SPEECH ENHANCEMENT SYSTEM

PROJECT SYNOPSIS

OF MINOR PROJECT

BACHELOR OF TECHNOLOGY

Computer Science & Engineering

SUBMITTED BY

Veerpal

Ishpreet Singh

Paras Kalyan

GURU NANAK DEV ENGINEERING COLLEGE,

LUDHIANA

Introduction

Speech is one of the most natural forms of communication for human beings and poses an effective tool to exchange ideas or to express needs and emotions. Due to technical advances, speech communication is no longer restricted to face-to-face conversations but is also performed over long distances, e.g., in the form of telecommunications, or is even used as a natural way for human-machine interaction.

In many speech processing applications, one or more microphones are used to capture the voice of the targeted speaker. As the microphones are often placed at a considerable distance from the target speaker, e.g., in hearing aids or hands-free telephony, the received signal does not only contain the sound of the target speaker, but possibly also sounds of other speakers or background noises.

To improve the quality and, if possible, also the intelligibility of a noisy speech signal, speech enhancement algorithms are employed.

Speech enhancement is concerned with improving some perceptual aspect of speech that has been degraded by additive noise. In most applications, the aim of speech enhancement is to improve the quality and intelligibility of degraded speech. The improvement in quality is highly desirable as it can reduce listener fatigue particularly in situations where the listener is exposed to high levels of noise for long periods of time (e.g., manufacturing). Speech enhancement algorithms reduce or suppress the background noise to some degree and are sometimes referred to as noise suppression algorithms. There are a wide variety of scenarios in which it is desired to enhance speech. Voice communication, for instance, over cellular telephone systems typically suffers from background noise present in the car, restaurant, etc. at the transmitting end. Speech enhancement algorithms can therefore be used to improve the quality of speech at the receiving end. That is, they can be used as a preprocessor in speech coding systems employed in cellular phone standards. If the cellular phone is equipped with a speech recognition system for voice dialing, then recognition accuracy will likely suffer in the presence of noise. In this case, the noisy speech signal can be preprocessed by a speech enhancement algorithm before being fed to the speech recognizer.

It is possible to reduce the background noise, but at the expense of introducing speech distortion, which in turn may impair speech intelligibility. Hence, the main challenge in designing effective speech enhancement algorithms is to suppress noise without introducing any perceptible distortion in the signal. Thus far, most speech enhancement algorithms have been found to improve only the quality of speech.

Need of Speech Enhancement system

Speech Enhancement system is intended to supress or remove background noises from audios and through this, people will be capable to have a telephonic conversation, consequently they will be farther from the telephone network. In those cases, speech enhancement will be required as well, as the concept is rather similar as of speaker and microphone standing in a long distance, and consequently it is probable to have the problems of background noise and echo.

Echo elimination is also needed for speech samples collected from big halls or house. The speech can pick some echo up when the distance between the speaker and microphone is large.

Objectives

1. To understand speech processing and enhancement fundamentals
2. To remove background noise from speech
3. To successfully implement speech enhancement system.

Hardware requirement

CPU:- i3 5th gen or above

RAM:- 2GB or more

Disk Space:- 50GB or more

Literature Review

Techniques for Speech Enhancement

Spectral-subtractive algorithm

The spectral-subtractive algorithm is historically one of the first algorithms proposed for noise reduction. More papers have been written describing variations of this algorithm than any other algorithm. It is based on a simple principle. Assuming additive noise, one can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated and updated, during periods when the signal is absent. The assumption made is that noise is stationary or a slowly varying process and that the noise spectrum does not change significantly between the updating periods. The enhanced signal is obtained by computing the inverse discrete Fourier transform of the estimated signal spectrum using the phase of the noisy signal. The algorithm is computationally simple as it only involves a forward and an inverse Fourier transform.

WIENER FILTER

The input signal goes through a linear and time-invariant system to produce an output signal y(n). We are to design the system in such a way that the output signal, $\hat{d}(n)$, is as close (in some sense) to the desired signal, d(n), as possible. This can be done by computing the estimation error, e(n), and making it as small as possible. The optimal filter that minimizes the estimation error is called the Wiener filter, named after the mathematician Norbert Wiener [1], who first formulated and solved this filtering problem in the continuous domain. It should be noted that one of the constraints placed on the filter is that it is lin- ear, thus making the analysis easy to handle. In principle, the filter could be finite impulse response (FIR) or infinite impulse response (IIR), but often FIR filters are used because (1) they are inherently stable and (2) the resulting solution is linear and computationally easy to evaluate.

Subspace Algorithms

These algorithms are based on the principle that the clean signal might be con- fined to a subspace of the noisy Euclidean space. Consequently, given a method for decomposing the vector space of the noisy signal into a subspace that is occupied primarily by the clean signal and a subspace occupied primarily by the noise signal, we could estimate the clean signal simply by nulling the component of the noisy vector residing in the “noise subspace.” The decomposition of the vector space of the noisy signal into “signal” and “noise” subspaces can be done using well-known orthogonal matrix factorization techniques from linear algebra and, in particular, the singular value decomposition (SVD) or eigenvector–eigenvalue factorizations.

By computing the SVD of a matrix, for instance, we can obtain orthonormal bases for four fundamental subspaces: the column space of the matrix (in the context of speech enhancement, this corresponds to the signal subspace), the left nullspace, the row space, and the nullspace.