The handsfree mode is convenient when using a mobile phone in a vehicle. This mode brings the microphone away from the talker’s mouth. As a consequence the speech signal picked up has low Signal to Noise Ratio. Thus, to improve the SNR a noise reduction method is employed. The proposed method is an improvement of the well-known spectral subtraction method which uses one SNR-dependent gain function to suppress the noise. The proposed method reduces the residual noise artifacts by using an adaptive averaging of the gain function. The adaptation is controlled by a spectral discrepancy measure, which is obtained by comparing the averaged noise spectrum and the current input signal spectrum. Experiments show a noise reduction of 10 dB with good noise quality and significantly low residual noise distortion. Sound results are presented at http://www.its.hk-r.se/research

1. INTRODUCTION

Mobile phones are commonly used today. People tend to use them in all surroundings, vehicles being no exception. Traffic safety authorities encourage people to use handsfree accessories, thus having both hands available for driving. Accordingly, this emphasizes the need for good sound characteristics even when the microphone is situated up to a meter from the talker’s mouth, a situation often resulting in a low SNR in the microphone signal. Hence, it is necessary to apply a noise reduction procedure. The suggested method is an extension, or a further development of the conventional spectral subtraction algorithm [1].

Spectral subtraction employs estimates of the noise spectrum and the noisy speech spectrum to form one SNR-based gain function. The gain function is multiplied by the input signal spectrum and will suppress frequencies with low SNR. The main disadvantage in using conventional spectral subtraction algorithms is the resulting “musical tones” which disturb not only the listener but also speech coding algorithms. The proposed algorithm reduces this disturbance to a low level by using an adaptive averaging of the gain function.

The adaptive exponential averaging of the gain function is dependent upon the discrepancy between the current signal spectrum estimate and the noise spectrum estimate. The algorithm is applied such that when the input signal is stationary, thus the discrepancy low, the gain function is heavily averaged resulting in low amplitude musical tones. A reduced averaging occurs when a high energy speech spectrum is present, since the discrepancy becomes large. Even though the averaging is reduced, the high energy speech will mask the musical tones so that good sound quality remains. One of the advantages of using this method is that the high variance of the input signal spectrum estimate results in a gain function with a reduced variance when required. Another known method to remove the musical tones is to use an over-subtraction factor, i.e. subtract a factor greater than one [2]. This method has the disadvantage of degrading the speech when musical tones are sufficiently removed. The currently proposed method removes the musical tones to such a degree that it is possible to use under-subtraction of the noise spectrum which in turn results in significantly less speech degradation.

2. SPECTRAL SUBTRACTION

Spectral subtraction assumes that the noise signal magnitude spectrum is almost constant during noisy speech and noise periods. Thus, an estimate of the noise magnitude spectrum is also valid during a noisy speech period. The noise is further considered to be additive to the speech signal. Let \( s(n) \), \( w(n) \) and \( x(n) \) represent speech, noise, and noisy speech signals, respectively, so that

\[
x(n) = s(n) + w(n)
\]

(1)

The corresponding power spectral density is defined as

\[
R_x(f) = R_s(f) + R_w(f)
\]

(2)

where \( f \in [0, N - 1] \) is a discrete variable corresponding to one frequency bin, and \( R_s(f) \) denotes the power spectral density of the signal. The short-time spectral density is estimated by using a periodogram

\[
R_{x,N}(f, i) = \frac{1}{N} \left| \mathcal{F}\{x_N^i(i)\} \right|^2
\]

(3)
where \( x_N(i) \) is the \( i \)th block of \( N \) data samples, \( w \) is a Hanning window, and \( F \) is an FFT. For convenience, the magnitude spectrum is defined as \( P_{x,N}(f,i) = |X_{N}(f,i)| \). The short-time noise magnitude spectrum can be estimated during speech pauses by

\[
P_{w,N}(f,i) = \frac{\mu P_{w,N}(f,i-1) + (1 - \mu) P_{x,N}(f,i)_{\text{noise}}}{P_{w,N}(f,i-1)}
\]

where \( \mu \) is the exponential averaging time constant.

The spectral subtraction operation corresponds to a time varying filtering, such that

\[
Y_N(f,i) = G_N(f,i)X_N(f,i)
\]

over a block of samples where a capital letter denotes the FFT of the \( i \)th block. The expression for the SNR-dependent gain function is given by

\[
G_N(f,i) = \left( 1 - k \cdot \frac{P_{w,N}(f,i)}{P_{x,N}(f,i)} \right)^{\frac{1}{\alpha}}
\]

where \( k \) is the subtraction factor and \( \alpha \) determines whether magnitude or power spectral subtraction should be used. The combination of \( k \) and \( \alpha \) controls the amount of noise reduction. The noise reduced spectrum, \( Y_N(f,i) \), is transformed to the time-domain. Consecutive time-blocks are overlapped to compensate for the previously applied Hanning window.

### 3. Adaptive Averaging of Gain Function

In conventional spectral subtraction the gain function varies highly between blocks, giving rise to unnatural background sound artifacts known as "musical tones." The artifacts in the output signal are due to high variance in the spectrum estimates used in the calculation of the gain function, \( G_N(f,i) \), which consequently is a highly varying function. The variations of the gain function can be decreased by using an adaptive exponential averaging of the this function, such that

\[
\bar{G}_N(f,i) = (1 - \beta(i)) \cdot \bar{G}_N(f,i-1) + \beta(i) \cdot G_N(f,i)
\]

where \( \beta(i) \) is an averaging time parameter derived from the spectral discrepancy measure. Furthermore, the spectral discrepancy measure for block \( i \) is defined as

\[
\beta(i) = \min \left\{ \frac{1}{N-1} \sum_{j=0}^{N-1} |P_{w,N}(f,j) - P_{w,N}(f,i)|, 1 \right\}
\]

A small discrepancy should yield a longer averaging time of the gain function, \( G_N(f,i) \), thus a smaller \( \beta(i) \). This corresponds to a stationary background noise situation. A large discrepancy should result in a shorter averaging time, or zero averaging of the gain function, \( G_N(f,i) \), and thus a \( \beta(i) \) close to 1. Consequently, a situation exists where speech or highly varying background noise is present. Figure 1 illustrates how the controlled exponential averaging is included in the total algorithm.

![Figure 1](image_url)

Figure 1: (a) Total structure of the new spectral subtraction algorithm. (b) A detailed view of the gain function in (a), consisting of the parts facilitating the adaptive averaging controlled by the spectral discrepancy.

Precaution must be taken when the input signal goes from a noisy speech period to a noise period, where the discrepancy measure, \( \beta(i) \), decreases so that the adaptive averaging of the gain function, \( G_N(f,i) \), increases according to equation (7). However a direct increase of the averaging time would result in an audible "shadow voice," since the gain function suited for a speech spectrum would linger for a long period. Accordingly, the av-
eraging time should only be allowed to increase slowly, allowing the gain function to adapt to the noise input. Thus, by averaging the discrepancy measure, $\beta(i)$, and using $\beta(i)$ in equation (7), the shadow voice is reduced. This introduced averaging time parameter, $\beta(i)$, is described by

$$\beta(i) = \gamma \cdot \beta(i - 1) + (1 - \gamma) \cdot \beta(i)$$

(9)

where

$$\gamma = \begin{cases} 0, & \beta(i - 1) < \beta(i) \\ \gamma_c, & \beta(i - 1) \geq \beta(i), \quad 0 < \gamma_c < 1 \end{cases}$$

(10)

An increase in spectral discrepancy, $i(i)$, increases the averaging time directly. In a situation where the discrepancy decreases, an exponential averaging - with the time constant $\gamma_c$ - is used. Consequently, we use $\beta(i)$ in equation (7).

Figure 2 presents the spectral discrepancy, $\beta(i)$, and the averaged spectral discrepancy, $\beta(i)$. The figure shows that the averaged spectral discrepancy decreases more slowly and thus will reduce the shadow voices.

Figure 3 presents one frequency bin of the gain function using averaging. As seen in the figure, the variability of the gain function is lower during noise periods and also for low energy speech periods, when the adaptive averaging is employed. When employing $\beta(i)$ the method differs after high energy speech periods by having a lower gain compared to that when $\beta(i)$ is used.

![Figure 2: Parameter $\beta(i)$, $\beta(i)$ vs. block $i$. Solid line: $\beta(i)$; dashed line: $\beta(i)$.](image)

![Figure 3: Frequency bin 56 of the gain function vs. block $i$ is shown. Solid line: using $\beta(i)$; dashed line: using $\beta(i)$; dotted line: using no averaging of the gain function. The parameter $\gamma_c = 0.8$. Right after a high energy speech block, e.g. $i = 170-180$, the gain function obtained using $\beta(i)$ or $\beta(i)$ differs, and since $\beta(i)$ will be larger there will be less averaging of the gain function. Same signal sequence used as in Figure 2.](image)

The new variability reduction of $G_N(f, i)$ can be interpreted for different input signal conditions as follows. During noise periods the gain function can be averaged to decrease the variance, as long as the noise spectrum has a steady mean value for each frequency. Noise level changes will create an increased spectral discrepancy. The adaptive exponential averaging method will thus decrease the gain function averaging until the noise level has settled to a new level. This behavior enables prompt response to noise level changes and gives a decrease in variance during stationary noise periods. High energy speech often exhibits time-varying spectral peaks. For speech signals a high energy spectrum is often very short-time stationary. Thus, the exponential averaging should be kept to a minimum during high energy speech periods. Since the spectral discrepancy is large during speech, no exponential averaging of the gain function is performed. The musical tones would then increase but since the high energy speech masks them the output will sound good.

4. RESULTS

The results presented are based on experiments with real-world recorded input signals. The signals used in the experiments were combined using separate recordings of speech and noise recorded in a car. The speech recording was performed in a quiet stationary car using a handsfree microphone, a telephone bandwidth filter and an 8 kHz sampling rate. The noise sequences were recorded using the same equipment in a car driving at a highway speed of 110 km/h.

The inputs and results are presented as sound files at our web-site: [http://www.its.hk-r.se/research](http://www.its.hk-r.se/research).

The frame length, $N$, is chosen as a power of 2 to enable Fast Fourier Transformation, and also to correspond to the assumed maximum time for stationarity of speech signals, i.e. 10–30 ms, $N = 256$. The amount of noise reduction is controlled by the parameters $a$ and $k$. A noise reduction with improvement in sound quality can be achieved with $a = 1$ and $k = 0.8$.

The parameter choices favor good sound quality preferably to large noise reduction. Typically, the proposed method yields a noise reduction in the vicinity of 10 dB,
as can be seen in Figures 4 and 5 where the input and output signal magnitudes are shown, in dB.

Figure 4: Input signal magnitude $|x(n)|$.

Figure 5: Noise reduced output signal magnitude $|y(n)|$.

The exponential averaging of the gain function yields lower variance when the signal is stationary. The main advantage is the reduction of musical tones and residual noise. The gain function with and without exponential averaging is presented in Figures 6 and 7. As can be seen in the figures, the variability of the gain function is lower during noise periods and also for low energy speech periods, when exponential averaging is employed. The lower variability of the gain function results in less noticeable tonal artifacts in the output signal.

5. CONCLUSIONS

A further developed spectral subtraction algorithm has been presented, incorporating a reduced variance gain function. The proposed gain function gives less artificial residual noise in the processed signal when compared with conventional spectral subtraction. A spectral discrepancy measure is used to control the adaptive averaging. The method is successful since it estimates both the level of stationarity of the present input signal and the time-point when high energy speech components may mask the musical tones. Experiments show that a 10 dB noise reduction with improved sound quality is achieved using signals recorded in a normal passenger vehicle driving at a highway speed of 110 km/h. These results offer the possibility of using noise reduction in handsfree mobile phones.

Figure 6: Absolute value of the gain function, $G_N(f,i)$, with the exponential averaging "on."

Figure 7: Absolute value of the gain function, $G_N(f,i)$, with the exponential averaging "off."

50 100 150 200

20 40 60 80

100 120

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

frequency bin $f$

block $i$

$G_N$ level

6. REFERENCES
