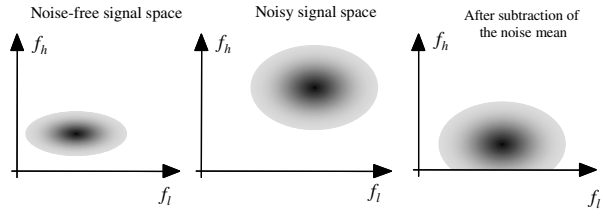
Advanced Digital Signal Processing and Noise Reduction, Second Edition.
Saeed V. Vaseghi

ISBNs: 0-471-62692-9 (Hardback): 0-470-84162-1 (Electronic)

11



SPECTRAL SUBTRACTION

Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated, and updated, from the periods when the signal is absent and only the noise is present. The assumption is that the noise is a stationary or a slowly varying process, and that the noise spectrum does not change significantly in-between the update periods. For restoration of time-domain signals, an estimate of the instantaneous magnitude spectrum is combined with the phase of the noisy signal, and then transformed via an inverse discrete Fourier transform to the time domain. In terms of computational complexity, spectral subtraction is relatively inexpensive. However, owing to random variations of noise, spectral subtraction can result in negative estimates of the short-time magnitude or power spectrum. The magnitude and power spectrum are non-negative variables, and any negative estimates of these variables should be mapped into non-negative values. This non-linear rectification process distorts the distribution of the restored signal. The processing distortion becomes more noticeable as the signal-to-noise ratio decreases. In this chapter, we study spectral subtraction, and the different methods of reducing and removing the processing distortions.

11.1 Spectral Subtraction

In applications where, in addition to the noisy signal, the noise is accessible on a separate channel, it may be possible to retrieve the signal by subtracting an estimate of the noise from the noisy signal. For example, the adaptive noise canceller of Section 1.3.1 takes as the inputs the noise and the noisy signal, and outputs an estimate of the clean signal. However, in many applications, such as at the receiver of a noisy communication channel, the only signal that is available is the noisy signal. In these situations, it is not possible to cancel out the random noise, but it may be possible to reduce the *average effects* of the noise on the signal spectrum. The effect of additive noise on the magnitude spectrum of a signal is to increase the mean and the variance of the spectrum as illustrated in Figure 11.1. The increase in the variance of the signal spectrum results from the random fluctuations of the noise, and cannot be cancelled out. The increase in the mean of the signal spectrum can be removed by subtraction of an estimate of the mean of the noise spectrum from the noisy signal spectrum. The noisy signal model in the time domain is given by

$$y(m)=x(m)+n(m) \tag{11.1}$$

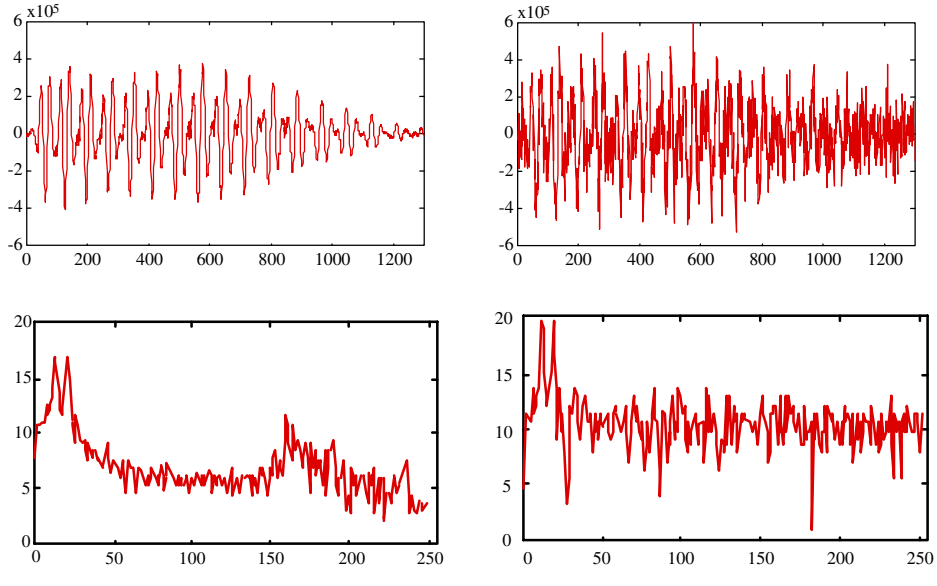


Figure 11.1 Illustrations of the effect of noise on a signal in the time and the frequency domains.

where $y(m)$, $x(m)$ and $n(m)$ are the signal, the additive noise and the noisy signal respectively, and m is the discrete time index. In the frequency domain, the noisy signal model of Equation (11.1) is expressed as

$$Y(f)=X(f)+N(f) \tag{11.2}$$

where $Y(f)$, $X(f)$ and $N(f)$ are the Fourier transforms of the noisy signal $y(m)$, the original signal $x(m)$ and the noise $n(m)$ respectively, and f is the frequency variable. In spectral subtraction, the incoming signal $x(m)$ is buffered and divided into segments of N samples length. Each segment is windowed, using a Hanning or a Hamming window, and then transformed via discrete Fourier transform (DFT) to N spectral samples. The windows alleviate the effects of the discontinuities at the endpoints of each segment. The windowed signal is given by

$$\begin{aligned} y_w(m) &= w(m)y(m) \\ &= w(m)[x(m)+n(m)] \\ &= x_w(m)+n_w(m) \end{aligned} \tag{11.3}$$

The windowing operation can be expressed in the frequency domain as

$$\begin{aligned} Y_w(f) &= W(f)*Y(f) \\ &= X_w(f)+N_w(f) \end{aligned} \tag{11.4}$$

where the operator * denotes convolution. Throughout this chapter, it is assumed that the signals are windowed, and hence for simplicity we drop the use of the subscript w for windowed signals.
Figure 11.2 illustrates a block diagram configuration of the spectral subtraction method. A more detailed implementation is described in Section 11.4. The equation describing spectral subtraction may be expressed as

$$\left|\hat{X}(f)\right|^b = |Y(f)|^b - \alpha\overline{|N(f)|^b} \tag{11.5}$$

where $|\hat{X}(f)|^b$ is an estimate of the original signal spectrum $|X(f)|^b$ and $\overline{|N(f)|^b}$ is the time-averaged noise spectra. It is assumed that the noise is a wide-sense stationary random process. For magnitude spectral subtraction, the exponent b=1, and for power spectral subtraction, b=2. The parameter α

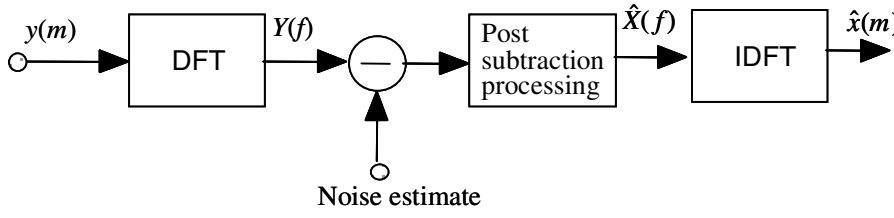


Figure 11.2 A block diagram illustration of spectral subtraction.

in Equation (11.5) controls the amount of noise subtracted from the noisy signal. For full noise subtraction, α=1 and for over-subtraction α>1. The time-averaged noise spectrum is obtained from the periods when the signal is absent and only the noise is present as

$$\overline{|N(f)|^b} = \frac{1}{K}\sum_{i=0}^{K-1}|N_i(f)|^b \tag{11.6}$$

In Equation (11.6), $|N_i(f)|$ is the spectrum of the i^{th} noise frame, and it is assumed that there are K frames in a noise-only period, where K is a variable. Alternatively, the averaged noise spectrum can be obtained as the output of a first order digital low-pass filter as

$$\overline{|N_i(f)|^b} = \rho\,\overline{|N_{i-1}(f)|^b} + (1-\rho)|N_i(f)|^b \tag{11.7}$$

where the low-pass filter coefficient ρ is typically set between 0.85 and 0.99. For restoration of a time-domain signal, the magnitude spectrum estimate $|\hat{X}(f)|$ is combined with the phase of the noisy signal, and then transformed into the time domain via the inverse discrete Fourier transform as

$$\hat{x}(m) = \sum_{k=0}^{N-1}|\hat{X}(k)|e^{j\theta_Y(k)}e^{-j\frac{2\pi}{N}km} \tag{11.8}$$

where $\theta_Y(k)$ is the phase of the noisy signal frequency $Y(k)$. The signal restoration equation (11.8) is based on the assumption that the audible noise is mainly due to the distortion of the magnitude spectrum, and that the phase distortion is largely inaudible. Evaluations of the perceptual effects of simulated phase distortions validate this assumption.

Owing to the variations of the noise spectrum, spectral subtraction may result in negative estimates of the power or the magnitude spectrum. This outcome is more probable as the signal-to-noise ratio (SNR) decreases. To avoid negative magnitude estimates the spectral subtraction output is post-processed using a mapping function $T[\cdot]$ of the form

$$T[|\hat{X}(f)|]=\begin{cases}|\hat{X}(f)| & \text{if } |\hat{X}(f)|>\beta|Y(f)| \\ \text{fn}[|Y(f)|] & \text{otherwise}\end{cases} \tag{11.9}$$

For example, we may chose a rule such that if the estimate $|\hat{X}(f)|>0.01|Y(f)|$ (in magnitude spectrum 0.01 is equivalent to –40 dB) then $|\hat{X}(f)|$ should be set to some function of the noisy signal fn[$Y(f)$]. In its simplest form, fn[$Y(f)$]=*noise floor*, where the noise floor is a positive constant. An alternative choice is fn[$|Y(f)|$]=$\beta|Y(f)|$. In this case,

$$T[|\hat{X}(f)|]=\begin{cases}|\hat{X}(f)| & \text{if } |\hat{X}(f)|>\beta|Y(f)| \\ \beta|Y(f)| & \text{otherwise}\end{cases} \tag{11.10}$$

Spectral subtraction may be implemented in the power or the magnitude spectral domains. The two methods are similar, although theoretically they result in somewhat different expected performance.

11.1.1 Power Spectrum Subtraction

The power spectrum subtraction, or squared-magnitude spectrum subtraction, is defined by the following equation:

$$|\hat{X}(f)|^2=|Y(f)|^2-\overline{|N(f)|^2} \tag{11.11}$$

where it is assumed that α, the subtraction factor in Equation (11.5), is unity. We denote the power spectrum by $\mathcal{E}[|X(f)|^2]$, the time-averaged power spectrum by $\overline{|X(f)|^2}$ and the *instantaneous* power spectrum by $|X(f)|^2$. By expanding the instantaneous power spectrum of the noisy

signal $|Y(f)|^2$, and grouping the appropriate terms, Equation (11.11) may be rewritten as

$$|\hat{X}(f)|^2 = |X(f)|^2 + \underbrace{\left(|N(f)|^2 - \overline{|N(f)|^2}\right)}_{\text{Noise variations}} + \underbrace{X^*(f)N(f) + X(f)N^*(f)}_{\text{Cross products}} \tag{11.12}$$

Taking the expectations of both sides of Equation (11.12), and assuming that the signal and the noise are uncorrelated ergodic processes, we have

$$\mathcal{E}[|\hat{X}(f)|^2] = \mathcal{E}[|X(f)|^2] \tag{11.13}$$

From Equation (11.13), the average of the estimate of the instantaneous power spectrum converges to the power spectrum of the noise-free signal. However, it must be noted that for non-stationary signals, such as speech, the objective is to recover the *instantaneous* or the short-time spectrum, and only a relatively small amount of averaging can be applied. Too much averaging will smear and obscure the temporal evolution of the spectral events. Note that in deriving Equation (11.13), we have not considered non-linear rectification of the negative estimates of the squared magnitude spectrum.

11.1.2 Magnitude Spectrum Subtraction

The magnitude spectrum subtraction is defined as

$$|\hat{X}(f)| = |Y(f)| - \overline{|N(f)|} \tag{11.14}$$

where $\overline{|N(f)|}$ is the time-averaged magnitude spectrum of the noise. Taking the expectation of Equation (11.14), we have

$$\begin{aligned} \mathcal{E}[|\hat{X}(f)|] &= \mathcal{E}[|Y(f)|] - \mathcal{E}[\overline{|N(f)|}] \\ &= \mathcal{E}[|X(f) + N(f)|] - \mathcal{E}[\overline{|N(f)|}] \\ &\approx \mathcal{E}[|X(f)|] \end{aligned} \tag{11.15}$$

For signal restoration the magnitude estimate is combined with the phase of the noisy signal and then transformed into the time domain using Equation (11.8).

11.1.3 Spectral Subtraction Filter: Relation to Wiener Filters

The spectral subtraction equation can be expressed as the product of the noisy signal spectrum and the frequency response of a spectral subtraction filter as

$$\begin{aligned} |\hat{X}(f)|^2 &= |Y(f)|^2 - \overline{|N(f)|^2} \\ &= H(f)|Y(f)|^2 \end{aligned} \tag{11.16}$$

where $H(f)$, the frequency response of the spectral subtraction filter, is defined as

$$\begin{aligned} H(f) &= 1 - \frac{\overline{|N(f)|^2}}{|Y(f)|^2} \\ &= \frac{|Y(f)|^2 - \overline{|N(f)|^2}}{|Y(f)|^2} \end{aligned} \tag{11.17}$$

The spectral subtraction filter $H(f)$ is a zero-phase filter, with its magnitude response in the range $0 \geq H(f) \geq 1$. The filter acts as a SNR-dependent attenuator. The attenuation at each frequency increases with the decreasing SNR, and conversely decreases with the increasing SNR.

The least mean square error linear filter for noise removal is the Wiener filter covered in chapter 6. Implementation of a Wiener filter requires the power spectra (or equivalently the correlation functions) of the signal and the noise process, as discussed in Chapter 6. Spectral subtraction is used as a substitute for the Wiener filter when the signal power spectrum is not available. In this section, we discuss the close relation between the Wiener filter and spectral subtraction. For restoration of a signal observed in uncorrelated additive noise, the equation describing the frequency response of the Wiener filter was derived in Chapter 6 as

$$W(f) = \frac{\mathcal{E}[|Y(f)|^2] - \mathcal{E}[|N(f)|^2]}{\mathcal{E}[|Y(f)|^2]} \tag{11.18}$$

A comparison of $W(f)$ and $H(f)$, from Equations (11.18) and (11.17), shows that the Wiener filter is based on the *ensemble-average* spectra of the signal and the noise, whereas the spectral subtraction filter uses the instantaneous spectra of the noisy signal and the *time-averaged* spectra of the noise. In spectral subtraction, we only have access to a single realisation of the process. However, assuming that the signal and noise are wide-sense stationary ergodic processes, we may replace the instantaneous noisy signal spectrum $|Y(f)|^2$ in the spectral subtraction equation (11.18) with the time-averaged spectrum $\overline{|Y(f)|^2}$, to obtain

$$H(f) = \frac{\overline{|Y(f)|^2} - \overline{|N(f)|^2}}{\overline{|Y(f)|^2}} \tag{11.19}$$

For an ergodic process, as the length of the time over which the signals are averaged increases, the time-averaged spectrum approaches the ensemble-averaged spectrum, and in the limit, the spectral subtraction filter of Equation (11.19) approaches the Wiener filter equation (11.18). In practice, many signals, such as speech and music, are non-stationary, and only a limited degree of beneficial time-averaging of the spectral parameters can be expected.

11.2 Processing Distortions

The main problem in spectral subtraction is the non-linear processing distortions caused by the random variations of the noise spectrum. From Equation (11.12) and the constraint that the magnitude spectrum must have a non-negative value, we may identify three sources of distortions of the instantaneous estimate of the magnitude or power spectrum as:

(a) the variations of the instantaneous noise power spectrum about the mean;
(b) the signal and noise cross-product terms;
(c) the non-linear mapping of the spectral estimates that fall below a threshold.

The same sources of distortions appear in both the magnitude and the power spectrum subtraction methods. Of the three sources of distortions listed

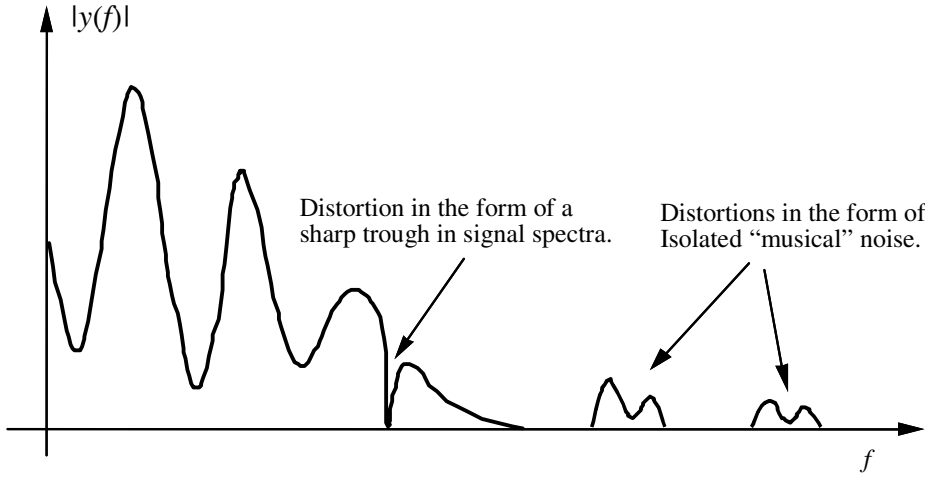


Figure 11.3 Illustration of distortions that may result from spectral subtraction.

above, the dominant distortion is often due to the non-linear mapping of the negative, or small-valued, spectral estimates. This distortion produces a metallic sounding noise, known as "*musical tone noise*" due to their narrow-band spectrum and the tin-like sound. The success of spectral subtraction depends on the ability of the algorithm to reduce the noise variations and to remove the processing distortions. In its worst, and not uncommon, case the residual noise can have the following two forms:

(a) a sharp trough or peak in the signal spectra;
(b) isolated narrow bands of frequencies.

In the vicinity of a high amplitude signal frequency, the noise-induced trough or peak is often masked, and made inaudible, by the high signal energy. The main cause of audible degradations is the isolated frequency components also known as *musical tones* or musical noise illustrated in Figure 11.3. The musical noise is characterised as short-lived narrow bands of frequencies surrounded by relatively low-level frequency components. In audio signal restoration, the distortion caused by spectral subtraction can result in a significant deterioration of the signal quality. This is particularly true at low signal-to-noise ratios. The effects of a bad implementation of subtraction algorithm can result in a signal that is of a lower perceived quality, and lower information content, than the original noisy signal.

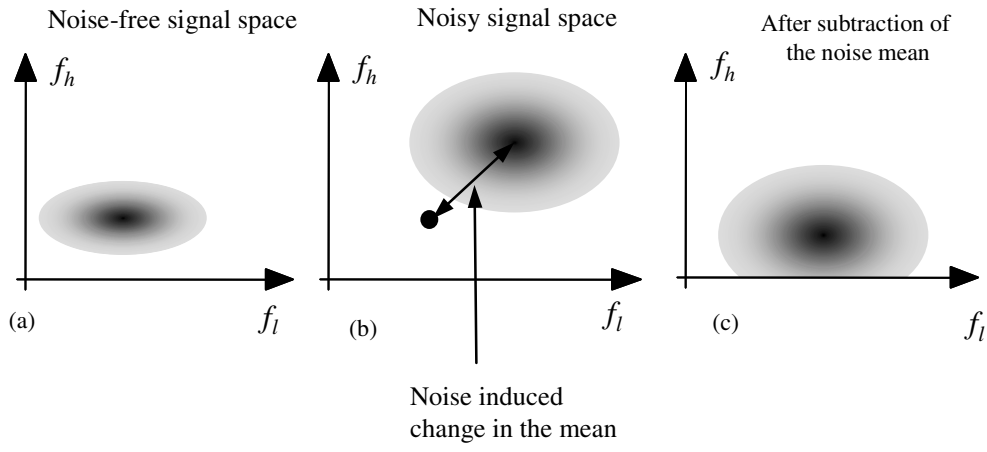


Figure 11.4 Illustration of the distorting effect of spectral subtraction on the space of the magnitude spectrum of a signal.

11.2.1 Effect of Spectral Subtraction on Signal Distribution

Figure 11.4 is an illustration of the distorting effect of spectral subtraction on the distribution of the magnitude spectrum of a signal. In this figure, we have considered the simple case where the spectrum of a signal is divided into two parts; a low-frequency band f_l and a high-frequency band f_h. Each point in Figure 11.4 is a plot of the high-frequency spectrum versus the low-frequency spectrum, in a two-dimensional signal space. Figure 11.4(a) shows an assumed distribution of the spectral samples of a signal in the two-dimensional magnitude–frequency space. The effect of the random noise, shown in Figure 11.4(b), is an increase in the mean and the variance of the spectrum, by an amount that depends on the mean and the variance of the magnitude spectrum of the noise. The increase in the variance constitutes an irrevocable distortion. The increase in the mean of the magnitude spectrum can be removed through spectral subtraction. Figure 11.4(c) illustrates the distorting effect of spectral subtraction on the distribution of the signal spectrum. As shown, owing to the noise-induced increase in the variance of the signal spectrum, after subtraction of the average noise spectrum, a proportion of the signal population, particularly those with a low SNR, become negative and have to be mapped to non-negative values. As shown this process distorts the distribution of the low-SNR part of the signal spectrum.

11.2.2 Reducing the Noise Variance

The distortions that result from spectral subtraction are due to the variations of the noise spectrum. In Section 9.2 we considered the methods of reducing the variance of the estimate of a power spectrum. For a white noise process with variance σ_n^2, it can be shown that the variance of the DFT spectrum of the noise $N(f)$ is given by

$$\mathrm{Var}[|N(f)|^2] \approx P_{NN}^2(f) = \sigma_n^4 \tag{11.20}$$

and the variance of the running average of K independent spectral components is

$$\mathrm{Var}\left[\frac{1}{K}\sum_{i=0}^{K-1}|N_i(f)|^2\right] \approx \frac{1}{K}P_{NN}^2(f) \approx \frac{1}{K}\sigma_n^4 \tag{11.21}$$

From Equation (11.21), the noise variations can be reduced by time-averaging of the noisy signal frequency components. The fundamental limitation is that the averaging process, in addition to reducing the noise variance, also has the undesirable effect of smearing and blurring the time variations of the signal spectrum. Therefore an averaging process should reflect a compromise between the conflicting requirements of reducing the noise variance and of retaining the time resolution of the non-stationary spectral events. This is important because time resolution plays an important part in both the quality and the intelligibility of audio signals.

In spectral subtraction, the noisy signal $y(m)$ is segmented into blocks of N samples. Each signal block is then transformed via a DFT into a block of N spectral samples $Y(f)$. Successive blocks of spectral samples form a two-dimensional frequency–time matrix denoted by $Y(f,t)$ where the variable t is the segment index and denotes the time dimension. The signal $Y(f,t)$ can be considered as a band-pass channel f that contains a time-varying signal $X(f,t)$ plus a random noise component $N(f,t)$. One method for reducing the noise variations is to low-pass filter the magnitude spectrum at each frequency. A simple recursive first-order digital low-pass filter is given by

$$|Y_{LP}(f,t)| = \rho|Y_{LP}(f,t-1)| + (1-\rho)|Y(f,t)| \tag{11.22}$$

where the subscript LP denotes the output of the low-pass filter, and the smoothing coefficient ρ controls the bandwidth and the time constant of the low-pass filter.

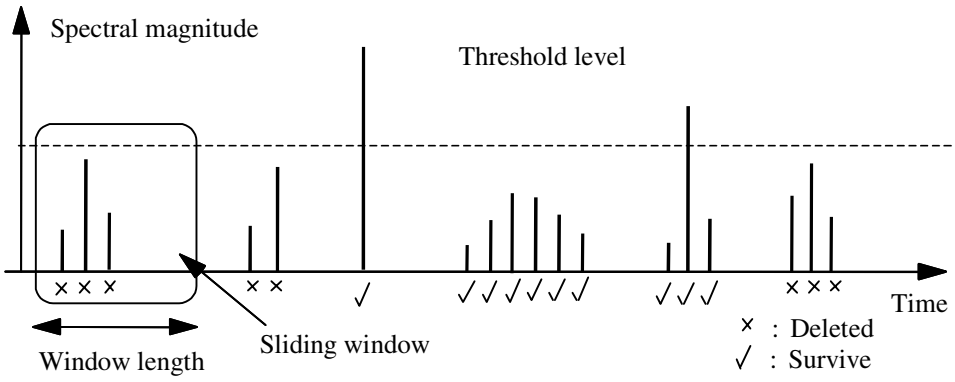


Figure 11.5 Illustration of a method for identification and filtering of "musical noise".

11.2.3 Filtering Out the Processing Distortions

Audio signals, such as speech and music, are composed of sequences of non-stationary acoustic events. The acoustic events are "born", have a varying lifetime, disappear, and then reappear with a different intensity and spectral composition. The time–varying nature of audio signals plays an important role in conveying information, sensation and quality. The musical tone noise, introduced as an undesirable by-product of spectral subtraction, is also time-varying. However, there are significant differences between the characteristics of most audio signals and so-called musical noise. The characteristic differences may be used to identify and remove some of the more annoying distortions. Identification of musical noise may be achieved by examining the variations of the signal in the time and frequency domains. The main characteristics of musical noise are that it tends to be relatively short-lived random isolated bursts of narrow band signals, with relatively small amplitudes.

Using a DFT block size of 128 samples, at a sampling rate of 20 kHz, experiments indicate that the great majority of musical noise tends to last no more than three frames, whereas genuine signal frequencies have a considerably longer duration. This observation was used as the basis of an effective "musical noise" suppression system. Figure 11.5 demonstrates a method for the identification of musical noise. Each DFT channel is examined to identify short-lived frequency events. If a frequency component has a duration shorter than a pre-selected time window, and an amplitude smaller than a threshold, and is not masked by signal components in the adjacent frequency bins, then it is classified as distortion and deleted.

11.3 Non-Linear Spectral Subtraction

The use of spectral subtraction in its basic form of Equation (11.5) may cause deterioration in the quality and the information content of a signal. For example, in audio signal restoration, the musical noise can cause degradation in the perceived quality of the signal, and in speech recognition the basic spectral subtraction can result in deterioration of the recognition accuracy. In the literature, there are a number of variants of spectral subtraction that aim to provide consistent performance improvement across a range of SNRs. These methods differ in their approach to estimation of the noise spectrum, in their method of averaging the noisy signal spectrum, and in their post processing method for the removal of processing distortions.
Non-linear spectral subtraction methods are heuristic methods that utilise estimates of the local SNR, and the observation that at a low SNR over-subtraction can produce improved results. For an explanation of the improvement that can result from over-subtraction, consider the following expression of the basic spectral subtraction equation:

$$\begin{aligned} |\hat{X}(f)| &= |Y(f)| - \overline{|N(f)|} \\ &\approx |X(f)| + |N(f)| - \overline{|N(f)|} \\ &\approx |X(f)| + V_N(f) \end{aligned} \tag{11.23}$$

where $V_N(f)$ is the zero-mean random component of the noise spectrum. If $V_N(f)$ is well above the signal $X(f)$ then the signal may be considered as lost to noise. In this case, over-subtraction, followed by non-linear processing of the negative estimates, results in a higher overall attenuation of the noise. This argument explains why subtracting more than the noise average can sometimes produce better results. The non-linear variants of spectral subtraction may be described by the following equation:

$$|\hat{X}(f)| = Y(f) - \alpha\left(SNR(f)\right)\overline{|N(f)|}_{NL} \tag{11.24}$$

where $\alpha(SNR(f))$ is an SNR-dependent subtraction factor and $\overline{|N(f)|}_{NL}$ is a non-linear estimate of the noise spectrum. The spectral estimate is further processed to avoid negative estimates as

$$|\hat{X}(f)| = \begin{cases} |\hat{X}(f)| & \text{if } |\hat{X}(f)| > |\beta Y(f)| \\ |\beta Y(f)| & \text{otherwise} \end{cases} \tag{11.25}$$

One form of an SNR-dependent subtraction factor for Equation (11.24) is given by

$$\alpha\left(SNR(f)\right) = 1 + \frac{sd\left(|N(f)|\right)}{\overline{|N(f)|}} \tag{11.26}$$

where the function $sd(|N(f)|$ is the standard deviation of the noise at frequency f. For white noise, $sd(|N(f)|=\sigma_n$, where σ_n^2 is the noise variance. Substitution of Equation (11.26) in Equation (11.24) yields

$$|\hat{X}(f)| = |Y(f)| - \left[1 + \frac{sd\left(|N(f)|\right)}{|N(f)|}\right]\overline{|N(f)|} \tag{11.27}$$

In Equation (11.27) the subtraction factor depends on the mean and the variance of the noise. Note that the amount over-subtracted is the standard deviation of the noise. This heuristic formula is appealing because at one extreme for deterministic noise with a zero variance, such as a sine wave, $\alpha(SNR(f))=1$, and at the other extreme for white noise $\alpha(SNR(f))=2$. In application of spectral subtraction to speech recognition, it is found that the best subtraction factor is usually between 1 and 2.

In the non-linear spectral subtraction method of Lockwood and Boudy, the spectral subtraction filter is obtained from

$$H(f) = \frac{\overline{|Y(f)|^2} - \overline{|N(f)|_{NL}^2}}{\overline{|Y(f)|^2}} \tag{11.28}$$

Lockwood and Boudy suggested the following function as a non-linear estimator of the noise spectrum:

$$\overline{|N(f)|^2}_{NL} = \Phi\left(\max_{\text{over } M \text{ frames}}\left(|N(f)|^2\right), SNR(f), \overline{|N(f)|^2}\right) \tag{11.29}$$

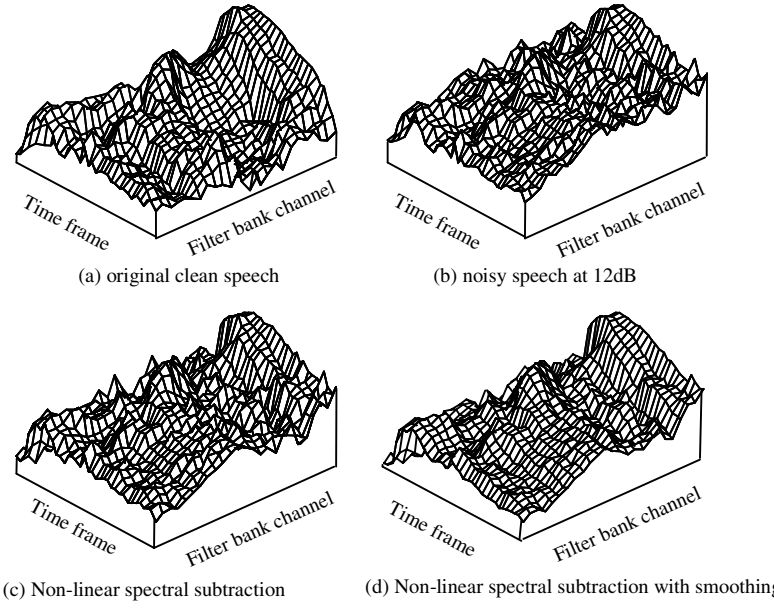


Figure 11.6 Illustration of the effects of non-linear spectral subtraction.

The estimate of the noise spectrum is a function of the maximum value of noise spectrum over M frames, and the signal-to-noise ratio. One form for the non-linear function $\Phi(\cdot)$ is given by the following equation:

$$\Phi\left(\max_{\text{over } M \text{ frames}}\left(|N(f)|^2\right), SNR(f)\right) = \frac{\max_{\text{Over } M \text{ frames}}\left(|N(f)|^2\right)}{1+\gamma\, SNR(f)} \tag{11.30}$$

where γ is a design parameter. From Equation (11.30) as the SNR decreases the output of the non-linear estimator $\Phi(\cdot)$ approaches $\max(|N(f)|^2)$, and as the SNR increases it approaches zero. For over-subtraction, the noise estimate is forced to be an over-estimation by using the following limiting function:

$$\overline{|N(f)|^2} \le \Phi\left(\max_{\text{over } M \text{ frames}}\left(|N(f)|^2\right), SNR(f), \overline{|N(f)|^2}\right) \le 3\overline{|N(f)|^2} \tag{11.31}$$

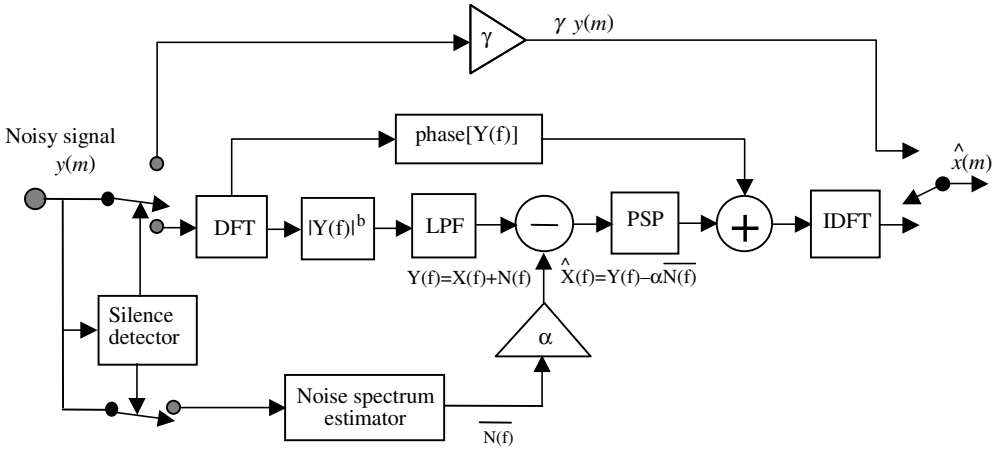


Figure 11.7 Block diagram configuration of a spectral subtraction system. PSP = post spectral subtraction processing.

The maximum attenuation of the spectral subtraction filter is limited to $H(f) \geq \beta$, where usually the lower bound $\beta \geq 0.01$. Figure 11.6 illustrates the effects of non-linear spectral subtraction and smoothing in restoration of the spectrum of a speech signal.

11.4 Implementation of Spectral Subtraction

Figure 11.7 is a block diagram illustration of a spectral subtraction system. It includes the following subsystems:

(a) a silence detector for detection of the periods of signal inactivity; the noise spectra is updated during these periods;
(b) a discrete Fourier transformer (DFT) for transforming the time domain signal to the frequency domain; the DFT is followed by a magnitude operator;
(c) a lowpass filter (LPF) for reducing the noise variance; the purpose of the LPF is to reduce the processing distortions due to noise variations;
(d) a post-processor for removing the processing distortions introduced by spectral subtraction.;
(e) an inverse discrete Fourier transform (IDFT) for transforming the processed signal to the time domain.
(f) an attenuator γ for attenuation of the noise during silent periods.

The DFT-based spectral subtraction is a block processing algorithm. The incoming audio signal is buffered and divided into overlapping blocks of N samples as shown in Figure 11.7. Each block is Hanning (or Hamming) windowed, and then transformed via a DFT to the frequency domain. After spectral subtraction, the magnitude spectrum is combined with the phase of the noisy signal, and transformed back to the time domain. Each signal block is then overlapped and added to the preceding and succeeding blocks to form the final output.

The choice of the block length for spectral analysis is a compromise between the conflicting requirements of the time resolution and the spectral resolution. Typically a block length of 5–50 milliseconds is used. At a sampling rate of say 20 kHz, this translates to a value for N in the range of 100–1000 samples. The frequency resolution of the spectrum is directly proportional to the number of samples, N. A larger value of N produces a better estimate of the spectrum. This is particularly true for the lower part of the frequency spectrum, since low-frequency components vary slowly with the time, and require a larger window for a stable estimate. The conflicting requirement is that, owing to the non-stationary nature of audio signals, the window length should not be too large, so that short-duration events are not obscured.

The main function of the window and the overlap operations (Figure 11.8) is to alleviate discontinuities at the endpoints of each output block. Although there are a number of useful windows with different frequency/time characteristics, in most implementations of the spectral subtraction, a Hanning window is used. In removing distortions introduced by spectral subtraction, the post-processor algorithm makes use of such information as the correlation of each frequency channel from one block to the next, and the durations of the signal events and the distortions. The

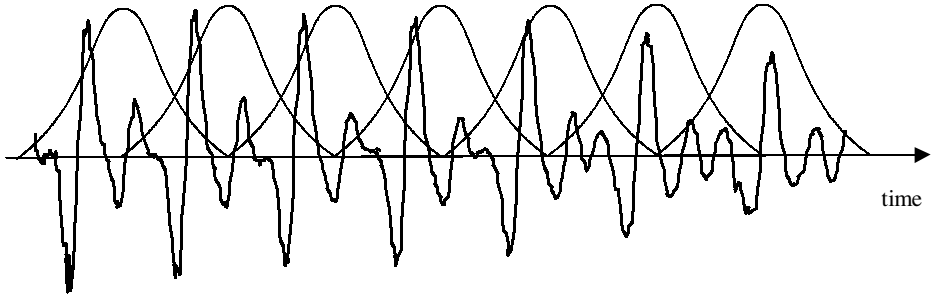


Figure 11.8 Illustration of the window and overlap process in spectral subtraction.

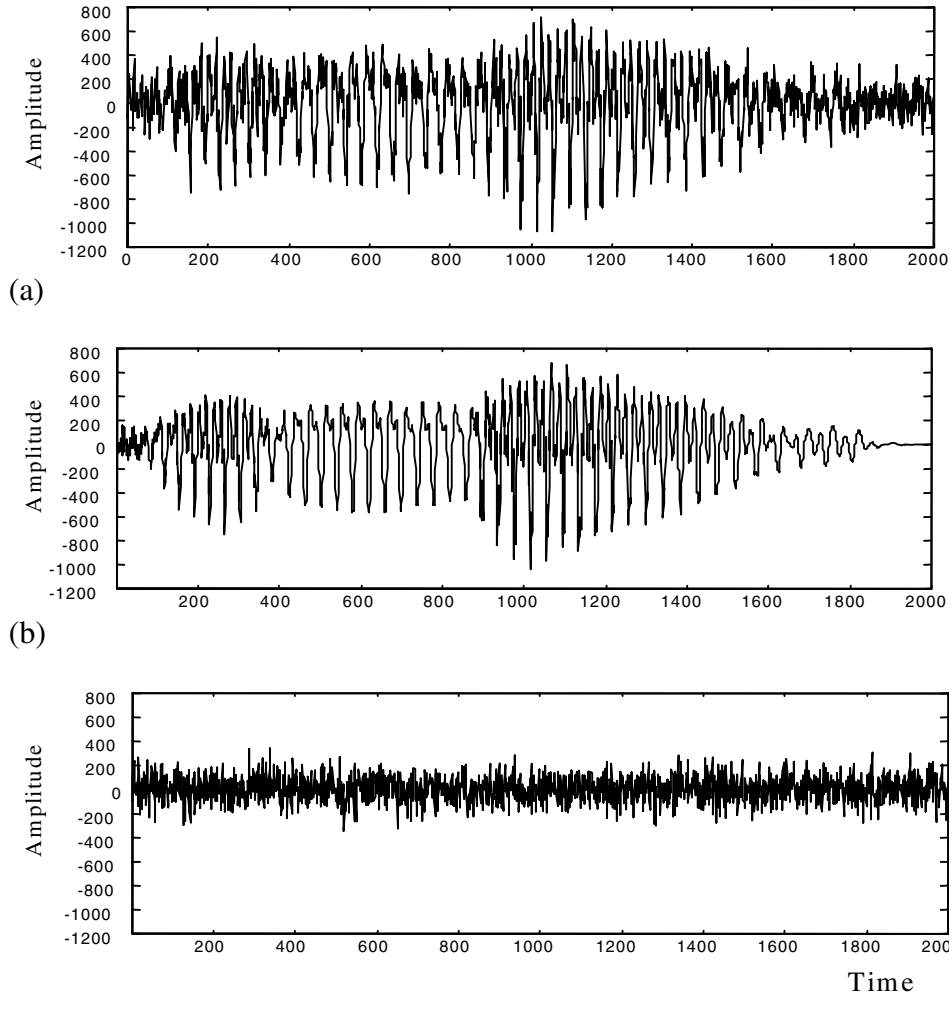


Figure 11.9 (a) A noisy signal. (b) Restored signal after spectral subtraction. (c) Noise estimate obtained by subtracting (b) from (a).

correlation of the signal spectral components, along the time dimension, can be partially controlled by the choice of the window length and the overlap. The correlation of spectral components along the time domain increases with decreasing window length and increasing overlap. However, increasing the overlap can also increase the correlation of noise frequencies along the time dimension.

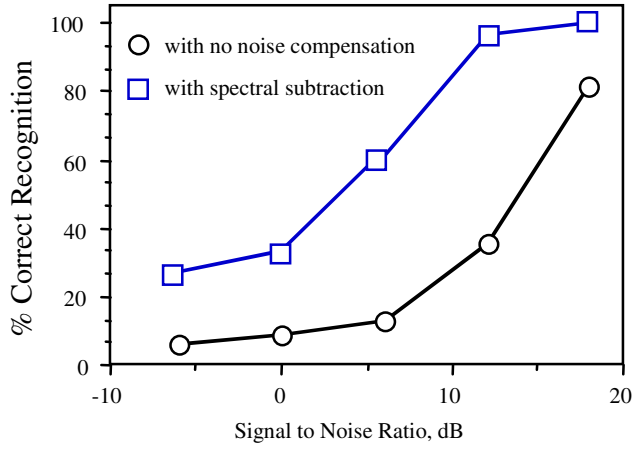


Figure 11.10 The effect of spectral subtraction in improving speech recognition (for a spoken digit data base) in the presence of helicopter noise.

11.4.1 Application to Speech Restoration and Recognition

In speech restoration, the objective is to estimate the instantaneous signal spectrum $X(f)$. The restored magnitude spectrum is combined with the phase of the noisy signal to form the restored speech signal. In contrast, speech recognition systems are more concerned with the restoration of the envelope of the short-time spectrum than the detailed structure of the spectrum. Averaged values, such as the envelope of a spectrum, can often be estimated with more accuracy than the instantaneous values. However, in speech recognition, as in signal restoration, the processing distortion due to the negative spectral estimates can cause substantial deterioration in performance. A careful implementation of spectral subtraction can result in a significant improvement in the recognition performance.

Figure 11.9 illustrates the effects of spectral subtraction in restoring a section of a speech signal contaminated with white noise. Figure 11.10 illustrates the improvement that can be obtained from application of spectral subtraction to recognition of noisy speech contaminated by a helicopter noise. The recognition results were obtained for a hidden Markov model-based spoken digit recognition.

11.5 Summary

This chapter began with an introduction to spectral subtraction and its relation to Wiener filters. The main attraction of spectral subtraction is its relative simplicity, in that it only requires an estimate of the noise power spectrum. However, this can also be viewed as a fundamental limitation in that spectral subtraction does not utilise the statistics and the distributions of the signal process. The main problem in spectral subtraction is the presence of processing distortions caused by the random variations of the noise. The estimates of the magnitude and power spectral variables, that owing to noise variations, are negative, have to be mapped into non-negative values. In Section 11.2, we considered the processing distortions, and illustrated the effects of rectification of negative estimates on the distribution of the signal spectrum. In Section 11.3, a number of non-linear variants of the spectral subtraction method were considered. In signal restoration and in applications of spectral subtraction to speech recognition it is found that over-subtraction, which is subtracting more than the average noise value, can lead to improved results; if a frequency component is immersed in noise then over-subtraction can cause further attenuation of the noise. A formula is proposed in which the over-subtraction factor is made dependent on the noise variance. As mentioned earlier, the fundamental problem with spectral subtraction is that it employs relatively too little prior information, and for this reason it is outperformed by Wiener filters and Bayesian statistical restoration methods.

Bibliography

BOLL S.F (1979) Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Tran. on Acoustics, Speech and Signal Processing **ASSP-27**, **2**, pp. 113–120.

BROUTI M., SCHWARTZ R. and MAKHOUL J. (1979) Enhancement of Speech Corrupted by Acoustic Noise. Proc. IEEE, Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-79, pp. 208–211.

CAPPE O. (1994) Elimination of Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor. IEEE Trans. Speech and Audio Processing, **2**, **2**, pp. 345–349.

CROZIER P.M. *et al* (1993) The Use of Linear Prediction and Spectral Scaling For Improving Speech Enhancement. EuroSpeech-93, pp. 231-234.

EPHRAIM Y. (1992) Statistical Model Based Speech Enhancement systems. Proc. IEEE, **80**, **10**, pp. 1526–1555.

EPHRAIM Y. and VAN TREES H.L. (1993) A Signal Subspace Approach for Speech Enhancement. Proc. IEEE, Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-93, pp. 355–58.

EPHRAIM Y. and MALAH D. (1984) Speech Enhancement Using a Minimum Mean-Square Error Short-Time Amplitude Estimator. IEEE Trans. Acoustics, Speech and Signal Processing. **ASSP-32**, **6**, pp. 1109–1121.

JUANG B.H. and RABINER L.R. (1987) Signal Restoration by Spectral Mapping. Proc. IEEE, Int. Conf. on Acoustics. Speech and Signal Processing, ICASSP-87 Texas.

KOBAYASHI T. *et al* (1993) Speech Recognition Under the Non-Stationary Noise Based on the Noise Hidden Markov Model and Spectral Subtraction. EuroSpeech-93, pp. 833–837.

LIM J.S. (1978) Evaluations of Correlation Subtraction Method for Enhancing Speech Degraded by Additive White Noise. IEEE Trans. Acoustics, Speech and Signal Processing, **ASSP-26**, **5**, pp. 471–472.

LINHARD K. and KLEMM H. (1997) Noise Reduction with Spectral Subtraction and Median Filtering for Suppression of Musical Tones. Proc. ECSA-NATO Workshop on Robust Speech Recognition, pp. 159–162.

LOCKWOOD P. and BOUDY J. (1992) Experiments with a Non-linear Spectral Subtractor (NSS) Hidden Markov Models and the Projection, for Robust Speech Recognition in Car, Speech Communications. Elsevier, pp. 215–228.

LOCKWOOD P. *et al* (1992) Non-Linear Spectral Subtraction and Hidden Markov Models for Robust Speech Recognition in Car Noise Environments. ICASSP-92, pp. 265–268.

MILNER B.P. (1995) Speech Recognition in Adverse Environments. Ph.D. Thesis, University of East Anglia, UK.

MCAULAY R.J. and MALPASS M.L. (1980) Speech Enhancement Using A Soft-Decision Noise Suppression Filter. IEEE Trans. **ASSP-28**, **2**, pp. 137–145, April.

NOLAZCO-FLORES J.A. and YOUNG S.J. (1994) Adapting a HMM-based Recogniser for Noisy Speech Enhanced by Spectral Subtraction. Proc. IEEE, Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP–94 Adelaide.

PORTER J.E. and BOLL S.F. (1984) Optimal Estimators for Spectral Restoration of Noisy Speech. Proc. IEEE, Int. Conf. on Acoustics. Speech and Signal Processing, ICASSP-84, pp. 18A.2.1–18A.2.4.

O'SHAUGHNESSY D. (1989) Enhancing Speech Degraded by Additive Noise or Interfering Speakers. IEEE Commun. Mag. pp. 46–52.

POLLAK P. *et al* (1993) Noise Suppression System For A Car. EuroSpeech-93, pp. 1073–1076.

SORENSON H.B. (1993) Robust Speaker Independent Speech Recognition Using Non-Linear Spectral Subtraction Based IMELDA. EuroSpeech-93, pp. 235–238.

SONDHI M.M., SCHMIDT C.E. and RABINER R. (1981) Improving the Quality of a Noisy Speech Signal. Bell Syst. Tech. J., **60**, **8**, pp. 1847–1859.

VAN COMPERNOLLE D. (1989) Noise Adaptation in a Hidden Markov Model Speech Recognition System. Computer Speech and Language, **3**, pp. 151–167.

VASEGHI S.V. and FRAYLING-CORCK R. (1993) Restoration of Archived Gramophone Records, Journal of Audio Engineering Society.

XIE F.(1993) Speech Enhancement by Non-Linear Spectral Estimation a Unifying Approach. EuroSpeech-93, pp. 617–620.

ZWICKER E. and FASTEL H. (1999) Psychoacoustics, Facts and Models, 2nd Ed. Springer.