THE SELECTION OF SPECTRAL MAGNITUDE EXPONENTS FOR SEPARATING TWO SOURCES IS DOMINATED BY PHASE DISTRIBUTION NOT MAGNITUDE DISTRIBUTION

Stephen Voran

Institute for Telecommunication Sciences
325 Broadway, Boulder, Colorado, 80305, USA, svoran@its.bldrdoc.gov

ABSTRACT

Separating an acoustic signal into desired and undesired components is an important and well-established problem. It is commonly addressed by decomposing spectral magnitudes after exponentiation and the choice of exponent has been studied from numerous perspectives. We present this exponent selection problem as an approximation to the actual underlying geometric situation. This approach makes apparent numerous basic facts and some of these have been ignored or violated in related work efforts. We show that exponent selection is dominated by the phase distribution and that magnitude distributions have almost no influence. We also show that exponents can be much more effectively selected in the estimated-source domain, rather than in the domain of the combined sources. Finally we describe the mechanism that causes exponents slightly above 1.0 to be preferred in many cases, completely independent of source distributions.

Index Terms— noise reduction, source separation, spectral magnitude exponent, spectral subtraction, speech enhancement

1. INTRODUCTION

Separating acoustic signals into two or more contributing sources is an important problem that has received much attention. A common and important application is separation into a desired portion (often speech) and an undesired portion (often other speech or environmental sounds). For this large and vital class of problems it suffices to view the signal as having just two sources. This paper addresses the two-source problem only.

The problem is commonly addressed in the frequency-domain by invoking an additive model for the two sources. This then leads to scaling down (Wiener filtering) or subtracting from (spectral subtraction) the frequency-domain representation of the combined sources to produce a representation for a single source.

Early researchers successfully treated both spectral magnitudes [1],[2] and squared-spectral magnitudes [3] as additive. Both approaches have been extensively studied, adapted, enhanced, and applied over the years and a very small but broad sampling of these efforts can be found in [4]-[8] and the citations therein. The approaches have also been extended to a domain where spectral magnitudes have been raised to an arbitrary positive real power α, thus generalizing the cases $\alpha = 1$ (magnitude) and $\alpha = 2$ (squared magnitude or power). Selection of exponents has been studied from numerous perspectives driven both by specific applications and by the quest for more general insight into the issue [9]-[16].

For the two-source problem, we present the selection of the spectral magnitude exponent α as a geometric approximation problem. This view leads to some basic yet helpful insights, including the fact that the approximation is strongly dominated by phase distributions and magnitude distributions are of very little consequence. Following previous research efforts, we apply established measures of spectral divergence in the combined-sources domain and show that they do not inform the selection of α as one might expect. We then demonstrate that simple and intuitive spectral error measures in the estimated-source domain provide meaningful results and also illuminate the true reason that spectral magnitude exponents slightly above 1.0 often perform best.

2. SELECTING A SPECTRAL MAGNITUDE EXPONENT AS A GEOMETRIC APPROXIMATION PROBLEM

Consider two independent acoustic sources $x_1(t)$ and $x_2(t)$. We sample, window, and apply the Discrete Fourier Transform (DFT) to produce complex spectral representations of these signals. Example complex-valued results are $z_1 = r_1 e^{j\phi_1}$ and $z_2 = r_2 e^{j\phi_2}$ where r_1 and r_2 are the real magnitudes of z_1 and z_2 respectively (time and frequency indices are not required for this development). Acoustic waveforms are additive so the two sources combine to produce $x(t) = x_1(t) + x_2(t)$. Because all processing steps are linear, the corresponding complex time-frequency sample of $x(t)$ is

$$\begin{aligned} z = z_1 + z_2 = r_1 e^{j\phi_1} + r_2 e^{j\phi_2} = r e^{j\phi}, \\ r = \mathcal{A}(r_1, r_2, \theta) = \sqrt{r_1^2 + r_2^2 + 2 r_1 r_2 \cos(\theta)}, \\ \theta = |\phi_2 - \phi_1|, \;\; 0 \le \theta \le \pi. \end{aligned} \tag{1}$$

(The phase difference is wrapped into $[-\pi, \pi]$ before the absolute value is taken, resulting in $0 \le \theta \le \pi$.) Note that $\mathcal{A}(r_1, r_2, \theta)$ is the function that produces the *actual* magnitude of z. An example is shown in Fig. 1.

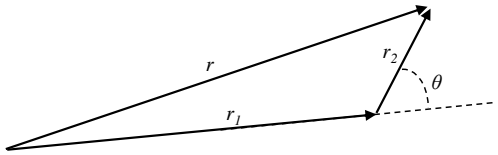

Figure 1: Example geometry.

In the two-source problem we observe $x(t)$ and calculate $re^{j\phi}$. If estimates of r_2 and ϕ_2 were available we could readily calculate corresponding estimates of r_1 and ϕ_1. Typically an estimate of r_2 (e.g. noise magnitude) is available but no estimate of ϕ_2 is available. A popular and effective way forward is to assume that magnitude additivity can be approximately satisfied in the r^α domain for some real α, $(0 < \alpha)$,

$$r^\alpha \approx r_1^\alpha + r_2^\alpha. \tag{2}$$

This paper was presented at and will appear in the Proceedings of the
2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 15-18, 2017.

This leads directly to an estimate of r_1 that uses the observation r, an estimate of r_2, and a small positive value of ζ:

$$\hat{r_1} = (\max(\zeta, r^\alpha - \hat{r_2}^\alpha))^{\frac{1}{\alpha}}. \quad (3)$$

Rearranging (2) yields a *model* $\mathcal{M}(r_1, r_2, \alpha)$ for magnitude of z,

$$r \approx \mathcal{M}(r_1, r_2, \alpha) = (r_1^\alpha + r_2^\alpha)^{\frac{1}{\alpha}}. \quad (4)$$

This model does not depend on θ. This makes it a practical model but also a very rough model.

Both $\mathcal{A}$ and $\mathcal{M}$ are symmetric with respect to r_1 and r_2,

$$\mathcal{A}(r_1, r_2, \theta) = \mathcal{A}(r_2, r_1, \theta),\ \mathcal{M}(r_1, r_2, \alpha) = \mathcal{M}(r_2, r_1, \alpha), \quad (5)$$

so we can assume $r_1 \geq r_2$ without loss of generality. The case $r_1 = r_2 = 0$ is trivial so we further assume $r_1 > 0$. $\mathcal{A}$ and $\mathcal{M}$ also share a scaling property:

$$\mathcal{A}(r_1, r_2, \theta) = r_1 \mathcal{A}(1, \frac{r_2}{r_1}, \theta),$$
$$\mathcal{M}(r_1, r_2, \alpha) = r_1 \mathcal{M}(1, \frac{r_2}{r_1}, \alpha). \quad (6)$$

This motivates us to define the magnitude ratio $\gamma = \frac{r_2}{r_1}, 0 \leq \gamma \leq 1$. The model $\mathcal{M}$ should approximate the actual result $\mathcal{A}$,

$$\mathcal{M}(r_1, r_2, \alpha) \approx \mathcal{A}(r_1, r_2, \theta) \implies$$
$$(r_1^\alpha + r_2^\alpha)^{\frac{1}{\alpha}} \approx \sqrt{r_1^2 + r_2^2 + 2 r_1 r_2 \cos(\theta)}. \quad (7)$$

This result is key and the basis for all that follows. It provides a reminder of the true goal in exponent selection. It reminds us not to spend undue effort to match signal models to $\mathcal{M}$ when in fact $\mathcal{M}$ is just a very rough approximation to the underlying truth expressed by $\mathcal{A}$. Equation (7) leads directly to the following eleven observations that may help one more fully appreciate the true approximation problem. Mathematical developments underpinning research in this area sometimes overlook or even violate observations 9, 10, or 11.

1. If $r_2 = 0$ we have sparsity in that time-frequency bin. The approximation problem disappears and (7) becomes an equality for any choice of α. Thus we assume $r_2 > 0$ (and $\gamma > 0$) without loss of generality.
2. If $r_1 \gg r_2$ we have approximate relative-sparsity in that time-frequency bin, the approximation (7) is nearly exact and shows minimal sensitivity to α.
3. For all r_1 and r_2 $\mathcal{A}$ decreases monotonically with θ ($0 \leq \theta \leq \pi$) and has the range $r_1 - r_2 \leq \mathcal{A}(r_1, r_2, \theta) \leq r_1 + r_2$.
4. For all r_1 and r_2 $\mathcal{M}$ decreases monotonically with α ($0 < \alpha$) and has the range $\max(r_1, r_2) = r_1 < \mathcal{M}(r_1, r_2, \alpha) < \infty$.
5. When $\theta = 0$ (matched phases), $\alpha = 1$ (magnitude addition) produces equality in (7) for all r_1 and r_2.
6. When $\theta = \frac{\pi}{2}$ (orthogonal phases), $\alpha = 2$ (power addition) produces equality in (7) for all r_1 and r_2.
7. Using (6) in (7) and canceling r_1 shows that exact solutions to (7) depend on the magnitude ratio γ and θ. For every value of γ in $(0, 1]$, if θ satisfies $0 \leq \theta \leq \theta_m(\gamma)$ there is a value of α that produces equality in (7). As θ moves through this range the value of α that produces equality moves from 1 toward ∞ as shown in Fig. 2. $\theta_m(\gamma) = \pi - \arccos\left(\frac{\gamma}{2}\right)$ so θ_m increases from $\frac{\pi}{2}$ to $\frac{2\pi}{3}$ as γ runs from 0 to 1.
8. For every value of γ in $(0, 1]$, if $\theta_m(\gamma) < \theta \leq \pi$, no value of α can produce equality in (7). For these angles the resulting cancellation cannot be emulated by $\mathcal{M}$.
9. The sources $x_1(t)$ and $x_2(t)$ are independent and thus uncorrelated but this does *not* force the phases in any time-frequency bin to be orthogonal. The expected value of θ over time or frequency is $\frac{\pi}{2}$ but $\mathbf{E}(\mathcal{A}) \neq \sqrt{r_1^2 + r_2^2}$ and general optimality for $\alpha = 2$ does not follow (see Sections 3 and 4).
10. Assuming Gaussian sources has no bearing on the orthogonality of the phases. It is true that when independent Gaussian random variables are added their variances add. If one is focused on $\mathcal{M}$ alone this may steer one to $\alpha = 2$. But Gaussian sources and $\alpha = 2$ do not lead to a better approximation in (7).
11. Assuming Cauchy sources has no bearing on the orthogonality of the phases. It is true that when independent Cauchy random variables are added their scale factors add. If one is focused on $\mathcal{M}$ alone this may steer one to $\alpha = 1$. But Cauchy sources and $\alpha = 1$ do not lead to a better approximation in (7). The same is true for other α-stable distributions and their associated α values.

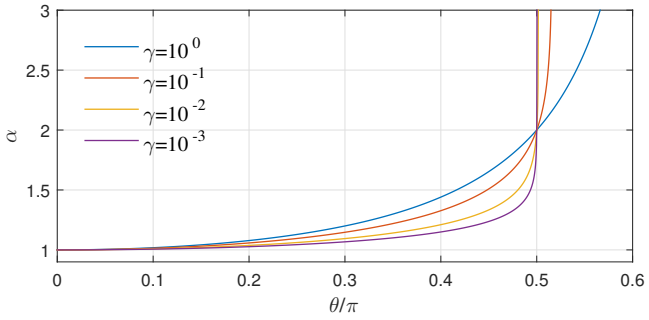

Figure 2: Values of α that produce equality in (7) for four magnitude ratios $\gamma = \frac{r_2}{r_1}$.

In any practical application θ will be unknown and solving for individual α values (observations 5-7) will not be an option. Our analysis of phase values produced by application of the windowed DFT to music and speech shows they are uniformly distributed on $[-\pi, \pi]$ and their differences follow the triangular distribution on $[-2\pi, 2\pi]$. Wrapping this distribution into $[-\pi, \pi]$ gives the uniform distribution again. It follows that θ is uniform on $[0, \pi]$.

The approximation problem is to select a spectral magnitude exponent α that is most effective for a given separation task in spite of unknown absolute phase differences θ drawn from the uniform distribution. A further difficulty is that when the absolute phase difference θ exceeds θ_m, the resulting cancellation produces actual magnitudes in $[r_1 - r_2, r_1]$ and magnitudes in this region cannot be achieved by $\mathcal{M}$ for any α (see observations 3, 4, and 8). The remaining range of $\mathcal{A}$ must often be approximated by $\mathcal{M}$ using a single value of α. Observations 5, 6, and 7 suggest that a compromise value may be in the range $1 \leq \alpha \leq \infty$.

The distribution of the absolute phase differences θ dominates this approximation problem and the distributions of the magnitudes r_1 and r_2 are much less significant. The high sensitivity of α to θ is clear from Fig. 2. In addition, depending on θ there can be a single solution α for all γ, an individual solution for each γ, or

This paper was presented at and will appear in the Proceedings of the
2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 15-18, 2017.

no solution. The variation in α across three orders of magnitude of γ is much smaller. Formal mathematical demonstrations of the low sensitivity to magnitude distributions follow in Sections 3 and 4 below.

It is true that α can be fine-tuned based on frequency or adapted over time based on r or estimates of r_2 [11],[15]. Such adaptation can address changing magnitude distributions but it cannot address the more significant issue of phase differences and thus can provide modest advantages at best.

3. EVALUATING APPROXIMATIONS IN THE COMBINED-SOURCES DOMAIN

It is natural to seek to quantify the approximation in (7) as it is written in the combined-sources domain. In [14],[16] exponents are evaluated through three different measures that are applied in the combined-sources, exponentiated-magnitude domain. These measures are Kullback-Liebler divergence (KL) [17], Itakura-Saito divergence (IS) [17] and α-dispersion (D_α). KL and IS have both been established as useful ways to compare spectra in numerous applications. In [14],[16] they are applied in the exponentiated magnitude domain:

$$KL(r_1, r_2, \theta) = \mathcal{A}^\alpha(\cdot) \log\left(\frac{\mathcal{A}^\alpha(\cdot)}{\mathcal{M}^\alpha(\cdot)}\right) - \mathcal{A}^\alpha(\cdot) + \mathcal{M}^\alpha(\cdot), \quad (8)$$

$$IS(r_1, r_2, \theta) = \frac{\mathcal{A}^\alpha(\cdot)}{\mathcal{M}^\alpha(\cdot)} - \log\left(\frac{\mathcal{A}^\alpha(\cdot)}{\mathcal{M}^\alpha(\cdot)}\right) - 1. \quad (9)$$

In [14],[16] D_α is used to compute absolute error between $\mathcal{A}$ and $\mathcal{M}$ in the exponentiated magnitude domain, then map the result back to the magnitude domain:

$$D_\alpha(r_1, r_2, \theta, \alpha) = \left|\mathcal{A}(r_1, r_2, \theta)^\alpha - \mathcal{M}(r_1, r_2, \alpha)^\alpha\right|^{\frac{1}{\alpha}}. \quad (10)$$

To capture this error directly in the magnitude domain we now introduce another measure: absolute approximation error (AAE),

$$AAE(r_1, r_2, \theta, \alpha) = \left|\mathcal{A}(r_1, r_2, \theta) - \mathcal{M}(r_1, r_2, \alpha)\right|^p, \quad (11)$$

with $p = 1$. Squared approximation error (SAE), also defined by (11) but with $p = 2$, provides another measure.

We have applied all five measures to six different signal classes. For maximum generality these signal classes cover actual audio signals and statistical models for audio signalsd . We created $x_1(t)$ and $x_2(t)$ to use in $\mathcal{M}$ and $x(t) = x_1(t) + x_2(t)$ (0 dB mixing ratio) which exactly produces $\mathcal{A}$. For the first two signal classes we processed a wide variety of music and speech signals. For the next three classes we used white sequences of random variables. For each signal class we used 10 minutes of signals sampled at $f_s = 48kHz$ for a total of $N \approx 3 \times 10^7$ time-domain samples. We used the Hann window and a DFT frame length of 512 samples (10.6 ms) with 50% frame overlap. Note that the Gaussian and Laplacian time-domain sequences both produce complex DFT results with independent Gaussian real and imaginary parts, resulting in Rayleigh magnitude distributions. For the sixth signal class (motivated by III-B of [14]) we generated independent frequency-domain real and imaginary parts from the Cauchy distribution. In every case we averaged results over all time-frequency samples except DC and Nyquist. We searched the range $0 < \alpha \leq 3.2$ and show minimizing values of α in Table 1.

Signal	KL	IS	D_α	AAE	SAE
Music	0.470	0.830	1.026	2.010	2.684
Speech	0.629	0.826	1.023	2.011	2.771
Gauss	≈ 0.000	0.668	1.045	2.000	2.719
Laplace	≈ 0.000	0.668	1.045	2.000	2.721
Cauchy (TD)	1.018	0.861	1.017	1.998	2.660
Cauchy (FD)	2.050	0.795	1.003	2.001	3.128

Table 1: Minimizing values of α for six signal classes and five measures, applied in the combined-sources domain.

These results demonstrate empirically the arguments in Section 2. Across the five measures there is no basis for linking particular values of α to particular source distributions. There is no measurement that even leans toward suggesting that Gaussian sources are better served by $\alpha = 2$ and and Cauchy sources by $\alpha = 1$. This is not a paradox nor is it counter-intuitive, it is simply a reflection of the underlying geometry of the true approximation problem.

In [14],[16] KL and IS are applied in the exponentiated magnitude domain. It is also natural to apply them in the magnitude domain since that might be more indicative of a real application. We replaced $\mathcal{A}^\alpha$ with $\mathcal{A}$ and $\mathcal{M}^\alpha$ with $\mathcal{M}$ everywhere in (8) and (9). The resulting minimizing values of α range from 2.7 to 2.9 for KL and from 2.7 to 2.8 for IS. Once again we see no indications of relationships between α values and source distributions.

Table 1 also shows that D_α is almost completely invariant to magnitude distributions. We conclude this section with a mathematical explanation for this near-invariance. The expected value of D_α requires probability density functions (pdfs) for r_1, r_2, and θ given here by f_{r_1}, f_{r_2}, and f_θ. We have documented that f_θ is uniform and independent of r_1 and r_2. Since r_1 and r_2 are also independent of each other, the expectation can be factored:

$$\mathop{\mathbf{E}}_{r_1, r_2, \theta}(D_\alpha(r_1, r_2, \theta, \alpha)) = \int_0^\infty f_{r_1}(r_1) \int_0^\infty f_{r_2}(r_2) \left[\int_0^\pi f_\theta(\theta) D_\alpha(r_1, r_2, \theta, \alpha) \mathrm{d}\theta\right] \mathrm{d}r_2 \, \mathrm{d}r_1. \quad (12)$$

The bracketed portion of (12) is

$$\int_0^\pi f_\theta(\theta) D_\alpha(r_1, r_2, \theta, \alpha) \mathrm{d}\theta = \frac{r_1}{\pi} \int_0^\pi D_\alpha(1, \frac{r_2}{r_1}, \theta, \alpha) \mathrm{d}\theta = \frac{r_1}{\pi} \int_0^\pi \left|(1 + \gamma^2 + 2\gamma\cos(\theta))^{\frac{\alpha}{2}} - (1 + \gamma^\alpha)\right|^{\frac{1}{\alpha}} \mathrm{d}\theta = \frac{r_1}{\pi} \overline{D}_\alpha(\gamma, \alpha). \quad (13)$$

$\overline{D}_\alpha(\gamma, \alpha)$ represents an integral that has no closed form solution but numeric integration is straight-forward and reveals that for any γ, $\overline{D}_\alpha(\gamma, \alpha)$ is always minimized by some α in $(1.000, 1.074)$.

Substituting $\overline{D}_\alpha(\gamma, \alpha)$ back into (12) shows that the expected value of α-dispersion is found by scaling $\overline{D}_\alpha$ with positive values and adding (integrating). Given that $\overline{D}_\alpha$ is always minimized by α in $(1.000, 1.074)$ and in light of (12) it is not hard to show that the expected value of α-dispersion will also always be minimized by an α value in $(1.000, 1.074)$ regardless of the pdfs of the sources' spectral magnitudes.

This paper was presented at and will appear in the Proceedings of the
2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 15-18, 2017.

4. EVALUATING APPROXIMATIONS IN THE ESTIMATED-SOURCE DOMAIN PROVIDES INSIGHT

Section 3 demonstrates the difficulty and ambiguity associated with evaluating magnitude exponents in the combined-sources domain. It is much more effective to evaluate magnitude exponents in the estimated-source domain, ideally invoking measurements that are consistent with human perception and judgment. Several investigations [13]-[16] have followed this approach, covering five different applications and invoking five different measurement tools [18]-[23]. Optimal values of α in $[1.0, 1.4]$ are commonly reported. For the case of spectral subtraction (3) we now offer a more general way to evaluate magnitude exponents in the estimated-source domain. This approach reproduces and illuminates the preference for the range of alpha values near and slightly above 1.0. It shows that this range is a consequence of geometry and is driven by the phase distribution.

The extremely straightforward and well-motivated way to evaluate just the approximation (7) in the context of spectral subtraction is to consider the error between r_1 and its estimate $\hat{r_1}$, as generated by (3) using the observation r and the oracle value of r_2 in place of the estimate $\hat{r_2}$:

$$\Delta(r_1, r_2, \theta, \alpha) = \left| r_1 - (\max(\zeta, \mathcal{A}(r_1, r_2, \theta)^\alpha - r_2^\alpha))^{\frac{1}{\alpha}} \right|. \quad (14)$$

(For the remainder of this paper r_1 and r_2 are not interchangeable and we do not require $r_1 \geq r_2$.) By using oracle values of r_2, Δ captures the absolute error in the estimated magnitude of r_1 due only to the approximation (7). Both Δ and Δ^2 are largely insensitive to the choice of ζ and this is consistent with real applications. In this application both KL and IS would include division by ζ making them very sensitive to the choice of ζ and thus poor choices for comparing r_1 and $\hat{r_1}$.

We calculated empirical expected values for Δ and Δ^2 using the same signal classes and processes described in Section 3. The resulting minimizing values of α (for ζ=0) are given in Table 2. As

Signal	$\mathbf{E}(\Delta)$	$\mathbf{E}(\Delta^2)$
Music	1.180	1.331
Speech	1.162	1.327
Gauss	1.182	1.337
Laplace	1.182	1.338
Cauchy (TD)	1.171	1.242
Cauchy (FD)	1.146	1.215

Table 2: Minimizing values of α for six signal classes and two measures, applied in the estimated-source domain.

expected, Table 2 shows again that source distributions are of little consequence. The range of minimizing α values shown in the table is consistent with those identified in [13]-[15] by considering fidelity of the separated sources. In addition, we can now understand why this range is best. Six observations follow directly from (14), the geometry of the problem, and observations in Section 2:

1. For any r_1 and r_2, even though θ is uniformly distributed, $\sqrt{r_1^2 + r_2^2 + 2r_1 r_2 \cos(\theta)}$ is *not* uniformly distributed and always has higher probability density at the top of its range $(r_1 + r_2)$ than at the bottom of its range $(|r_1 - r_2|)$.
2. Estimates of r_1 increase with α.
3. $\alpha < 1$ produces only underestimations of r_1.
4. As α increases from 1 to 2 the faction of underestimations falls rapidly from 100% to 50% (overestimations increase from 0 to 50%). The changes are rapid due to the nonuniform pdf noted in observation 1.
5. As α increases above 2, the fraction of underestimations slowly reaches an asymptotic value in the range 41 to 45% depending on the signal class (overestimations reach 55 to 59%).
6. Magnitudes of underestimations are limited by the max function in (14) but the magnitudes of overestimations are unbounded.

As shown in Fig. 3, the result is that as α increases above 1.0 the total contribution of underestimations decreases ever more slowly while the total contribution of overestimations increases from zero at a nearly constant rate. Thus the balance between underestimations and overestimations that minimizes average absolute error ($\mathbf{E}(\Delta)$) is found not far above $\alpha = 1.0$ and consideration of average squared error ($\mathbf{E}(\Delta^2)$) shifts that balance point to a slightly higher value of α. Figure 3 is based on 10 minutes of speech signals and results for the other five signal classes are nearly identical.

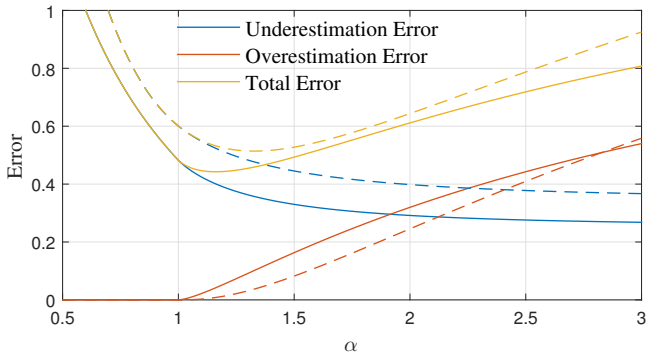

Figure 3: Solid lines show $\mathbf{E}(\Delta)$ and its components, dashed lines show $\mathbf{E}(\Delta^2)$ and its components (scaled down for display).

On average, added time-frequency components are more perceptually significant than missing time-frequency components (see equation (4) in [24]). One reason is that they can produce new and separate auditory precepts often described as artifacts. This may motivate placing a larger cost on overestimation than on underestimation and doing so will shift the balance point down, but never below 1.0, because at $\alpha = 1.0$ overestimation has been completely eliminated. This observation is consistent with reports that larger α values produce more artifacts [13],[16].

5. CONCLUSION

Manipulating spectral magnitudes raised to some power α is a common, effective, and practical technique for separating sources. Those who wish to optimize α must acknowledge the underlying approximation inherent in this approach and shift their attention from the distribution of magnitudes to the distribution of phases because it is the phase distribution that dominates the approximation. The approximation is best evaluated in the estimated-source domain rather than the combined-sources domain. Our evaluation in the estimated-source domain provides an intuitive, mathematical, and perceptually-consistent explanation for the optimality of α at or slightly above 1.0.

This paper was presented at and will appear in the Proceedings of the
2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 15-18, 2017.

6. REFERENCES

[1] S. Boll, "Suppression of noise in speech using the SABER method," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Apr. 1978, pp. 606–609.

[2] ——, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, Apr. 1979, pp. 208–211.

[4] J. Laroche, "Removing preechoes from audio recordings," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1995, pp. 147–150.

[5] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 328–337, Jul. 1998.

[6] V. Schless and F. Class, "SNR-dependent flooring and noise overestimation for joint application of spectral subtraction and model combination," in *International Conference on Spoken Language Processing*, 1998, pp. 721–725.

[7] L.P. Yang and Q.J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *The Journal of the Acoust. Society of America*, vol. 117, 2005.

[8] P. Loizou, *Speech Enhancement, Theory and Practice*. Boca Raton, Florida: CRC Press, 2013.

[9] S. Voran, "Observations on auditory excitation and masking patterns," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1995, pp. 206–209.

[10] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki, "Noise reduction based on adaptive β-order generalized spectral subtraction for speech enhancement," in *Interspeech 2007*, Aug. 2007, pp. 802–805.

[11] J. Li, Q.J. Fu, H. Jiang, and M. Akagi, "Psychoacoustically-motivated adaptive β-order generalized spectral subtraction for cochlear implant patients," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2009, pp. 4665–4668.

[12] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1770–1779, Aug. 2011.

[13] B. King, C. Fevotte, 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*, Sep. 2012.

[14] A. Liutkus and R. Badeau, "Generalized Wiener filtering with fractional power spectrograms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2015, pp. 266–270.

[15] S. Voran, "Exploration of the additivity approximation for spectral magnitudes," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2015.

[16] C. Dittmar, J. Driedger, M. Müller, and J. Paulus, "An experimental approach to generalized Wiener filtering in music source separation," in *24th European Signal Processing Conference*, Aug. 2016, pp. 1743–1747.

[17] D. FitzGerald, M. Cranitch, and E. Coyle, "On the use of the beta divergence for musical source separation," in *IET Irish Signals and Systems Conference*, June 2009.

[18] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, July 2006.

[19] R. Huber and B. Kollmeier, "PEMO-Q — A new method for objective audio quality assessment using a model of auditory perception," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006.

[20] ITU-T Recommendation P.862.2, "*Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs*," Geneva, 2007.

[21] ITU-R Recommendation BS.1387, "*Method for Objective Measurements of Perceived Audio Quality*," Geneva, 2001.

[22] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," McGill University, Tech. Rep., Dec. 2003. http://www.tsp.ece.mcgill.ca

[23] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2046–2057, Sep. 2011.

[24] S. Voran, "A simplified version of the ITU algorithm for objective measurement of speech codec quality," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, May 1998, pp. 537–540.

This paper was presented at and will appear in the Proceedings of the
2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 15-18, 2017.