

Wiener filters for the reconditioning of old speech recordings under uncertainty

Dedicated to the anniversary of 100 years from the Grand Union of Romania

Horia-Nicolai L. Teodorescu

¹ ‘Gheorghe Asachi’ Technical University of Iasi, Iasi, Romania

² Romanian Academy, Iasi Branch, Iasi, Romania

Abstract – The article revisits the design of optimal Wiener filters under uncertainty, specifically under limited information and imprecision of the available data on the noise and signal spectra, as in case of old speech recordings. The derivation of the filters is based on a few reasonable assumptions regarding the noise measurement and on the variability of the noise. The filters use several properties of the speech signal in the choice of the filter structure and optimization procedure.

Keywords- *optimality, imprecision, limited information, old speech recordings*

I. INTRODUCTION

The problem solved is the design of a modified, adaptive Wiener filter for the case when there is uncertain knowledge of a specified number of spectral components of the signal and of the noise. The main contributions are (i) to propose a short-memory type of adaptive filter, designed for taking into account the stated uncertainty; (ii) to reduce uncertainty by averaging over the recent noise spectra; (iii) to predict the change in the signal spectrum; (iv) to separate the signal in two frequency sub-bands such that to allow the use of windows of different widths for determining the fast varying spectra, and (v) the use the predicted noise spectrum variation for correcting the filter initial design. The proposed filter exploits several specificities of the voice signal and of other biological signals for better determining approximations of the non-stationary speech/signal and noise instantaneous spectra. These features of the speech include the presence of pauses in the signal, the faster variability of the speech spectrum than that of the noise (different time scales of non-stationarity) and the great spectral difference between fricative consonants and the other consonants. The filters were conceived primarily for recovering and enhancements of old speech recordings from the period around WWI (about 1913-1921).

While the focus is on filters for speech, we propose a treatment of the general class of filtering problems where the uncertainty and the limited knowledge play a role in the design of Wiener filters, but information on the noise and signal spectra can be obtained with approximation at certain time moments. Such cases include speech signals and certain biological signals, including ECG signals. We also show how to describe the uncertainty in the discussed problem in terms of fuzzy rules.

The layout of the paper is as follows. Section 2 is a review of several progresses in optimal filtering. In Section 3, we describe the general setting of the

discussion. The frequencies sampling of the spectra of the noise and of the noisy signals are analyzed, specifying the uncertainty and errors assumed in the problem and the specifics of recorded speech filtering. In Section 4, we introduce the design, including interpolation aspects, for the filter. The last section summarizes the conclusions.

II. LITERATURE REVIEW

The optimality of the Wiener filter is limited to the case of precisely known power spectra densities for the signal and noise. Extensions to various cases of uncertainty in the system and signals, with bounded uncertainty, have been dealt with in numerous studies, e.g., by Kassam and Lim [1], Poor [2], Vastola and Poor [3], and more recently by Petersen and McFarlane [4], Shen and Deng [5], Chen et al. [6], and Corrêa and Sales [7]. In [2], Poor used a minimax approach for robust Wiener filters under uncertainty on signal and noise spectra; then, he compared the results with the minmax criterion with those obtained for the least-favorable spectral pairs.

Vastola and Poor have shown in [3] that the performance of Wiener filters under spectral uncertainty is poor even under “small deviations from those signal and noise spectral densities which were used to design the filter” and concluded that the robust filter in the sense of the design for best possible performance under worst uncertainty works better than the basic Wiener filter in most cases. To a similar conclusion arrived Kassam and Lim [1] who dealt with robust Wiener filters for the case when the signal and noise power spectral densities (PSDs) are known only approximately. In a series of papers [8-11], Zadeh made a critical analysis of the optimality concept, including Wiener optimal filtering, and studied various optimality problems. Years later, he also suggested the use of fuzzy logic as a tool to deal with both optimality and uncertainty [12]. He has not, however, provided more than an intuitive discussion of the subject of using fuzzy logic and fuzzy systems in optimality.

Shen and Deng [5] proposed a H_∞ filter for speech signals; their design uses no a priori knowledge of the additive noise statistics. The design is based on the criterion of minimizing the worst error signals (including modelling errors and noise). In their approach, they consider the typical speech generation model based on a source signal w_k and an autoregressive model $x_k = \sum_{j=1}^M a_j x_{k-j} + w_k$ [5] plus an additive noise v_k . As in other applications related to speech, these authors [5] recommend performing a

segmentation of the speech, “*In speech enhancement, the noisy speech is usually divided into number of frames and the length of each frame is within 10 to 30 ms.*” While departing from their model, we will use a similar initial strategy, based on the observation that the speech spectrum changes from one phoneme to the next, requiring adaptation on short time frames.

Chen et al. [6], addressing speech signals, stress that noise is very much dependent on the application and that, in most cases, it is non-stationary. This is in fact the situation with other signals, such as bio-electric signals. Their solution takes into account the number of channels (microphones) used; they assume additive noise independent of the speech (no echo or reverberation). The hypotheses of additive, independent noise are also used in this paper. In a similar line, Gibson et al. [13] analyzed a version of optimal filtering aimed for speech signals in high level, colored noise.

While there are numerous software packages for ‘cleaning’ audio files, many of them leave to the user the (manual) choice of the filter and only a few (e.g. Audacity, <http://www.instructables.com/id/How-to-Improve-Vocal-Quality-in-Audacity/>) suggest that they use of optimal filtering.

III. PROBLEM SETTING AND ANALYSIS

The hypotheses of the design and the idea flow in this paper are as follows.

In many applications, one searches an adaptive solution for the filter, which is optimal or sub-optimal after adaptation. We are interested in the design problem for adaptive Wiener optimal filters that should be adapted to a specified signal, typically speech, under noise conditions that are partly and imprecisely known at the adaptation stage. The noise is assumed to be additive; the signal and the noise are assumed uncorrelated.

A key issue of the design is that it exploits the different time scales of the non-stationarity for speech and noise; precisely, noise in speech recordings varies much slower than the speech signal, which allows one to determine the noise spectrum at sparse moments, while the signal spectrum is determined at small intervals, compatible with the phoneme duration.

We assume that the signal has brief pauses that can be identified. Pause detectors are well-known in speech processing [14, 15]; such detectors are included, for example, in cellular phones. During the pauses, the noise values can be measured and the noise spectra at time moments n' are estimated at the frequencies f_j . The noise and signal spectra are determined for the set of frequencies $\{f_j\}_{j=1,\dots,m}$. Once the partial (in terms of number of frequencies in the spectrum) and imperfect, imprecise (in terms of lack of values for the noise spectral density at the given frequencies at the moment of adaptation) noise and noisy signal spectra are collected, one may proceed to design a filter that is optimal, in the limits of the collected knowledge. Notice that the sampling of the noise is largely random; While the random sampling has disadvantages, it is largely used in classification, when regular sampling is not possible (see [16] Broumand).

Throughout the paper, the following notations are used: $s(t), v(t)$ signal and noise as function of time t ; $f_j, j = 1, \dots, n$ frequencies where the noise is known with an assumed maximal error, ϵ_j ; $S_s^2(\omega)$ is the power spectral density of the noiseless signal, unknown; $S_v^2(\omega)$ true power spectral density of the noise, unknown; $S_{j,s}^2(n), S_{j,v}^2(n)$ spectral components j of the true signal and of the noise, at time moment t_n ; unknown, estimated; $S_{j,v}^2(n')$ spectral components j of the noise, measured at time moment $t_{n'}$, where $t_{n'}$ is during a signal pause; $\epsilon(\omega)$ is the error in the determination or estimation of the noise at ω .

IV. PROPOSED FILTERS

Wiener optimal filters are linear filters and are defined by the transfer function [17]

$$H^2(\omega) = \frac{S_s^2(\omega)}{S_s^2(\omega) + S_v^2(\omega)}, \quad (1)$$

where $S_s^2(\omega), S_v^2(\omega)$ are assumed stationary and known in Wiener’s original treatment; moreover, the signal and the noise are assumed independent, and the noise additive $s'(t) = s(t) + v(t)$. Wiener optimal filter is not necessarily realizable with standard circuits. However, this aspect is disregarded here, because the filtering is performed on recorded signals, not in real time.

The concept of the proposed adaptive filter is as follows. Consider the Wiener optimal filters with the knowledge of noise spectrum limited to a set of frequencies $f_j, j = 1 \dots n$, and affected by imprecision in the sense explained in the subsequent paragraphs. First, sample the frequency space and consider using one optimal filter for each band $[f_j, f_{j+1}]$.

Use a pause detector and determine at each frequency, f_j , during pauses, the noise component power (i.e., determine the power spectrum of the noise, e.g., using narrow bandpass filters or FFT). The sampling of the noise is irregular because the occurrence of the pauses is random. For speech signals, the pauses may occur during words, that is at intervals of several tens or even hundreds of ms. When the pauses are small, several pause intervals can be concatenated to better characterize the noise.

Next, the power spectrum component at f_j is determined, for the signal plus noise, $(S_s^2 + S_v^2)(\omega_j) = S_{j,s}^2 + S_{j,v}^2$, during the signal presence intervals. Then, one can subtract for each f_j the noise from the signal plus noise to estimate the signal spectrum component $S_{j,s}^2$ at f_j . With these values, approximate the noise and signal spectrum in the bands $[f_j, f_{j+1}]$ and define as in (1) the optimal filter for $[f_j, f_{j+1}]$. Summing the outputs of these filters, one obtains the overall (approximated, see below) optimally filtered signal.

The subsequent qualitative analysis of the signal is essential in the design. Because the spectrum of the speech signal changes with every phoneme (and even inside a phoneme), that is every 5-30 ms, the above operations should be repeated accordingly. This creates a problem, because some of the frequencies in the speech signal are low; the pitch for male subjects

may be as low as 70 Hz, requiring a window of more than 10 ms for analysis. However, the pitch changes at a slower pace. For avoiding the contradictory requirement of small window for fast varying spectra and of large windows for detection of slower pace variation of the spectra at low frequencies, the signal should be filtered out in two components, one up to 500 Hz and the second above 500 Hz. The filtering procedure is applied separately, with different windows, for the two components. The subsequent explanations are for a single component, with the understanding that it is applied to both components and the results summed after processing.

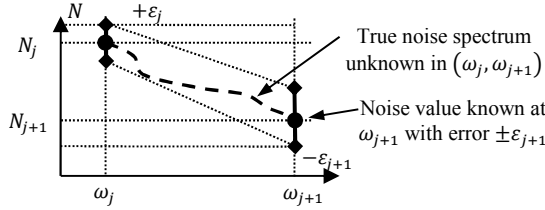


Figure 1. An example of estimation of the spectra at two subsequent frequencies and the approximation errors.

As mentioned, the imprecision in the knowledge of the noise spectrum may be due to the variability (non-ergodicity) of the noise affecting under various conditions the signal, to the nonstationarity of the noise, or to errors in the measurement of the noise.

In other words, in this problem setting, the power spectrum of the signal, $S_s^2(\omega_j)$ is incompletely and imprecisely known; also, the spectrum of the noise is known only through samples j in the frequency domain, $S_v^2(\omega_j)$, at different moments than the measurement of $S_s^2(\omega) + S_v^2(\omega_j)$ is performed. The situation is illustrated schematically in Fig. 1 for an interval between two successive values of frequencies where the noise spectral power density is determined.

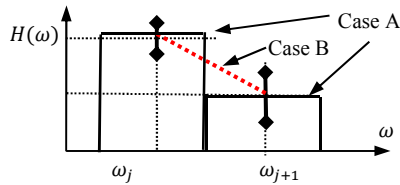


Figure 2. Two methods of defining the filter: with constant $H(\omega)$ around the measuring frequencies, and by linear interpolation (thick dotted line).

A first key assumption, visually suggested in Fig. 1, is that the noise power density remains in the interval

$S_v^2(\omega) \in [\min(S_{j,v}^2(n), S_{j+1,v}^2(n)), \max(S_{j,v}^2(n), S_{j+1,v}^2(n))]$ for the interval $\omega \in [\omega_j, \omega_{j+1}]$. This allows us to approximate the spectra either on an interval around the measurement frequencies, or between those frequencies, see Fig. 2. The block diagram of the adaptive filter is sketched in Fig. 3.

In one basic version of the filter, the adaptive filter operation follows the algorithmic steps:

Step 1. The input signal, with the spectrum $S_s^2(\omega) + S_v^2(\omega)$, is analyzed (with a specified

sampling frequency) to determine breaks in the actual signal, when only the noise $S_v^2(\omega)$ is present.

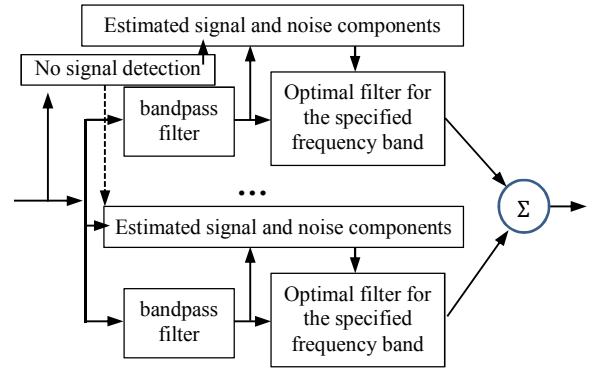


Figure 3. Operation diagram of the adaptive Wiener filter proposed

Step 2. The signal is bandpass filtered (narrow bands) and a sequence of values of the power spectral density (PSD) components are determined for frequencies f_j delimiting the bands, at moments t_n for the signal with additive noise, $S_{j,s}^2(n) + S_{j,v}^2(n)$. Similarly, one determines at moments $t_{n'}$ for the noise, the PSD components $S_{j,v}^2(n')$ during signal pauses. Alternatively, these values are obtained by FFT analysis. An illustration of the time moments when noise is determined is shown in Fig. 4.

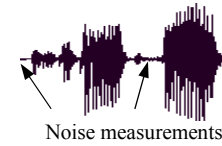


Figure 4. Hypothetical regions of noise (lack of signal) where noise spectra are determined.

Step 3. The spectral components of the noise are estimated at time moments t_n , based on the most recent three time moments $n', n' - 1, n' - 2$ when the noise was measured during breaks, assigning to the noise at the current moment t_n the average of the three values,

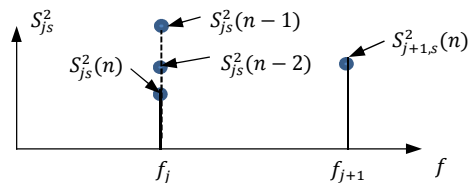
$$\hat{S}_{j,v}^2(n) = (S_{j,v}^2(n' - 2) + S_{j,v}^2(n' - 1) + S_{j,v}^2(n'))/3.$$


Figure 5. Illustration of the estimations of two successive components of the signal and noise spectra

Step 4. The spectral component of the signal at time moment t_n is determined as the measured component of the noisy signal minus the estimated noise component, $S_{j,s}^2(n) = (S_{j,s}^2(n) + S_{j,v}^2(n)) - \hat{S}_{j,v}^2(n)$.

Step 5. Correct the estimated value of the signal component by averaging over the last three estimated values,

$$\hat{S}_{j,s}^2(n) =$$

$(\hat{S}_{j,s}^2(n-2) + \hat{S}_{j,s}^2(n-1) + S_{j,s}^2(n))/3$. See Fig. 4 for an illustration.

Step 6. Further correct the estimated value $\hat{S}_{j,s}^2(n)$ in a predictive (recursive) manner as:

$$\hat{S}_{j,s}^2(n) \leftarrow \hat{S}_{j,s}^2(n) + \alpha_j (\hat{S}_{j,s}^2(n) - \hat{S}_{j,s}^2(n-1)),$$

using $\alpha_j = 1$ at this stage.

Step 7. For each frequency band $[f_j, f_{j+1}]$ and time moment n , compute the averages

$$S_{A,j,v}^2(n) = (S_{j,v}^2(n), S_{j+1,v}^2(n))/2 \quad \text{and} \quad S_{A,j,s}^2(n) = (S_{j,s}^2(n), S_{j+1,s}^2(n))/2$$

and build the Wiener filter for that band, according to (1), then sum all the partial signal results for all j values. Repeat from Step 2 for t_{n+1} .

In other variants of the filter (Case A in Fig. 2), at Step 7, divide the band $[f_j, f_{j+1}]$ in two halves and use optimal filters according to (1) for each of the halves, computed with $S_{j,v}^2(n)$ and $S_{j,s}^2(n)$, respectively with $S_{j+1,v}^2(n)$ and $S_{j+1,s}^2(n)$. (Case B in Fig. 2 is dealt with similarly.)

Step 8. At this stage, the residual noise defined by $\hat{S}_{j,v}^2(n)$ may include some component due to the signal because the noise between the frequencies f_j and f_{j+1} is only estimated, not measured, and because the estimation of the signal spectral components is imperfect. Therefore, there is a correlation between the spectrum of the estimated signal, $\hat{S}_s^2(f)$, and the residual noise $(S_s^2 + S_v^2)(f) - \hat{S}_s^2(f)$ at frequencies $(f_j + f_{j+1})/2$. For details, see Fig. 4 in [6]. Adjust the parameters α_j to reduce the correlation.

The method presented can be further improved using supplementary knowledge of the signal and noise, for example combining it with the method shown in [18].

V. DISCUSSION AND CONCLUSIONS

The improvement obtainable with the proposed procedure compared with the standard method (used for instance by the software Audacity™) is dependent on the variability of the noise during the speech, where the variation is determined with respect with the measured pre-speech noise. The quality of pause detector is crucial in this application, because it determines the intervals where the noise spectrum is estimated. During speech, if the noise is comparable in amplitude with the fricative consonants and the consonants are detected as pauses, the filter may perform poorly, eliminating consonants and distorting the spectrum of the signal. Especially the fricative consonants are prone to false detection as pause, as their energy is small and their spectrum close to the white noise. If the noise is very large, comparable in amplitude with the useful signal, e.g. vowel signal, the detection of the pauses again can be erroneous and the filters described will work very poorly.

The method and related optimal filters are adaptive in the sense that they track the noise and signal variations in time, largely in real-time. The filtering is

sub-optimal in the sense that noise is determined at different moments than the signal spectrum and both signal and noise spectra determinations are somewhat averages over time. The filtering is making a tradeoff between estimations based on current values and tendencies in the evolution of the noise and signal spectra. The method would not work for the cases when pauses in the useful signal do not occur.

It would be worth to extend the method to other signals. However, other signals may have no proper pauses, but may exhibit almost constant amplitude on brief durations; the electrocardiographic signal falls in this category and may allow the reconstruction of the noise spectra on such intervals.

REFERENCES

- [1] S.A. Kassam, T.L. Lim, "Robust Wiener filters." J Franklin Inst, Vol. 304, nos. 4-5, Oct-Nov 1977, pp. 171-185.
- [2] H. Poor, "On robust Wiener filtering," IEEE Trans. Automatic Control, Vol. 25, no. 3, Jun 1980, pp. 531-536, Jun 1980.
- [3] K.S. Vastola, H.V. Poor, "An analysis of the effects of spectral uncertainty on Wiener filtering." Automatica, Vol. 19, no. 3, May 1983, pp. 289-293.
- [4] I. R. Petersen, McFarlane, D. C., "Optimal guaranteed cost filtering for uncertain discrete-time linear systems." Int. J. Robust Nonlinear Control, vol. 6, 1996, 267-280.
- [5] X. Shen, L. Deng, "A dynamic system approach to speech enhancement using the filtering algorithm." IEEE Trans. Speech and Audio Processing, Vol. 7, No. 4, Jul 1999 pp. 391-399.
- [6] J. Chen, J. Benesty, Y. (Arden) Huang, S. Doclo, "New insights into the noise reduction Wiener filter." IEEE Trans. Audio, Speech, and Language Processing, Vol. 14, No. 4, July 2006, pp. 1218- 1234.
- [7] G.O. Corrêa, D.M. Sales, "Discrete-time, robust Wiener filtering with non-parametric spectral uncertainty." European J. Control, Vol. 18, no. 4, 2012, pp. 378-390.
- [8] L.A. Zadeh, J. Ragazzini, "An extension of Wiener's theory of prediction." J. Appl. Phys. 21, no. 7, Jul 1950, pp. 645-655.
- [9] LA Zadeh, JR Ragazzini, "Optimum filters for the detection of signals in noise." Proc. IRE 40 no. 10, pp. 1223-31, 1952.
- [10] L.A. Zadeh, "Optimum nonlinear filters." Journal of Applied Physics 24, no. 4, 396-404, 1953.
- [11] L.A. Zadeh, "What is optimal?" IRE Trans. Information Theory, IT4, p. 3, 1958.
- [12] H.N.L. Teodorescu, "On Zadeh's contribution to the optimality problem: four points of view on optimality - logic, set theory, theory of uncertainty, and analytic." Romanian J. Information Science and Technology, Vol. 20, no. 3, pp. 179-184, 2017.
- [13] J.D. Gibson, B. Koo, S.D. Gray, "Filtering of colored noise for speech enhancement and coding." IEEE Trans. Signal Processing. Vol. 39. No. 8, Aug 1991, pp. 1732-1742.
- [14] M. Marzinzik, B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics." IEEE Trans. Speech & Audio Proc., Vol. 10, No. 2, 2002, 109-118.
- [15] V. Stejskal, N. Bourbakis, A. Esposito, "Empty speech pause detection algorithms' comparison." J. Advanced Comp Intel Intel Informatics, Vol 2, 1, pp.145-160, July, 2010.
- [16] A Broumand, M.S. Esfahani, B.-J. Yoon, E.R. Dougherty, "Discrete optimal Bayesian classification with error-conditioned sequential sampling." Pattern Recognition vol. 48 (2015), pp. 3766-3782.
- [17] B.D.O. Anderson, J.B. Moore, Optimal Filtering. Prentice-Hall, Englewood Cliffs, 1979.
- [18] H.N.L. Teodorescu, "A set of nonlinear adaptive filters." Proc. 18th International Conference on Applied Electronics, Pilsen, Czech Republic, Sep 10-12, 2013, pp. 305-308, 2013.