BLIND BROADBAND SOURCE LOCALIZATION AND SEPARATION IN MINIATURE SENSOR ARRAYS

Gert Cauwenberghs[1], Milutin Stanacevic[1], and George Zweig[2,3]

[1]ECE/CLSP, Johns Hopkins University, Baltimore, MD 21218, USA
[2]RLE, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]LANL, Theory Division, Los Alamos, NM 87545, USA
E-mail: gert@jhu.edu; miki@jhu.edu; zweig@mit.edu

ABSTRACT

We study the ability of a sensor array to blindly separate and localize broadband traveling waves impinging on the array, with additive sensor noise. We consider arrays smaller than the shortest wavelength in the sources, such as MEMS acoustic arrays or VLSI arrays of RF receivers. A series expansion about the center of the array of the time-delayed signals emanating from the sources reduces the problem of separating and localizing the delayed sources to that of separating instantaneous signal mixtures using conventional tools of Independent Component Analysis. The covariance of the noise in the estimated sources is expressed in terms of the covariance of the sensor noise and the angular direction of the sources. Physical simulations demonstrate separation and localization of three non-coplanar speech sources using a planar array of four sensors within a 1 mm radius.

1. INTRODUCTION

Wavefront sensing in space for localizing sound has been in practice since the pioneering work by Blumlein in the 1930s [1], a precursor to the advances in binaural signal processing that we know today. The direction of a traveling wave can be inferred directly by sensing spatial differentials on a sub-wavelength scale, a principle exploited in biology, *e.g.,* for localizing prey emitting or reflecting high-frequency sound [2, 3], and implemented in biomimetic MEMS systems [4].

Super-resolution spectral methods are commonly used to localize multiple narrowband sources [5]. Yet little is known about the problem of localizing and separating multiple broadband sources. Separating mixtures of delayed sources has been addressed with Independent Component Analysis (ICA), *e.g.,* [6, 7], but requires a large number of parameters to obtain sufficient temporal resolution for precise localization. Adapting delays [8] reduces the number of parameters, but is prone to local optima.

The approach we present here avoids the problem of separating delayed mixtures by expanding the delayed signals in a series of terms that only contain the instantaneous signals and their time derivatives. By taking spatial derivatives, we are able to isolate terms of various orders in the temporal derivatives, all indirectly contributing observations of linearly independent instantaneous mixtures of the source signals. This formulation is equivalent to

that of standard ICA, and a number of approaches exist for such blind separation, some utilizing VLSI hardware [9]. The mixing coefficients obtained from ICA yield the angles of the incoming waves. Therefore our method can be seen as an combination of the wave sensing idea by Blumlein, and ICA, performing at once blind separation *and* localization of traveling waves.

In a physical setting with an array of sensors, the spatial derivatives are estimated on a grid, and the sensor observations contribute additive noise. We analyze the noise characteristics of the estimated source signals by ICA, and present simulated results that show successful separation and localization of speech sources using a planar geometry of sensors much smaller than the shortest wavelength present in the speech. The technique can be used to separate multiple signals with miniature distributed sensors or sensor arrays that are integrated on a single MEMS or VLSI chip, *e.g.,* [4].

2. PHYSICAL MODELING

2.1. Wave Propagation

We consider linear mixtures of traveling waves emitted by sources at various locations, and observed over a distribution of sensors in space. The distribution of sensors could be continuous or discrete. In what follows we assume an array of discrete sensors, but the theory applies as well to sensors distributed continuously in space. However, the sources are assumed to be discrete.

2.2. Instantaneous Series Expansion

The usual approach to wideband separation tries to find the sources by combining the received signals at multiple delayed times. This is computationally expensive. The approach proposed here reduces the problem of separating mixtures of delayed sources to that of separating an instantaneous mixture of signals related to the sources through a succession of time derivatives of increasing order.

Let the coordinate system $\mathbf{r}$ be centered in the array so that the origin coincides with the "center of mass" of the sensor distribution. We define $\tau(\mathbf{r})$ as the time lag between the wavefront at point $\mathbf{r}$ and the wavefront at the center of the array, *i.e.,* the propagation time $\tau(\mathbf{r})$ is referenced to the center of the array. Then the field $s(t + \tau(\mathbf{r}))$ can be expanded about the center of the array in the power series expansion,

$$s(t + \tau(\mathbf{r})) = s(t) + \tau(\mathbf{r})\dot{s}(t) + \tfrac{1}{2}\tau(\mathbf{r})^2 \ddot{s}(t) + \ldots \quad (1)$$
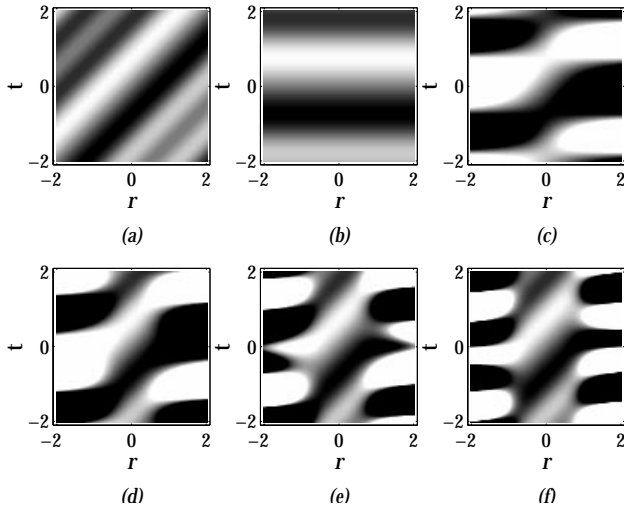
Figure 1: (a)*: Far-field wave traveling to the right* $(u = -1, c = 1)$*. (b) through (f): Series expansions of order 0 through 4 around the origin* $(r = 0)$*.*

2.3. Limiting Cases of Interest

Two types of approximations are in order, each with a different physical origin.

2.3.1. The far field

In the *far-field* approximation, the distance from the source is much larger than the dimensions of the sensor array. This is a sensible approximation for an integrated MEMS or VLSI array with dimensions typically smaller than 1 cm. Then the wavefront delay $\tau(\mathbf{r})$ is approximately linear in the projection of $\mathbf{r}$ on the unit vector $\mathbf{u}$ pointing towards the source,

$$\tau(\mathbf{r}) \approx \frac{1}{c}\mathbf{r} \cdot \mathbf{u} \qquad (2)$$

where c is the speed of (acoustic or electromagnetic) wave propagation. An example traveling wave in the far field, and its series expansion, are illustrated in Figure 1.

2.3.2. Wave resolution

The second condition, which we term *wave resolution*, pertains to the relative size of terms in the series expansion, or the scale on which the signals are sampled. In fact, there are two conditions that need to be satisfied simultaneously, bounding the dimensions of the array from below and above.

The Fourier transform of (1) in time corresponds to a frequency-domain transfer function with series approximation

$$\exp(+j\omega\tau(\mathbf{r})) = 1 + j\omega\tau(\mathbf{r}) - \frac{1}{2}\big(\omega\tau(\mathbf{r})\big)^2 + \dots . \qquad (3)$$

The series (3) converges uniformly for all ω, but the number of terms needed increases rapidly with $|\omega\tau(\mathbf{r})|$. On the other hand, for purposes of *separating* the sources from physical observations of individual terms in the series, the ratio of consecutive terms in the series cannot be too small. A trade-off between approximation and resolution power is thus obtained when $\epsilon \leq |\omega\tau(\mathbf{r})| \leq \alpha$

where ϵ is sufficiently above the observation noise floor, and α is not too large.

In the far-field approximation, $\omega\tau(\mathbf{r}) = -\mathbf{k} \cdot \mathbf{r}$ where $\mathbf{k}$ is the wave vector, $\mathbf{k} = -\frac{\omega}{c}\mathbf{u} = -\frac{2\pi}{\lambda}\mathbf{u}$ and the *wave resolving* condition translates to

$$\epsilon\lambda_{\max} \leq |\mathbf{r}| \leq \alpha\lambda_{\min}, \qquad (4)$$

where $\lambda_{\min} = \frac{c}{f_{\max}}$ is the shortest wavelength present in the spectrum of the traveling wave and $\lambda_{\max} = \frac{c}{f_{\min}}$ is the largest. In other words, we consider array dimensions not much larger than the smallest wavelength, but not too small compared with the largest wavelength. Note that α in the upper limit depends on the number of terms in the expansion, as illustrated in Figure 1.

3. STATISTICAL MODELING

To separate a mixture of signals, it is necessary to specify assumptions on the signals and the way they are mixed and observed.

3.1. Signal Model

We assume that the sources are statistically independent so that their joint probability density function factors:

$$\Pr(\mathbf{s}) = \prod_{\ell=1}^{L} \varphi^{\ell}(s^{\ell}). \qquad (5)$$

This allows us to apply ICA to the instantaneous mixture problem that follows.

3.2. Mixing and Acquisition Model

Let $x(\mathbf{r}, t)$ be the signal mixture picked up by a sensor at position $\mathbf{r}$. As one special case we will consider a two-dimensional array of sensors, with position coordinates p and q so that $\mathbf{r}_{pq} = p\mathbf{r}_1 + q\mathbf{r}_2$ with orthogonal vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ in the sensor plane.

In the *far-field* approximation (2), each source signal s^{ℓ} contributing to x_{pq} is advanced in time by $\tau_{pq}^{\ell} = p\tau_1^{\ell} + q\tau_2^{\ell}$, where

$$\tau_1^{\ell} = \frac{1}{c}\mathbf{r}_1 \cdot \mathbf{u}^{\ell}$$
$$\tau_2^{\ell} = \frac{1}{c}\mathbf{r}_2 \cdot \mathbf{u}^{\ell} \qquad (6)$$

are the inter-time differences (ITD) of source ℓ between adjacent sensors on the grid along the p and q place coordinates, respectively. Knowledge of the *angle coordinates* τ_1^{ℓ} and τ_2^{ℓ} uniquely determines, through (6), the direction vector $\mathbf{u}^{\ell}$ along which source s^{ℓ} impinges the array, in reference to the $\{p, q\}$ plane[1].

Under the *wave resolving* condition, the series expansion (1) for each source yields

$$x_{pq}(t) = \sum_{\ell=1}^{L} s^{\ell}(t) + \tau_{pq}^{\ell}\dot{s}^{\ell}(t) + \frac{1}{2}(\tau_{pq}^{\ell})^2\ddot{s}^{\ell}(t) + \dots + n_{pq}(t) \quad (7)$$

where $n_{pq}(t)$ represents additive noise in the sensor observations. Although not essential, we will assume that the observation noise

---

[1]We assume that the sources impinge on top, not on bottom, of the array. This is a reasonable assumption for an integrated MEMS or VLSI array since the substrate masks any source impinging from beneath.

is independent across sensors, and follows a univariate Gaussian distribution $n_{pq}(t) \propto \mathcal{N}(0, \sigma)$.

In what follows we will concentrate on the first two terms in the series expansion (7), linear in the space coordinates:

$$x_{pq}(t) \approx \sum_{\ell=1}^{L} s^{\ell}(t) + (p\tau_1^{\ell} + q\tau_2^{\ell})\dot{s}^{\ell}(t) + n_{pq}(t) . \qquad (8)$$

4. BLIND SEPARATION AND LOCALIZATION

Based on the models above, we are now ready to address the problem of inferring the unknown source signals and their unknown angular coordinates. In principle we could consider MAP (*Maximum A Posteriori*) estimation based on a generative model of the signals (5) and the mixing model (8). Since the mixing model contains time derivatives of various order in the signals, the MAP estimates of the signals $s^{\ell}(t)$ lead to a set of coupled differential equations that are computationally involved.

4.1. Spatial gradients and ICA

We proceed with an alternative, *gradient*-based method, that isolates time derivatives of the linearly combined signals by taking spatial gradients of x along p and q. The advantage of this technique is that it effectively reduces the problem of estimating $s^{\ell}(t)$ and τ_i^{ℓ} to that of separating instantaneous mixtures of the independent source signals.

Under *wave-resolving* conditions, individual terms in the series expansion (7) can be resolved. Different linear combinations[2] of the signals s^l are thus obtained by taking spatial derivatives of various orders i and j along the position coordinates p and q, around the origin $p = q = 0$:

$$
\begin{aligned}
\xi_{ij}(t) &\equiv \left. \frac{\partial^{i+j}}{\partial^i p \partial^j q} x_{pq}(t) \right|_{p=q=0} \\
&= \sum_{\ell} (\tau_1^{\ell})^i (\tau_2^{\ell})^j \frac{d^{i+j}}{d^{i+j}t} s^{\ell}(t) + \nu_{ij}(t), \qquad (9)
\end{aligned}
$$

where ν_{ij} are the corresponding spatial derivatives of the sensor noise n_{pq} around the center. The point here is that all signals s^l in (9) are differentiated to the *same* order $i + j$ in time. Therefore, taking spatial derivatives ξ_{ij} of order $i + j \leq k$, and differentiating ξ_{ij} to order $k - (i + j)$ in time yields a number of different linear observations in the kth-order time derivatives of the signals s^{ℓ}.

In practice, spatial derivatives (9) are approximated by discrete sampling on the grid $x_{pq}(t)$. Finding the proper sampling coefficients on a grid to approximate derivatives is a well studied problem in digital signal processing [12]. For dense sensor arrays, an alternative is to approximate the derivatives using moments over the sensor distribution, giving estimates that are more robust to noise. Distributed sensors that acquire spatial gradients directly are also implementable in MEMS technology [4].

As an example, consider the first-order case $k = 1$, corresponding to (8):

$$\xi_{00}(t) = \sum_{\ell} s^{\ell}(t) + \nu_{00}(t),$$

---

[2]The issue of linear independence will be revisited when we consider the geometry of the source angles relative to that of the sensors in Section 4.2.

$$
\begin{aligned}
\xi_{10}(t) &= \sum_{\ell} \tau_1^{\ell} \dot{s}^{\ell}(t) + \nu_{10}(t), \qquad (10) \\
\xi_{01}(t) &= \sum_{\ell} \tau_2^{\ell} \dot{s}^{\ell}(t) + \nu_{01}(t).
\end{aligned}
$$

Estimates of ξ_{00}, ξ_{10} and ξ_{01} are obtained with just four sensors x_{pq}:

$$
\begin{aligned}
\xi_{00} &\approx \tfrac{1}{4}\left(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1} \right) \\
\xi_{10} &\approx \tfrac{1}{2}\left(x_{1,0} - x_{-1,0} \right) \qquad (11) \\
\xi_{01} &\approx \tfrac{1}{2}\left(x_{0,1} - x_{0,-1} \right)
\end{aligned}
$$

Taking the time derivative of ξ_{00}, we thus obtain from the sensors a linear instantaneous mixture of the time-differentiated source signals,

$$
\begin{bmatrix} \dot{\xi}_{00} \\ \xi_{10} \\ \xi_{01} \end{bmatrix} \approx
\begin{bmatrix} 1 & \cdots & 1 \\ \tau_1^1 & \cdots & \tau_1^{\ell} \\ \tau_2^1 & \cdots & \tau_2^{\ell} \end{bmatrix}
\begin{bmatrix} \dot{s}^1 \\ \vdots \\ \dot{s}^{\ell} \end{bmatrix} +
\begin{bmatrix} \dot{\nu}_{00} \\ \nu_{10} \\ \nu_{01} \end{bmatrix}, \qquad (12)
$$

an equation in the standard form $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$, where $\mathbf{x}$ is given and the mixing matrix $\mathbf{A}$ and sources $\mathbf{s}$ are unknown. Ignoring for now the noise term $\mathbf{n}$ (and for a square matrix, $\ell = 3$) this problem setting is standard in ICA, with an independence assumption (5) on the sources $\mathbf{s}$. ICA produces, at best, an estimate $\hat{\mathbf{s}}$ that recovers the original sources $\mathbf{s}$ up to arbitrary scaling and permutation. The direction cosines τ_i^{ℓ} are found from the ICA estimate of $\mathbf{A}$, after first normalizing each column (*i.e.,* , each source estimate) so that the first row of the estimate $\hat{\mathbf{A}}$, like the real $\mathbf{A}$ according to (12), contains all ones. This simple procedure together with (6) yields estimates of the direction vectors $\hat{\mathbf{u}}^{\ell}$ along with the source estimates $\hat{s}^{\ell}(t)$, which are obtained by integrating the components of $\hat{\mathbf{s}}$ over time and removing the DC components.

It is interesting to note the similarity between (12), with $\ell = 1$, and *optical flow* for constraint-solving velocity estimation in a visual scene [13].

4.2. Noise Characteristics

The presence of the noise term $\mathbf{n}$ complicates the estimation of $\mathbf{s}$ and $\mathbf{A}$. Assume for now a standard formulation of ICA (*e.g.,* [7]) that attempts to linearly unmix the observations $\mathbf{x}$:

$$\hat{\mathbf{s}} = \hat{\mathbf{A}}^{-1}\mathbf{x} = \hat{\mathbf{A}}^{-1}\mathbf{A}\mathbf{s} - \hat{\mathbf{A}}^{-1}\mathbf{n}, \qquad (13)$$

where $\mathbf{A}$ is square and invertible. Assume also a reasonable ICA estimate $\hat{\mathbf{A}}$ so that (13) reduces to $\hat{\mathbf{s}} \approx \mathbf{s} - \mathbf{A}^{-1}\mathbf{n}$, disregarding arbitrary permutation and scaling in the source estimates. The error term $\mathbf{e} \equiv -\mathbf{A}^{-1}\mathbf{n}$ contributes *variance* to the estimate $\hat{\mathbf{s}}$; in general the noise $\mathbf{n}$ will also affect the estimate $\hat{\mathbf{A}}$ and produce a *bias* term in $\hat{\mathbf{s}}$ according to (13).

The functional form of the error $\mathbf{e}$ allows us to estimate the noise characteristics of the source estimates, without considering details on how ICA obtained these estimates. The covariance of the estimation error is $E[\mathbf{e}\mathbf{e}^T] = \mathbf{A}^{-1}E[\mathbf{n}\mathbf{n}^T](\mathbf{A}^{-1})^T$. In other words, the error covariance depends on the covariance of the sensor noise, the geometry of the sensor array, and the orientation of the sources $\mathbf{u}^{\ell}$ as determined by the mixing matrix $\mathbf{A}$.

For example, consider the case $k = 1$, suitable for a miniature array. $\mathbf{A}$ in (12) is square when $\ell = 3$. The determinant of $\mathbf{A}$ can be geometrically interpreted as the volume of the polyhedron spanned by the three source direction vectors $\mathbf{u}^{\ell}$. The error covariance is minimum when the vectors are orthogonal, and
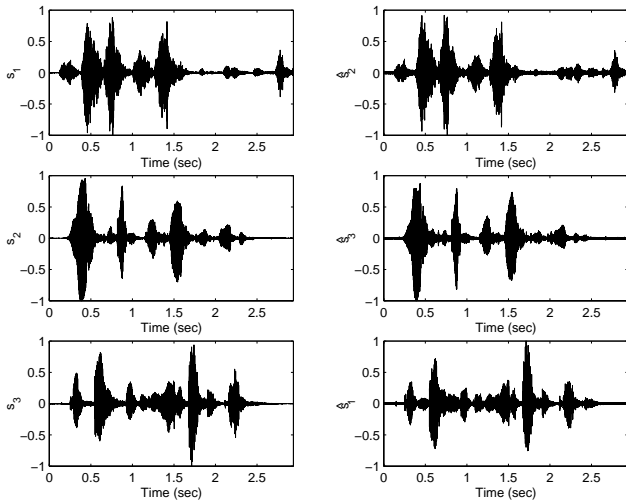
Figure 2: *Blind identification and localization of three speech sources using a simulated array of four co-planar sensors within a 1 mm radius.* Left: *source signals* $s_i(t)$. Right: *reconstructed signals* $\hat{s}_i(t)$*, normalized.*

the estimates of $\mathbf{s}$ and $\mathbf{A}$ become unreliable as the source direction vectors $\mathbf{u}^\ell$ approach the same plane. Therefore, to first order ($k = 1$) at most three *non-coplanar* sources can be separated and localized with a planar array of sensors.

For arrays of larger dimensions, *wave-resolving* conditions support a larger number of terms k, and thus a larger number of sources ℓ, given by the number of mixture observations up to order k in (9). The condition that $\mathbf{A}$ as determined by (9) be full rank amounts to constraints on the geometry of the source direction vectors $\mathbf{u}^l$. For instance, only one source can lie along any given direction $\mathbf{u}$.

When the number of sources present is greater than the number of gradient observations $\ell_{\max}$ in (9), separation and localization is still possible, but requires an informative *prior* on the sources (5). In particular, a sparse ICA decomposition is obtained in the overcomplete case $\ell > \ell_{\max}$ by using a Laplacian prior on the sources [10]. For example, overcomplete ICA could be directly applied on the mixture (12) to separate more than three *sparse* sources.

4.3. Results

To demonstrate that the proposed algorithms both localize and separate, we simulated the signals received by a hypothetical array of four sensors in the plane as specified by (11), with radius $|\mathbf{r}_1| = |\mathbf{r}_2| = 1$ mm. Three speech waveforms, sampled at 16 kHz, were used for the source signals With these parameters, the time delays between sensors amount to about 10 % of the sampling time, satisfying the *wave-resolving* conditions. Gaussian noise was added to the sensor signals, with a signal-to-noise ratio (SNR) of 50 dB for each sensor.

Results of the simulated experiment are given in Figure 2. The reconstructed sources $\hat{s}_i(t)$ are affected by additive noise, but the SNR is comparable to that of the sensors, and no cross-talk is audible. The extracted direction vectors from $\hat{\mathbf{A}}$ are indistinguishable from the simulated directions $\mathbf{u}^\ell$.

5. CONCLUSIONS

A method for localizing and separating broadband sources in space by measuring spatial and temporal derivatives of the field over a sensor array has been described. The essential point is that by making a power series expansion of the field about the center of the array and identifying the terms in the series with measurements of the spatial gradients, a set of instantaneous equations in the time derivatives of the source signals is obtained. These instantaneous equations can then be solved with standard ICA methods. The ICA solution yields estimates of the sources, along with the cosines of the angles of the direction vectors of the sources with respect to the coordinate axes of the array.

The requirement that the terms in the power series be amenable to separation places upper and lower bounds on the size of the array, smaller than the shortest wavelength but not too small, as given by (4). The number of sources that can be extracted depends strongly on the number of resolvable terms in the series. With just zero and first order terms in the expansion, three independent and non-coplanar sources can be extracted with as few as four planar sensors.

6. REFERENCES

[1] A.D. Blumlein, "Improvements in and Relating to Sound-Transmission, Sound-Recording and Sound-Reproducing Systems," British Patent, 394325, 1933.

[2] D. Robert, R.N. Miles, and R.R. Hoy, "Tympanal Hearing in the Sarcophagid Parasitoid Fly Emblemasoma sp.: the Biomechanics of Directional Hearing," *J. Experimental Biology*, vol. 202, pp 1865-1876, 1999.

[3] J.A. Simmons, M.J. Ferragamo and C.F. Moss, "Echo-Delay Resolution in Sonar Images of the Big Brown Bat, Eptesicus-Fuscus," *Proc. Nat. Academy Sciences USA,* vol. **95** (21), pp 12647-12652, Oct. 1998.

[4] A.G. Andreou, D.H. Goldberg, E. Culurciello, M. Stanacevic and G. Cauwenberghs, "Heterogeneous Integration of Biomimetic Acoustic Microsystems," *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'2001),* Sydney, Australia, May 6-9, 2001.

[5] S. Haykin, *Adaptive Filter Theory,* 2nd ed., Prentice-Hall, 1991.

[6] S. Li and T. Sejnowski, "Adaptive Separation of Mixed Broadband Sound Sources with Delays by a Beamforming Herault-Jutten Network," *IEEE J. Oceanic Eng.,* vol. **20** (1), pp 73-79, Jan. 1995.

[7] A.J. Bell and T.J. Sejnowski, "An Information Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Comp.*, vol. **7** (6), pp 1129-1159, Nov 1995.

[8] K. Torkkola, "Blind Separation of Delayed Sources Based on Information Maximization," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'96), Atlanta GA, 1996.

[9] M.H. Cohen and A.G. Andreou, "Current-Mode Subthreshold MOS Implementation of Herault-Jutten Autoadaptive Network," *IEEE J. Solid-State Circuits,* vol. **27**, pp 714-727, May 1992.

[10] M.S. Lewicki and T.J. Sejnowski, "Learning Overcomplete Representations." *Neural Computation*, vol **12**, pp 337-365, 2000.

[11] T. Lee, A. Bell and R. Lambert, "Blind Separation of Delayed and Convolved Sources," *Adv. Neural Information Processing Systems*, vol. **9**, Cambridge MA: MIT Press, pp 758-764, 1997.

[12] L.R. Rabiner and R.W. Schafer, "On the Behavior of Minimax Relative Error FIR Digital Differentiators," *Bell System Technical Journal*, vol. **53**, pp 333-361, 1974.

[13] J.L. Barron, D.J. Fleet and S.S. Beauchemin, "Performance of Optical Flow Techniques," *International Journal of Computer Vision,* vol. **12** (1), pp 43-77, 1994.