Introduction to the Wiener Filter for Noise Reduction

Jacob Benesty[[1]](#footnote-1), Jingdong Chen[[2]](#footnote-2), Yiteng (Arden) Huang, and Simon Doclo[[3]](#footnote-3)

Abstract

The problem of noise reduction has attracted a considerable amount of research attention over the past several decades. Numerous techniques were developed, and among them is the optimal Wiener filter, which is the most fundamental approach, and has been delineated in different forms and adopted in diversified applications. It is not a secret that the Wiener filter achieves noise reduction with some integrity loss of the speech signal. However, few efforts have been reported to show the inherent relationship between noise reduction and speech distortion. By defining a speech-distortion index and a noise-reduction factor, this chapter studies the quantitative performance behavior of the Wiener filter in the context of noise reduction. We show that for a single-channel Wiener filter, the amount of noise attenuation is in general proportionate to the amount of speech degradation. In other words, the more the noise is reduced, the more the speech is distorted. This may seem discouraging as we always expect an algorithm to have maximal noise attenuation without much speech distortion. Fortunately, we show that the speech distortion can be better managed by properly manipulating the Wiener filter, or by considering some knowledge of the speech signal. The former leads to a sub-optimal Wiener filter where a parameter is introduced to control the tradeoff between speech distortion and noise reduction, and the latter leads to the well-known parametric model-based noise reduction technique. We also show that speech distortion can even be avoided if we have multiple realizations of the speech signal.

2.1 Introduction

The existence of noise is inevitable in real-world applications of speech processing.In a voice communication system, for example, a desired speech signal, when propagating through an acoustic channel and picked up by a microphone sensor, is corrupted by unwanted noise, which may result in appreciable or even significant degradation in the quality and intelligibility of the recorded speech. Therefore, it is essential for such systems that we can have some effective noise reduction/speech enhancement techniques to extract the desired speech signal from its corrupted observations.

The noise reduction technique has a broad range of applications, from hearing aids, cellular phones, voice-controlling systems, teleconferencing and multiparty teleconferencing, to automatic speech recognition (ASR) systems. The difference between two systems using and not using such techniques can be significant; therefore, the choice can have a great impact on the functioning of the system.

Research on noise reduction/speech enhancement can be traced back to 40 years ago with 2 patents by Schroeder, where an analog implementation of the spectral magnitude subtraction method was described. Since then, it has become an area of active research. Over the past several decades, researchers and engineers have approached this challenging problem by exploiting different facets of the properties of the speech and noise signals. A variety of approaches have been developed, including Wiener filter, spectral restoration, signal subspace method,parametric-model-based approach, statistical-model-based method, and spatio-temporal filtering.

Most of these algorithms were developed independently of each other and their performance on noise reduction were evaluated mostly by assessing the improvement of signal-to-noise ratio (SNR) or subjective speech quality when the methods were formulated. It has been noticed that these algorithms,almost with no exception, achieve noise reduction by some integrity loss of the speech signal. Some algorithms are even formulated explicitly based on the tradeoff between noise reduction and speech distortion, such as the subspace method. However, so far, few efforts have been devoted to analyzing such a tradeoff behavior even though it is a very important issue. In this chapter, we attempt to provide an analysis about the compromise between noise reduction and speech distortion. On the one hand, such a study may offer us some insight into the range of the existing algorithms that can be employed in practical noisy environments. On the other hand, a good understanding may help us to find new algorithms that can work more effectively than the existing ones.

Since there are so many algorithms in the literature, it is extremely difficult if not impossible to find a universal analytical tool that can be applied to any algorithm. In this study, we choose the Wiener filter as the basis since it is the most fundamental approach, and many algorithms are closely connected to this technique. For example, the minimum-mean-square-error (MMSE) estimator presented in [15], which belongs to the category of spectral restoration, converges to the Wiener filter at a high SNR. Also it is widely known that the Kalman filter is tightly related to the Wiener filter.

Starting from the optimal Wiener filtering theory, we introduce two new concepts: the speech-distortion index and the noise-reduction factor. We then show that for a single-channel Wiener filter, the amount of noise attenuation is in general proportionate to the amount of speech degradation. In other words, the more the noise is attenuated, the more the speech is distorted. This observation may seem quite discouraging as we always expect an algorithm to have maximal noise attenuation without much speech distortion. Fortunately, we show that the compromise between noise reduction and speech distortion can be better managed by properly manipulating the Wiener filter, or by considering some knowledge of the speech signal. The former leads to a suboptimal Wiener filter where, like in the spectral subtraction, a parameter is introduced to control the tradeoff between speech distortion and noise reduction, and the latter leads to the well-known parametric-model-based noise-reduction technique. We also discuss the possibility to avoid speech distortion by using an array of microphones.

2.2 Estimation of the Clean Speech Samples

We consider a zero-mean clean speech signalcontaminated by a zeromean noise process , so that the noisy speech signal, at the discrete time sample , is

Define the error signal between the clean speech sample at time n and its estimate:

where superscript T denotes transpose of a vector or a matrix,

is an FIR filter of length L, and

is a vector containing the L most recent samples of the observation signal

.

We now can write the mean-square error (MSE) criterion:

where denotes mathematical expectation. The optimal estimate of the clean speech sample tends to contain less noise than the observation sample , and the optimal filter that forms is the Wiener filter which is obtained as follows,

Consider the particular filter,

This means that the observed signal will pass this filter unaltered (no noise reduction), thus the corresponding MSE is,

In principle, for the optimal filter, we should have.

In other words, the Wiener filter will be able to reduce the level of noise in

the noisy speech signal .

We easily find the Wiener-Hopf equation:

where

is the correlation matrix of the observed signal and

is the cross-correlation vector between the noisy and clean speech signals. However, is unobservable; as a result, an estimation of may seem difficult to obtain. But,

=

=

Now p depends on the correlation vectors and. The vector (which is also the first column of ) can be easily estimated during speech and noise periods while rv can be estimated during noise-only intervals assuming that the statistics of the noise do not change much with time.

Using the last formula and the fact that ，we obtain the optimal filter:

=

=

where

is the signal-to-noise ratio, I is the identity matrix, and

We have,

The minimum MSE (MMSE) is:

We see clearly from the previous expression that ; therefore,noise reduction is possible.

The normalized MMSE is

And

基于维纳滤波器的消噪算法

Jacob Benesty[[4]](#footnote-4), Jingdong Chen[[5]](#footnote-5), Yiteng (Arden) Huang, and Simon Doclo[[6]](#footnote-6)

摘要

最近几十年，噪声消除的问题得到了研究者很大的关注。这期间产生了很多消噪方法，它们中最基础的方法是理想维纳滤波法，维纳滤波法以不同的形式应用在了各种场合和程序中。 维纳滤波算法以语音的一定程度的失真为代价得到了噪声的消除。然而，人们已经为了搞清噪声消除和语音失真的内在关系做了大量的努力。通过定义一个语音失真参数和噪声消除参数，本文研究了用于语音增强的维纳滤波算法的效果。对于单通道维纳滤波器，我们发现，噪声消除的程度与语音失真的程度在一定程度上是成正比的，也就是说，噪声消除的越多，语音失真越严重。这看起来可能有点让人失望，因为我们总是希望有一种算法可以在最大程度的消除噪音的同时并不伴随着语音的失真。 幸运的是，我们发现，只要使用合适的维纳滤波算法或者考虑一下信号的特性就可以很好的控制语音的失真程度。前者称为次佳维纳滤波器，它包含一个可以控制语音失真程度和噪声消除程度的参数。后者统称为基于参数的噪声消除算法。

2.1 引言

噪声广泛存在于真实世界的各种语音处理系统中。比如，在语音通信系统中，有用信号通过声音信道传播，并被话筒拾取到，但是在这个过程中信号不可避免的被背景噪声所污染。这造成信号在可懂度和质量上的很严重的失真。因此，对于这些系统，探索一种有效的消除噪声或者语音增强技术用来从带噪信号中分离出有用信号是非常必要的。噪声消除技术在实际生活中具有广泛的应用前景，从助听器，手机，声控系统，远程会议，多媒体远程会议，到自动语音识别系统，它们中都非常有必要加入消除噪声的装置。在这些系统中，使不使用消噪装置带来的结果截然不同，这个选择很大程度上会影响整个系统的表现。对于噪声消除和语音增强技术的研究可以追溯到40多年前，那时，Schroeder 发明了在模拟域使用谱减法来进行语音增强的两个专利。从那时起，这个语音增强就成了很热门的研究领域。在过去的几十年，研究者和工程师利用语音和噪声在本质特征上的区别攻克了这个领域的很多难题。很多种方法被提出来，这其中就包括了维纳滤波法，谱减法，基于信号子空间分解的方法，基于参数的方法，以及基于统计模型的方法。

这其中的大多数算法都是独立发展而来的，与其他方法无关。我们在大多数情况下使用信噪比的增益或者主管的听觉舒适度提升来评估这些算法的性能。我们发现，这些算法中绝大多数都毫无例外的以语音信号的失真作为代价来换取噪声的消除。一些算法甚至就是基于这种语音失真和噪声消除之间的矛盾的折中而构造的，比如，子空间分解法。然而，到目前为止，还没有人考虑过这种折中造成的影响，尽管这是一个非常重要的问题。本文尝试分析降噪和语音失真之间这种造成的影响。一方面，这个研究会使我们更加深入的理解现在已有的算法应用在实际中出现的问题。另一方面，当我们很好的理解了这个问题之后也许会产生新的想法，发明出新的比现在已有算法更加有效的算法。

由于现在已经有了大量的语音增强降噪算法，所以，如果不找到一个适用于所有算法的分析工具那么分析起来将是非常困难的。本文中，我们使用维纳滤波法作为基础，因为它是这个领域基本的方法，有不少算法都与维纳滤波法有着紧密的联系。比如，最小均方误差估计法，它属于短时谱分析法，在高信噪比的情况下会回归到维纳滤波法。另外，广为人知的卡尔曼滤波法也和维纳滤波法紧密联系。

从最佳维纳滤波器开始，我们在这里介绍两个概念：语音失真参数和噪声消除参数。之后我们将会表明，对于单信道维纳滤波法，噪声的衰减通常来说是和语音的失真成正比的。换句话说，噪声消除的越多，语音失真越严重。这个结论看起来有点令人失望，因为我们希望有算法在最大程度的抑制噪声的同时也能避免语音的失真。幸运的是，我们发现，噪声抑制与语音失真之间的折中程度是我们可以控制的，使用维纳滤波器或者考虑语音信号的特性就可以完成这一功能。前者称为次佳维纳滤波器，它包含一个可以控制语音失真程度和噪声消除程度的参数。后者就是著名的基于参数的噪声消除算法。

2.2纯净语音估计

考虑一个零均值纯净语音信号被一个零均值噪音s所污染，所以在时刻，带噪语音信号为：

定义在 时刻纯净语音和纯净语言之估计之间的误差为：

其中，上标T表示矩阵或者向量的转置

是一个长度为L的FIR滤波器。

是一个包含了当前时间点之前的L个采样的向量。

现在，我们可以写出均方误差的定义式：

其中表示数学期望。

纯净信号 的最佳估计 比带噪信号 含有更少的噪声，输出 的最佳滤波器是满足下面式子的维纳滤波器：

考虑一个特殊的滤波器

带噪信号 通过此滤波器，输出的信号保持不变（噪声没有消除掉），因此相对应的均方误差为：

大体上对于理想滤波器，我们有：

换句话说，维纳滤波器可以降低带噪信号中的噪声水平。

从上面的几个式子我们可以很快的推导出Wiener-Hopf 方程：

其中

是带噪信号 的相关系数矩阵

是带噪信号和纯净语音的互相关矩阵。但是， 未知；所以，看起来 的值是很难估计到的。但是

=

=

现在我们发现 由互相关向量 和所决定。 向量 (也是的第一列) 可以在语音和噪声的间隙很容易的估计出来而 也可以在只有噪声的间隙估计出来（这里假设噪声性质不随时间变化）。根据上面的式子以及，我们就得到了理想滤波器：

=

=

其中

是信噪比，I是单位矩阵，

我们可以推导出

最小均方误差是：

我们可以从上式很清楚的看到; 因此，噪声的消除是完全可能的。

规范化的最小均方误差为：

其中， 。

1. Universite du Quebec, INRS-EMTMontreal, QC H5A 1K6, Canada [↑](#footnote-ref-1)
2. Bell Laboratories, Lucent Technologies Murray Hill, NJ 07974，USA [↑](#footnote-ref-2)
3. Katholieke Universiteit Leuven, Dept, of Electrical Engineering (ESAT-SCD) Leuven 3001, Belgium [↑](#footnote-ref-3)
4. Universite du Quebec, INRS-EMTMontreal, QC H5A 1K6, Can

   ada [↑](#footnote-ref-4)
5. Bell Laboratories, Lucent Technologies Murray Hill, NJ 07974，USA [↑](#footnote-ref-5)
6. Katholieke Universiteit Leuven, Dept, of Electrical Engineering (ESAT-SCD) Leuven 3001, Belgium [↑](#footnote-ref-6)