ON THE ESTIMATION OF THE SHORT-TIME PHASE IN SPEECH ENHANCEMENT SYSTEMS
Y. Ephraim and D. Malah
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

Abstract

The idea of improving speech enhancement, by combining a better estimate of the short-time phase than the commonly used noisy phase, with an independently derived spectral amplitude estimator, is examined. It is shown under the MMSE criteria and Gaussian assumptions, that improving the estimation of the complex exponential of the phase, affects the spectral amplitude estimation. On the other hand, the optimal complex exponential estimator which does not affect the spectral amplitude estimation, is the complex exponential of the noisy phase. Better results in enhancing speech were obtained when the complex exponential of the noisy phase, rather than the optimal complex exponential estimator, is combined with an optimum amplitude estimator.

I. Introduction

Speech enhancement systems which capitalize on the major importance of the short-time spectral amplitude in speech perception, are known to be the most successful ones [1]. In these systems, the short-time spectral amplitude of the speech signal is estimated, and combined with the short-time phase of the degraded speech, for constructing the enhanced speech signal. The "spectral subtraction" algorithm [1], is a well known example in this class of speech enhancement systems. Another example is the recently developed algorithm which utilizes an optimal minimum mean square error (MMSE) short-time spectral amplitude estimator [2,3].

The use of the noisy phase as an estimate of the speech short-time phase, follows from the relative unimportance of the latter in the perception of speech signals. Recently [4,5], a question has been raised, whether a further improvement in speech enhancement could be achieved, if a better estimate than the noisy phase is employed. In [4,5], an experimental approach is taken to answer this question, but the estimation problem has been ignored. The "better estimate" of the short-time phase is achieved there, from a higher signal to noise ratio (SNR) speech signal, than the degraded speech to be enhanced.

In this paper we address the above problem, by examining the estimation of the spectral amplitude and phase, under the MMSE criterion and Gaussian assumption. Rather than looking for an estimator of the short-time phase, we derive an estimator of its complex exponential, which actually is needed in the construction of the enhanced signal.

II. Optimal Complex Exponential Estimator

The estimation problem of the complex exponential of the short-time phase, is formulated similarly to the estimation problem of the short-time spectral amplitude in [3], and is based on the same assumptions. Specifically, it is the problem of estimating the complex exponential of the phase, of each Fourier expansion coefficient of the signal \( z(t), 0 \leq t \leq T \), given the degraded signal \( y(t) = z(t) + d(t) \), where the signal \( z(t) \) and the noise \( d(t) \) are assumed to be quasistationary uncorrelated zero mean random processes.

The Fourier expansion coefficients of the speech signal, as well as of the noise process, are modeled as statistically independent Gaussian random variables. The Gaussian model is commonly used when the estimation is done in the frequency domain, and is motivated by the Central Limit Theorem.

Derivation of Optimal Estimator

Let \( X_k \triangleq A_k e^{j\phi_k}, A_k, \) and \( Y_k \triangleq R_{k} e^{j\phi_k}, \) denote the \( k \)-th Fourier expansion coefficient of the speech signal, the noise process, and the noisy observations, respectively. Under the above statistical model, the estimation problem can be reduced to that of estimating \( e^{j\phi_k} \) from the spectral components \( Y_k \) of the noisy observations \( y_k(t), 0 \leq t \leq T \) [6, Appendix B]. Based on this observation and the statistical independence assumption, the optimal MMSE estimator of \( e^{j\phi_k} \) given \( Y_k \) is given.

\[ 
\phi_{\text{opt}} = \arg \min \left\{ E[|e^{j\phi_k} - \phi_{\text{est}}|^2] \right\}
\]

\[ 
\hat{\phi}_{\text{opt}} = \arg \min \left\{ E[|e^{j\phi_k} - \phi_{\text{est}}|^2] \right\}
\]

The research was supported by the Technion V.P.R. Fund & Neilkim Fund for Electrical Engineering, Research.
by:
\[
\hat{\sigma}^2_k = E[e^{j\theta_k} | Y_k] \\
= E[e^{j\theta_k} | Y_k] \\
= E[e^{-j\theta_k} Y_k] e^{j\theta_k} \\
\hat{A} \cos \hat{\phi}_k - j \sin \hat{\phi}_k e^{j\theta_k}
\]
where \( \theta_k \) is the phase error which is defined by \( \theta_k = \hat{\theta}_k - \theta_k \), and \( \hat{\phi}_k \) is the noisy phase. 
\( \cos \hat{\phi}_k \hat{A} = E[\cos \phi_k | Y_k] \) and similarly \( \sin \hat{\phi}_k \hat{A} = E[\sin \phi_k | Y_k] \). \( \cos \hat{\phi}_k \hat{A} \) and \( \sin \hat{\phi}_k \hat{A} \) can be calculated, once the conditional probability density function (PDF) \( p(\hat{A}_k | Y_k) \) is determined. By the Bayes Theorem, \( p(\hat{A}_k | Y_k) \) is given by:
\[
p(\hat{A}_k | Y_k) = \frac{\int p(Y_k | \hat{A}_k, \theta_k) p(\hat{A}_k, \theta_k) d\hat{A}_k}{\int p(Y_k | \hat{A}_k, \theta_k) p(\hat{A}_k, \theta_k) d\hat{A}_k d\theta_k}
\]
On the basis of the Gaussian assumptions we made, \( Y_k | \hat{A}_k, \theta_k \) and \( p(\hat{A}_k, \theta_k) \) are given by:
\[
p(Y_k | \hat{A}_k, \theta_k) = \frac{1}{\pi \lambda_4(k)} \exp \left\{ -\frac{1}{\lambda_4(k)} |Y_k - \hat{A}_k e^{j\theta_k}|^2 \right\}
\]
\[
p(\hat{A}_k, \theta_k) = \frac{\lambda_3(k) \hat{A}}{\pi \lambda_4(k)} \exp \left\{ -\frac{\lambda_3(k) \hat{A}^2}{\lambda_4(k)} \right\}
\]
where \( \lambda_4(k) = E[|Y_k|^2] \), and \( \lambda_3(k) \hat{A} = E[|A|^2] \). From (1-4) we obtain:
\[
\hat{\phi}_k = 0
\]
and
\[
\hat{\sigma}^2_k = \cos \hat{\phi}_k e^{j\theta_k}
\]
where \( \Gamma(1.5) \) is the Gamma function with \( \Gamma(1.5) = \sqrt{\pi}/2 \), \( l_1(\cdot) \) and \( l_4(\cdot) \) are the modified Bessel functions of zero and first order respectively; \( \nu_k \) is defined by:
\[
\nu_k = \frac{\lambda_3(k)}{\lambda_4(k)}
\]
where \( \xi_k \) and \( \gamma_k \) are defined by:
\[
\xi_k = \frac{\xi_4(k)}{\lambda_4(k)}
\]
\[
y_k = \frac{\xi_4(k)}{\lambda_4(k)}
\]
now that the resulting estimator is nearly equivalent to the "Wiener spectral estimator" \( \hat{X}_k \), which is given by [2,3]:
\[
\hat{X}_k = \frac{\nu_k}{\nu_k^2 + 1} \hat{X}_k
\]
On the one hand, this fact implies that combining the optimal complex exponential estimator, rather than the complex exponential of the noisy phase, with the optimal amplitude estimator, improves the signal waveform estimation. On the other hand, it enables to estimate the degradation in the spectral amplitude estimation, by using the error analysis presented in [3]. The above conclusions are based on the fact that the Wiener estimator is the optimal MMSE estimator of the signal waveform, but not of its spectral amplitude.

To show that \( \hat{X}_k \) and \( \hat{X}_k \) are nearly equivalent, we need to examine the modulus of each of the two estimators only. To do so, we substitute \( \cos \hat{\phi}_k \) from (6), and \( \hat{A}_k \) which is given by [2,3]:
\[
\hat{A}_k = l_1(\nu_k) \sqrt{\nu_k/\xi_4(k)} (1 + \nu_k l_2(\nu_k/2) + \nu_k l_4(\nu_k/2)) \hat{A}_k
\]
into (10). \( |\hat{X}_k| \) can then be written as:
\[
|\hat{X}_k| = G(\xi_k, \gamma_k) \hat{A}_k
\]
where \( G(\xi_k, \gamma_k) \) is interpreted as a multiplicative non-linear gain function, which depends on the a-priori and a-posteriori SNR, \( \xi_k \) and \( \gamma_k \), respectively. \( |\hat{X}_k| \) can also be described in a similar way, where from (11) we obtain:
\[
G(\xi_k, \gamma_k) = \frac{\xi_k}{\xi_k^2 + 1}
\]
This representation enables a convenient comparison between \( |\hat{X}_k| \) and \( |\hat{X}_k| \), by comparing their corresponding gain functions. Fig. 1 describes two sets of parametric gain curves, which result from (13) and (14). The similarity of the gain curves in each pair, corresponding to the same value of \( \xi_k \), implies that the two estimators \( \hat{X}_k \) and \( \hat{X}_k \) are nearly equivalent.

Due to the major importance of the spectral amplitude in speech perception, it is of interest to derive an optimal estimator of the complex exponential of the phase, which does not affect the amplitude estimation.

To derive the above estimator, which we denote by \( \hat{a}^2_k \), the following constrained optimization problem should be solved:
\[
\min_{\hat{a}^2_k} E[|a_{k} e^{j\hat{\theta}_k} - \hat{a}^2_k|^2]
\]
subject to \( |\hat{a}^2_k| = 1 \)
Using the Lagrange multipliers method, we got:
\[
\hat{a}^2_k = e^{j\theta_k}
\]
That is, the complex exponential of the noisy phase, is the optimal estimator which does not affect the amplitude estimation.

The combination of the complex exponential estimator (19) with an amplitude estimator \( \hat{A}_k \), results in the following estimator \( \hat{X}_k \), for the k-th spectral component:
\[
\hat{X}_k = \hat{A}_k e^{j\hat{\theta}_k}
\]
Optimal Phase Estimator

By passing, it is of interest to derive the optimal estimator of the principle value of the phase. Its complex exponential provides an estimator which does not affect the amplitude estimation, although obviously, it cannot be a better estimator than (10). However, it may result in a better estimator for the principle value of the phase, than the noisy phase. Since it is unknown which one, the phase or its complex exponential, is more important in speech perception, the optimal estimators of both of them should be examined.

We derive the optimal estimator of the principle value of the phase, \( \tilde{\alpha}_k \), by minimizing the following criterion [7]:

\[
E[\{1 - \cos(\alpha_k - \tilde{\alpha}_k)\}] \tag{16}
\]

This criterion is invariant under a modulo 2\pi transformation of the phase \( \alpha_k \), the estimated phase \( \tilde{\alpha}_k \), and the estimation error \( \alpha_k - \tilde{\alpha}_k \). For small estimation errors, \( \alpha_k - \tilde{\alpha}_k \), \( \alpha_k \) is a type of least squares criterion, since \( 1 - \cos \theta = \theta^2 / 2 \), for \( \theta \ll 1 \).

The optimal estimator \( \tilde{\alpha}_k \) which minimize (18) is easily shown to satisfy:

\[
tg \tilde{\alpha}_k = \frac{E[\sin \alpha_k | Y_k]}{E[\cos \alpha_k | Y_k]} \tag{19}
\]

By using \( \alpha_k = \phi_k - \tilde{\phi}_k \), and \( \sin \phi_k = 0 \) (see (5)), it is easy to see that:

\[
E[\sin \alpha_k | Y_k] = \sin \tilde{\phi}_k \cos \phi_k \tag{20}
\]

\[
E[\cos \alpha_k | Y_k] = \cos \tilde{\phi}_k \cos \phi_k \tag{21}
\]

By substituting (20) and (21) into (19) we get:

\[
tg \tilde{\alpha}_k = tg \tilde{\phi}_k \tag{22}
\]

or alternatively, the principle values of the phases are equal, i.e.,

\[
\tilde{\alpha}_k = \tilde{\phi}_k \tag{23}
\]

We see that the complex exponential of the optimal phase estimator, is in fact the estimator (16).

IV. Summary and Conclusions

In this paper we examine the idea of improving speech enhancement results by using a better estimate of the speech short-time phase than the commonly used noisy phase.

We address the problem by both theoretical and experimental approaches. It is shown under the MMSE criterion and Gaussian assumption, that improving the estimation of the complex exponential of the phase, by using its optimal estimator, affects the spectral amplitude estimation. On the other hand, the optimal estimator of the complex exponential of the phase, which does not affect the spectral amplitude estimation, is the complex exponential of the noisy phase.

We also show that combining the optimal estimator of the complex exponential of the phase, with the optimal amplitude estimator, results in a spectral estimator which is nearly equivalent to the Wiener estimator. This fact implies that using the optimal complex exponential estimator, rather than the complex exponential of the noisy phase, improves the signal wavefrom estimation, but degrades the spectral amplitude estimation as well [2, 3].

The two spectral estimators, which utilize the optimal amplitude estimator, but differ in the complex exponential estimator, are examined in speech enhancement. As judged by informal listening, the spectral estimator which utilizes the complex exponential of the noisy phase, performs better than the one which utilizes the optimal complex exponential estimator.

Acknowledgement

The authors wish to thank Mr. A. Schatzberg and Mr. S. Shitz for their critical reading of the manuscript and their helpful comments.

References


Figure Captions

Fig. 1: Parametric gain curves of the gain functions:

(a) \( \hat{G}(\tau_k, \gamma_k) \) (full lines).

(b) \( G_k(\tau_k, \gamma_k) \) (dashed lines).

![Graph showing parametric gain curves](image-url)