A Detailed Study of Different Speech Enhancement Techniques

Raktim Pal and Mikhail Fedarau

Abstract

Speech enhancement is a topic of great importance and there are numerous existing methods to achieve the goal of enhancement. In this project, we study several speech enhancement techniques and compare them on the basis of enhancement quality, algorithm complexity, and overall performance. However, due to the different types of noise, lack of common speech database, and inconsistencies in testing methods, it is problematic to make comparisons between the performance of different algorithms. Hence, one of the goals of this project is to determine the best ways to assess quality of speech enhancement.

1 Introduction

It is hard to imagine our everyday lives without the digital assistants like Apple’s Siri, Microsoft’s Cortana, Google Now, etc. We rely on them to tell us the weather for tomorrow, give us directions to nearby Starbucks, set a reminder, or quickly reply to a text. The easiest and the most intuitive way to interact with them is to do it as one would do it with his human personal assistant—just tell them. It works most of the time—usually in a quiet space when you are the only one who’s talking. However, the voice detection by digital assistants in real-life environment, with the background noise and multiple speakers, is far from perfect. Since we rely increasingly more on our gadgets it is paramount to establish reliable and intuitive way of communication between a person and his digital assistant and speech enhancement is an integral part of that process. However, speech enhancement is not limited to communication with one’s personal assistants and voice communication. It is a powerful preprocessing tool that is useful and sometimes plain necessary for speech compression, keyword spotting, speaker verification etc.
2 Enhancement Techniques

Speech enhancement is widely used nowadays not only in telecommunications to improve quality of speech passing through a network, but in various noisy environments as a preprocessing technique. Its importance spurred the interest in that area and led to development of numerous methods of speech enhancement that operate differently, used for different purposes, and yield different results. Because of that variety of goals, implementations algorithms, and types of noise they operate with, it is very hard to properly compare them. In this project we decided to choose the most prominent speech enhancement techniques and design or implement methods to quantitatively compare their performance. To achieve such complicated and nontrivial task, we first had to conduct an extensive literature survey to determine the main areas of focus and state of the art enhancement techniques as well as the methods to compare them. According, to [1] there are four main classes of speech enhancement techniques: Statistical-model Based, methods based on Spectral Subtraction, Wiener type algorithms, and Deep Learning techniques.

![Enhancement Techniques](image1)

First three groups of speech enhancement techniques are based on fundamental principles of signal processing while the last one is a machine learning based method making use of Deep Neural Networks. Since, we wanted to incorporate our knowledge from the class and in some instances dive deeper into the subject of speech processing, we decided to focus on the three DSP - based speech enhancement techniques. Apart from the four fundamental and popular methods discussed above, a few other recent methods of speech enhancement that we came across are briefly mentioned below:
Seon M. Kim et al. [2] presented a method for target speech estimation by considering the spatial cues in noisy environments. D. P. K. Lun et al. [3] presented an improved speech enhancement algorithm based on a novel expectation maximization (EM) framework. This method uses the sparsity of speeches in the cepstral domain and performs well when the speech is distorted by the non-stationary noises. R. Bendoumia et al. [4] introduced two new two channel VSS- FB algorithms for speech enhancement and noise reduction. The two proposed methods are based on optimal step-size estimation with the use of decorrelation criteria. T. Mellahi et al. [5] proposed a new iterative Kalman filtering scheme for speech enhancement distorted by AWGN and colored noises. Although, these methods sound promising, but due to limited time we have abstained from exploring these techniques in detail.

**Data:** In this project we have made use of the NOIZEUS database for testing all the speech enhancement algorithms and their quality comparison. NOIZEUS is a noisy speech corpus recorded in UT Dallas to facilitate comparison of speech enhancement algorithms among research groups. The noisy database contains 30 IEEE sentences (IEEE Subcommittee, 1969) produced by three male and three female speakers, and was corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, multi-talker babble, car, exhibition hall, restaurant, street, airport and train-station noise.

MATLAB was used for implementing all the speech enhancement techniques and their comparative studies.

### 2.1 Method of Spectral Subtraction

Spectral subtraction [6] offers a computationally efficient, processor independent approach to effective digital speech enhancement. This approach estimates the magnitude frequency spectrum of the underlying clean speech by subtracting the noise magnitude spectrum from the noisy speech spectrum. This estimator requires an estimate of the current noise spectrum. Rather than obtain this noise estimate from a second microphone source, it is approximated using the average noise magnitude measured during non speech activity. Secondary procedures are then applied to attenuate the residual noise left after subtraction. Since the algorithm resynthesizes a speech waveform, it can be used as a preprocessor to narrow-band voice communications systems, speech recognition systems, or speaker authentication systems.

Speech, suitably low-pass filtered and digitized, is analyzed by windowing the data. The magnitude spectra of the windowed data are calculated and the spectral noise bias calculated during non speech activity is subtracted off. Resulting negative amplitudes are then zeroed out. Secondary residual noise suppression such as magnitude averaging is then applied. A time waveform is recalculated from the modified magnitude. This waveform is then overlap added to the previous data to generate the output speech.

Assume that a windowed noise signal $n(k)$ has been added to a windowed speech signal $s(k)$, with their sum denoted by $x(k)$. Then,
\[ x(k) = s(k) + n(k) \]

Taking the Fourier Transform gives,

\[ X(e^{j\omega}) = S(e^{j\omega}) + N(e^{j\omega}) \]

The spectral subtraction filter \( H(e^{j\omega}) \) is calculated by replacing the noise spectrum \( N(e^{j\omega}) \) with spectra which can be readily measured. The magnitude \( |N(e^{j\omega})| \) of \( N(e^{j\omega}) \) is replaced by its average value \( \mu(e^{j\omega}) \) taken during nonspeech activity, and the phase \( \theta_N(e^{j\omega}) \) of \( N(e^{j\omega}) \) is replaced by the phase \( \theta_X(e^{j\omega}) \) of \( X(e^{j\omega}) \). These substitutions result in the spectral subtraction estimator \( \hat{S}(e^{j\omega}) \):

\[
\hat{S}(e^{j\omega}) = [|X(e^{j\omega})| - \mu(e^{j\omega})]e^{j\theta_X(e^{j\omega})}
\]

or

\[
\hat{S}(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})
\]

with

\[
H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|}
\]

\[
\mu(e^{j\omega}) = E\{|N(e^{j\omega})|\}
\]

The spectral error \( \epsilon(e^{j\omega}) \) resulting from this estimator is given by:

\[
\epsilon(e^{j\omega}) = \hat{S}(e^{j\omega}) - S(e^{j\omega}) = N(e^{j\omega}) - \mu(e^{j\omega})e^{j\theta_X}
\]

A number of simple modifications are available to reduce the auditory effects of this spectral error. One of these include magnitude averaging.

Magnitude Averaging: Since the spectral error equals the difference between the noise spectrum \( N \) and its mean \( \mu \), local averaging of spectral magnitudes can be used to reduce the error. Replacing \( |X(e^{j\omega})| \) with \( \overline{|X(e^{j\omega})|} \) where,

\[
\frac{1}{M} \sum_{i=0}^{M-1} |X_i(e^{j\omega})|
\]

where, \( X_i(e^{j\omega}) = i^{th} \) time-windowed transform of \( x(k) \). The rationale behind averaging is that the spectral error becomes approximately \( \epsilon(e^{j\omega}) \approx |N| - \mu \), where, \( |N(e^{j\omega})| = \frac{1}{M} \sum_{i=0}^{M-1} |N_i(e^{j\omega})| \). Thus, the sample mean of \( |N_i(e^{j\omega})| \) will converge to \( \mu(e^{j\omega}) \) as a longer average is taken. The obvious problem with this modification is that the speech is non-stationary, and therefore only limited time averaging is allowed. The algorithm was implemented in MATLAB and tested on a noisy speech waveform taken from "NOIZEUS" database. The results are shown in the figures that follow.
Figure 2: Spectrogram of noisy speech

Figure 3: Spectrogram of enhanced speech obtained using Spectral Subtraction
2.2 Weiner Noise Suppressor with TSNR and HRNR algorithms

Scalart and Vieira Filho presented in [7] a unified view of the main single microphone noise reduction techniques where the noise reduction process relies on the estimation of a short-time spectral gain, which is a function of the a priori signal-to-noise ratio (SNR) and/or the a posteriori SNR. They also emphasize the interest of estimating the a priori SNR thanks to the decision-directed (DD) approach proposed by Ephram and Malah in [8]. [9] proposes a method, called two-step noise reduction (TSNR), to refine the estimation of the a priori SNR which removes the drawbacks of the DD approach while maintaining its advantage, i.e., highly reduced musical noise level. The major advantage of this approach is the suppression of the frame delay bias leading to the cancellation of the annoying reverberation effect characteristic of the DD approach. Furthermore, one major limitation that exists in classic short-time suppression techniques, including the TSNR, is that some harmonics are considered as noise only components and consequently are suppressed by the noise reduction process. This is inherent to the errors introduced by the noise spectrum estimation which is a very difficult task for single channel noise reduction techniques. Note that in most spoken languages, voiced sounds represent a large amount (around 80%) of the pronounced sounds. Then it is very interesting to overcome this limitation. For that purpose, [10] proposes a method, called harmonic regeneration noise reduction (HRNR), that takes into account the harmonic characteristic of speech. In this approach, the output
signal of any classic noise reduction technique (with missing or degraded harmonics) is further processed to create an artificial signal where the missing harmonics have been automatically regenerated. This artificial signal helps to refine the a priori SNR used to compute a spectral gain able to preserve the harmonics of the speech signal. The algorithms TSNR and HRNR were implemented in MATLAB and tested on a noisy speech waveform taken from \textit{NOIZEUS} database. The results are shown in the figures that follow.

![Spectrogram - noisy speech](image)

Figure 5: Spectrogram of Noisy Speech
Figure 6: Spectrogram of enhanced speech using TSNR

Figure 7: Spectrogram of enhanced speech using HRNR
2.3 Minimum Mean Square Error Log-Spectral Amplitude Estimator

In [8], Ephraim and Malah proposed an algorithm for enhancing speech degraded by uncorrelated additive noise when the noisy speech alone is available. This algorithm capitalizes on the major importance of the short-time spectral amplitude (STSA) of the speech signal in its perception, and utilizes a minimum mean-square error (MMSE) STSA estimator for enhancing the noisy speech.

While the distortion measure of mean-square error of the spectra (i.e., the original STSA and its estimator) used in [8] is mathematically tractable, and leads also to good results, it is not the most subjectively meaningful one. It is well known that a distortion measure which is based on the mean-square error of the log-spectra is more suitable for speech processing (e.g., see [11]). Such a distortion measure is therefore extensively used for speech analysis and recognition. For this reason, it is of great interest to examine the STSA estimator which minimizes the mean-square error of the log-spectra in enhancing noisy speech [12].

Specifically, the estimation problem of the STSA is formulated as that of estimating the amplitude of each Fourier expansion coefficient of the speech signal \( \{x(t), 0 \leq t \leq T\} \), given the noisy process \( \{y(t), 0 \leq t \leq T\} \). The Fourier expansion coefficients of the speech process, as well as of the noise process, are modeled as statistically independent Gaussian random variables. This model utilizes asymptotic statistical properties (as \( T \to \infty \)) of spectral components. In particular, the Gaussian model is motivated by the central limit theorem, as each Fourier expansion coefficient is after all a weighted sum of random variables. In addition, the statistical independence assumption is motivated by the fact that the correlation between the spectral components reduces as the analysis interval length increases. A detailed discussion concerning the above statistical model
is given in [8].

Let $X_k = A_k e^{j\omega_k}$, $D_k$, and $Y_k = R_k e^{j\nu_k}$, denote the kth Fourier expansion coefficient of the speech signal, the noise process, and the noisy observations, respectively, in the analysis interval $[0, T]$. According to the formulation of the estimation problem given above, we are looking for the estimator $\hat{A}_k$, which minimizes the following distortion measure: $E[(\log A_k - \log \hat{A}_k)^2]$ given the noisy observations $\{y(t), 0 \leq t \leq T\}$. This estimator is easily shown to be:

$$\hat{A}_k = \exp\{E[\log A_k | y(t), 0 \leq t \leq T]\}$$

As noted in [8], under the assumed statistical model, the expected value of $A_k$ given $\{y(t), 0 \leq t \leq T\}$ equals to the expected value of $A_k$, given $Y_k$ only. Since this statement remains true when $A_k$ is replaced by $\ln A_k$, the estimator equals:

$$\hat{A}_k = \exp\{E[\ln A_k | Y_k]\}$$

The evaluation of $E[\ln A_k | Y_k]$ for the Gaussian model assumed here is conveniently done by utilizing the moment generating function of $\ln A_k$ given $Y_k$. The details of the derivation can be found in [12]. The final expression for $\hat{A}_k$ can be obtained as:

$$\hat{A}_k = \frac{\xi_k}{1 + \xi_k} \exp\left\{\frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt \right\} R_k$$

where $\xi_k$ is defined as the a priori signal to noise ratio (SNR); $v_k$ is defined as

$$v_k \triangleq \frac{\xi_k}{1 + \xi_k} \lambda_k$$

where,

$$\frac{1}{\lambda_k} = \frac{1}{E\{|X_k|^2\}} + \frac{1}{E\{|D_k|^2\}}$$

The algorithm was implemented in MATLAB and tested on a noisy speech waveform taken from NOIZEUS database. The results are shown in the figures that follow.
Figure 9: Spectrogram of Noisy Speech

Figure 10: Spectrogram of enhanced speech using log-MMSE
3 Quality Assessment

Having the enhanced files we had to decide on how to rate the performance of each speech enhancement method. However, due to the different types of noise, lack of common speech database, and inconsistencies in testing methods, it is problematic to make comparisons between the performance of different algorithms. After conducting an extensive research, we came to the conclusion that the best way was to incorporate several Quality Assessment techniques while making some adjustments to the proposed concepts so that the results were uniformly applicable.

3.1 Logarithmic Spectral Distortion (LSD)

The first measure that we implemented to assess the quality of the enhanced speech was Logarithmic Spectral Distance (LSD).

\[
D_{LC} = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(10\log_{10} \frac{P(\omega)}{\hat{P}(\omega)}\right)^2 d\omega}
\]

where \(P(\omega)\) and \(\hat{P}(\omega)\) are power spectra of speech signals being compared. The LSD method operates in frequency domain on frame by frame basis.
3.2 PESQ

Perceptual Evaluation of Speech Quality (PESQ) is an industry standard and the most widely used method of speech quality assessment. It is standardized as ITU-T recommendation P.862 (02/01) and is universally used for objective voice quality testing used by phone manufacturers, network equipment vendors and telecom operators. PESQ was developed to model subjective tests commonly used in telecommunications (e.g. ITU-T P.800) to assess the voice quality by human beings.

![Figure 12: PESQ Algorithm](image)

The enhanced signal and the reference signal are individually level aligned and filtered with the transfer characteristics of a receiving end. The algorithms utilizes two different filter functions, a narrow band (IRS) filter and a wide band filter, for every input. Then there’s signal alignment stage in order to compensate for small time shifts that can occur. Since not all distortions are perceived by human ear equally and some distortions are not even perceived at all, PESQ method operates only on the distortions that could be detected by a listener. Thus, in order to account for the distortions that are actually perceived by a human listener the model transforms the two signals that are now aligned and filtered from the time-amplitude domain into a frequency-loudness domain (auditory transform) [13].

After that it estimates the amplitude differences by subtracting the two signal representations. The audible differences are accumulated over time while they are weighted differently depending on whether a distortion was added to the signal or if parts of the signal were missing after the transmission (cognitive model). Audible differences can be represented via the average disturbance value $D_{ind}$ and the average asymmetrical disturbance values $A_{ind}$:

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind}$$

At the end, after the analysis is done, a single Mean Opinion Score (MOS) is generated. The MOS is a quantitative measurement of speech quality and is commonly used to
describe the voice quality on a scale from 1 (bad) to 5 (good). Another measure is MOS-LQ (MOS Listening Quality). We used both in this project. Since PESQ was specifically designed to assess the quality of a speech passing through a telecommunication networks and not for speech enhancement, the PESQ algorithms as it is was not very suitable for our purposes. Therefore, we had to modify it. As was mentioned above, PESQ is computed as a linear combination of the average disturbance value $D_{\text{ind}}$ and the average asymmetrical disturbance values $A_{\text{ind}}$ weighted by coefficients: $a_0, a_1, a_2$. We run extensive tests to properly choose those coefficients by treating $a_0, a_1, and a_2$ as parameters that need to be optimized for each of the three rating scales: speech distortion, noise distortion, and overall quality.

3.3 Segmental SNR

The last metric that we designed to assess the quality of speech enhancement algorithms was Segmental Signal-to Noise Ratio (SegSNR). Segmental SNR is a version of general SNR taken on frame-by-frame basis. The (general) SNR can be expressed as following:

$$SNR = 10 \log_{10} \frac{\sum_n s^2(n)}{\sum_n (s(n) - \hat{s}(n))^2}$$

The segmental SNR (SegSNR) is defined as averages of measurements of SNR over short, "good frames". [14]

The "good frames" are frames for which the SNR is above a lower threshold, and is lower than an upper threshold. In this project the range was form -10 dB to 25 dB. Typically these frames are of 15-25 ms. The SegSNR can be written as:

$$SNR = \frac{1}{N} \sum_{k=1}^{N} 10 \log_{10} \left[ \frac{\sum_{n \in \text{frame}_k} |s(n)|^2}{\sum_{n \in \text{frame}_k} |s(n) - \hat{s}(n)|^2} \right]$$

where: $N$ the number of "good frames" and $n \in \text{frame}_k$ the time instances $n$ that are in the time window of the $k^{th}$ frame. We implemented this and other two algorithms in MATLAB for testing purposes.

3.4 Comparison

For testing purposes, we chose several types on noise, different speakers (male and female), and different signal-to-noise ratios. The data used in this project is from the "NOIZEUS". We organized the quality assessments of four algorithms designed in this project in tables for clarity. For each speech enhancement method we analyzed its performance with four techniques: LSD, MOS, MOS-LQ, and Segmental SNR. In Tables 1 and 2, we can see the performance of designed algorithms for a female speaker in a presence of strong train noise at SNR=0 dB (SNR=10 dB for Table 2). As it can be seen from the Tables 1 and 2, Spectral Subtraction method performs the best with lowest LSD and highest MOS and Segmental SNR. It is followed by LogMMSE method.
which perform similarly. Finally, TSNR and HRNR performance is significantly worse than that of SS and LogMMSE algorithms. Quality of LogMMSE and Spectral Subtraction method are superior to Wiener type algorithms. LSD parameters for Weiner type algorithms are lower while MOS and Segmental SNR is higher. Similar results are obtained for male speaker in different type of noise (street noise) for SNR=0 dB and SNR=10 dB. The results are shown on Tables 3 and 4.

**Table 1: Train Noise, Female, SNR 0**

<table>
<thead>
<tr>
<th></th>
<th>LogMMSE</th>
<th>TSNR</th>
<th>HRNR</th>
<th>Spectral Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD</td>
<td>2.0069</td>
<td>2.2597</td>
<td>2.323</td>
<td>2.0075</td>
</tr>
<tr>
<td>MOS</td>
<td>1.609</td>
<td>1.092</td>
<td>1.055</td>
<td>1.593</td>
</tr>
<tr>
<td>MOS-LQ</td>
<td>1.378</td>
<td>1.184</td>
<td>1.174</td>
<td>1.370</td>
</tr>
<tr>
<td>SegSNR</td>
<td>-0.7450</td>
<td>-2.9166</td>
<td>-2.1459</td>
<td>-0.6018</td>
</tr>
</tbody>
</table>

**Table 2: Train Noise, Female, SNR 10**

<table>
<thead>
<tr>
<th></th>
<th>LogMMSE</th>
<th>TSNR</th>
<th>HRNR</th>
<th>Spectral Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD</td>
<td>1.5547</td>
<td>3.04</td>
<td>3.4210</td>
<td>1.5089</td>
</tr>
<tr>
<td>MOS</td>
<td>2.441</td>
<td>1.707</td>
<td>1.646</td>
<td>2.349</td>
</tr>
<tr>
<td>MOS-LQ</td>
<td>2.065</td>
<td>1.432</td>
<td>1.398</td>
<td>1.961</td>
</tr>
<tr>
<td>SegSNR</td>
<td>1.7408</td>
<td>1.3636</td>
<td>1.2583</td>
<td>4.0985</td>
</tr>
</tbody>
</table>

**Table 3: Street Noise, Male, SNR 0**

<table>
<thead>
<tr>
<th></th>
<th>LogMMSE</th>
<th>TSNR</th>
<th>HRNR</th>
<th>Spectral Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD</td>
<td>1.8936</td>
<td>2.6825</td>
<td>2.8903</td>
<td>1.8336</td>
</tr>
<tr>
<td>MOS</td>
<td>1.7777</td>
<td>1.532</td>
<td>1.393</td>
<td>1.649</td>
</tr>
<tr>
<td>MOS-LQ</td>
<td>1.474</td>
<td>1.341</td>
<td>1.281</td>
<td>1.399</td>
</tr>
<tr>
<td>SegSNR</td>
<td>-0.1105</td>
<td>-0.2754</td>
<td>-0.8155</td>
<td>-0.0054</td>
</tr>
</tbody>
</table>

**Table 4: Street Noise, Male, SNR 10**

<table>
<thead>
<tr>
<th></th>
<th>LogMMSE</th>
<th>TSNR</th>
<th>HRNR</th>
<th>Spectral Subtraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD</td>
<td>1.5572</td>
<td>2.0618</td>
<td>2.0775</td>
<td>1.5699</td>
</tr>
<tr>
<td>MOS</td>
<td>2.301</td>
<td>2.356</td>
<td>2.405</td>
<td>2.296</td>
</tr>
<tr>
<td>MOS-LQ</td>
<td>2.023</td>
<td>1.909</td>
<td>1.904</td>
<td>1.968</td>
</tr>
<tr>
<td>SegSNR</td>
<td>2.5076</td>
<td>2.3914</td>
<td>2.2946</td>
<td>3.2272</td>
</tr>
</tbody>
</table>

Similarly with the previous test, quality of LogMMSE and Spectral Subtraction method are superior to Wiener type algorithms. LSD parameters for them are lower.
while MOS and Segmental SNR is higher. Finally, we tested the performance of speech enhancement algorithms that we designed on the multi-speaker speech (Babble Noise). The results in this case (Tables 5 and 6) were different. Multi-speaker environment turned out to be not a trivial case. As we can see in Tables 5 and 6, there is no statistical difference in performance of all four methods. Unlike in previous scenarios, bubble noise is harder to clear out with any enhancement technique. No algorithm produced significant quality improvement in multi-talker babble, i.e., in highly non-stationary environments. However, HSNR did relatively well in multi-speaker environment.

<table>
<thead>
<tr>
<th>Table 5: Babble Noise, Male, SNR 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD</td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td>MOS</td>
</tr>
<tr>
<td>MOS-LQ</td>
</tr>
<tr>
<td>SegSNR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6: Babble Noise, Male, SNR 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSD</td>
</tr>
<tr>
<td>-------------------------------------</td>
</tr>
<tr>
<td>MOS</td>
</tr>
<tr>
<td>MOS-LQ</td>
</tr>
<tr>
<td>SegSNR</td>
</tr>
</tbody>
</table>

4 Conclusion

The present project studied DSP-based speech enhancement techniques and reported on the objective evaluation of the four speech enhancement algorithms using four different parameters of three evaluation methods. Based on the quantitative analysis of overall quality and speech and noise distortion, we came to the following conclusions:

a) At high SNR, LogMMSE yields lowest speech distortion and highest overall quality followed by Spectral Subtraction. While Weiner methods perform poorly. b) At low SNR, LogMMSE and Spectral Subtraction methods have very close performance. c) No algorithm produced significant quality improvement in multi-talker babble, i.e., in highly non-stationary environments. d) HRNR tends to do well in Babble noise. Probably this is due the harmonic regeneration ability of the HRNR. HRNR takes into account the harmonic characteristics of speech and after processing it reconstructs the missing harmonics. This allows for further reduction of noise since we can reconstruct the speech. e) Based on the overall complexity of the algorithms Spectral Subtraction method is the least complex in implementation. At the same time, Spectral Subtraction algorithm shows good performance mostly on par with LogMMSE.
References


