Waveform Audio

CHAPTER 1

INTRODUCTION

1.1   Introduction

Speech is one of the fundamental means of communication. The goal in speech enhancement is to enhance the quality of speech signals. Speech enhancement can contribute for the improvement of technology especially in telephony, hearing aids, voice over IP, and automatic speech recognizers. The crucial method for enhancing speech is the removal of background noise. A noisy environment can reduce the ability to communicate between speaker and listeners. The impact of this problem can be reduced with performance of speech enhancement to obtain a much clearer speech. However, clear speech is difficult to achieve in the real world environment.

Over the past decades, researchers have been proposed a numerous numbers of algorithms to improve the performance of speech enhancement. Algorithms are usually generated based on the applications (Banchor et al, 2013). This is because the design of algorithm defers application to application and the performance of the algorithms can also be different for each application. It is very difficult for the researchers to find algorithms that really work in various environments in order to enhance the speech in an audio (Maher, 2010). There is no single algorithm that can work to solve the entire problems.

Audio recordings which suffer from unwanted noise, interfering sounds, and other signal processing are the obstacles for audio forensics investigators as these can prevent smooth analysis. Examiners encounter with many issues and the most common enhancement issue is the secret surveillance recording of conversation via a microphone (Maher, 2009). The hidden nature of the recording system often leads to poor microphone placement for examples, interference from wind and other environmental sounds. It is better to improve the signal to noise ratio before playback when a recorded audio signal contains unwanted noise. All of the enhancement processes are prerequisite to be performed using a copy of the original recording to avoid from the evidence being altered.

Another problem in speech enhancement is complicated subjective evaluation of speech enhancement algorithm (Loizou & Kim, 2011). The quality judgements from the audience are not convinced whether they are based on the signal distortion, noise distortion or both. It is difficult for researchers to compare their algorithm objective performance with other related works as there is no access to common speech database (Hu & Loizou, 2006). As the background noise level is rapidly changing, thus it is remains unknown as to which algorithm performs better in real situation (Loizou, 2006). This makes it hard to have fair and reliable comparisons between algorithms.

As the technology evolution keep on changing rapidly over the time, the speech enhancement researchers also need to keep their pace with the technology. The importance of speech enhancement algorithms has made the researchers in the field keenly developing new enhancement algorithms in this area to help solves many communication issues and devices related with sounds. There are many types of noises present in the surrounding and a standard algorithm alone is insufficient to recover all of these noises and the search for new and better methods continues until today (Nabi et al, 2016). In 2017 researchers, Soni and Vaghela have claimed that it is better to combine two or more types of algorithm in order to get a better result of speech enhancement as stated in their paper for International Research Journal of Engineering and Technology (IRJET).

1.2   Problem Background

Spectral subtraction is a well-known frequency domain method to reduce the effect of additive noise in a signal (Kamath & Loizou, 2017). Among all of the traditional spectral subtraction algorithms Boll, Berouti and Multiband Spectral Subtractions are usually being used for enhancement as they are easier to conduct. These three are commonly being compared to find which algorithm works better than another. However, one single algorithm cannot solve the entire problem encountered in the speech enhancement. As a result, researchers concluded that a single algorithm cannot perform in all situations.

Over the past decades, researchers have been proposed a numerous numbers of algorithms in speech enhancement to improve the performance of audio enhancement. Many researchers have proposed that hybrid of several algorithms can be used for speech enhancement purposes (Dash & Solanki, 2017). Most of researchers who conducted hybrid experiments have claimed that hybrid approach results are very promising. However there are not many comparative studies being made to compare these works especially regarding spectral subtraction method. Most of the works are comparing one single algorithm or focusing more on new proposed works. Thus, it is still remains unclear as to which spectral subtraction algorithms perform the best.

Evaluation performance is very important to figure which algorithms works best. The most accurate method for evaluating speech quality is through user's subjective listening tests however it is very costly and time consuming (Hu & Loizou, 2008). Therefore a reliable and most accurate evaluation performance is needed to conduct an unbiased and reliable experiment.

1.3   Problem Statement

Speech enhancement recordings are often involves in non-ideal surroundings with non-optimal microphone placement and these recordings are often corrupted by environmental noises such as street or machinery noise. Over the decades various algorithms to enhance noisy speech in an audio have been proposed, however the speech quality and intelligibility comes into question as many of these algorithms do not achieve great improvements when the audio is corrupted with real environmental noise (Doire et al, 2017). Therefore, many researchers have combined algorithms and proposed new methods for improvement in audio performance. Spectral subtraction algorithms are chosen for this research as they are the most easiest to conduct especially for single channel algorithm. However, the comparison between these chosen hybrid algorithms are not yet been done by other researchers.

1.4   Research Question

Below show some questions which related to this research. The questions to support the goal of this research are as follows:

   i.   How to reduce the background noise in an audio by using spectral subtraction algorithms?
   ii.   How a hybridization of the spectral subtraction algorithm give a better result than a single algorithm?
   iii. How to evaluate the performance of speech quality of the enhanced audio?

1.5   Aim of the Study

This study is aim to compare three types of hybrid speech enhancement algorithms in terms of effectiveness and performance of speech quality by reducing the level of background noise so that the best among three audio enhancement technique can be determined. All three speech enhancement algorithms evaluated is based on spectral subtraction method. The speech quality of the enhanced speech produced by the three hybrid enhancement methods will be evaluated and benchmarked based on Perceptual Evaluation of Speech Quality (PESQ).

1.6   Objectives of the Study

The objectives of the project are as the following:

i.   To reduce the background noise in an audio by using Boll, Berouti and Multiband Spectral Subtraction algorithms.

ii.   To reduce the background noise in an audio by combining two of the selected spectral subtraction method algorithms.

iii.   To evaluate the performance of speech quality of the enhanced audio by using Perceptual Evaluation of Speech Quality (PESQ).

1.7   Project Scope

The research scope and project scope is a very important part which limited the area of field. This acts as guidelines for the project. The scopes of this study are listed as the following:

i. The domain of this research focuses on three single channel spectral subtraction speech enhancement techniques.

ii. The audio recorded used in this study is a noisy speech corpus (NOIZEUS) with 30 Institute of Electrical and Electronics Engineers (IEEE) sentences.

iii. The dataset are in Waveform Audio File Format (WAV).

iv. Four types of noise at 3 different signal-to-noise ratios (SNRs) were taken from AURORA database which are street, train, babble and car are used for this research.

1.8   Significance of the Project

The purpose of this research is to study and compare three hybridization types of spectral subtraction method speech enhancement algorithms and to find which works much better in performance. Speech enhancement field can be used for improvements in many applications such as mobile communication systems, Voice over Internet Protocol (VoIP), teleconferencing systems, speech recognition, hearing aids and many more. Speech enhancement algorithms are important in human-to-human communication systems as well as human-to-machine communication systems. The highest recognizable contribution that speech enhancement has made is towards the law enforcement agencies. Speech enhancement is used in audio forensics enhancement in removal real environmental noise from noisy audio recording signals. Speech enhancement method has been a crucial step in audio forensics to produce a more accurate and meaningful transcript of a recorded conversation to be used in court as evidence.

1.9   Project Organization

This research is organized as follows. Chapter 2 described in details about background information of speech enhancement including its definition, concerned issues found regarding speech enhancement and in-depth discussion of three techniques being compare for this research. Further, the overview of literature on various existing speech enhancement algorithms and various research groups working with Boll, Berouti and Multiband Spectral Subtraction algorithms are also discussed.

Chapter 3 explained the research methodology framework for this research. Research framework is discussed and each of three phases for this research is outlined in details for further understanding.

Chapter 4 explained the design and implementation of this research. The experimental studies based on objective and subjective quality measure are performed based on the phases in research methodology. Performance evaluation with existing method is also described and conducted using specific measurement technique.

In Chapter 5, the PESQ scores of the enhanced speech performed by the three speech enhancements being studied is analysed. The analysed experimental results from three algorithms are compared to find which one performs better.

Lastly, Chapter 6 summarized the work presented in this research, highlights the main contributions of the work, draws the conclusion and provides suggestions for future work.

REFERENCES

Aparna U.R. and Shaiju Paul, (2016), Audio Forensic examination: authenticity, enhancement and interpretation. (IC-GET), 2016 Online International Conference.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Modeling and signal processing of acoustic gunshot recordings..

Barak, A. (2005). Subjective comparison of Speech enhancement algorithms, 23(1) 77-92

Breitinger, C., Gipp, B., Langer S., (2015). Research-paper recommender systems: a literature survey. International Journal on Digital Libraries. 17 (4): 305–338.

Cawley G. and Talbot N., (2010), Over-fitting in model selection and subsequent selection bias in performance evaluation, Journal of Machine Learning Research. Research, vol. 11, pp. 2079-2107

Celli F., Pianesi F., Stillwell D., Kosinski M., (2013), Workshop on computational speech processing (shared task), In Proceedings of WCPR13, in conjunction with ICWSM-13, 2013.

Danah M., Nicole B., (2007), Speech Enhancement: Definition, History, and Scholarship, Journal of Computer-Mediated Communication, Vol. 13.

E. Leopold, J. Kindermann, (2002), Speech enhancement using a minimum mean square error log spectral amplitude estimator. Mach. Learn. 46 (1–3) 423–444

F. Sebastiani, (2002), Machine learning in automated text categorization, ACM Comput. Surv. 34 (1) (2002) 1–47.