

GRADIENT FLOW ADAPTIVE BEAMFORMING AND SIGNAL SEPARATION IN A MINIATURE MICROPHONE ARRAY

Milutin Stanacevic, Gert Cauwenberghs*

ECE Dept. and CLSP
Johns Hopkins University
Baltimore, MD 21218
{miki,gert}@jhu.edu

George Zweig

RLE, MIT, Cambridge, MA 02139;
LANL, Los Alamos, NM 87545
gzweig@mit.edu

ABSTRACT

Gradient flow converts the problem of separating unknown delayed mixtures of sources, from traveling waves impinging on an array of sensors, into a simpler problem of separating unknown instantaneous mixtures of the time-differentiated sources, obtained by acquiring or computing spatial and temporal derivatives on the array. The linear coefficients in the instantaneous mixture directly represent the delays, which in turn determine the direction angles of the sources. This formulation is attractive, since it allows to separate and localize waves of broadband signals using standard tools of independent component analysis (ICA), yielding the sources along with their direction angles. The technique is suited for arrays of small aperture, with dimensions shorter than the coherence length of the waves. We present gradient flow experiments on an array of four hearing aid microphones placed within a 5 mm radius, yielding 20 dB separation of joint speech in outdoors acoustic environments, and 10 dB separation indoors under mild reverberant conditions. These results suggest applications of gradient flow miniature microphone arrays to intelligent hearing aids with adaptive suppression of interfering signals and nonstationary noise.

1. INTRODUCTION

The human auditory system performs remarkably well in separating multiple acoustic streams in the presence of noise, such as babble noise in a cafeteria. It does so partly by resolving time and amplitude differences between the sound entering both ears, which are different for each source component. The impaired ear seems to lose most of the ability to function in the presence of interfering noise, and even the most advanced hearing aids with directional microphones and noise suppression have fallen short of compensating adequately for this loss.

Because of the wave properties of sound, the problem of separating acoustic sources is tightly coupled to the problem of localizing them by adaptive beamforming. For speech and other sound sources this problem is generally considered a hard and mostly unsolved problem, with mixed degree of success in practical realizations. The delays in the source mixture observations render instantaneous mixing models, as commonly assumed in independent component analysis (ICA), inadequate. The conventional approach is to blindly estimate multiple time delays in the wave propagation between sources and microphones, besides blindly esti-

imating the sources themselves [1, 2, 3]. Several systems based on this approach have been demonstrated on real speech, *e.g.*, [4, 5, 6, 7, 8, 9].

Like optical flow for motion estimation [10], the direction of sound propagation can be inferred directly from sensing spatial and temporal gradients of the wave signal, on a sub-wavelength scale. This principle is exploited in biology [11], and implemented in biomimetic MEMS systems [12]. Wavefront sensing in space for localizing sound has been in practice since the pioneering work by Blumlein in the 1930s [13], a precursor to the advances in binocular signal processing that we know today.

Gradient flow converts the problem of separating delayed mixtures into a simpler problem of separating instantaneous mixtures of spatial and temporal gradients of the wave signals. This yields a formulation equivalent to that of static ICA, and a number of approaches exist for such blind separation, some utilizing VLSI hardware [14]. The mixing coefficients obtained from ICA directly yield the angles of the incoming waves. Therefore our method can be seen as a broadband beamforming extension to static ICA, performing at once blind separation *and* localization of traveling waves.

Gradient flow acoustic source separation offers the possibility of selectively amplifying the source of interest and suppressing interfering sources. Furthermore, in an extended ICA framework that accounts for a convolutive mixing model [15, 16, 17], this solution allows cancellation of reverberation which further aids in the clarity of speech under difficult acoustic conditions.

In what follows we review gradient flow, and present experimental results that demonstrate real-world separation and localization of speech sources using a planar geometry of miniature microphones (Knowles IM-3268) on a spatial scale significantly smaller than the shortest wavelength present in the speech. Finally we discuss extensions of the principle that can be used in the design of intelligent hearing aids with adaptive beam steering and interference noise suppression.

2. MODELS

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.

*This work was partly supported by ONR N00014-99-1-0612 and ONR/DARPA N00014-00-C-0315 and N00014-00-1-0838.

2.1. Instantaneous Series Expansion

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,

$$s(t + \tau(\mathbf{r})) = s(t) + \tau(\mathbf{r})\dot{s}(t) + \frac{1}{2}\tau(\mathbf{r})^2\ddot{s}(t) + \dots \quad (1)$$

2.2. 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,

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

where c is the speed of (acoustic or electromagnetic) wave propagation.

2.3. Array dimensions

The ratio in amplitude of successive terms in the series should not be too large so that it converges properly, nor too small so an adequate number of terms in the series can be resolved to identify the sources. This implies that the dimensions of the sensor array should be smaller than that of the *coherence length* of the sources, but not much smaller. For a narrowband signal, the coherence length can be defined as the wavelength. For a narrowband signal, it corresponds to the shortest wavelength present in the spectrum of the signal. For speech with a bandwidth of roughly 8 kHz, an array of dimensions in the lower cm range gives best results. More details on the conditions for proper wave resolution are given in [18].

2.4. Signal Model

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

$$\Pr(\mathbf{s}) = \prod_{\ell=1}^{\mathcal{L}} \varphi^{\ell}(s^{\ell}). \quad (3)$$

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

2.5. 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

$$\begin{aligned} \tau_1^{\ell} &= \frac{1}{c} \mathbf{r}_1 \cdot \mathbf{u}^{\ell} \\ \tau_2^{\ell} &= \frac{1}{c} \mathbf{r}_2 \cdot \mathbf{u}^{\ell} \end{aligned} \quad (4)$$

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 (4), the direction vector \mathbf{u}^{ℓ} along which source s^{ℓ} impinges the array, in reference to the $\{p, q\}$ plane¹.

The series expansion (1) for each source yields

$$x_{pq}(t) = \sum_{\ell=1}^{\mathcal{L}} s^{\ell}(t) + \tau_{pq}^{\ell} \dot{s}^{\ell}(t) + \frac{1}{2}(\tau_{pq}^{\ell})^2 \ddot{s}^{\ell}(t) + \dots + n_{pq}(t) \quad (5)$$

where $n_{pq}(t)$ represents additive noise in the sensor observations. Although not essential, we will assume that the observation noise 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 (5), linear in the space coordinates:

$$x_{pq}(t) \approx \sum_{\ell=1}^{\mathcal{L}} s^{\ell}(t) + (p\tau_1^{\ell} + q\tau_2^{\ell})\dot{s}^{\ell}(t) + n_{pq}(t). \quad (6)$$

3. GRADIENT FLOW ICA

A gradient flow formulation is obtained by isolating 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 gradient components of the independent source signals.

For array dimensions smaller than but comparable to the largest wavelength (*wave-resolving* conditions [18]), individual terms in the series expansion (5) can be resolved. Different linear combinations² 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 $p = q = 0$:

$$\begin{aligned} \xi_{ij}(t) &\equiv \left. \frac{\partial^{i+j}}{\partial^i p \partial^j q} x_{pq}(t) \right|_{p=q=0} \\ &= \sum_{\ell} (\tau_1^{\ell})^i (\tau_2^{\ell})^j \frac{d^{i+j}}{dt^{i+j}} s^{\ell}(t) + \nu_{ij}(t), \end{aligned} \quad (7)$$

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 (7) 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 k th-order time derivatives of the signals s^{ℓ} .

As an example, consider the first-order case $k = 1$, corresponding to (6):

$$\begin{aligned} \xi_{00}(t) &= \sum_{\ell} s^{\ell}(t) + \nu_{00}(t), \\ \xi_{10}(t) &= \sum_{\ell} \tau_1^{\ell} \dot{s}^{\ell}(t) + \nu_{10}(t), \\ \xi_{01}(t) &= \sum_{\ell} \tau_2^{\ell} \dot{s}^{\ell}(t) + \nu_{01}(t). \end{aligned} \quad (8)$$

¹We 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.

²The 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.

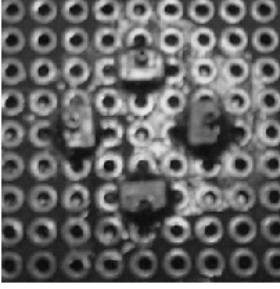


Fig. 1. Geometry of four co-planar sensors computing first-order gradient flow in two dimensions according to (9). Using four Knowles IM-3268 microphones, the array is contained within a 5 mm radius.

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

$$\begin{aligned}\xi_{00} &\approx \frac{1}{4}(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}) \\ \xi_{10} &\approx \frac{1}{2}(x_{1,0} - x_{-1,0}) \\ \xi_{01} &\approx \frac{1}{2}(x_{0,1} - x_{0,-1})\end{aligned}\quad (9)$$

A practical realization of this configuration, using miniature microphones, is illustrated in Figure 1.

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

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

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, $\mathcal{L} = 3$) this problem setting is standard in ICA, with an independence assumption (3) 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 (10), contains all ones. This simple procedure together with (4) 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 functional similarity between (10), with $\mathcal{L} = 1$, and *optical flow* for constraint-solving velocity estimation in a visual scene [10].

4. NOISE CHARACTERISTICS

The presence of the noise term \mathbf{n} complicates the estimation of \mathbf{s} and \mathbf{A} . For localization of a *single* source, simple expressions can be obtained for the Cramer-Rao lower bound on the variance of the τ_i estimates, assuming second-order (Gaussian) statistics. As in [19], this bound depends on the aperture, *i.e.*, the dimensions of the array relative to one wavelength. However, the bound becomes independent of aperture if the noise is dominated by interference

from other sources, rather than intrinsic (sensor) noise. For miniature arrays, it is therefore important to boost the sensor signal-to-noise ratio to compensate for the loss in aperture. The critical factor here is high sensitivity of gradient acquisition, which can be attained by a differential sensor design either through mechanical coupling [11, 12] or differential amplification at the sensor level. In the present implementation, improved differential amplification is obtained by adaptive LMS cancelation of the common mode component ξ_{00} present in the estimate of the gradients ξ_{10} and ξ_{01} due to gain mismatch in the observations x_{ij} in (8).

General Cramer-Rao bounds on the direction cosines τ_i^ℓ for joint localization and separation are harder to come by, since it requires assumptions on the higher-order statistics of the signals. The variance component of noise in the estimated sources can be readily estimated assuming nominal (unbiased) values of the direction cosines. A detailed analysis, expressing the covariance of the error in terms of the covariance of the sensor noise and the geometry of the sources relative to the array, is presented in [18].

5. EXPERIMENTAL SETUP AND RESULTS

To demonstrate source separation and localization in a real environment, we used the experimental setup of Figure 1, with four omnidirectional miniature microphones (Knowles IM-3268) as specified by (9) with radius $|\mathbf{r}_1| = |\mathbf{r}_2| = 1$ cm.

We conducted experiments in two distinct environments: outdoors surrounded by buildings at a 30 m distance in the presence of wind and other ambient noise, and indoors in a reverberant room with dimensions 8, 4, and 2.7 m with office furniture. Two male speakers were both 50cm from the microphone array and from each other. The recorded signals ξ_{00} , ξ_{10} , ξ_{01} , $\dot{\xi}_{00}$, along with the reconstructed sources \hat{s}_1 and \hat{s}_2 , for the outdoors case are shown in Figure 2. Listening tests and visual inspection of the waveforms reveal that the cross-talk in the unmixed signals \hat{s}_1 and \hat{s}_2 is less than -20 dB in the outdoors case, and -12 dB in the indoors case. This performance is adequate for most hearing aid applications. Further gains can be obtained by extending the traveling wave assumption to multipath, and assuming a convolutive mixing model solved by means of deconvolutive ICA [15, 16, 17].

6. CONCLUDING REMARKS

Conventional adaptive beamforming assumes arrays of large dimensions to obtain sufficient spatial diversity to resolve time delays between source observations. In gradient flow, arrays of small dimension are suitable since time delays have been converted to relative amplitude by taking derivatives. Improved differential sensitivity of gradient sensing allows to shrink the aperture of the sensor array without degrading signal-to-noise ratio.

A microphone array with differential gain enhancement would fit comfortably in a behind-the-ear, or even in-the-canal, hearing aid, and would deliver a performance similar to much larger arrays that are worn, for instance, around the neck [20]. Gradient sensitive acoustic sensors with differential mechanical coupling in MEMS [12] technology, with inspiration in biology [11], are especially promising candidates for miniaturization.

Gradient flow therefore offers an opportunity to design intelligent hearing aids with integrated adaptive suppression of unwanted speech and other sources of nonstationary noise and interfering sounds. With *auto-adaptive* directionality of beamforming in the gradient flow ICA framework, a miniature array is capable of

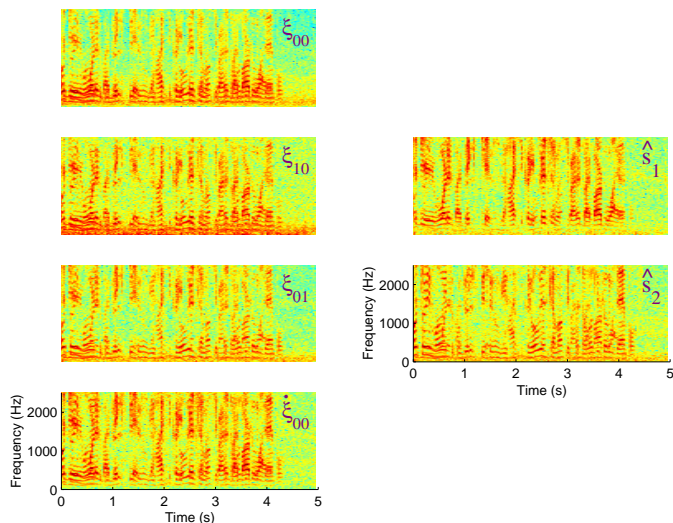


Fig. 2. Experimental separation of speech from two male speakers in an outdoors setting. Left: Components ξ_{00} , ξ_{10} , ξ_{01} and its time derivative $\dot{\xi}_{00}$, measured from the microphone array in Fig. 1. Right: Reconstructed sources \hat{s}_1 and \hat{s}_2 using static ICA on ξ_{10} , ξ_{01} and ξ_{00} .

blindly separating several sources. The separated sources can be reconstituted to the user with different weighting factors for improved intelligibility. For instance, the estimated direction cosines from gradient flow ICA can be used to identify and preferentially emphasize a frontal source among the separated components. With a binaural link between a pair of hearing aids, this weighting could be coordinated to preserve the relative time delays between ears for each source. In more demanding environments, such as the infamous “cocktail party,” a menu-driven interface could allow the user to switch emphasis between speakers of interest. Individual speakers could be identified over an extended time window based on distinct properties of speech such as pitch and timbre. These capabilities by far outpace what is available on the hearing aid market today.

7. REFERENCES

- [1] 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, no. 1, pp. 73–79, Jan. 1995.
- [2] A.J. Bell and T.J. Sejnowski, “An information maximization approach to blind separation and blind deconvolution,” *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov 1995.
- [3] K. Torkkola, “Blind separation of delayed sources based on information maximization,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP’96)*, Atlanta GA, 1996.
- [4] J.L. Flanagan, D.A. Berkley, G.W. Elko, J.E. West, and M.M. Sondhi, “Autodirective microphone systems,” *ACUSTICA*, vol. 73, no. 1, pp. 58–91, 1991.
- [5] G. Elko and A. Pong, “A steerable and variable first-order differential microphone array,” in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP’97)*, Munich, Germany, April 1997, pp. 223–226.
- [6] T-W. Lee, A. Bell, and R. Oerglmeister, “Blind source separation of real world signals,” in *Proceedings of International Conference on Neural Networks*, Houston, 1997.
- [7] L. Parra and C. Spence, “Convolutional blind separation of non-stationary sources,” *IEEE Transactions Speech and Audio Processing*, pp. 320–327, 2000.
- [8] M. Brandstein and H. Silverman, “A practical methodology for speech source localization with microphone arrays,” *Computer, Speech and Language*, vol. 11, no. 2, pp. 91–126, 1997.
- [9] S. Affes and Y. Granier, “A signal subspace tracking algorithm for microphone array processing of speech,” *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 5, pp. 425–437, 1997.
- [10] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, “Performance of optical flow techniques,” *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [11] D. Robert, R.N. Miles, and R.R. Hoy, “Tympanal hearing in the sarcophagid parasitoid fly *emblemiasoma* sp.: the biomechanics of directional hearing,” *J. Experimental Biology*, vol. 202, pp. 1865–1876, 1999.
- [12] A.G. Andreou, D.H. Goldberg, E. Culurciello, M. Stanacevic, and G. Cauwenberghs, “Heterogeneous integration of biomimetic acoustic microsystems,” in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS’2001)*, Sydney, Australia, May 6-9 2001.
- [13] A.D. Blumlein, “Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems,” *British Patent*, no. 394325, 1933.
- [14] 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.
- [15] T. Lee, A. Bell, and R. Lambert, “Blind separation of delayed and convolved sources,” in *Adv. Neural Information Processing Systems*, Cambridge MA, 1997, vol. 9, pp. 758–764, MIT Press.
- [16] S. Amari, S. Douglas, A. Cichocki, and H.H. Yang, “Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient algorithm,” in *Proc. of 11th IFAC Symposium on System Identification*, Kitakyushu City, Japan, July 1997, pp. 1057–1062.
- [17] M. Cohen and G. Cauwenberghs, “Blind separation of linear convolutional mixtures through parallel stochastic optimization,” in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS’1998)*, 1998, vol. 3, pp. 17–20.
- [18] G. Cauwenberghs, M. Stanacevic, and G. Zweig, “Blind broadband source localization and separation in miniature sensor arrays,” in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS’2001)*, Sydney, Australia, 2001, vol. 3, pp. 193–196.
- [19] B. Friedlander, “On the cramer-rao bound for time delay and doppler estimation,” *IEEE Transactions on Information Theory*, vol. 30, no. 3, pp. 575–580, 1984.
- [20] B. Widrow, “A microphone array for hearing aids,” *IEEE Circuits and Systems Mag.*, vol. 1, no. 2, pp. 26–32, 2001.