SOURCE LOCALIZATION AND ROOM MAPPING USING INFORMATION DERIVED FROM INDEPENDENT COMPONENT ANALYSIS

Todd K. Moon, Michael Schena, and Jacob H. Gunther

Electrical and Computer Engineering Department and Information Dynamics Laboratory
Utah State University
todd.moon@usu.edu, jake.gunther@usu.edu, michael.schena@aggiemail.usu.edu

ABSTRACT

Convolutional independent component analysis (ICA) algorithms, which have proven capable at separating convolutionally mixed signals, can also provide information about the geometry of the setting. This geometry includes source location information and wall locations (room shape). From the multi-input/multi-output impulse responses learned from convolutional ICA, peaks indicating direct path or reflection delays are extracted. The location of sensors (or sources) is obtained using hyperbolic geometry based on direct path time delays. Delays from reflected paths learned from the impulses responses correspond geometrically to ellipses, which are tangent at the reflecting point. Ellipse tangent directions are clustered to determine wall locations. Following a summary of the method, experiments are presented on actual room measurements.

1. INTRODUCTION

Independent component analysis (ICA), also known as blind source separation (BSS), is a well-researched technique (see, e.g., [1, 2, 3, 4]) for separating linearly mixed sources. Existing research has focused on developing and characterizing separating algorithms. In addition to providing separation, however, the matrix of impulse responses which performs convolutional source separation contains information which can be used to localize sources and measure room shape. In this paper we propose a new technique which exploits ICA-produced information to determine room shape and dimension, as well as source (or sensor) location, using conventional acoustic sources, such as conversations taking place in the room. Thus, it is not necessary to perform explicit impulse response measurements. It is only necessary that the number of sensors be greater than or equal to the number of sources. For simplicity, we assume here that the number of sources is equal to the number of sensors.

In contrast to this paper, in [5] room shape is estimated using explicitly measured room acoustic impulse responses, with information acquired from sources and sensors at known locations. (See also [6, 7, 8]). Impulse responses are measured between a single source/sensor pair at a time using conventional impulse response measurement techniques [9]. In these papers, measurements are done between one source/sensor pair at a time, in which all sources can be on at the same time. Following extraction of peaks from the impulse responses, they are processed into distance measurements using multidimensional scaling [10].

Figure 1 shows a representative measurement scenario with three sources and three sensors. Between source \( s_i(t) \) there is a transfer function which represents the direct and reflected transmission paths. In terms of the direct and first reflected path, this can be expressed as

\[
A_{ji}(z) = a_{ji, \tau_{ji1}}z^{-\tau_{ji1}} + a_{ji, \tau_{ji2}}z^{-\tau_{ji2}} + \text{noninteger delay effects + high order reflections.}
\]

Delays \( \tau_{ji} \) are taken to be integer multiples of the sampling time (actual delay is rounded to the nearest integer multiple). The transfer functions are estimated using a convolutional source separation algorithm. The relationship between sources and sensors can be expressed as \( X(z) = A(z)S(z) \) or, in the time domain, \( x[t] = \sum_{p=-\infty}^{\infty} A_{ji}(t-p) s_p(t) \) for \( m \times m \) matrices \( A_p \).

Let \( A_{ji}(z) \) denote the transfer function from source \( s_i(t) \) to sensor \( x_j(t) \). From strong peaks in this transfer function, the delay due to the direct path \( \tau_{ji1} \) and the delay due to the first reflection \( \tau_{ji2} \) can be extracted.

The algorithm is outlined as follows, with the steps portrayed in figure 2. If the microphone positions are known, the direct path time differences can be used to identify the speaker positions (or vice versa) by time delay of arrival (TDOA) methods, employing hyperbolic geometry (see, e.g., [11]). For the first reflection (fig. 2(b)), the locus of points that cause a delay of \( \tau_{ji2} \) is an ellipse that is tangent to the reflecting surface (wall). The ellipses from all first reflections are illustrated in figure 2(c). Tangents to ellipses for source/sensor paths which reflect off the same wall are colinear — they share tangents. These tangent segments are candidates for wall positions. Tangent segments are illustrated in figure 2(d), where examples of shared tangent vectors are shown for representative pairs of ellipses. In the center of figure 2(d), the entire set of tangent segments from all source/sensor pairs are shown.
This large set of tangent segments is clustered via a clustering algorithm to determine wall segments which may be most consistent with the tangent segments. Figure 2(e) shows the clustered segments, which give estimates of the wall positions. If desired, these wall segments can be extended to points of intersection, giving nearly complete estimates of the walls, as shown in figure 2(f).

The following sections provide detail on the steps just outlined.

2. ICA TECHNIQUE

Of many possible convolutive ICA methods that could be used, the ICA used in our experiments is due to Torkkola [12, Section 8.4.3]. This technique uses a feedback network with FIR filters adapted for separation. The separation filters are

\[
 u_1(n) = \sum_{k=0}^{L/1} w_{1k}x_1(n-k) + \sum_{k=1}^{L/2} w_{1k2}u_2(n-k) \tag{1}
\]

and

\[
 u_1(n) = \sum_{k=0}^{L/2} w_{2k1}x_2(n-k) + \sum_{k=1}^{L/2} w_{2k1}u_1(n-k) \tag{2}
\]

where \( u_i \) are the separated signals, \( x_i \) are the convolutively mixed signals, and \( w_{ij} \) are the weights of the adaptive FIR filter. The learning rules used to adapt the FIR filter weights are given are

\[
 \Delta w_{i0i} \propto \hat{y}_i x_i(n) + 1/w_{i0i} \tag{3}
\]

\[
 \Delta w_{ikj} \propto \hat{y}_i x_j(n-k) \tag{4}
\]

\[
 \Delta w_{ikj} \propto \hat{y}_j x_i(n-k) \tag{5}
\]

Strong local peaks in the transfer function indicate path time differences of the direct paths between source/sensor pairs, and path time differences of reflected paths.

3. ELLIPSES AND TANGENTS TO ELLIPSES

Knowing the sensor position at a point \( s_1 = (s_1, s_2) \), and a source position at a point \( r_1 = (r_1, r_2) \), and a speaker-to-microphone transition time \( t_1 \), for a single reflection, as determined from the transfer function learned from ICA, there is an ellipse of possible reflection loci. The corresponding distance is \( a_{11} = s_1 t_1 \), for a signal speed \( s \). The ellipse can be described by

\[
 ax^2 + by^2 + cxy + dx + ey + f = 0, \quad \text{where}
\]

\[
 a = [(r_1 - s_2)^2 - d_{11}^2] \quad b = [(r_2 - s_2)^2 - d_{11}^2] \\
 c = 2(r_1 - s_1)(r_2 - s_2) \quad d = 2\alpha(r_1 - s_1) + 2d_{11}r_1 \\
 e = 2\alpha(r_2 - s_2) + 2d_{11}r_2 \quad f = \alpha^2 - d_{11}^2(r_1^2 + r_2^2) \\
 \alpha = \frac{1}{2}(s_1^2 + s_2^2 - d_{11}^2 - r_1^2 - r_2^2)
\]

A tangent line segment \( [(x_1, y_1), (x_2, y_2)] \) to the ellipses

\[
 a_1x_1^2 + b_1y_1^2 + c_1x_1y_1 + d_1x_1 + e_1y_1 + f_1 = 0 \tag{6}
\]

\[
 a_2x_2^2 + b_2y_2^2 + c_2x_2y_2 + d_2x_2 + e_2y_2 + f_2 = 0 \tag{7}
\]

occurs when the slope is equal to the implicit derivative at \((x_1, y_1)\) and \((x_2, y_2)\):

\[
 \frac{y_2 - y_1}{x_2 - x_1} = \frac{-2a_1x_1 - c_1y_1 - d_1}{c_1x_1 + 2b_1y_1 + e_1}
\]

which simplifies to equations

\[
 2a_1x_1^2 + 2b_1y_1^2 + 2c_1x_1y_1 - c_1x_1y_1 - c_1x_2y_1 - c_1x_2y_1 = 0 \tag{8}
\]

\[
 -2a_1x_1x_2 + d_1x_1 - d_1x_2 - 2c_1x_1y_2 + e_1y_1 - c_1y_2 = 0
\]

\[
 -2a_2x_2^2 - 2b_2y_2^2 - c_2x_2 - y_2 - d_2x_2 + 2c_2x_1x_2 +
 c_1x_1y_2 + d_1x_1 + c_2x_2y_2 + 2b_1y_1 + e_2y_2 + c_2y_2 = 0 \tag{9}
\]

The four equations (6) – (9) can be efficiently solved by fixing a point \((x_1, y_1)\) on the first ellipse and solving for a new point \((x_2, y_2)\) on the second ellipse using (7) and (9) by Newton’s method, then fixing \((x_2, y_2)\) and solving for a point \((x_1, y_1)\) using (6) and (8). There are generically four tangents to a pair of ellipses. The multiple tangents are obtained by initializing the Newton process at the four extrema on each ellipse. Once the four tangents are found, crossing tangents are eliminated.

4. CLUSTERING THE SEGMENTS

The large number of tangent segments can be clustered first on the basis of direction (angle), and then on the basis of spatial locality. Let \( C \) denote the set of clusters, where each cluster contains direction information (an angle in \([0, \pi]\)), a number of associated points, and a list of tangent segments associated with the cluster.

Each new tangent segment is compared with each existing cluster. If the segment direction is sufficiently close to the direction of the cluster, then it is verified that the tangent segment lies close to the space of the cluster, to avoid the problem of tangents which have the same direction, but are spatially distant. This is done as follows. Let the cluster tangent endpoints be \( e = [e_1, e_2] \) and \( d = [d_1, d_2] \), and let the endpoints of the tangent segment be \( a \) and \( b \). The extension of the cluster segment is given by

\[
 T(\lambda; e, d) = (1 - \lambda)[e_1, e_2] + \lambda[d_1, d_2]
\]

Consider the case that the extension \( T(\lambda; e, d) \) matches the \( x \) coordinate of the tangent segment point \( a \). Then

\[
 (1 - \lambda)c_1 + \lambda d_1 = a_1,
\]

from which \( \lambda = (a_1 - c_1)/(d_1 - c_1) \). For this value of \( \lambda \), the corresponding \( y \) value is computed, and compared to the tangent segment \( a_2 \). If they are sufficiently close, the tangent segment is deemed to belong to the cluster. There are now four cases to consider: comparing the extension in the \( x \) coordinate to both \( a \) and \( b \), and comparing the extension in the \( y \) coordinate to both \( a \) and \( b \). If the smallest of the four distances is sufficiently small, the tangent segment is added to the cluster.

Following clustering, those clusters associated with the most tangent segments are used as estimates of the wall positions, possibly after averaging. The extrema of the tangent segments associated with these clusters give the wall estimates. These segments may be extended, to find points of intersection. If the intersections are sufficiently close to the segments, the intersection is interpreted as a valid corner.

5. EXPERIMENTAL RESULTS

Acoustic data were collected in the acoustic anechoic chamber at Utah State University, which was equipped with artificial walls, as pictured in figure 3. We present here an example of the processing, in which two or three flat reflecting walls were
Fig. 2. Wall reconstruction steps: (a) Sensor locations; (b) Delays; (c) Ellipses; (d) Ellipse tangent segments; (e) clustering; and (f) Extension.
Fig. 3. Anechoic chamber with two walls

Fig. 4. Impulse response matrix for three sources, three mikes

placed in the anechoic chamber. Three white noise sources at “unknown” locations were played simultaneously and recorded using microphones at known locations using an NI PXI-4496 board in a PXI-1031 DC chassis at 2250 samples/second at 16 bits/samples. Figure 4 portrays the matrix impulse response functions extracted using blind source separation. From these impulse responses, the direct path (the strong first pulse) and other reflections from the walls are evident. From the peaks extracted from these impulse responses, for the direct paths, the speaker positions can be extracted, as shown in figure 5.

Figure 6 shows the processing with a two-wall scenario. Figure 7 shows the processing with a three-wall scenario. In these, the red segments are tangent segments. The cyan identifies the clusters. These are averaged and extended to form the black wall estimates. The black line shows the extension of the estimated walls. Green walls are true.

6. CONCLUSIONS

Despite the large amount of research into ICA, there has been very little reported on the geometric information inherent in the convolutive separation matrix. Here, the ICA separating matrix has been shown to represent geometric information about the source/sensor locations and room geometry. We have demonstrated that locations can be extracted. Furthermore, reflections at walls correspond to delay ellipses, whose tangents can be clustered to find wall locations. (The multidimensional scaling approach of [5] could also be used.) Preliminary experimental results show that room measurements can be made.

Fig. 5. Determining microphone locations

Fig. 6. Processing with two walls

Fig. 7. Ellipse tangent segments (red); clustered positions (cyan); true walls (green)
7. REFERENCES


