Explaining the Parameterized Wiener Filter with Alpha-Stable Processes

Mathieu Fontaine, Antoine Liutkus, Laurent Girin, Roland Badeau

EXPLAINING THE PARAMETERIZED WIENER FILTER WITH ALPHA-STABLE PROCESSES

Mathieu Fontaine,[1] *Antoine Liutkus,*[1] *Laurent Girin,*[2] *Roland Badeau*[3]

[1]Inria, speech processing team, Villers-lès-Nancy, France.
[2]Univ. Grenoble Alpes, GIPSA-Lab, 38400 Saint-Martin d'Hères, France.
[3]LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France.

ABSTRACT

This paper introduces a new method for single-channel denoising that sheds new light on classical early developments on this topic that occurred in the 70's and 80's with Wiener filtering and spectral subtraction. Operating both in the short-time Fourier transform domain, these methods consist in estimating the power spectral density (PSD) of the noise without speech. Then, the clean speech signal is obtained by manipulating the corrupted time-frequency bins thanks to these noise PSD estimates. Theoretically grounded when using power spectra, these methods were subsequently generalized to magnitude spectra, or shown to yield better performance by weighting the PSDs in the so-called *parameterized Wiener filter*. Both these strategies were long considered ad-hoc. To the best of our knowledge, while we recently proposed an interpretation of magnitude processing, there is still no theoretical result that would justify the better performance of parameterized Wiener filters. Here, we show how the α-stable probabilistic model for waveforms naturally leads to these weighted filters and we provide a grounded and fast algorithm to enhance corrupted audio that compares favorably with classical denoising methods.

***Index Terms*—denoising, Wiener filtering, α-stable processes, probability theory**

I. INTRODUCTION

Single-channel speech enhancement consists in estimating a speech signal $s(t)$ corrupted by an additive noise $n(t)$ from the noisy observation $x(t) = s(t) + n(t)$. Early foundations to solve this problem date back to the late 70's and early 80's, with a series of works on spectral subtraction and Wiener filtering [9], [2], [1], [14], [8] (see also [13] for a more recent overview). Both techniques work in the short-time Fourier transform (STFT) domain. Therefore, let $v(f,t)$ denote the STFT coefficients of a signal $v(t)$ at frequency f and time frame t. Let $\sigma_v^2(f,t) = \mathbb{E}(|v(f,t)|^2)$ denote its power spectral density (PSD), assuming local wide-sense stationarity. In the STFT domain we have:

$$\forall (f,t),\ x(f,t) = s(f,t) + n(f,t). \tag{1}$$

We now describe spectral subtraction, that consists of the following. An estimate $\hat{\sigma}_n^2(f,t)$ of $\sigma_n^2(f,t)$ is first calculated on segments of $x(t)$ where speech is assumed to be absent[1], assuming that the noise does not change too rapidly and that some prior automatic voice activity detection has been applied. Then, an estimate of the speech short-time spectral amplitude (STSA) $|\hat{s}(f,t)|$ is obtained as:[2]

$$|\hat{s}(f,t)| = \left(|x(f,t)|^\alpha - \hat{\sigma}_n^\alpha(f,t)\right)^{1/\alpha}, \tag{2}$$

with $\alpha \in (0,2]$. Finally, the STSA estimate $|\hat{s}(f,t)|$ is combined with the phase of the noisy signal to provide an estimate of the speech STFT coefficients, and the estimated speech signal is obtained using inverse STFT. For $\alpha = 2$, (2) corresponds to *power spectral subtraction*, which is theoretically supported by i) the assumption of signal and noise decorrelation, and thus summation of their PSDs into the PSD of $x(t)$, and ii) the fact that $|\hat{s}(f,t)|^2$ is the maximum likelihood (ML) estimator of $\sigma_s^2(f,t)$ under the Gaussian assumption [14]. Setting $\alpha = 1$ corresponds to *magnitude spectral subtraction* and was the original choice in [2]. It was experimentally justified based on good noise rejection and enhanced speech intelligibility, but, to our knowledge, it has never been justified theoretically.

As another filtering method, Wiener filtering provides the minimum mean squared error (MMSE) linear estimator of $s(f,t)$ as:

$$\hat{s}(f,t) = \frac{\sigma_s^2(f,t)}{\sigma_s^2(f,t) + \sigma_n^2(f,t)} x(f,t). \tag{3}$$

Eq. (3) is a theoretical result and in practice estimates of $\sigma_s^2(f,t)$ and $\sigma_n^2(f,t)$ must be used. Since we can use the estimate of $\sigma_n^2(f,t)$ discussed above and we can use the square of (2) with $\alpha = 2$ as an estimate of $\sigma_s^2(f,t)$, and since in turn the spectral subtraction can be reshaped into a filtering process [9], there are strong links between power spectral subtraction and Wiener filtering.

Extending this baseline spectral subtraction method, further studies propose to use [1], [9]:

$$|\hat{s}(f,t)|^\alpha = |x(f,t)|^\alpha - k\ \hat{\sigma}_n^\alpha(f,t), \tag{4}$$

with $0 < k < 1$ being a parameter tempering the spectral subtraction. Again, although this formula is now part of "signal processing classics", to our knowledge, it has never been properly theoretically justified. Instead, it has been presented as an ad-hoc improvement leading to better enhanced speech quality or intelligibility, depending on the nature of the noise.

Similarly, the following *parameterized Wiener filter* has been proposed in the "classical" literature [9]:

$$\hat{s}(f,t) = \left(\frac{\sigma_s^2(f,t)}{\sigma_s^2(f,t) + k\ \sigma_n^2(f,t)}\right)^\beta x(f,t), \tag{5}$$

This work was partly supported by the research programme KAMoulox (ANR-15-CE38-0003-01) funded by ANR, the French State agency for research.

[1]In full generality, some studies consider estimation of σ_n^α with $\alpha \in (0,2]$.

[2]In fact, half-wave (or full-wave) rectification is applied to the term in parentheses to ensure its positivity. Other post-processes can be applied. Here and in the following, we omit their description for clarity, since they are poorly relevant to the discussion.

where $\beta > 0$. Again, the true PSDs must be replaced in practice with their corresponding estimates. Following that spirit, we can think of alternatives such as:

$$\hat{s}(f,t) = \frac{\sigma_s^{\alpha}(f,t)}{\sigma_s^{\alpha}(f,t) + k\ \sigma_n^{\alpha}(f,t)} x(f,t). \tag{6}$$

In the same vein, the renowned MMSE STSA estimator of [3] can be interpreted as a sophisticated Wiener filter, parameterized by both a prior and a posterior signal-to-noise ratio (SNR). Independently from the speech enhancement problem, and many years later, (6) with $\alpha = 1$ (and $k = 1$) has been largely considered for audio source separation [20], [15], [7]. Yet, using (6) with $\alpha \neq 2$ was for a long time justified only heuristically as a spectral *soft-weighting* scheme that gives good performance in practice, rather than from grounded theoretical foundations.

A theoretical interpretation of (6) with $\alpha \neq 2$ has been proposed recently in [10] as naturally arising when the signals are assumed α-*harmonizable*. Skipping details, it may be shown that (6) with $k = 1$ can be understood as the computation of $\mathbb{E}[s(f,t) \mid x]$ under this model, hence the name *generalized Wiener filter* given to the procedure. Furthermore, a remarkable fact is that the so-called *fractional* α-PSD $\sigma^{\alpha}(f,t)$ of additive α-harmonizable sources add up to form the α-PSD of their sum, generalizing the Gaussian case $\alpha = 2$. Although it has never been mentioned in the literature, this fact is a strong theoretical support to the general form (2) of spectral subtraction, for $\alpha \in (0,2]$ and $k = 1$. See also [11], [19] for recent works on α-stable audio modeling.

In the present study, we make a further step in this attempt to revisit the above signal processing classics, by providing grounded theoretical justification for the parameterized forms (4) and (6) of spectral subtraction and Wiener filtering. In the above-mentioned studies, we indeed always had $k = 1$. Here, we focus on this parameter k, and we show that it can be understood in a theoretically grounded way through α-stable modeling. Interestingly, this comes with an automatic way to tune k as a consequence of the signals distribution, rather than through empirical tuning. In short, a parameter $k \neq 1$ arises whenever we are considering the separation of α-stable sources with different characteristic exponents α, which is a typical situation in denoising. Interestingly, the same line of thought may be used for devising parametric Wiener filters for an arbitrary number of signals, as in source separation. As a result of this discussion, we propose the Multi-Alpha Denoising (MAD) algorithm, that is very simple to implement and fully exploits this renewed understanding of parameterized Wiener filtering.

II. THEORETICAL BACKGROUND & MODEL

II-A. α-harmonizable processes

The α-harmonizable model introduced in [10] generalizes the classical Wide-Sense Stationary (WSS) model for an audio signal. It is the probabilistic model that obeys the following three fundamental assumptions:

1) The signal should be *infinitely divisible*, meaning that it is possible to write it as a sum of an arbitrary large number of independent and identically distributed (i.i.d.) contributions [21]. In the case of audio, we may think of the infinitely many vibrating elements present in a sound source.
2) For convenience of analytical derivation, it is desirable that the sum of these i.i.d components has a distribution that belongs to the same family, up to some scaling. These first two assumptions lead us to model each source as an α-stable process [18], [21].
3) When observed on a sufficiently short time-scale, the signal may be considered locally stationary, meaning that its distribution does not depend on the time origin.

While WSS signals obey these three assumptions, they are only a small subset of the set of all models that do so, which may be proved to coincide with the α-stable harmonizable processes [18], coined in as α-harmonizable in [10]. Skipping irrelevant technical details that can be found in [10], the main features of an α-harmonizable signal $v(t)$ are twofold:

1) All the entries $v(f,t)$ of its STFT are independent;
2) Each entry is distributed as:

$$v(f,t) \sim S\alpha S_c\left(\sigma_v(f,t)\right), \tag{7}$$

where $S\alpha S_c$ denotes the centered isotropic complex α-stable distribution [18]. This distribution for complex numbers is rotationally invariant and fully characterized by a *scale parameter* $\sigma_v \geq 0$ and a *characteristic exponent* $\alpha \in (0,2]$ that controls the thickness of the tails: the smaller α, the most likely the extreme values. $\alpha = 2$ corresponds to the Gaussian case, for which σ_v is called the *standard deviation*. In practice, a noticeable feature of these processes is that they can account for the sparsity, high impulsiveness and large dynamics of audio signals [4], [6], while allowing for effective processing in the STFT domain.

II-B. Conditional Gaussianity of α-harmonizable processes

A well documented feature of $S\alpha S_c$ random variables is that they are conditionally Gaussian [5], [19], [18]. Intuitively, it means we can understand $v \sim S\alpha S_c(\sigma_v)$ as a mixture of complex isotropic Gaussians, whose variances are distributed w.r.t. a random *impulse* variable ϕ_v:

$$v|\phi_v, \sigma_v \sim \mathcal{N}_c\left(0, \phi_v \sigma_v^2\right), \tag{8}$$

where $\mathcal{N}_c(\mu, \sigma^2)$ denotes the proper complex Gaussian distribution of mean μ and variance σ^2, and $\phi_v \geq 0$ follows a positive $\frac{\alpha}{2}$-stable distribution,[3] written $P\frac{\alpha}{2}S$, whose parameters depend only on α:

$$\phi_v \sim P\frac{\alpha}{2}S\left(2\cos\left(\frac{\pi\alpha}{4}\right)^{2/\alpha}\right). \tag{9}$$

The impulsive nature of v appears in the fact that this variance can take very large values since ϕ_v has a heavy tail distribution.

II-C. Multi-α Wiener filtering

We now address the single-channel speech enhancement problem and thus focus on the case of only two sources: speech and noise. In the present study, we extend the previous work on α-harmonizable models [10] by assuming that speech and noise do *not* share the same characteristic exponent. In the STFT domain, this gives:

$$\forall (f,t),\, s(f,t) \sim S\alpha_s S_c\left(\sigma_s(f,t)\right), \tag{10}$$

$$\forall (f,t),\, n(f,t) \sim S\alpha_n S_c\left(\sigma_n(f,t)\right), \tag{11}$$

where all variables in (10) and (11) are mutually independent. Allowing each source to have its own characteristic exponent

[3] This particular subset of the α-stable distributions has support on $\mathbb{R}_+$ and exhibits very heavy tails. A noticeable example is the Levy distribution [18].

accounts for different general "impulsiveness" for different sources. On the other hand, the time-varying spectral content of the signals is encoded into the scale parameters $\sigma_s(f,t)$ and $\sigma_n(f,t)$.

Applying conditional Gaussianity (8)-(9) independently to (10) and to (11) leads to define impulse variables $\phi_s(f,t)$ and $\phi_n(f,t)$ for target signal and noise respectively. We thus obtain:

$$s(f,t)|\phi_s(f,t),\sigma_s(f,t) \sim \mathcal{N}_c\left(0,\, \phi_s(f,t)\sigma_s^2(f,t)\right), \tag{12}$$

$$n(f,t)|\phi_n(f,t),\sigma_n(f,t) \sim \mathcal{N}_c\left(0,\, \phi_n(f,t)\sigma_n^2(f,t)\right). \tag{13}$$

Assuming from now on that we know all parameters σ and ϕ, we can estimate the speech signal through classical Wiener filtering:

$$\mathbb{E}\left[s(f,t)\,|x,\sigma,\phi\right] = w(f,t)x(f,t), \tag{14}$$

where

$$w(f,t) = \frac{\phi_s(f,t)\,\sigma_s^2(f,t)}{\phi_s(f,t)\,\sigma_s^2(f,t) + \phi_n(f,t)\,\sigma_n^2(f,t)}. \tag{15}$$

Eq. (14) provides a form of the Wiener filter which adds degrees of freedom compared to the classical form (3). As can be seen, it establishes a theoretically grounded justification to the generalized form (5), provided we set $\beta = 1$ and identify $k = \phi_n(f,t)/\phi_s(f,t)$.

We recall that (5) was an ad-hoc proposition in the signal processing literature of the 80's. To the best of our knowledge, no theoretical justification for parameter k was given to date. Of course, although the α-harmonizable model comes with the nice property of bringing some explanations for parameterized Wiener filters, we now face the important issue of estimating not only the scale parameters σ but also the impulse variables ϕ for this multi-α-Wiener filter (14). We address this problem in the next section.

III. PARAMETERS ESTIMATION AND SIGNAL DENOISING

In this section, we present an iterative algorithm designed to estimate the parameters of the multi-α Wiener filter (14). Target signal estimation is embedded as a subtask into this algorithm, and is thus implicitly solved. First, assuming knowledge of the scale parameters σ_s and σ_n, we propose to marginalize out the impulse variables to get a simple expression for the filter. Then, assuming knowledge of the separated signals, we show how the scale parameters can easily be estimated. The resulting Multi-Alpha Denoising (MAD) algorithm is depicted in Fig. 1 and summarized in the algorithm box.

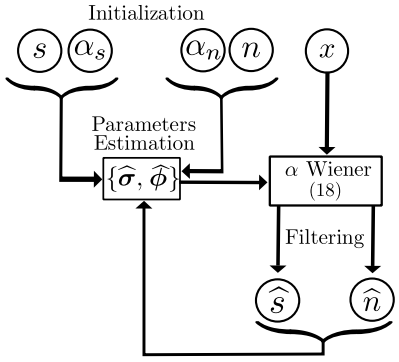

Fig. 1. Outline of the proposed denoising algorithm.

Algorithm 1 MultiAlpha Denoising (MAD) algorithm

1) **Input**
   - STFT x of the single-channel signal.
   - Number of iterations.
   - Characteristic exponents α_s and α_n.
   - Fixed-size horizons Δ_s and Δ_n for log-spectrogram smoothing.
2) **Initialization**
   - $\forall (f,t)\,,\ \hat{s}(f,t) = \hat{n}(f,t) = x(f,t)/2$.
   - Compute $\hat{\phi}_s$ and $\hat{\phi}_n$ as the empirical median values under distribution (9).
3) **Parameter estimation**
   - Compute $\ln|\hat{s}(f,t)|$ and $\ln|\hat{n}(f,t)|$. Smooth them on Δ_s and Δ_n. This yields $\mathbb{E}\ln(|\hat{s}(f,t)|)$ and $\mathbb{E}\ln(|\hat{n}(f,t)|)$.
   - Compute $\hat{\sigma}_s(f,t)$ and $\hat{\sigma}_n(f,t)$ using (20) and (21).
4) **Filtering**
   - Apply multi-α Wiener filtering (18) to get $\widehat{s}(f,t)$ and $\widehat{n}(f,t)$.
   - Go back to 3) if needed.

III-A. Marginalizing out the impulse variables

Let us first assume we know the scale parameters σ_s and σ_n. For the sake of clarity, the TF indexes (f,t) will be omitted in the following derivations. Considering (14), we see that the knowledge of the impulse variables ϕ_s and ϕ_n is required to perform filtering. One solution is to marginalize out these variables, so as to compute:

$$\begin{aligned}\mathbb{E}(s\,|\,x,\sigma) &= \mathbb{E}_\phi\left(\mathbb{E}(s\,|\,x,\sigma,\phi)\right)\\ &= \int_\phi \mathbb{E}(s\,|\,x,\sigma,\phi)\,\mathbb{P}(\phi\,|\,x,\sigma)\,d\phi,\end{aligned} \tag{16}$$

where $\phi = \{\phi_s,\phi_n\}$. Unfortunately, the posterior distribution $\mathbb{P}(\phi\,|\,x,\sigma)$ in the integral in (16) has no analytical form. One solution is to run a Markov Chain Monte Carlo (MCMC) estimation of this integral, as done in [19]. Still, this MCMC strategy is computationally very demanding. A quite simplifying alternative we consider in the present study is to replace the distribution $\mathbb{P}(\phi\,|\,x,\sigma)$ in (16) by a degenerate Dirac distribution, with unit mass at the most probable value for ϕ. In other words, we assume:

$$\int_\phi \mathbb{E}(s\,|\,x,\sigma,\phi)\,\mathbb{P}(\phi|x,\sigma)\,d\phi \approx \mathbb{E}\left(s\,|\,x,\sigma,\hat{\phi}\right), \tag{17}$$

where $\hat{\phi}$ is a point estimate for the impulse variable. Now let us remind that the impulse variable distribution (9) does not depend on the scale parameter, but only on the characteristic exponent, i.e. either α_s or α_n in the present problem. Hence, the distribution is independent of (f,t), and when it comes to picking a point estimate for ϕ in (17), there is no particular reason to choose a different one for different TF bins, since all impulse variables for a given source are i.i.d. Eventually, we found out that replacing $\phi_s(f,t)$ and $\phi_n(f,t)$ in (15) by their median $\mathbb{M}(\phi_s)$ and $\mathbb{M}(\phi_n)$ under their prior distribution (9) gives a very satisfying fit to the optimal multi-α Wiener filter (15). The resulting parameterized filter becomes:

$$\mathbb{E}\left[s(f,t)\,|x,\sigma\right] \approx \widehat{w}(f,t)\,x(f,t), \tag{18}$$

where

$$\widehat{w}(f,t) = \frac{\sigma_s^2(f,t)\,\mathbb{M}(\phi_s)}{\sigma_s^2(f,t)\,\mathbb{M}(\phi_s) + \sigma_n^2(f,t)\,\mathbb{M}(\phi_n)}. \tag{19}$$

In practice, the median values are computed from i.i.d. sampled random variables following distribution (9) using the method discussed in [22], [5]. We set $\alpha_s = 1.2$ and $\alpha_n = 1.89$, since these two values fit well the impulsiveness of speech and the near-Gaussianity of non-stationary noise, respectively (see [19]). Note that this computation needs only to be done once. In Fig. 2 (up), we display the ground-truth Wiener masks for both target signal (up-left) and noise (up-right). The ground-truth masks were obtained for each "fixed" couple (σ_s, σ_n) by averaging (15) over many realizations drawn according to the generative model (9). We can see that the approximation (19) plotted in Fig. 2 (down), fits the ground-truth very well. The average relative error is as small as 1.8%. For completeness, we mention that using the mode instead of the median for $\hat{\phi}$ induces a greater average error ($\simeq 3.1\%$).

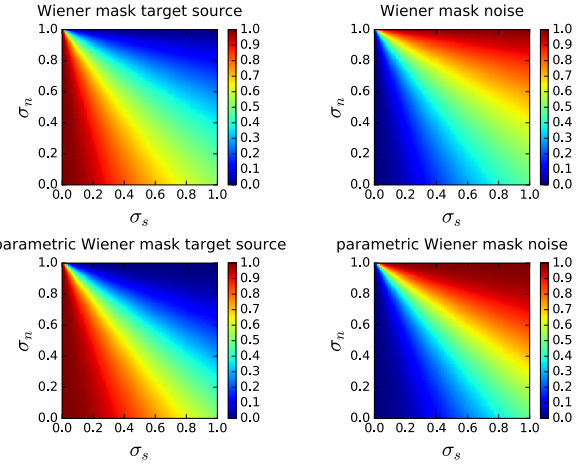

Fig. 2. Illustration of the ground-truth Wiener mask (up) and the approximated mask (down) for several couples $(\sigma_s, \sigma_n) \in [0,1]^2$.

III-B. Estimation of scale parameters σ

Now, assume that for each TF-bin (f,t) the target signal $s(f,t)$ and the noise $n(f,t)$ have been separated at the previous step of the algorithm (using (18) for the speech signal). We propose to build the estimation of $\sigma_s(f,t)$ and $\sigma_n(f,t)$ on the following relations, taken from [16, p. 69]:

$$\forall (f,t), \mathbb{E}\ln(|s(f,t)|) = \gamma\left(\frac{1}{\alpha_s} - 1\right) + \alpha_s \ln(\sigma_s(f,t)), \quad (20)$$

$$\forall (f,t), \mathbb{E}\ln(|n(f,t)|) = \gamma\left(\frac{1}{\alpha_n} - 1\right) + \alpha_n \ln(\sigma_n(f,t)), \quad (21)$$

where $\gamma \approx 0.5772156$ is the Euler constant. In practice, we assume that several adjacent TF bins share the same scale parameters σ_s and an effective strategy for their estimation is to average $\ln(|s(f,t)|)$ over those TF bins with a moving average and then use (20) to obtain a local estimate of σ_s. σ_n is estimated in the same way from (21). The method is mostly parameterized by the length Δ_s and Δ_n of those smoothing filters for speech and noise. This strategy is reminiscent of Kernel Additive Modeling (KAM, [12]) in the sense that separation is alternated with spectrogram filtering. The difference is that the filtering performed here is linear w.r.t. log-spectrograms, instead of the median filter considered for KAM. A nice feature is that the proposed MAD method appears as theoretically grounded in terms of probabilistic models.

In practice, we noticed that using longer averaging horizons for the noise source (approximatively 0.16 s) than for the target source (0.09 s) leads to the best results.

IV. EVALUATION

To evaluate the proposed algorithm, we used a corpus of 30, 16-bit, 3-s long speech signals sampled at 8-kHz from the NOIZEUS dataset[4]. Each excerpt was corrupted by additive noise at different signal-to-noise ratios (SNR). Three types of noise were used: car engine, babble noise and airport environment. We used a 1024-sample long Hamming STFT window with 85%-overlap. For the evaluation metrics, we used the perceptual evaluation of speech quality (PESQ) improvement score [17], ranging between -0.5 and 4.5 respectively corresponding to the worst and best denoising performance.

The MAD algorithm described in Fig. 1 and in Algorithm 1 was compared with three baseline methods: the magnitude spectral subtraction (MSS) defined in (2) with $\alpha = 1$, the generalized spectral subtraction (GSS) defined in (4) with $\alpha = 1.2$ and $k = 0.8$, and the MMSE-STSA estimator method of [3] with $\alpha = 0.92$ and $\beta = 2$. In this assessment, the parameters for the MAD algorithm are $\alpha_s = 1.3$, $\alpha_n = 1.8$, $\Delta_s = 0.09$ s, $\Delta_n = 0.16$ s (we assume that speech is slightly less stationary than noise) and a fixed number of 4 iterations. Note that all those parameters for the baseline and proposed methods were selected to yield the best PESQ scores.

The PESQ scores are illustrated in Table I. Basically, the MAD algorithm outperforms all baseline methods for every SNR value and every type of noise. In a general manner, the MMSE-STSA method has the best scores among the baseline methods. However, when listening the denoised speech signals, a more significant residual noise is present with MMSE-STSA compared to MAD.

Noise	SNR	MAD	MSS	GSS	MMSE
Car	0 dB	**1.312**	1.183	1.201	1.178
	5 dB	**1.558**	1.300	1.343	1.282
	10 dB	**1.913**	1.405	1.560	1.515
	15 dB	**2.208**	1.504	1.755	1.889
Airport	0 dB	**1.308**	1.174	1.192	1.208
	5 dB	**1.561**	1.301	1.390	1.368
	10 dB	**1.844**	1.389	1.556	1.642
	15 dB	**2.102**	1.507	1.755	2.000
Babble	0 dB	**1.272**	1.170	1.200	1.199
	5 dB	**1.519**	1.272	1.351	1.357
	10 dB	**1.816**	1.390	1.576	1.603
	15 dB	**2.121**	1.494	1.771	2.042

Table I. PESQ scores for each algorithm. Bold font indicates best scores.

V. CONCLUSION

In this paper, we proposed a new denoising algorithm which sets the classical parameterized Wiener filtering (and spectral subtraction method) on solid theoretical grounds. We demonstrated how the α-stable theory naturally leads to a strategy for automatically tuning the parameters of the Wiener filter, that depend on different tail behaviors for the distribution of the different sources, with also a robust initialization. The resulting multi-alpha denoising (MAD) algorithm was shown to give very good performance compared to several baseline methods. Future works may include multichannel and multi-source denoising scenarios.

[4]NOIZEUS http://ecs.utdallas.edu/loizou/speech/noizeus/

VI. REFERENCES

[1] M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 208–211. IEEE, 1979.

[2] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, Signal Process.*, 27(2):113–120, 1979.

[3] Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Process.*, 32(6):1109–1121, 1984.

[4] P.G. Georgiou, P. Tsakalides, and C. Kyriakakis. Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE Trans. Multimedia*, 1(3):291–301, 1999.

[5] S. Godsill and E. Kuruoglu. Bayesian inference for time series with heavy-tailed symmetric α-stable noise processes. *Proc. Applications of heavy tailed distributions in economics, engineering and statistics*, 1999.

[6] P. Kidmose. Alpha-stable distributions in signal processing of audio signals. In *41st Conference on Simulation and Modelling*, pages 87–94, 2000.

[7] B. King, C. Févotte, and P. Smaragdis. Optimal cost function and magnitude power for NMF-based speech separation and music interpolation. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2012.

[8] J.S. Lim. *Speech enhancement*. Prentice-Hall, 1983.

[9] J.S. Lim and A.V. Oppenheim. Enhancement and bandwidth compression of noisy speech. *Proc. IEEE*, 67(12):1586–1604, 1979.

[10] A. Liutkus and R. Badeau. Generalized Wiener filtering with fractional power spectrograms. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 266–270. IEEE, 2015.

[11] A. Liutkus, D. Fitzgerald, and R. Badeau. Cauchy nonnegative matrix factorization. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–5. IEEE, 2015.

[12] A. Liutkus, Z. Rafii, B. Pardo, D. Fitzgerald, and L. Daudet. Kernel spectrogram models for source separation. In *4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, pages 6–10. IEEE, 2014.

[13] P.C. Loizou. *Speech enhancement: theory and practice*. CRC press, 2013.

[14] R. McAulay and M. Malpass. Speech enhancement using a soft-decision noise suppression filter. *IEEE Trans. Acoust., Speech, Signal Process.*, 28(2):137–145, 1980.

[15] N. Mohammadiha, P. Smaragdis, and A. Leijon. Supervised and unsupervised speech enhancement using nonnegative matrix factorization. *IEEE Trans. Acoust., Speech, Signal Process.*, 21(10):2140–2151, 2013.

[16] C.L. Nikias and M. Shao. *Signal processing with alpha-stable distributions and applications*. Adaptive and learning systems for signal processing, communications, and control. Wiley, 1995.

[17] A. Rix, J. Beerends, M. Hollier, and A. Hekstra. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 749–752. IEEE, 2001.

[18] G. Samoradnitsky and M. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC Press, 1994.

[19] U. Şimşekli, A. Liutkus, and A.T. Cemgil. Alpha-stable matrix factorization. *IEEE Signal Process. Lett.*, 22(12):2289–2293, 2015.

[20] P. Smaragdis, B. Raj, and M. Shashanka. A probabilistic latent variable model for acoustic modeling. *Advances in models for acoustic processing, NIPS*, 148:8–1, 2006.

[21] M. Unser and P. Tafti. *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.

[22] R. Weron. On the chambers-mallows-stuck method for simulating skewed stable random variables. *Statistics & probability letters*, 28(2):165–171, 1996.