of having an array with large physical dimensions. In Section 5.5 we discuss improvements made to the algorithm when operating with human speakers in acoustically challenging environments. The work in this chapter is largely described in Reference [64].

When developing an algorithm for isolating different sources, known as the blind source separation (BSS) problem, one of the principal challenges in a real room is the unpredictability of the acoustic path. Not only is the path affected by reverberations with surfaces and objects in the room, but human sources can move. To solve BSS one approach is beamforming, which leverages the spatial filtering capability of a microphone array to isolate sources. Unfortunately, classical delay-sum beamforming is not well-suited to a practical room. This is because it uses pre-defined time delays between the microphones that are independent of frequency, with the aim of constructively adding the signal from a target source and destructively adding the signals from all interfering sources [65]. An alternative approach to BSS is to use algorithms based on a frequency domain implementation of independent component analysis (ICA), which typically exploit statistical independencies of the signals [47]. However, there are concerns about the robustness of these algorithms, especially in a reverberant room. This is due to the inherent permutation ambiguity of this approach, where after separation independently at each frequency, the components must further be assigned to the correct source. This necessitates an additional decision step [66].

Weinstein et al. [67] were able to isolate speech signals using conventional delay-sum beamforming, but had to utilize an array with over 1000 microphones to obtain acceptable results. Levi et al. continued to use conventional delay-sum beamforming, but incorporated a spectral subtraction step after beamforming, enabling an array with just 16 microphones [68]. Unfortunately, this approach is not blind, since it requires the location of the sources and microphones.

In this work we propose and demonstrate a beamforming-based algorithm for BSS, with the following main contributions:

1. We also use delay-sum beamforming, but unlike prior work we do not use a single time delay across all frequencies for a given microphone-source pair. Rather we use frequency dependent time delays. This is needed since reverberations from the surfaces in a practical room lead to multipath propagation, for which a linear phase model as a function of frequency is inadequate [69].

2. We crucially differ from other beamforming attempts by being blind, requiring no prior information about the location of the sources or microphones. The only information our algorithm needs about the environment is the number of sources. Thus, we avoid time consuming and technically challenging location measurements [70]. Furthermore, we make no assumptions about the propagation of sound in a room. Rather, we extract time delays for each microphone-source pair on the fly from the sound mixture of simultaneous sources. This enables our algorithm to adapt to the unique acoustic properties of each room (e.g., size, reverberation time, placement of objects) and a change in location of the sources. We use k-means clustering, an unsupervised classification technique, to identify a short (64 ms) frame at the beginning of the sound mixture in which only a single source is prominent, making such a frame well-suited for time delay extraction.

3. We apply our algorithm to experimental data from two adjacent linear arrays, measured in a conference room: (1) an array of commercial electret capsules, and (2) an array of LAE microphones, which are fabricated in-house [52]. The LAE microphones have non-idealities (e.g. non-flat frequency response, large variation across elements) compared to electret microphones, which arise due to fabrication in a large-area, thin, and flexible form factor. For both arrays we

achieved high-quality separation results. Our algorithm outperformed simple beamforming and was competitive with Independent Vector Analysis (IVA) BSS, a modern frequency-domain ICA-based algorithm [71], while avoiding the associated permutation problem.

5.1 Audio Signal Processing: Background Concepts

This section provides a brief background about some of the basic audio signal processing concepts and terms used when describing our algorithm.

5.1.1 Conventional Delay-and-sum Beamforming

Delay-and-sum beamforming is a signal processing technique that uses a microphone array to "focus" on a given source. Therefore, it can be used to enhance a target sound source, and cancel undesired sources, known as interferers. Inversely, a recording from an array for a single source can be used to determine the location of this source.

As shown in Fig. 5.1, since the array can be treated as a linear system, each microphone receives a summation of the target signal and the interferer. In an anechoic room (e.g. with no reverberations), the difference in the time it takes for the sound from a given source to reach two microphones, known as the time difference of arrival (TDOA), is the same for all frequencies. Assuming the source is in the far-field, so sound can be treated as a planar wave, the difference in TDOAs between two microphones for a source, s, can be calculated using a simple geometric expression (see Fig. 5.2):

$$D_{ms} = \frac{nd\sin(\theta)}{c} \tag{5.1}$$

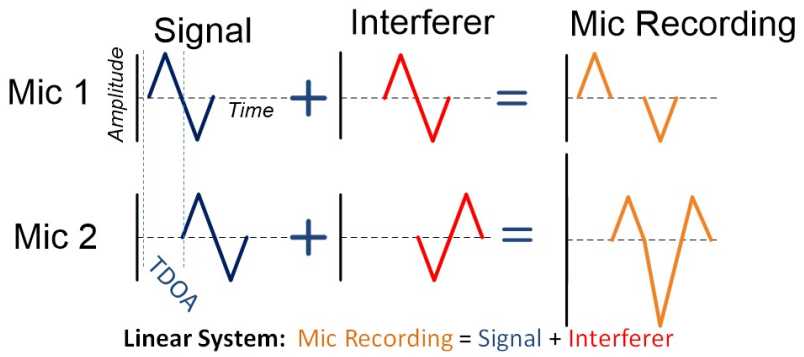

Figure 5.1: Illustration showing how each microphone in an array receives a signal, consisting of the sum of the time delayed target and the interferer.

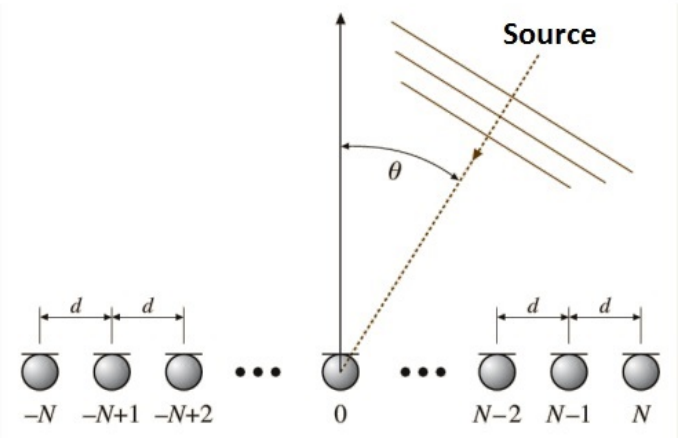

Figure 5.2: Illustration showing how TDOAs can be geometrically calculated for a linear microphone array in an anechoic room with a source in the far-field [65, pp. 1024].

where d is the microphone spacing, c is the speed of sound in air, and is θ is the angle between a microphone and the source. n is the number of microphones between a given microphone, m, and another microphone in the array designated as a reference.

In conventional delay-and-sum beamforming, the already known TDOAs are used to time align the target signal across all microphones in the array, so they add constructively. At the same time the interferer will preferably be attenuated, due to it adding destructively. The operation of a delay-and-sum beamformer can be mathe-

matically expressed as [65]:

$$\widehat{X'}^{s}(t) = \sum_{m=1}^{M} Y_m(t - D_{ms}) \tag{5.2}$$

where Y_m is the signal recorded at each microphone, m, and consists of the time delayed sum of the target signal and the interferer. $\widehat{X'}^{s}(t)$ is the enhanced output of the delay-and-sum beamformer when "focusing" on a source, s.

5.1.2 k-means Cluster Analysis

k-means is an algorithm used to identify groups of objects, known as clusters, in a dataset. It is an example of an unsupervised machine-learning algorithm, since prior to running the algorithm, the feature vectors in a dataset are not labelled as belonging to a given cluster. In other words, unlike a supervised algorithm, such as a support vector machine (SVM), there is no distinct "training" data. To help visualization, Fig. 5.3 illustrates the concept of k-means cluster analysis using a two-dimensional feature vector, although it should be noted that k-means can be applied to a feature vector with any number of dimensions.

The k-means algorithm consists of three steps [72]:

1. Each feature vector, $\mathcal{F}$, is assigned to a given cluster, k. This is done by considering the value of a distance function, $\mathcal{D}$, between a given feature vector and a centroid, μ_k. The feature vector is assigned to the cluster with the smallest value for $\mathcal{D}$. An example of a possible distance function is the Euclidean distance:

$$\mathcal{D} = \|\mathcal{F} - \mu_k\|^2 \tag{5.3}$$

2. The value of the centroid for each cluster is updated by calculating the arithmetic mean of all feature vector assigned to that cluster.

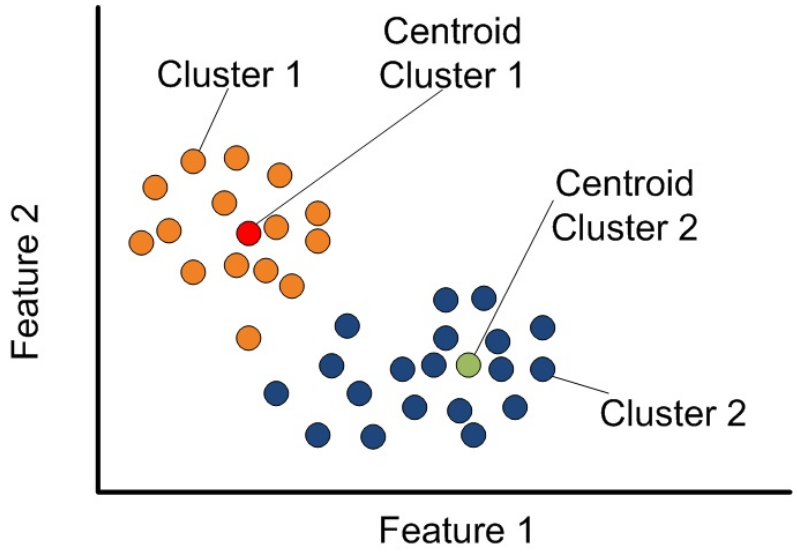

Figure 5.3: Illustration of using k-means to find two clusters. Each point corresponds to a two-dimensional feature vector.

3. Steps 1 and 2 are repeated until the feature vectors assigned to a given cluster no longer change in each successive iteration, meaning the algorithm has reached a local optimum.

It should be noted that prior to running k-means, the number of clusters, k, needs to be selected. Additionally, initialization values are required for the centroid. These can be obtained by randomly assigning each feature vector to a given cluster and calculating the arithmetic mean of all feature vectors assigned to that cluster.

5.1.3 Algorithm Implementation Considerations

A typical approach when implementing an audio signal processing algorithm in the frequency domain, is to divide a long, uniformly sampled signal in the time domain, $x(t)$, into a series of short time segments. These shorts segments, whether in the time-domain or frequency-domain, are referred to as frames. They are transformed to the frequency-domain for processing using an FFT (fast Fourier transform). $X(\omega, L)$ denotes a frame in the frequency domain with frame index L and frequency bins, w. The number of frequency bins is equal to the number of samples in a given frame.

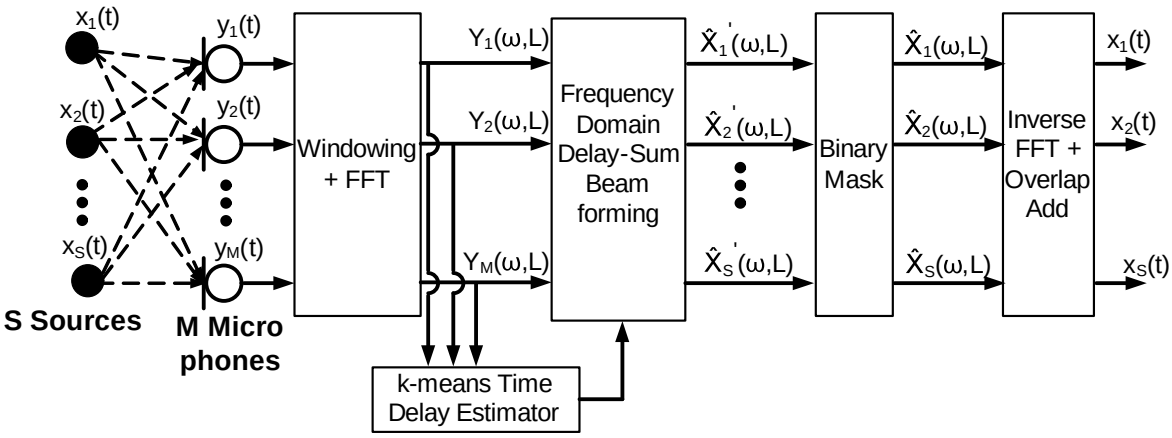

Figure 5.4: Block diagram of our proposed algorithm.

Once they have undergone processing in the frequency-domain, the frames can be brought back to the time-domain using an inverse FFT. Multiple time frames can be concatenated using the standard overlap-add method [61]. In this work each frame is overlapped by 75% with the preceding frame, so as to ensure it meets the constant overlap-add condition for the Hann windows used in order to mitigate artifacts [62].

5.2 Algorithm

Fig. 5.4 shows the block diagram of the proposed algorithm. The beamforming stage receives the convoluted mixture from all the sources in the room and carries out delay-sum beamforming with frequency-dependent time delays. These are provided by the k-means time-delay estimator, wherein an optimal segment for estimation is first identified. To further cancel out interfering sources, the beamformer is followed by a binary mask stage.

5.2.1 Problem Setup

The array consists of M microphones, which separate S simultaneous sound sources, $x_s(t)$. The sound recorded by each microphone, $y_m(t)$, is determined by the room

impulse, $h_{ms}(t)$, between each source and microphone:

$$y_m(t) = \sum_{s=1}^{S} x_s(t) * h_{ms}(t). \tag{5.4}$$

We designate one of the microphone channels as a reference, ref, and express the signal recorded in the time-frequency domain at this reference microphone, for frequency ω and frame L as:

$$Y_{ref}(\omega, L) = \sum_{s=1}^{S} X_s(\omega, L)|H_{ref\,s}(\omega)|e^{j\omega \mathrm{T}_{ref\,s}(\omega)} \tag{5.5}$$

where

$$H_{ref\,s}(\omega) = |H_{ref\,s}(\omega)|e^{j\omega \mathrm{T}_{ref\,s}(\omega)} \tag{5.6}$$

is the room impulse response in the frequency domain, and $\mathrm{T}_{ref\,s}(\omega)$ is the time delay between the reference microphone and a source s. Our objective is to recover each source s at the reference microphone, as if it were recorded with the other sources muted:

$$X^s_{ref}(\omega, L) = X_s(\omega, L)|H_{ref\,s}(\omega)|e^{j\omega \mathrm{T}_{ref\,s}(\omega)}. \tag{5.7}$$

5.2.2 Beamforming with Frequency Dependent Time Delays

The first step of our algorithm is delay-sum beamforming. During this step, for a given source we time align all microphone signals with respect to the reference microphone and sum them:

$$\widehat{X'}^s_{ref}(\omega, L) = \sum_{m=1}^{M} Y_m(\omega, L)e^{-j\omega \mathrm{D}_{ms}(\omega)} \tag{5.8}$$

where $D_{ms}(\omega)$ is the time delay between the reference and each microphone. In this way we constructively sum the contributions from the source we want recover over all microphones, and attenuate the other sources through destructive interference.

In classical delay-sum beamforming, D_{ms} is treated as a constant, frequency-invariant value, such as found in anechoic conditions [65]. Instead, this implementation takes into account multipath propagation of sound in a reverberant room, which has the effect of randomizing the phase spectrum of the room impulse response [69]. Our time delay estimator stage, described in Section 5.2.4, estimates $D_{ms}(\omega)$.

5.2.3 Binary Mask

To further suppress interfering sound sources, a binary mask, $M_s(\omega, L)$, is applied to the output of the delay-sum beamformer:

$$\widehat{X}^s_{ref}(\omega, L) = \widehat{X'}^s_{ref}(\omega, L) M_s(\omega, L). \tag{5.9}$$

The objective of a binary mask is to keep frequency bins that have a high signal-to-interferer ratio, and discard bins that have a low signal-to-interferer ratio [61]. When constructing a binary mask, frequency bins are assigned a value of 1 if they meet the following criterion, otherwise they are assigned a value of 0:

$$\frac{|\widehat{X'}^s_{ref}(\omega, L)|}{\max(|\widehat{X'}^1_{ref}(\omega, L)|, |\widehat{X'}^2_{ref}(\omega, L)|, \cdots, |\widehat{X'}^S_{ref}(\omega, L)|)} > \alpha \tag{5.10}$$

where α is a constant threshold value that is experimentally chosen, so as to provide high quality separation across a wide variety of experimental conditions, such as different number of sources or different sound pressure levels. After applying the binary mask, the inverse FFT is taken of each frame to recover the time domain signal, and successive frames are concatenated using the standard overlap-add method.

5.2.4 Time Delay Estimates Based on k-Means Clustering

Time delays between the reference and other microphones can be estimated by making each source play a test sound one-by-one in isolation. A frame from the test sound, such as speech or white noise with the desired spectral content, can be used to find the time delays:

$$D_{ms}(\omega) = T_{m\,s}(\omega, L) - T_{ref\,s}(\omega, L) = \frac{1}{2\pi f}(\angle X_m^s(\omega, L) - \angle X_{ref}^s(\omega, L)) = \frac{\phi_m(\omega, L)}{2\pi f} \quad (5.11)$$

where f is the frequency and $\angle X_m^s(\omega, L)$ is the phase of a frame from the desired source recorded at microphone m.

We replace this calibration procedure by estimating the time delays directly from the signal when all sources are playing simultaneously. We are able to achieve this by using a standard implementation of k-means clustering based on Euclidean distance [72]. We set the number of clusters, k, to be equal to the number of sources, S. A feature vector, $\phi_{m'}(\omega, L)$, is extracted for each frame, which consists of the phase difference between a given microphone, m', and the reference at the N frequencies of interest:

$$\phi_{m'}(\omega, L) = [\theta_{\omega 1}, \theta_{\omega 2}, \cdots, \theta_{\omega N}] \quad (5.12)$$

with θ taken to be in the range $[0, 2\pi)$.

Our intent is not just to classify each frame as belonging to a given source, since many frames have spectral content from multiple sources, which would lead to poor time delay estimates. Rather we want to identify the best possible frame from which to derive the time delays. To identify these frames we calculate the silhouette [73],

$s(L)$, for every feature vector, and choose the frame with the highest value:

$$s(L) = \frac{b(L) - a(L)}{\max(b(L), a(L))} \tag{5.13}$$

where $a(L)$ is the mean distance between the feature vector from the frame with index L and all other feature vectors assigned to the same cluster. Then, the mean distances to the feature vectors corresponding to all other clusters are also calculated, and the minimum among these is designated as $b(L)$. The value of $s(L)$ is bounded between $[-1, 1]$, and a larger value indicates it is more likely a feature vector has been assigned to an appropriate cluster.

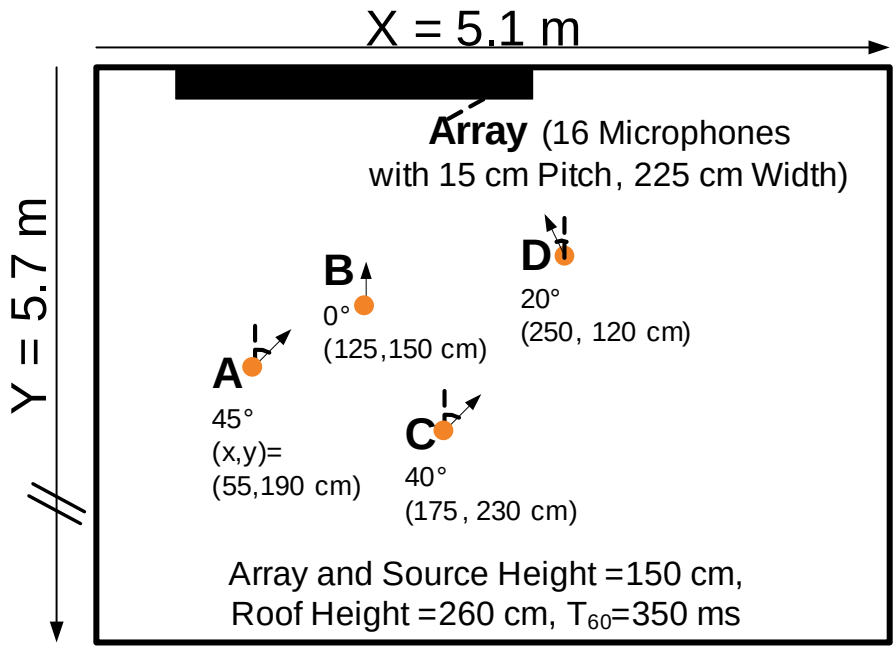

Figure 5.5: Experimental room setup (top view). Drawn to scale. A, B, C and D are the speakers located at coordinates (x,y), with speaker orientation given by the angle accompanying each source.

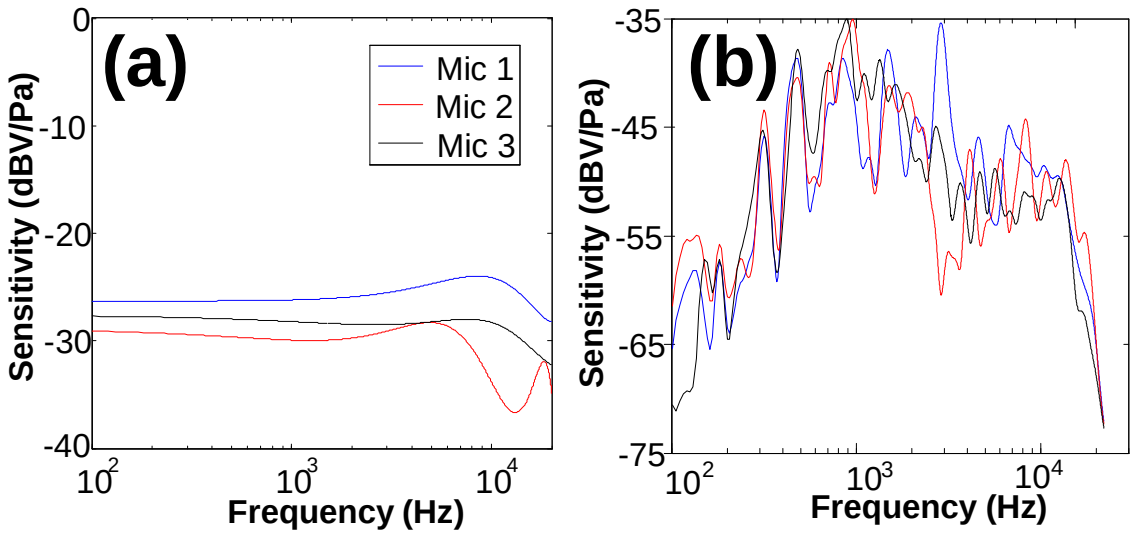

Figure 5.6: Microphone sensitivity measured in an anechoic chamber. (a) Omnidirectional electret microphone, (b) LAE microphone.

Number of Sources	S = 2 (B,C) and S = 4 (A, B, C, D)
Number of Microphones	M = 16
Microphone Pitch	15 cm (total array width = 2.25 m).
Source Signals	12 Harvard sentences from the TSP database[50] (Duration = 30 s).
Sampling Rate	16 kHz
Reverberation Times	T_{60} = 350 ms (set by acoustics of the room, measured).
Window Type	Hamming
STFT Length	1024 samples (64 ms)
STFT Frame Shift	256 samples (16 ms)
Reference Microphone	Located at center of linear array.
Threshold for Binary Mask	α = 1.4 (see Equation 5.10).

Table 5.1: Experimental and Signal Processing Parameters

5.3 Experimental results

5.3.1 Setup Conditions

Experiments were carried out in a conference room, as shown in Fig. 5.5, playing both two (B and C) and four (A, B, C and D) simultaneous sound sources, consisting of speech recordings from different speakers, played with loudspeakers (Altec ACS90). The sound pressure level of each speaker was $\sim$ 60 dB_{SPL}. Table 5.1 has a summary of experimental conditions. The two linear arrays were mounted horizontally, with a PVDF microphone approximately 3 cm above a corresponding electret microphone;

thus, allowing us to directly compare the performance of the two arrays. Each array used different elements: (1) Commercial omnidirectional electret capsules (Primo Microphone EM-172); (2) LAE microphones, which are based on a flexible piezoelectric polymer, PVDF, and are fabricated in-house. Fig. 5.6 shows the frequency response of both types of microphones, including the non-idealities of LAE microphones arising due to the fabrication methods which lead to their large-area, thin, and flexible form factor e.g. reduced sensitivity, a non-flat response and large variations across elements.

To assess the performance of our algorithm we used two metrics: (1) Signal-to-Interferer Ratio (SIR) calculated with the BSS_Eval Toolbox [49] [74]; (2) Perceptual evaluation of speech quality (PESQ) using the clean speech recording from the TSP database [50] as the reference signal. PESQ mean opinion scores (MOS) range from -0.5 (bad) to 4.5 (excellent) [75]. Further information about these metrics can be found in Appendix C.

The separation algorithm was implemented in Matlab, and used data that was previously recorded on the PC. Nevertheless, it should be noted that the algorithm is compatible with a real-time implementation, since it is based on short frames (64 ms), which do not depend on future frames (e.g. it is causal).

5.3.2 Time Delay Estimator Performance

We compared the performance of our algorithm using time delays extracted under two conditions: (1) White Gaussian noise, which was played by each speaker one-at-a-time, before the simultaneous recording, and (2) from a single frame of simultaneous speech that was selected by our k-means-based silhouette criterion. It should be noted that to improve the estimate when extracting the time delays from white noise, the phase difference in Equation 5.11 consisted of the circular mean [76] calculated from 50 successive frames.

To identify the best frames for time delay extraction, we implemented k-means with 312 features vectors. Each feature vector was extracted from a different frame (frame length = 64 ms, frame shift =16 ms) taken from the first 5 s of the recording with the simultaneous sources. We used a total of 160 features, corresponding to the phase difference between the closest adjacent microphone and the reference microphone for each frequency bin between 500 Hz and 3000 Hz.

After k-means, the silhouette was calculated for all 312 feature vectors in order to select a feature vector per source for extracting time delays. Fig. 5.7 validates the use of the silhouette as a metric for selecting a frame to use for time-delay extraction. In this figure each point corresponds to the SIR calculated over the entire 30 s speech segment (after the beamforming stage), using time delays extracted from different frames. Fig. 5.8 shows a comparison, for two representative microphones in the array, of the phase delays estimated using white noise played in isolation versus those estimated from frames selected based on the silhouette. Phase delays found using either method are closely matched. In the following section we also compare the performance of our algorithm when using time delays from white noise and k-means. In most experiments there is only a small performance degradation for k-means, highlighting its effectiveness for enabling BSS.

5.3.3 Overall Algorithm Performance

A lower limit on performance is given by calculating the SIR and PESQ at the reference microphone before any signal processing. An upper limit is given by the PESQ at the reference microphone when only a single source is playing, using as a reference signal the clean anechoic recording that was inputted into the loudspeaker. In Fig. 5.9 to 5.12, we show that for all configurations our algorithm successfully enhances speech, significantly increasing both SIR and PESQ. It also shows how our algo-

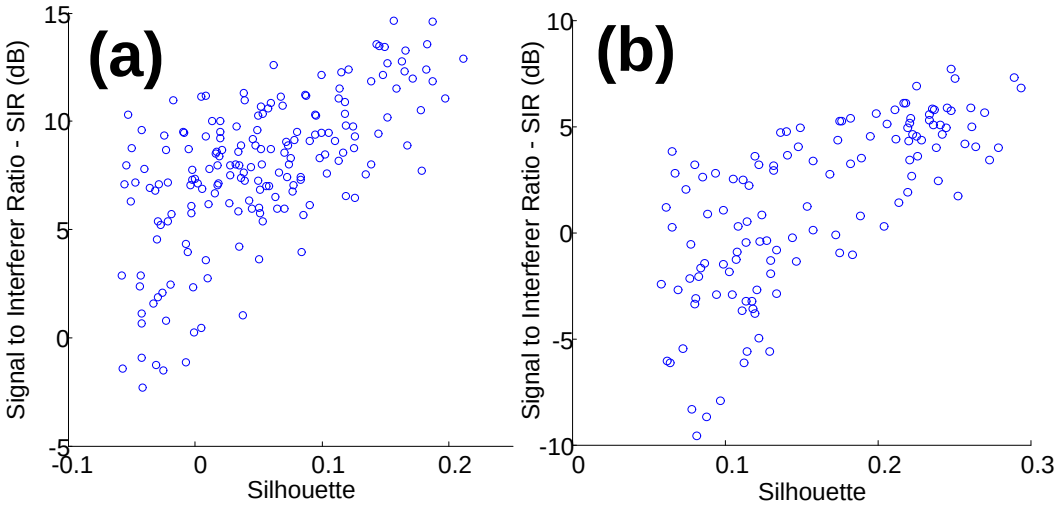

Figure 5.7: SIR for time delays extracted from different frames versus the silhouette of the frame for two sources (a)Source B (b)Source C. Each point corresponds to the SIR calculated over the entire speech segment (after the beamforming stage) using delays from a different frame.

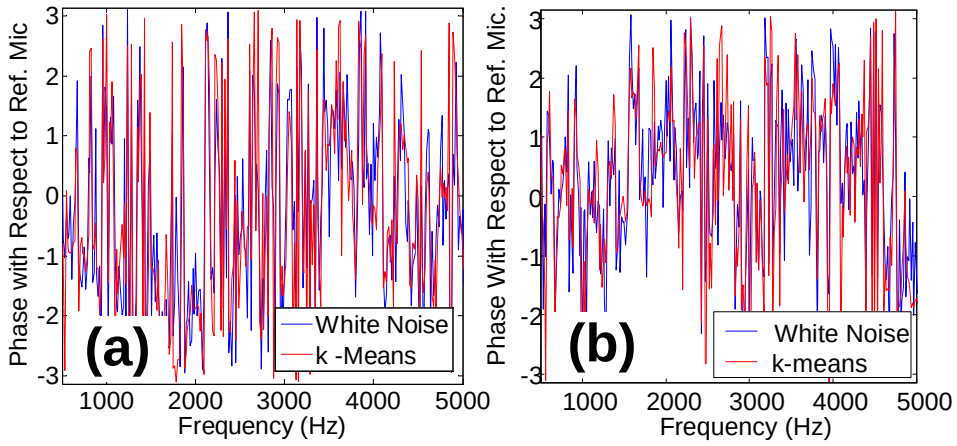

Figure 5.8: Comparison of phase for two representative microphones extracted from white noise and k-means (a) Microphone 4, Source B; (b) Microphone 12, Source C.

rithm, combining beamforming followed by a binary mask, outperforms using only the beamforming stage.

To compare the performance of our algorithm with a modern, conventional BSS algorithm, we chose IVA BSS [71]. When using the minimum number of microphones for IVA BSS (2 microphones for 2 sources, 4 microphones for 4 sources) our algorithm (using the entire 16-microphone array) outperforms by a wide margin. On the other

hand when using IVA BSS with the entire array and selecting the best channels from the 16 separated channels it outputted, IVA BSS and our algorithm perform at a similar level. For two sources IVA BSS performs slightly better than our algorithm, but for four sources it sometimes fails to significantly enhance certain sources.

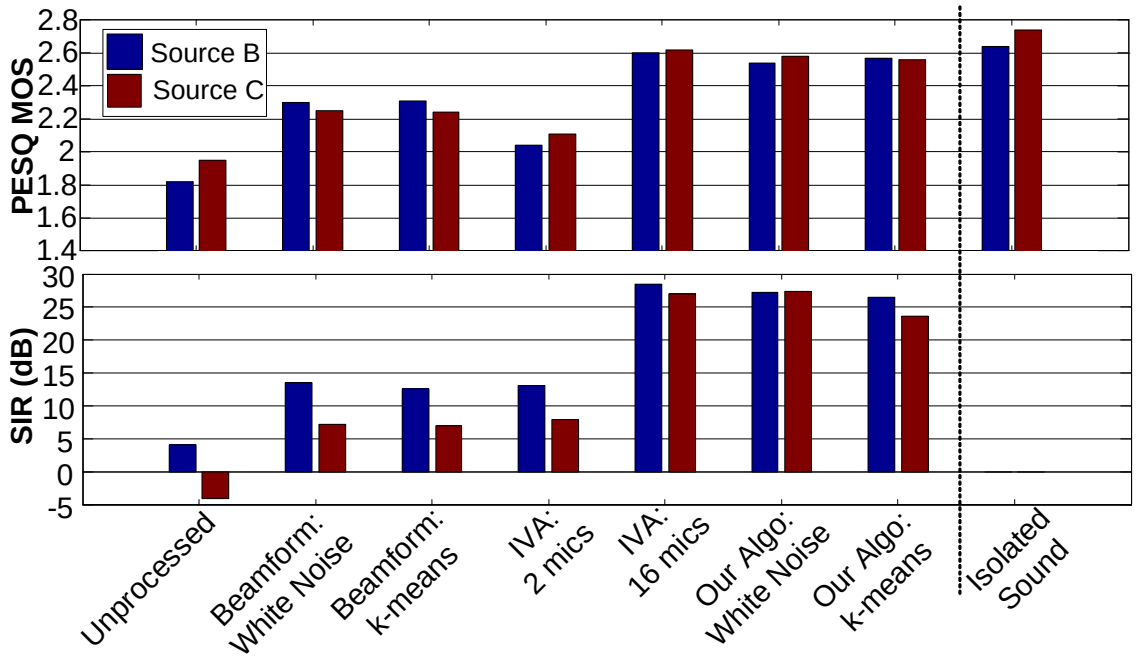

Figure 5.9: Separating two sources with an array of electret microphones.

In Figure 5.9, when using two sources and the array with electret capsules, the PESQ is nearly the same as the maximum possible (e.g. the sound played in isolation at the reference microphone), highlighting the effectiveness of our proposed algorithm. A mean PESQ improvement of 0.7 is obtained when comparing the blind algorithm (with k-means delays) to the unprocessed signal.

In Fig. 5.10, we repeat the same experiment with the LAE microphone array and find that the PESQ of the upper limit is lower, due to the reduced performance of the LAE microphones. In this case the PESQ from white noise is close to the upper limit, while the PESQ from k-means is lower. This suggests that the reduced sensitivity of the LAE microphones causes the time delay estimates, extracted from the mix with

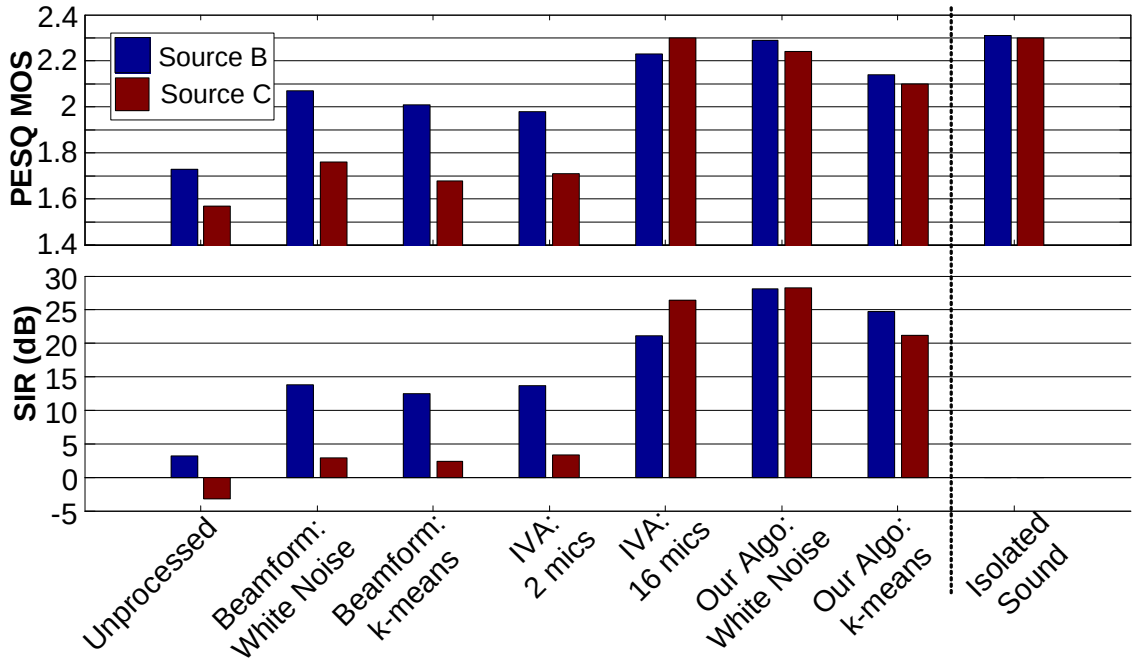

Figure 5.10: Separating two sources with an array of LAE (PVDF) microphones.

simultaneous sources, to degrade. Nevertheless, a mean PESQ improvement of 0.5 is still obtained.

In Fig. 5.11, we test the electret microphone array with four sources. The PESQ scores of our algorithm are no longer as close to the upper limit as with two sources, due to the initial lower PESQ and SIR of most of the unprocessed signals. Nevertheless, speech is significantly enhanced, with a mean PESQ MOS improvement of 0.6. In Fig. 5.12, we repeat the same experiment with the LAE microphone array and find the algorithm shows a larger degradation, with a mean PESQ improvement of 0.3. These results demonstrate how our algorithm can still provide improvements in speech quality even in settings where the unprocessed input signal has been severely degraded, due to non-ideal microphones and low initial SIR values.

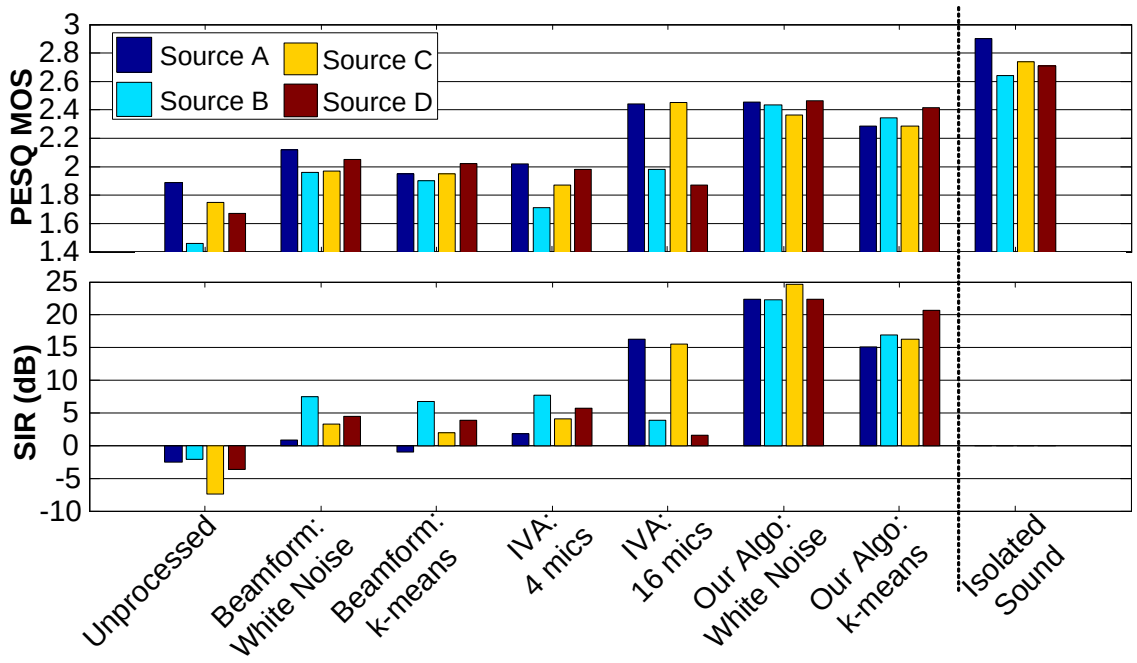

Figure 5.11: Separating four sources with an array of electret microphones

5.4 Array Design Considerations

When designing a linear array for our algorithm, the two principal parameters are the number of microphones and the spacing between the microphones. Fig. 5.13 shows the performance of our algorithm with two sources when using different numbers of electret microphones with a spacing of 15 cm. The same experimental setup and array was used as in Fig. 5.5. Microphones were incrementally added, starting from the center of the array and working towards the corners of the room. These results suggest that for our experimental setup, there is a logarithmic relationship between the number of microphones and the signal-to-interferer ratio, as predicted by beamforming theory [46] [77]. This relationship is due to the beamforming stage coherently adding the signal we want to recover from the different microphones, so the signal power after beamforming can be expressed (in dB) as:

$$P_{Signal} = 10 \log_{10}(MA)^2, \tag{5.14}$$

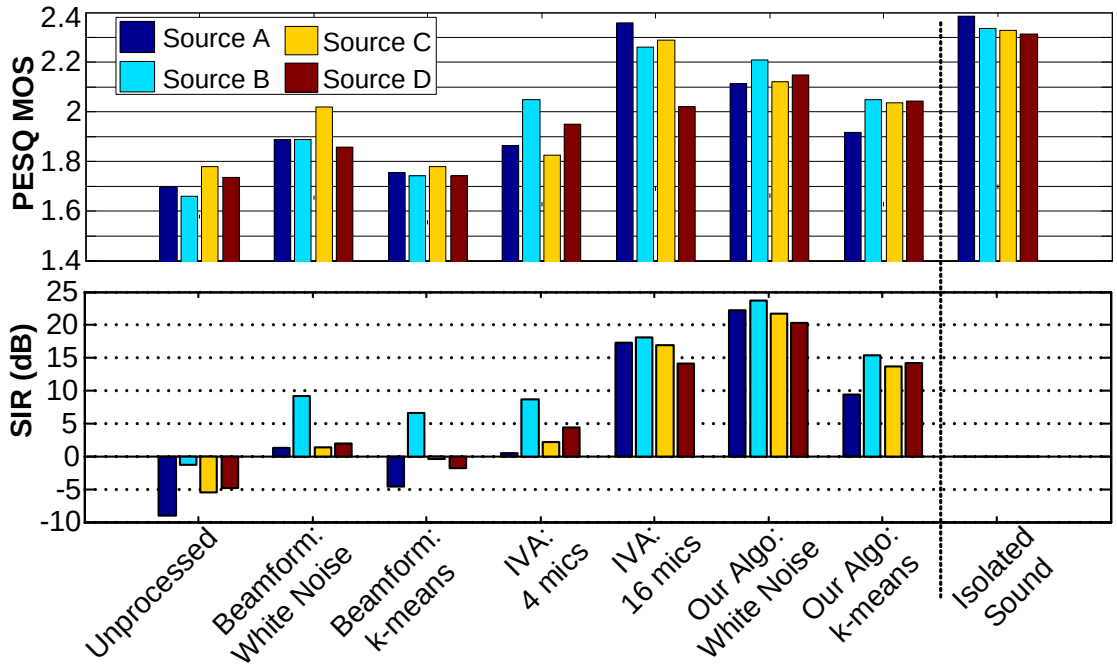

Figure 5.12: Separating four sources with an array of LAE (PVDF) microphones.

where M is the number of microphones and A is the amplitude of the signal. The interferers recorded at each microphone can be treated as uncorrelated, gaussian, white noise, w_m, so the interferer power after beamforming can be expressed as:

$$P_{Interferer} = 10\,log_{10}(M\sigma_m^2), \tag{5.15}$$

where σ_m^2 is the variance of w_m. Therefore the signal-to-interferer ratio can be expressed as:

$$SIR = \frac{P_{Signal}}{P_{Interferer}} = 10\log_{10}\left(\frac{MA^2}{\sigma_m^2}\right). \tag{5.16}$$

To determine a suitable spacing between microphones, we simulated a linear array [48] with the room configuration shown in Fig. 5.5, with 16 elements with different spacings. We did simulations and not experiments, because it would have been difficult to reposition the microphones in each case. As shown in Fig. 5.14, performance (especially PESQ) significantly deteriorates when the spacing is less than

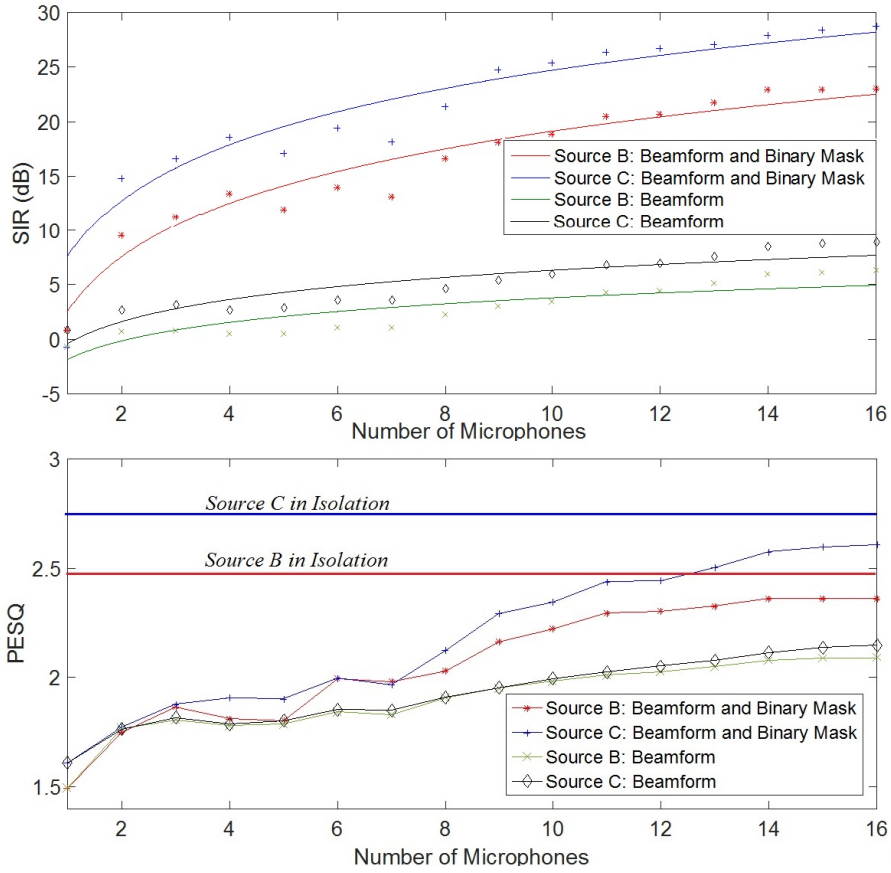

Figure 5.13: Algorithm performance as function of the number of microphones, experimentally measured with two sources and electret microphones. For SIR graph, points are data and lines are a logarithmic fit.

10 cm. A possible explanation is that for small microphone spacing, the interferers after beamforming are not uncorrelated, which is detrimental for the $SIR \propto \log_{10} M$ relationship.

5.5 Algorithm Improvements

When testing our algorithm with two stationary humans speakers in an exhibit room at a conference, we found that the time delay estimator stage occasionally failed to identify appropriate time segments from which two extract the time delays. This is not surprising considering the acoustically challenging conditions presents in the exhibit room, including background music and noise from other speakers in the distance,

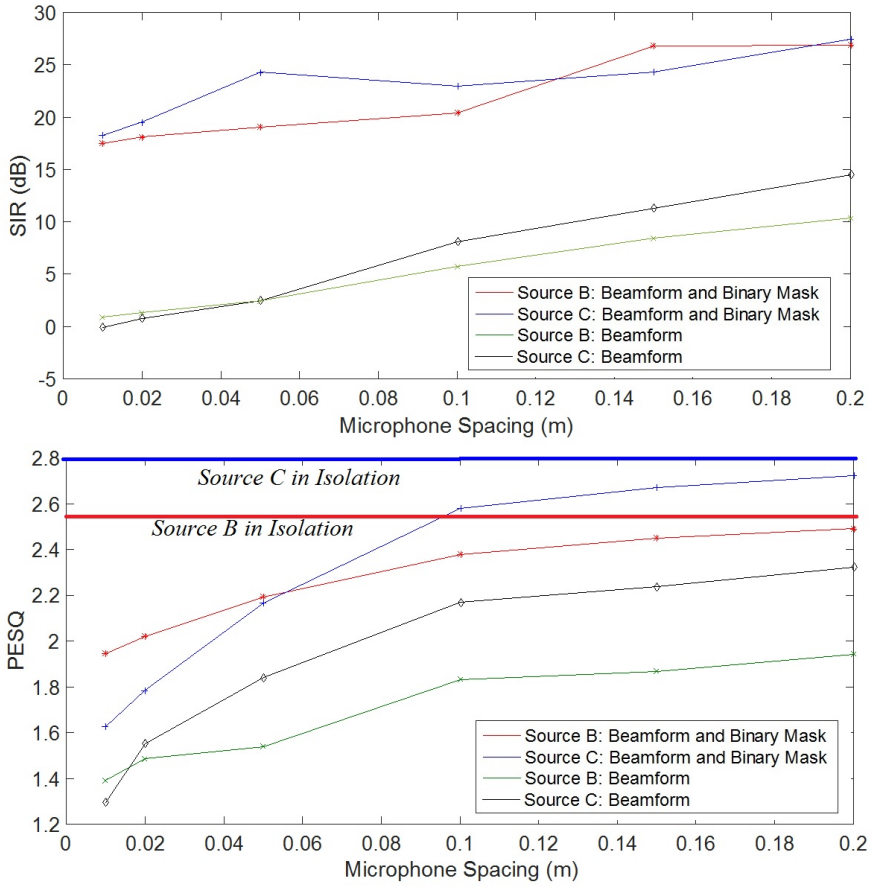

Figure 5.14: Algorithm performance as function of the spacing between microphones, simulated with two sources.

large differences in the initial (before processing) SIR levels of the speakers, and the human speakers not being fully stationary.

To improve the robustness of the time delay estimator under these challenging acoustic conditions, we modified the standard implementation of k-means and the silhouette (as used in Section 5.2.4) to take into account the circular nature of the angular feature vectors (e.g. due to phase wrapping, the difference between an angle of 2π rad and 0 rad, should be 0 rad and not 2π rad). This modification of k-means is based on [78]. In particular, for k-means we no longer use a Euclidean distance function between each feature vector and its centroid, as given by:

$$Distance_{kmeans}(L) = \|\phi_{m'}(\omega, L) - \mu_k(\omega)\|^2 \tag{5.17}$$

where $\mu_k(\omega)$ is the centroid of the cluster to which the feature vector, $\phi_{m'}(\omega, L)$, has been assigned. Instead, we use a circular distance function:

$$Distance_{kmeans\_circ} = \sum_{\omega=\omega 1}^{\omega=\omega N} (1 - cos(\phi_{m'}(\omega, L) - \mu_k(\omega)))^2 \tag{5.18}$$

where $1 - cos(\phi_{m'}(\omega, L) - \mu_k(\omega))$ is scaled between [0,2]. Also, we no longer calculate the centroid using the standard arithmetic mean of all the features vectors assigned to a given cluster:

$$\mu_k(\omega) = \frac{\sum_{L=1}^{L=\mathfrak{L}} R_{Lk}\phi_{m'}(\omega, L)}{\sum_{L=1}^{L=\mathfrak{L}} R_{Lk}} \tag{5.19}$$

where $R_{Lk} = 1$ is if a feature vector belongs to the cluster being considered and otherwise $R_{Lk} = 0$. Instead, to find the centroid we calculate the circular mean [76] of all feature vectors from a given cluster:

$$\mu_k(\omega) = \arctan\left(\frac{\sum_{L=1}^{L=\mathfrak{L}} R_{Lk} \sin \phi_{m'}(\omega, L)}{\sum_{L=1}^{L=\mathfrak{L}} R_{Lk} \cos \phi_{m'}(\omega, L)}\right) . \tag{5.20}$$

When calculating the silhouette we also no longer use a Euclidean distance between two features vectors. Instead, similarly to eqn. 5.20, we use a circular distance function:

$$Distance_{silhouette\_circ} = \sum_{\omega=\omega 1}^{\omega=\omega N} (1 - cos(\phi_{m'}(\omega, L) - \phi_{m'}(\omega, \mathfrak{L})))^2 \tag{5.21}$$

For the modified time delay estimator stage, when testing with the same stationary sources played through speakers as in Section 5.3.3 , we obtained the same or comparable results. This is expected since we identified the same or similar frames for time delay extraction. The main improvement was with the robustness of this stage when tested with two actual human speakers, meaning the modified time delay estimator stage failed to identify appropriate time segments for time delay extraction less frequently than before. Due to reference sources not being possible for human

speakers, we cannot report PESQ or SIR values for human speakers. However, qualitatively we were able to obtain high quality and high intelligibility separation results for human speakers in an exhibit room at a conference, despite it being an acoustically challenging environment.

5.6 Conclusion

We develop a beamforming algorithm for blind source separation using a large-aperture microphone array. The algorithm estimates time delays between each source and microphone from the sound mixture of simultaneous sources, by using k-means cluster analysis to identify suitable frames for the estimate. This enables our algorithm to be "blind", since we do not require the location of the microphones and sources, and can adapt to the acoustic properties of each room and a change in location of the sources. We tested the algorithm using both commercial electret and LAE microphone arrays, with two and four simultaneous sources, and in all cases we obtained significant improvements in speech quality, as measured with PESQ and SIR. This coincided with qualitative listening tests carried out by human subjects. These improvements, combined with the simplicity of our algorithm, makes it a strong potential candidate for a real-time implementation for an embedded system.

Chapter 6

Conclusion

The objective of this thesis is to develop the technological foundations to enable LAE sensing systems, based on vertically integrating materials and devices, circuits and signal processing. Only through innovation across the entire signal processing chain can the full potential of LAE to provide high-spatial resolution sensing over physically large-areas be realized. Specifically, we focus on the design of hybrid systems, exploiting the complementary strengths of LAE, which enables spatially distributed sensors, and CMOS ICs, which provide high-performance computing.

In Part I of this thesis, as a vehicle to explore the effect of devices on LAE system design, we developed thin-film diodes. Initially, we developed entirely a-Si Schottky diodes, but found that their relatively low current density and associated low cutoff frequencies limited our systems. In particular, their poor frequency performance prevented us from using using these diodes as high-frequency rectifiers for non-contact inductive links and RFID tags. To overcome these limitation we needed to develop a high performance diode, but due to the requirement that the diodes had to be compatible with a-Si TFT processing, we were constrained by having to use the same starting materials, deposition equipment and low processing temperature as for the entirely a-Si Schottky diode. We achieved this objective by developing hybrid a-Si /

nc-Si Schottky diodes, which leverages the higher conductivity of nc-Si, and the high quality Schottky barrier formed between a-Si and Cr.

In Part II, as a vehicle to explore the potential for co-designing devices, circuits and signal processing algorithms, we developed a system for separating the voices of multiple simultaneous speakers, which can ultimately be fed to a voice-command recognition engine for controlling electronic systems. On a device level, we developed flexible piezopolymer microphones, made from PVDF, which were used to create a large-area microphone array. On a circuit level we developed localized a-Si TFT amplifiers, which ensured the integrity of the signal measured by the microphones, and a custom CMOS IC, which carried out system control, sensor readout and digitization. On a signal processing level we developed a novel algorithm for blind source separation in a real, reverberant room with real speakers, based on beamforming and binary masking. The algorithm requires no prior knowledge about the location of the speakers or microphones. Instead, from the recording with multiple simultaneous speakers, it uses cluster analysis techniques to determine the time delays between each speaker and each microphone; thus, adapting to the unique acoustic environment of the room.

6.1 Acknowledgements

Developing LAE systems is a team endeavour and the author would like to extend his thanks to all the members of the LAE systems team. In particular, the author would like to acknowledge the following specific contributions. In Chapter 3, the thin-film diodes were integrated with non-contact inductive interfaces, as part a large-area structural health monitoring system, by Yingzhe Hu. In Chapter 4 the entirety of CMOS IC was designed and characterized by Liechao Huang. He and the author also closely collaborated when integrating the LAE amplifiers with the CMOS IC. The

PVDF microphone fabrication process was refined, and techniques for optimizing the the frequency response were developed by Richard Cheng. The large-area scanning circuit was designed and characterized by Tiffany Moy.

6.2 Future Work

6.2.1 Thin-Film Diodes

- Oxide semiconductors, such as ZnO or IGZO, have been reported to have mobilities typically more than 5× greater than nc-Si, while still supporting low temperature processing, so as to be compatible with deposition on plastic. This suggests they should be able to attain far higher current densities and cutoff frequencies than the hybrid a-Si / nc-Si diode presented in this thesis. IGZO diodes with a current density of 800 A/cm^2 and a cutoff frequency of ∼1 GHz have already been developed [79], but questions remain about these devices ability to withstand even intermediate reverse bias voltages (e.g. -5 V), such as typically found in many LAE system applications.

- Our hybrid a-Si / nc-Si hybrid diode have a higher cutoff frequncy than the f_t of a-Si TFTs. The superior frequency performance of the hybrid diodes could potentially be leveraged when developing RF circuits that are composed entirely of diodes and inductors, such as certain RF mixers and switch topologies [80].

- DC Voltage multipliers, such as a Cockcroft-Walton multiplier or a Dickson charge pump, could be composed from thin-film diodes, so as to provide different voltage levels to different LAE system components. It would be interesting to compare this approach to using an inductive link (with voltage step-up) followed by a rectifier.

- We developed a reliable Schottky barrier between a-Si and chromium. This suggests that MESFETs might be a viable device, which could potentially improve upon the frequency performance of standard a-Si TFTs.

6.2.2 Flexible Microphones

- Flexible microphones have been demonstrated based on piezoelectrically-poled, cellular polypropylene, which is commercially known as Emfit. These microphones have the advantage over PVDF that since they operate exclusively in d_{33} mode (e.g. a pressure acting across the top faces leads to a voltage across the top faces), they do not operate at resonance and have a flat frequency response. They also can be used as a flexible, flat structure that does not require supports, which facilitates manufacturing. Furthermore, films can be stacked to increase sensitivity, since there is a linear relationship between the number of stacked films and the sensitivity [81]. The principal challenge of using cellular polypropylene is that the d_{33} coefficient, and associated microphone sensitivity, has been observed to degrade rapidly after poling [82].

  A potential alternative class of flexible, piezoelectric materials are piezocomposites, which consist of inorganic piezoelectric cylinders or square pillars, embedded in a polymer matrix. They combine the large piezoelectric coefficient and stability of the of the inorganic piezoelectric material, while still being flexible. Some of these materials are now being offered commercially, such as Macro Fiber Composite (MFC) from Smart Materials, which consists of a PZT / polymer composite [83].

- In the author's opinion the most robust, and only commercially viable option for developing large-area microphone arrays on a flexible substrate are MEMS condenser microphones. They provide a combination of low variability, ade-

quate physical robustness, high SNR and a small form-factor factor (e.g. small surface-mount packages), which makes them well suited for system development. MEMS microphone arrays, based on mounting MEMS microphones on flexible PCBs, have already been deployed for carrying out acoustic measurements in wind tunnels [84]. In the future one could envision strips of flexible MEMS microphones, which use a similar packaging format to today's LED strips.

6.2.3 Algorithms for Large-Area Microphone Arrays

- To further improve the time delay estimator stage of our blind source separation algorithm, especially in cases where there are more than two speakers and some of the speakers have a very low initial signal-to-interferer ratio (SIR) before processing, it would be worthwhile to assess other clustering algorithms that are different from k-means. In particular, it would be worthwhile to test density based clustering methods, such as DBSCAN, since they are typically considered to be more robust to noise and outliers [85]. To evaluate the effectiveness of the clustering algorithm, the correlation between the silhouette and the SIR ratio for every feature vector could be used, as shown in Fig. 5.7.

- By utilizing directional microphones and selecting only a subset of the microphones, which have the highest SIR for a given source before any processing, improved source separation results can be obtained. In effect the microphones are now carrying out some of the interference cancellation that used to be carried out by the algorithm. The algorithmic challenge would be to blindly identify (e.g. while multiples sources are speaking simultaneously) which microphones have the highest SIR for a given source.

- The output of the blind speech separation algorithm needs to be interfaced with an automatic speech recognition (ASR) engine to be able to identify voice

commands. One of the principal challenges is that distant speech recognition (DSR), due to reduced SNR and reverberations, typically performs significantly worse than close-talking speech recognition; currently, it is an area of active research [86]. There have been reports of ASR systems being adapted to function with binary masked speech [87], which suggests our blind source separation algorithms is viable for ASR. Nevertheless, to obtain satisfactory results, the author expects extensive research will be required in order to determine suitable features and training data.

- Near-field acoustic holography would be a compelling algorithm to explore using large-area microphone arrays. This is a technique in which a two-dimensional array of microphones is used to measure a two-dimensional sound field. Subsequently, the two-dimensional sound field can be transformed to other surfaces, so as to generate a three-dimensional description of the sound field. This technique is widely used industrially to enable noise-control engineers to obtain a visual image of noise. For example, it can be used when studying noise inside the cabin of an aircraft [88]. To visualize noise over a large space, planar arrays with many microphones are required, leading to approaches such as using a robotic arm to scan a small microphone array over a large area. Not only would large-area microphones arrays be well suited for these applications, but their flexible nature means that instead of only being able to use a planar array, their shape could be adjusted so as to aid when visualizing sound sources with different geometries, such as a cylindrical source.

Appendix A

Thin-Film Diode Fabrication Details

A.1 a-Si Schottky Diode

1. Deposit a SiN_x layer on a glass slide. This layer has no device functionality, but has been observed to improve yield.

    - Cleaning Step: PECVD, Ar=50 sccm, 200 °C, 500 mTorr, 12 W, 3 mins.
    - SiN_x Deposition: PECVD, 150nm thickness, $SiH_4/NH_3/H_2$=6/60/200 sccm, 200 °C, 500 mTorr, 5 W, 30 mins.

2. Deposit the bottom Cr contact.

    - Cr Deposition: Angstrom Sputter, 200 nm thickness, Sturm_DC2_Cr recipe.

3. Pattern and etch the bottom Cr contact (Cr-7 wet etchant). This corresponds to the first mask.

4. Deposit the device stack (n+ a-Si, followed by intrinsic a-Si).

- Cleaning Step: PECVD, Ar=50 sccm, 200 °C, 500 mTorr, 12 W, 3 mins.
- n+ a-Si Deposition: PECVD, 50 nm thickness, SiH_4/PH_3 (1% concentration)=44/10 sccm, 200 °C, 500 mTorr, 4 W, 4 mins.
- Intrinsic a-Si Deposition: PECVD, 1000 nm thickness, SiH_4/H_2=20/20 sccm, 200 °C, 500 mTorr, 6 W, 110 mins.

5. Leave the sample on a hotplate in ambient air at 200 °C for approximately 24 hours. This step has been observed to improve yield by reducing the number of devices with shorts between the top and bottom contact.

6. Deposit the top Cr contact.

   - Cr Deposition: Angstrom Sputter, 100 nm thickness, Sturm_DC2_Cr recipe.

7. Pattern and etch the top Cr contact (Cr-7 wet etchant). This corresponds to the second mask.

8. Pattern the device stack by dry etching. The top Cr contact from step 7 acts as a hard mask for this step.

   - RIE Etch: Samco 800, Recipe 66, 300 cycles.

9. Deposit a SiN_x passivation layer.

   - Cleaning Step: PECVD, Ar=50 sccm, 200 °C, 500 mTorr, 12 W, 3 mins.
   - SiN_x Deposition: PECVD, 600 nm thickness, $SiH_4/NH_3/H_2$=6/60/200 sccm, 200 °C, 500 mTorr, 5 W, 120 mins.

10. Pattern and etch the SiN_x passivation layer. This corresponds to the third mask.

    - RIE Etch: PT 720, CF_4/O_2=70/10 sccm, 100 mTorr, 100 W, 8 mins.

11. Deposit a final metal contact layer for leads and to allow for soldering.

    - Cr Deposition: Angstrom Sputter, 200 nm thickness, Sturm_DC2_Cr recipe.
    - Au Deposition: Angstrom Sputter, 200 nm thickness, Sturm_DC1_Au recipe.

12. Pattern and etch the final metal contact layer (GE-6 wet etchant for Au, Cr-7 for Cr). This corresponds to the fourth mask.

A.2 Hybrid a-Si / nc-Si Schottky Diode

1. Deposit a SiN_x layer on a glass slide. This layer has no device functionality, but has been observed to improve yield.

    - Cleaning Step: PECVD, Ar=50 sccm, 200 °C, 500 mTorr, 12 W, 3 mins.
    - SiN_x Deposition: PECVD, 150nm thickness, $SiH_4/NH_3/H_2$=6/60/200 sccm, 200 °C, 500 mTorr, 5 W, 30 mins.

2. Deposit the bottom Cr contact. Do not pattern the Cr at this stage, since for good adhesion the device stack needs to be deposited on top of a blank layer of Cr.

    - Cr Deposition: Angstrom Sputter, 200 nm thickness, Sturm_DC2_Cr recipe.

3. Deposit the device stack (intrinsic nc-Si, followed by n+ nc-Si).

    - Cleaning Step: PECVD, Ar=50 sccm, 200 °C, 500 mTorr, 12 W, 3 mins.

- Intrinsic nc-Si Deposition: PECVD, 750 nm thickness, SiH_4 (1% concentration)/H_2=40/100 sccm, 200 °C, 500 mTorr, 27 W at 80 MHz, 170 mins.
- n+ nc-Si Deposition: PECVD, 120 nm thickness, SiH_4(1% concentration) /PH_3(1% concentration)/H_2=40/20/100 sccm, 200 °C, 500 mTorr, 27 W at 80 MHz, 30 mins.

4. Deposit the top Cr contact.
   - Cr Deposition: Angstrom Sputter, 100 nm thickness, Sturm_DC2_Cr recipe.
5. Pattern and etch the top Cr contact (Cr-7 wet etchant). This corresponds to the second mask.
6. Pattern and etch the nc-Si device stack by dry etching. This corresponds to the third mask. The pattern of the device stack should be slightly larger than the top electrode from the previous step. This prevents the top electrode from over-hanging in the event of over-etching the nc-Si stack, and potentially collapsing and shorting with the bottom electrode.
   - RIE Etch: Samco 800, Recipe 66, 300 cycles.
7. Pattern and etch the bottom Cr contact (Cr-7 wet etchant). This corresponds to the fourth mask.
8. Deposit a SiN_x passivation layer.
   - Cleaning Step: PECVD, Ar=50 sccm, 200 °C, 500 mTorr, 12 W, 3 mins.
   - SiN_x Deposition: PECVD, 300nm thickness, SiH_4/NH_3/H_2=6/60/200 sccm, 200 °C, 500 mTorr, 5 W, 60 mins.

9. Pattern and etch the SiN_x passivation layer. This corresponds to the fifth mask.

    - RIE Etch: PT 720, CF_4/O_2=70/10 sccm, 100 mTorr, 100 W, 4 mins.

10. Deposit a final metal contact layer for leads and to allow for soldering.

    - Cr Deposition: Angstrom Sputter, 200 nm thickness, Sturm_DC2_Cr recipe.
    - Au Deposition: Angstrom Sputter, 200 nm thickness, Sturm_DC1_Au recipe.

11. Pattern and etch the final metal contact layer (GE-6 wet etchant for Au, Cr-7 for Cr). This corresponds to the sixth mask.

Appendix B

Diode Spice Models

B.1 a-Si Diode Spice Model

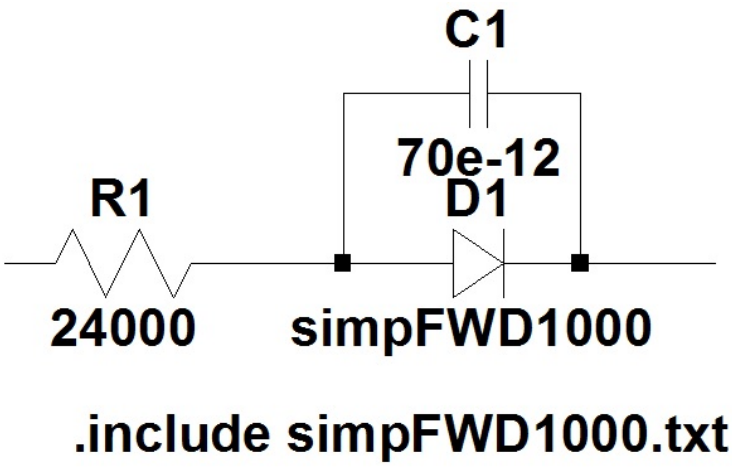

Figure B.1: Schematic of a-Si diode SPICE model.

.MODEL simpFWD1000 D (IS=5e-10 RS=0 TT=0 CJ0=0 VJ=0.7 m=0 N=1.39)

B.2 Hybrid a-Si / nc-Si Diode SPICE Model.

.MODEL simpFWD100 D (IS=5e-10 RS=0 TT=0 CJ0=0 VJ=0.7 N=1.31)

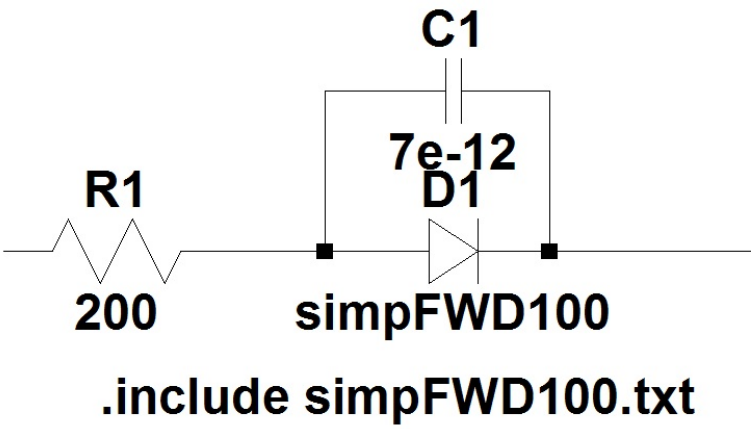

Figure B.2: Schematic of hybrid diode SPICE model.

Appendix C

Large-Area Microphone Arrays Measurements and Experimental Setup

C.1 Measuring the Microphone Frequency Response

To characterize the sensitivity of the microphone as a function of frequency, a reference microphone is required. As our calibrated reference microphone, we used a Dayton Audio EMM-6 microphone, which was characterized by Cross-Spectrum Labs (http://www.cross-spectrum.com). All microphone characterization measurements were carried out in the anechoic chamber of the 3D3A Lab at Princeton.

Microphones were characterized in the far field, with a distance greater than 1 m between the source and the microphone under test. The following procedure is used [89]:

1. Measure the frequency response of the calibrated reference microphone.

2. Measure the frequency response of the microphone under test.

3. Calculate the sensitivity (in mV/Pa) of the microphone under test based on the known sensitivity of the calibrated reference microphone:

$$Sensitivity_{mic\_under\_test}(f) = \frac{Sensitivity_{ref\_mic}(f)}{10^{(Response_{ref}(f) - Response_{mic\_under\_test}(f))/20}} \quad \text{(C.1)}$$

where the frequency response ($Response_{ref}(f)$ and $Response_{mic\_under\_test}(f)$) is in arbitrary dB units.

Particular care must be taken when measuring the frequency response to ensure the volume (sound pressure level) of the source is sufficiently high that the measurement has a high SNR. However, the sound pressure level should still be sufficiently low that minimal non-linear distortion occurs.

The frequency response of the microphone was characterized using an exponential swept sine wave that was emitted by the source [90]. This is the industry standard technique for measuring microphone frequency response, since it provides measurements that are rapid, have a high SNR and are robust to weak non-linearities. HOLMImpulse software from Holm Acoustics was used to carry out this frequency response measurement (http://www.holmacoustics.com/holmimpulse.php).

C.2 Measuring the Sound Pressure Level (SPL)

To measure the SPL of our frequency measurement we used a Galaxy CM-140 sound meter and played a 1 kHz tone. Alternatively, when the meter was not available, we connected the amplified output of the calibrated reference microphone (Dayton Audio EMM-6) to an oscilloscope. We used a 48 V phantom power (Behringer PS400) for biasing the reference microphone, since the sensitivity was calibrated by the manufacturer under these biasing conditions. We also used an amplifier (Stage Line MPA-102)

to ensure the signal was above the voltage floor of the oscilloscope. When calculating the SPL using the amplitude of the oscilloscope signal, it should be noted that the known sensitivity (in mV/Pa) of the calibration microphone is customarily quoted as an RMS voltage.

C.3 Simulating the Transfer Function Between a Source and a Microphone

To simulate the transfer function between a source and a microphone we used the image method [48]. Specifically, we used RIR Generator software from Audio Laboratories Erlangen (https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator) to carryout simulations. Sample Matlab source code for generating the room impulse response for a five microphone array is shown below:

```
c = 340; % Sound velocity (m/s)
fs = 16000; % Sample frequency
sLeft1 = [1.5 1 1.2]; % Source position [x y z] (m)
L = [5.0000001 5.000001 2.5]; % Room dimensions [x y z] (m)
beta = [0.8 0.8 0.8 0.8 0.8 0.8];
% Reverberation coefficients for each wall.
n = 2048; % Number of samples
mtype = 'omnidirectional'; % Type of microphone
order = -1; % -1 equals maximum reflection order
dim = 3; % Room dimensions
orientation = 0; % Microphone orientation (rad)
hp_filter = 0; % Enable high-pass filter
r= [5 5 5 5 5; 2.05 2.20 2.35 2.50 2.65; 1.20 1.20 1.20 1.20 1.20]' % Position of mics.
hLeft1 = rir_generator(c, fs, r, sLeft1, L, beta, n, mtype, order, dim, orientation,
hp_filter);% Generate room impulse response for each microphone-source pair.
MDO=fftfilt(hLeft1',MD);
% Convolute anechoic sound time series (MD) with room impulse response to generate
%sound in a reverberant room (MDO).
```

C.4 Characterizing the Performance of a Source Separation Algorithm

C.4.1 Signal-to-Interferer Ratio (SIR)

The signal-to-interferer ratio is a metric used to characterize the performance of a source separation or speech enhancement algorithm [49] [74]. It is defined by:

$$SIR = 10\ \log_{10}\left(\frac{\|S_{Target}(t)\|^2}{\|E_{Interferer}(t)\|^2}\right). \tag{C.2}$$

$S_{Target}(t)$ is the original sound source we wish to recover, while $E_{Interferer}(t)$ is the remaining component from the second source, which has not been fully removed by the separation algorithm. In this paper to calculate SIR we used the BSS_Eval Toolkit version 3.0 software package (http://bass-db.gforge.inria.fr/bss_eval) from INRIA (French Institute for Research in Computer Science and Automation). This requires the separated sound file, which is outputted by the algorithm. It also requires a reference sound file, which is obtained by playing the same source in isolation and taking a recording from one of the microphones in the array. In this work the microphone selected as the reference for SIR calculations was the same microphone that was used as a reference when measuring the time delays between microphones. This reference microphone for time delays was chosen to be the microphone at the center of the linear array. It should be noted that the time axis of the algorithm output and the reference sound should be aligned when calculating the SIR.

C.4.2 Perceptual Evaluation of Speech Quality (PESQ)

While SIR is a valuable tool when measuring the performance of source separation algorithm, it can sometimes fail to take into account certain artifacts that have a large impact on the speech quality as perceived by a human listener. For example, an artifact consisting of a single undesired tone could considerably reduce the speech quality, but it might have negligible detrimental impact on the SIR, since it has low power compared to the rest of the signal outputted by the algorithm. Therefore, when assessing the performance of a speech enhancement algorithm, it is beneficial to consider a psychoacoustic metric in addition to SIR.

PESQ (Perceptual evaluation of speech quality) is a widely used psychoacoustic model when assessing sound quality. It provides a PESQ MOS (mean opinion score) number, which ranges from -0.5 (low quality) to 4.5 (high quality) [75]. To calculate PESQ, the audio file outputted by the algorithm and a reference audio

file are required. In this work the reference audio file consists of a close-talking sound file recorded in an anechoic environment. It should be noted that the time axis of the algorithm output and the reference sound should be aligned when calculating PESQ. In this work PESQ calculations were carried out by the P. 862 Version 2 software (http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en) from the International Telecommunication Union Standardization Sector (ITU-T). To interface this software with Matlab we used the PESQ Matlab Driver by Arkadiy Prodeus (http://www.mathworks.com/matlabcentral/fileexchange/47333-pesq-matlab-driver).

C.5 PVDF Microphone Fabrication

The procedure for fabricating a PVDF microphone is illustrated in Fig. C.1 and described below:

1. The PVDF is cut into rectangular strips using a paper cutter. Special care must be taken to ensure that rectangular strips are cut lengthwise in the d_{31} poling direction, as labelled by the manufacturer.

   Already piezoelectrically poled PVDF sheets (28 um thick), coated with metal on both sides, were purchased directly from the Measurement Specialities Inc. (Part Number: 1007093-1, http://www.meas-spec.com/).

2. Two acrylic posts are glued to an acrylic backing plane. Cyanoacrylate glue is used throughout the entire microphone fabrication.

3. The rectangular PVDF strip is attached to to one of the posts.

   (a) A lead for electrically contacting the sheet is cut from a strip of flexible copper and glued to one of the acrylic posts.

(b) One end of the rectangular PVDF strip is glued to the same acrylic post as the copper lead. Care must be taken not to get glue on the top of the flexible copper lead, since it is electrically connected to the bottom electrode of the PVDF through metal to metal contact. The glue is left to dry for at least four hours.

4. The rectangular PVDF strip is attached to the other post.

   (a) A thin fishing line is glued to the other end of the PVDF strip, and a weight(20 g) is tied to the line. The line is passed through a smooth pulley.

   (b) Glue is applied to the acrylic post. The pulley, which is mounted on top of lab jack, is lowered until the PVDF comes into contact with the post. The glue is left to dry for at least four hours.

   (c) The weight is detached and the extra PVDF is cutoff.

   For applications where some microphone variance is acceptable, such as the array in Chapters 4 and 5, this step can be simplified by not using a pulley and weight. Instead, the PVDF can be tensioned approximately by hand.

5. A copper lead for electrically contacting the top electrode is attached to the PVDF. This lead is attached to PVDF in the region where it covers the acrylic post using conductive tape (3M Type 9713 XYZ axis electrically conductive, double sided tape).

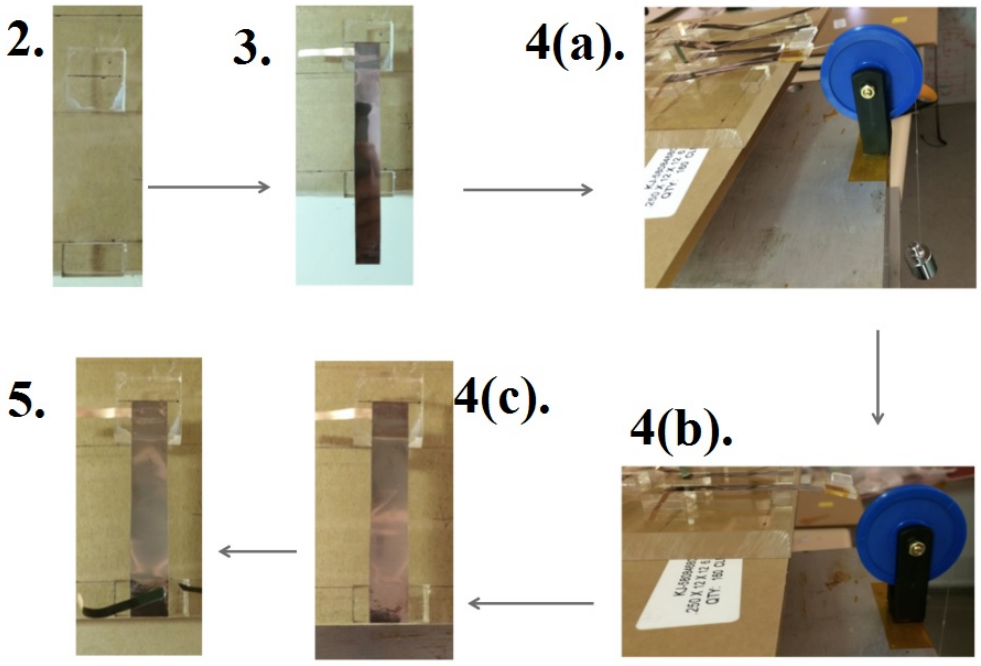

Figure C.1: Procedure for fabricating a PVDF microphone [55].

C.6 CMOS Amplifiers for Microphone Prototyping

C.6.1 Electret Condenser Microphones

The microphone consists of an EM172 electret, omnidirectional condenser capsule from Primo Microphones Inc. The amplifier consists of a non-inverting opamp configuration with a gain of 40 V/V. An LME49710 IC from Texas Instruments was chosen for the op-amp. The circuit schematic is shown in Fig. C.2.

C.6.2 Thin-film, PVDF Microphone

The PVDF microphone is connected differentially to an instrumentation amplifier with a gain of 1000 V/V. This gain is set by choosing a resistor value of $R_G = 10\,\Omega$, where R_G is the resistance set across the gain setting pins of the instrumentation am-

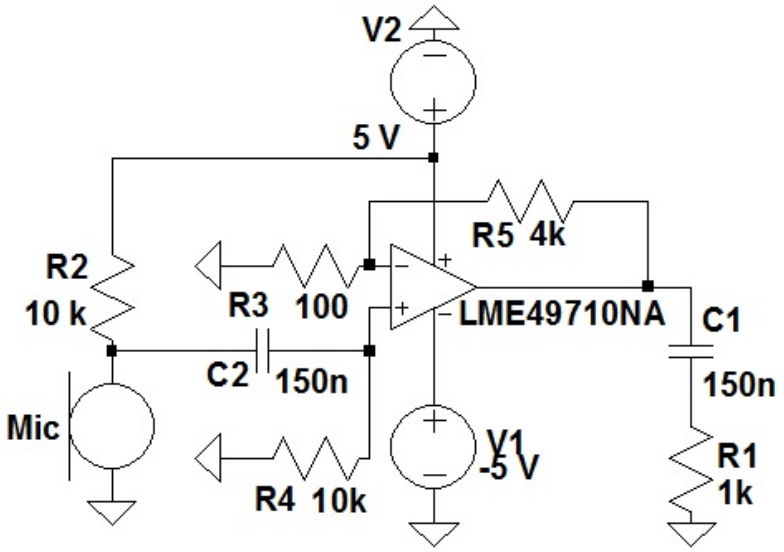

Figure C.2: Schematic of amplifier circuit for electret condenser microphone.

plifier IC. An INA217 IC from Texas Instruments was chosen for the instrumentation amplifier. The circuit schematic is shown in Fig. C.3.

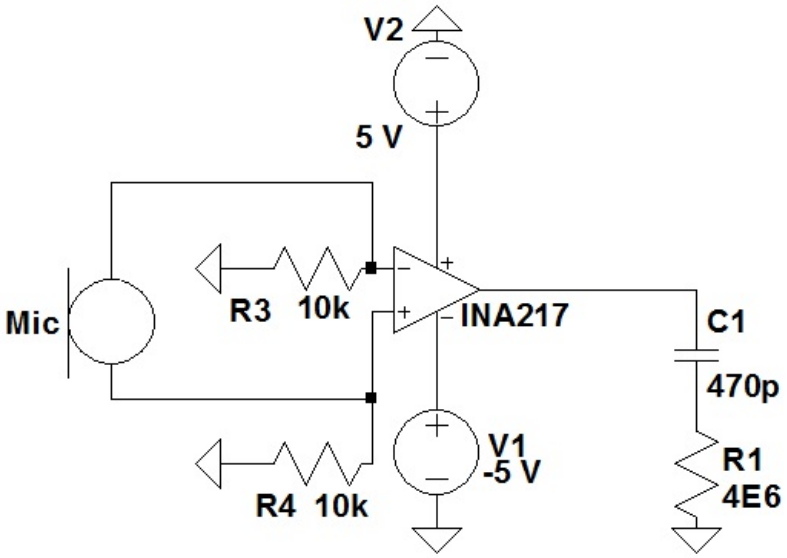

Figure C.3: Schematic of differential, instrumentation amplifier circuit for PVDF microphone.

C.7 Flexible Implementation of Combined Voice Separation and Gesture Sensing Sheet

To demonstrate the potential flexible nature of the technology, a demonstration system was constructed integrating together voice separation and gesture sensing, in collaboration with Honyang Jia. The system was implemented on a 72-inch-wide (183 cm) roll-up window shade purchased from a hardware store (see Fig. C.4). The gesture sensing sub-system consisted of a separate sheet with dimensions 40 cm width $\times$ 70 cm length, which had four sensing electrodes located around its perimeter. It was attached to the the right hand side of the windows shade (see Fig. C.5).The voice separation sub-system consisted a linear, uniformly spaced array of 16 electret condenser microphones (Primo Microphones Inc. EM172-Z1, with a 3 V operating voltage), with a microphone spacing of 11 cm (total width is 165 cm). For portability it was implemented on 3 separate PCBs, each with dimensions of 5 cm width $\times \sim 60$ cm length, attached with velcro to the top of the window shade. The window shade had circular openings for the microphone capsules.

An external PCB (attached with velcro to the board behind the window shade), with ICs for instrumentation circuitry, ADCs and microcontrollers (MSP430), received the signals from all the sensors (both gesture sensing electrodes and microphones). After these signals were digitalized on the PCB, they were sent to a PC via a USB FTDI connection. High-level algorithms for gesture labelling and speech separation were implemented on the PC using custom software.

The system was shown to be portable (it was packed in a ski bag and travelled from Princeton to California on a commercial flight for FLEX 2016 [91]), and demonstrated gesture sensing and voice separation in the field successfully.

Future work could involve: (a) physically integrating the ICs and connections onto the backplane, (b) wireless external connections, (c) on sheet power harvest-

ing and supply, (d) flexible microphones, and (e) integrating with automatic speech recognition.

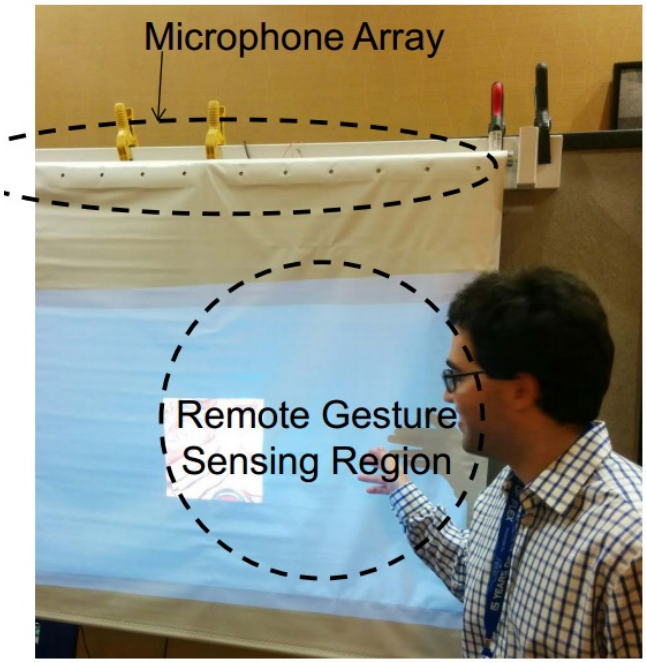

Figure C.4: Combined gesture sensing and speech separation demo implemented on a flexible window shade.

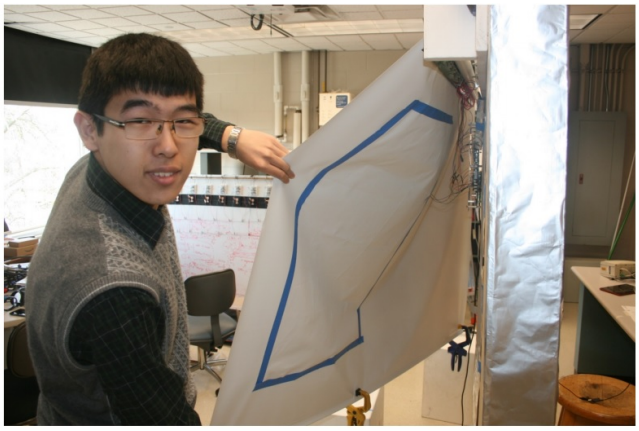

Figure C.5: Flexible gesture sensing sheet attached to the back of the window shade.

Appendix D

List of Publications and Patent Disclosures Resulting from this Thesis

D.1 Publications

J. Sanz-Robinson, L. Huang, T. Moy, W. Rieutort-Louis, Y. Hu, S. Wagner, J. C. Sturm, and N. Verma, "Robust Blind Source Separation in a Reverberant Room Based on Beamforming With a Large-Aperture Microphone Array," *to appear in Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April. 2016.

J. Sanz-Robinson, L. Huang, T. Moy, Y. Hu, W. Rieutort-Louis, S. Wagner, J. C. Sturm, and N. Verma, "Large-Area Microphone Array for Audio Source Separation Based on a Hybrid Architecture Exploiting Thin-Film Electronics and CMOS," *IEEE J. of Solid-State Circuits (JSSC),* vol. 51, no 4. pp. 979-991, April 2016. (invited)

W. Rieutort-Louis, J. Sanz-Robinson, T. Moy, L. Huang, Y. Hu, Y. Afsar, J. C. Sturm, N. Verma, S. Wagner, "Integrating and Interfacing Flexible Electronics in Hybrid Large-Area Systems," *IEEE Trans. Components, Packaging and Manufacturing Technology (TCPMT)*, vol. 5, no. 9, pp. 1219-1229, Sept. 2015. (invited)

L. Huang, W. Rieutort-Louis, A. Gualdino, L. Teagno, Y. Hu, J. Mouro, J. Sanz-Robinson, J. C. Sturm, S. Wagner, V. Chu, J. P. Conde, and N. Verma, "A System Based on Capactive Interfacing of CMOS with Post-processed Thin-film MEMS Resonators Employing Synchronous Readout for Parasitic Nulling," *IEEE J. Solid-State Circuits (JSSC)*, vol. 50, no. 4, pp. 1002-1015, April 2015. (invited)

N. Verma, Y. Hu, L. Huang, W. Rieutort-Louis, J. Sanz-Robinson, T. Moy, B. Glisic, S. Wagner, J. C. Sturm, "Enabling Scalable Hybrid Systems: architectures for exploiting large-area electronics in applications," *Proc. of IEEE*, vol. 103, no. 4, pp. 690-712, April 2015. (invited)

J. Sanz-Robinson, W. Rieutort-Louis, Y. Hu, L. Huang, N. Verma, S. Wagner, and J. C. Sturm, "Hybrid Amorphous / Nanocrystalline Silicon Schottky Diodes for High Frequency Rectification," *IEEE Electron Device Letters,* vol. 35, no. 4, pp. 425-427, April 2014.

Y. Hu, L. Huang, W. Rieutort-Louis, J. Sanz-Robinson, J. C. Sturm, S. Wagner, and N. Verma, "Self-powered System for Large-scale Strain Sensing by Combining CMOS ICs with Large-area Electronics," *IEEE J. of Solid-State Circuits (JSSC)*, vol. 49, no. 4, pp. 838-850, April 2014. (invited)

Hu, W. Rieutort-Louis, J. Sanz-Robinson, L. Huang, B. Glisic, J. C. Sturm, S. Wagner, and N. Verma, "Large-scale Sensing System Combining Large-area Electronics and CMOS ICs for Structural Health Monitoring," *IEEE J. of Solid-State Circuits (JSSC)*, vol. 49, no. 2, pp. 513-523, Feb. 2014.

W. Rieutort-Louis, L. Huang, Y. Hu, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "A Complete Fully Thin-film PV Harvesting and Power-management System on Plastic with On-sheet Battery Management and Wireles Power Delivery to Off-sheet Loads," *IEEE J. of Photovoltics (J-PV)*, vol. 4, no. 1, pp. 432-439, Oct. 2013. (invited)

D.2 Conferences

J. Sanz-Robinson, H. Jia, L. Huang, T. Moy, W. Rieutort-Louis, S. Wagner, N. Verma, and J. C. Sturm, "Hybrid Large-Area Microphone Array for Multi-User Voice Command Reconstruction", *Flexible and Printed Electronics Conference and Exhibition (FLEX)*, March 2016.

L. Huang, J. Sanz-Robinson, T. Moy, Y. Hu, W. Rieutort-Louis, S. Wagner, J. C. Sturm, and N. Verma, "Reconstruction of Multiple-user Voice Commands using a Hybrid System Based on Thin-film Electronics and CMOS," *VLSI Symp. on Circuits (VLSI)*, June 2015.

J. Sanz-Robinson, W. Rieutort-Louis, Y. Hu, L. Huang, T. Moy, S. Wagner, J. Sturm, and N. Verma, "Combining Thin Film Electronics and CMOS ICs: A Generalized Platform for Large-Area Sensing Systems", *Flexible and Printed Electronics Conference and Exhibition (FLEX)*, March 2015.

Y. Hu, T. Moy, L. Huang, W. Rieutort-Louis, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "3D Multi-gesture Sensing System for Large Areas based on Pixel Self-capacitance Readout using TFT Scanning and Frequency-conversion Circuits," *Custom Integrated Circuits Conference (CICC)*, Sept. 2014.

T. Moy, W. Rieutort-Louis, Y. Hu, L. Huang, J. Sanz-Robinson, J. C. Sturm, S. Wagner, and N. Verma, "Thin-film Circuits for Scalable Interfacing Between Large-area Electronics and CMOS ICs," *Device Research Conf. (DRC)*, June 2014.

W. Rieutort-Louis, L. Huang, Y. Hu, J. Sanz-Robinson, T. Moy, Y. Afsar, J. C. Sturm, N. Verma, and S. Wagner, "Characterization of Cutoff Frequency in Amourphous Silicon Thin-film Transistors," *Device Research Conf. (DRC)*, June 2014.

Y. Hu, L. Huang, W. Rieutort-Louis, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "3D Gesture Sensing System for Interactive Displays Based on Extended-range Capacitive Sensing," *Int'l Solid-State Circuits Conf. (ISSCC)*, Feb. 2014.

J. Sanz-Robinson, W. Rieutort-Louis, Y. Hu, L. Huang, N. Verma, S. Wagner, and J. C. Sturm, "High Current Density, Hybrid Nanocrystalline / Amorphous Silicon Schottky Diodes," *Int'l Conf. Amorphous and Nano-crystalline Semiconductors (ICANS)*, Aug. 2013.

Y. Hu, L. Huang, J. Sanz-Robinson, W. Rieutort-Louis, S. Wagner, J. C. Sturm, and N. Verma, "Energy Harvesting and Power Management for Sensing Skins in SHM

Applications," *Int'l Workshop on Structural Health Monitoring (IWSHM)*, Sept. 2013.

W. Rieutort-Louis, J. Sanz-Robinson, Y. Hu, L. Huang, J. C. Sturm, N. Verma, S. Wagner, "Readout from TFT Current-based Sensors Over Non-contact Interfaces using a TFT Gilbert-type Modulator," *Int'l Conf. on Amorphous and Nano-crystalline Semiconductors (ICANS)*, Aug. 2013.

W. Rieutort-Louis, L. Huang, Y. Hu, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "A Complete fully Thin-film PV Harvesting and Power-management System on Plastic with On-sheet Battery Management and Wireless Power delivery to Off-sheet Loads," *IEEE Photovoltaic Specialist Conf. (PVSC)*, June 2013.

Y. Hu, L. Huang, J. Sanz-Robinson, W. Rieutort-Louis, S. Wagner, J. C. Sturm, and N. Verma, "A Fully Self-powered Hybrid System Based on CMOS ICs and Large-area Electronics for Large-scale Strain Monitoring," *VLSI Symp. Circuits (VLSI)*, June 2013. (Best Student Paper Award)

W. Rieutort-Louis, Y. Hu, L. Huang, J. Sanz-Robinson, S. Wagner, N. Verma, and J. C. Sturm, "Effect of Low-Temperature TFT Processing on Power Delivery from Thin-Film Power Electronics on Flexible Substrates," *Materials Research Society Meeting (MRS)*, April 2013.

J. Sanz-Robinson, Y. Hu, W. Rieutort-Louis, L. Huang, N. Verma, S. Wagner, and J. C. Sturm, "Hybrid Nanocrystalline/Amorphous-silicon Schottky Diodes for Large-area Electronic Systems," *Materials Research Society Meeting (MRS)*, April 2013.

L. Huang, W. Rieutort-Louis, Y. Hu, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "A Super-regenerative Radio on Plastic Based on Thin-film Transistors and Antennas on Large Flexible Sheets for Distributed Communication Links," *Int. Solid State Circuits Conf. (ISSCC*), Feb. 2013.

J. Sanz-Robinson, W. Rieutort-Louis, N. Verma, S. Wagner, and J. C. Sturm, "A Full-wave Bridge Rectifier Based on Thin-film Amorphous-silicon Schottky Diodes for Wireless Power and Signal Transfer in Systems-on-plastic," *Materials Research Society Meeting (MRS)*, April 2012.

W. Rieutort-Louis, J. Sanz-Robinson, Y. Hu, L. Huang, J. C. Sturm, N. Verma, and S. Wagner, "Device Optimization for Integration of Thin-Film Power Electronics with Thin-film Energy-harvesting Devices to Create Power-delivery Systems on Plastic Sheets," *Int. Electron Device Meeting (IEDM)*, Dec. 2012.

Y. Hu, W. Rieutort-Louis, L. Huang, J. Sanz-Robinson, S. Wagner, J. Sturm, and N. Verma, "Flexible Solar-Energy Harvesting System on Plastic with Thin-film LC Oscillators Operating Above ft for Inductively-coupled Power Delivery," *Custom Integrated Circuits Conf. (CICC)*, Sept. 2012.

J. Sanz-Robinson, W. Rieutort-Louis, N. Verma, S. Wagner, and J. C. Sturm, "Frequency Dependence of Amorphous Silicon Schottky Diodes for Large-Area Rectification Applications" *Device Research Conf. (DRC)*, June 2012.

W. Rieutort-Louis, L. Huang, Y. Hu, J. Sanz-Robinson, S. Wagner, J. Sturm, and N. Verma, "Figure of Merit for Oscillator-based Thin-film Circuits on Plastic for

High-performance signaling, Energy Harvesting and Driving of Actuation Circuits," *Device Research Conf. (DRC)*, June 2012.

Y. Hu, W. Rieutort-Louis, J. Sanz-Robinson, K. Song, J. Sturm, S. Wagner, and N. Verma, "High-resolution Sensing Sheet for Structural-health Monitoring via Scalable Interfacing of Flexible Electronics with High-performance ICs," *VLSI Symp. Circuits (VLSI*), June 2012.

L. Huang, W. Rieutort-Louis, Y. Hu, J. Sanz-Robinson, S. Wagner, J. Sturm, and N. Verma, "Integrated All-silicon Thin-film Power Electronics on Flexible Sheets For Ubiquitous Wireless Charging Stations based on Solar-energy Harvesting, " *VLSI Symp. Circuits (VLSI)*, June 2012.

W. Rieutort-Louis, J. Sanz-Robinson, J. C. Sturm, S. Wagner, and N. Verma, "Thin-film Transistors and Circuit-design Styles for Scalable Control and Access Functionality over Sensor Arrays on Plastic," *Materials Research Society Meeting (MRS)*, April 2012.

D.3 Patent

Yingzhe Hu, Liechao Huang, Warren Rieutort-Louis, Josue Sanz-Robinson, Naveen Verma, Sigurd Wagner and James C. Sturm (2013). "System and Method for 3D Position and Gesture Sensing of Human Hand", Patent No. WO2014182824.

Bibliography

[1] W. Rieutort-Louis, J. Sanz-Robinson, T. Moy, L. Huang, Y. Hu, Y. Afsar, J. Sturm, N. Verma, and S. Wagner, "Integrating and interfacing flexible electronics in hybrid large-area systems," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 5, no. 9, pp. 1219 – 1229, Sep 2015.

[2] J. Viventi, D. Kim, L. Vigeland, E. Frechette, J. Blanco, Y. Kim, A. Avrin, V. Tiruvadi, S. Hwang, A. Vanleer, D. Wulsin, K. Davis, C. Gelber, L. Palmer, J. Spiegel, J. Wu, J. Xiao, Y. Huang, D. Contreras, J. Rogers, and B. Litt, "Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo," *Nature Neuroscience*, vol. 14, pp. 1599–1605, October 2011.

[3] K. Ishida, N. Masunaga, Z. Zhou, T. Yasufuku, T. Sekitani, U. Zschieschang, H. Klauk, M. Takamiya, T. Someya, and T. Sakurai, "A stretchable emi measurement sheet with 8x8 coil array, 2v organic cmos decoder, and -70dbm emi detection circuits in 0.18 m cmos," *International Solid-State Circuits Conference (ISSCC)*, pp. 472–474, Feb 2009.

[4] N. Verma, Y. Hu, L. Huang, W. Rieutort-Louis, J. Sanz-Robinson, T. Moy, B. Glisic, S. Wagner, and J. C. Sturm, "Enabling scalable hybrid systems: architectures for exploiting large-area electronics in applications," *Proceedings of IEEE*, vol. 103, pp. 690–712, April 2015.

[5] L. Zhou, S. Jung, E. Brandon, and T. N. Jackson, "Flexible substrate microcrystalline silicon and gated amorphous silicon strain sensors," *IEEE Transactions on Electron Devices*, vol. 53, pp. 380–385, Feb. 2006.

[6] H. Wang, L. Chen, J. Wang, Q. Sun, and Y. Zhao, "A micro oxygen sensor based on a nano sol-gel TiO_2 thin film," *Sensors*, vol. 14, pp. 16423–16433, Sept. 2014.

[7] C. Dagdeviren, Y. Su, P. Joe, R. Yona, Y. Liu, Y. Kim, Y.Huang, A. Damadoran, J. Xia, L. Martin, Y. Huang, and J. Rogers, "Conformable amplified lead zirconate titanate sensors with enhanced piezoelectric response for cutaneous pressure monitoring," *Nature Communications*, vol. 5, pp. 1–10, Aug. 2014.

[8] K. Ishida, T. C. Huang, K. Honda, Y. Shinozuka, H. Fuketa, T. Yokota, U. Zschieschang, H. Klauk, G. Tortissier, T. Sekitani, M. Takamiya, H. Toshiyoshi,

T. Someya, and T. Sakurai, "Insole pedometer with piezoelectric energy harvester and 2v organic digital and analog circuits," *IEEE Journal of Solid-State Circuits*, vol. 48, pp. 255–264, Jan. 2013.

[9] R. J. M. Vullers, R. van Schaijk, I. Doms, C. V. Hoof, and R. Mertens, "Micropower energy harvesting," *Solid-State Electronics*, vol. 53, no. 7, pp. 684–693, 2009.

[10] J. B. Bates, N. J. Dudney, B. Neudecker, A. Ueda, and C. D. Evans, "Thin-film lithium and lithium-ion batteries," *Solid State Ionics*, vol. 135, no. 1-4, pp. 33–45, 2000.

[11] J. Fjelstad, *Flexible Circuit Technology*. Delhi, India: BR Publishing, 3rd ed., 2006.

[12] K. Myny, S. V. Winckel, S. Steudel, P. Vicca, S. D. Jonge, M. J. Beenhakkers, C. W. Sele, N. A. J. M. van Aerle, G. H. Gelinck, J. Genoe, and P. Heremans, "An inductively-coupled 64b organic rfid tag operating at 13.56mhz with a data rate of 787b/s," *International Solid-State Circuits Conference (ISSCC)*, pp. 290–291, Feb. 2008.

[13] K. Myny, M. Rockel, A. Chasin, D. V. Pham, J. Steiger, S. Botnaras, D. Weber, B. Herold, J. Ficker, B. van der Putten, G. Gelinck, J. Genoe, W. Dehaene, and P. Heremans, "Bidirectional communication in an hf hybrid organic/solution-processed metal-oxide rfid tag," *International Solid-State Circuits Conference (ISSCC)*, pp. 312–313, Feb. 2012.

[14] Y. Hu, L. Huang, W. Rieutort-Louis, J. Sanz-Robinson, J. Sturm, S. Wagner, and N. Verma, "A self-powered system for large-scale strain sensing by combining CMOS ICs with large-area electronics," *Journal of Solid State Circuits (JSSC)*, vol. 49, Apr. 2014.

[15] Y. Hu, W. Rieutort-Louis, J. S. Robinson, L. Huang, B. Glisic, J. C. Sturm, S. Wagner, and N. Verma, "Large-scale sensing system combining large-area electronics and cmos ics for structural health monitoring," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 49, pp. 838–850, April 2014.

[16] W. Rieutort-Louis, L. Huang, Y. Hu, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "A complete fully thin-film pv harvesting and power-management system on plastic with on-sheet battery management and wireless power delivery to off-sheet loads," *IEEE Journal of Photovoltaics*, vol. 4, pp. 432–439, Jan. 2014.

[17] J. Sanz-Robinson, W. Rieutort-Louis, Y. Hu, L. Huang, N. Verma, S. Wagner, and J. C. Sturm, "Hybrid amorphous / nanocrystalline silicon schottky diodes for high frequency rectification," *IEEE Electron Device Letters*, vol. 35, pp. 425–427, April 2014.

[18] J. Sanz-Robinson, W. Rieutort-Louis, N. Verma, S. Wagner, and J. C. Sturm, "Frequency dependence of amorphous silicon schottky diodes for large-area rectification applications," *Device Research Conf. (DRC)*, pp. 135–136, June 2012.

[19] Y. Kuo, "Thin film transistor technology-past, present, and future," *Electrochemical Society Interface*, vol. 22, no. 1, pp. 55–61, 2013.

[20] R. Street, *Technology and Applications of Amorphous Silicon*. Springer-Verlag, 2000.

[21] R. Weisfield, M. Hartney, R. Street, and R. Apte, "New amorphous-silicon image sensor for x-ray diagnostic medical imaging applications," *Proc. SPIE 3336, Medical Imaging 1998: Physics of Medical Imaging*, vol. 3336, July 1998.

[22] C. R. Wronski and D. E. Carlson, "Surface states and barrier heights of metal-amorphous silicon schottky barriers," *Solid State Communications*, vol. 23, pp. 421–424, 1977.

[23] J. I. Pankove, *Semiconductor and Semimetals: Part C*, vol. 21. London, UK: Academic Press Inc., pp. 375-405, 1984.

[24] V. Chu, *Schottky barriers on amorphous silicon and amorphous silicon-germanium alloys*. PhD thesis, Princeton University, October 1989.

[25] I. Cheng, *Nanocrystalline silicon thin-film transistors on plastic substrates*. PhD thesis, Princeton University, June 2004.

[26] A. Shah, *Thin-Film Silicon Solar Cells*. Lausanne, Switzerland: EPFL Press, 2010.

[27] R. Street, *Hydrogenated Amorphous Silicon*. Cambridge University Press, 1991.

[28] T. Dylla, S. Reynolds, R. Carius, and F. Finger, "Electron and hole transport in microcrystalline silicon solar cells studied by time-of-flight photocurrent spectroscopy," *Journal of Non-Crystalline Solids*, vol. 352, pp. 1093–1096, June 2006.

[29] B. Yan, J. Yang, and S. Guha, "Amorphous and nanocrystalline silicon thin film photovoltaic technology on flexible substrates," *Journal of Vacuum Science & Technology A*, vol. 30, pp. 1–10, April 2012.

[30] Y. Huang, *Novel Approaches to Amorphous Silicon Thin Film Transistors for Large Area Electronics*. PhD thesis, Princeton University, Nov. 2011.

[31] Y. Kuo, *Thin Film Transistors: Polycrystalline silicon thin film transistors*. Kluwer Academic Publishers, pp. 428-430, 2004.

[32] F. Finger, J. Muller, C. Malten, R. Carius, and H. Wagner, "Electronic properties of microcrystalline silicon investigated by electron spin resonance and transport measurements," *Journal of Non-Crystalline Solids Volumes*, vol. 266, pp. 511–518, May 2000.

[33] P. Torres, J. Meier, R. Fluckiger, U. Kroll, J. A. Selvan, H. Keppner, and A. Shah, "Device grade microcrystalline silicon owing to reduced oxygen contamination," *Applied Physics Letters*, vol. 69, no. 10, pp. 1373–1375, 1996.

[34] U. Kroll, J. Meier, A. Shah, S. Mikhailov, and J. Weber, "Hydrogen in amorphous and microcrystalline silicon films prepared by hydrogen dilution," *Journal of Applied Physics*, vol. 80, no. 4971, 1996.

[35] B. Strahm, A. Howling, L. Sansonnens, and C. Hollenstein, "Plasma silane concentration as a determining factor for the transition from amorphous to microcrystalline silicon in sih4/h2 discharges," *Plasma Sources Science and Technology*, vol. 16, no. 1, pp. 80–89, 2006.

[36] R. Muller, T. Kamins, and M. Chan, *Device Electronics for Integrated Circuits*. John Wiley and Sons, 2002.

[37] B. Drevillon and F. Vaillant, "Oxidation of plasma-deposited hydrogenated amorphous silicon," *Journal of Thin Solid Films*, vol. 124, pp. 217–222, February 1985.

[38] S. Arimoto, H. Hasegawa, H. Yamamoto, and H. Ohno, "Anodic oxidation for enhancement of fabrication yield and efficiency of amorphous silicon solar cells," *Journal of The Electrochemical Society*, vol. 135, no. 2, pp. 431–436, 1988.

[39] D. Schroder, *Semiconductor Material and Device Characterization*. John Wiley and Sons, pp. 185 - 190, 2006.

[40] S. Ashok, A. Lester, and S. Fonash, "Evidence of space-charge-limited current in amorphous silicon schottky diodes," *IEEE Electron Device Letters*, vol. 1, pp. 200–202, October 1980.

[41] R. Street, W. Wong, and R. Lujan, "Low-temperature amorphous silicon pin photodiodes," *Physica Status Solidi B*, vol. 246, pp. 1854–1857, July 2009.

[42] I. Souleiman, K. Kandoussi, K. Belarbi, C. Simon, N. Coulon, S. Crand, and T.Mohammed-Brahim, "Schottky diode based on microcrystalline silicon deposited at 165c for rfid applications," *ECS Transactions*, vol. 33, pp. 227–236, Oct. 2010.

[43] T. Lee, *Planar Microwave Engineering: A Practical Guide to Theory, Measurement, and Circuits*. Cambridge University Press, pp. 275-280, 2004.

[44] J. Sanz-Robinson, L. Huang, T. Moy, Y. Hu, W. Rieutort-Louis, S. Wagner, J. C. Sturm, and N. Verma, "Large-area microphone array for audio source separation based on a hybrid architecture exploiting thin-film electronics and cmos," *IEEE Journal of Solid-State Circuits*, vol. 51, April 2016.

[45] E. Weinstein, E. Steele, K. Agarwal, and J. Glass, "A 1020-node modular microphone array and beamformer for intelligent computing spaces," tech. rep., MIT, M1T/LCS Technical Memo MIT-LCS-TM-642, 2004.

[46] H. V. Trees, *Optimum Array Processing: Part IV Detection, Estimation, and Modulation Theory*. New York: John Wiley and Sons, pp. 66, 2002.

[47] K. Kokkinakis and P. Loizou, *Advances in Modern Blind Signal Separation Algorithms: Theory and Applications*. Morgan and Claypool, pp. 13-19, 2010.

[48] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[49] 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, 2006.

[50] P. Kabal, "Tsp speech database," tech. rep., Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, Canada, September 2002.

[51] R. L. Freeman, *Fundamentals of Telecommunications*. John Wiley and Sons, pp. 90-91, 2005.

[52] L. Huang, J. Sanz-Robinson, T. Moy, Y. Hu, W. Rieutort-Louis, S. Wagner, J. C. Sturm, and N. Verma, "Reconstruction of multiple-user voice commands using a hybrid system based on thin-film electronics and cmos," *VLSI Symposium on Circuits (VLSIC), pp. 198 - 199*, no. JFS4-4, 2015.

[53] J. Spechler and C. Arnold, "Direct-write pulsed laser processed silver nanowire networks for transparent conducting electrodes," *Applied Physics A*, vol. 108, no. 1, pp. 25–28, July 2012.

[54] L. L. Beranek and T. J. Mellow, *Acoustics: Sound Fields and Transducers*. Elsevier, pp. 202-205, 2012.

[55] R. Cheng, *Polyvinylidene Fluoride (PVDF) Piezoelectric Microphones and Their Application to Sound Source Localization*. Undergraduate Thesis, Princeton University, April 2015.

[56] H. Gleskova and S. Wagner, "Amorphous silicon thin-film transistors on compliant polyimide foil substrates," *Electron Device Letters*, vol. 20, no. 9, pp. 473–475, 1999.

[57] L. Huang, W. Rieutort-Louis, Y. Hu, J. Sanz-Robinson, S. Wagner, J. Sturm, and N. Verma, "Integrated all-silicon thin-film power electronics on flexible sheets for ubiquitous wireless charging stations based on solar-energy harvesting," *VLSI Symp. Circuits (VLSI)*, pp. 198 – 199, June 2012.

[58] T. Moy, W. Rieutort-Louis, Y. Hu, L. Huang, J. Sanz-Robinson, J. C. Sturm, S. Wagner, and N. Verma, "Thin-film circuits for scalable interfacing between large-area electronics and cmos ics," *Device Research Conference*, pp. 271–272, June 2014.

[59] P. Sommen and C. Janse, "On the relationship between uniform and recurrent nonuniform discrete-time sampling schemes.," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5147–5156, 2008.

[60] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 2, pp. 109–116, 2003.

[61] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, Fl: CRC Press, pp. 38-40, 2013.

[62] C. Roads, *The Computer Music Tutorial*. Massachusetts Institute of Technology: MIT Press, pp. 553-555, 1996.

[63] I. Graz, M. Kaltenbrunner, C. Keplinger, R. Schwodiauer, S. Bauer, S. Lacour, and S. Wagner, "Flexible ferroelectret field-effect transistor for large-area sensor skins and microphones," *Applied Physics Letters*, vol. 89, 073501, 2006.

[64] J. Sanz-Robinson, L. Huang, T. Moy, W. Rieutort-Louis, Y. Hu, S. Wagner, J. C. Sturm, and N. Verma, "Robust blind source separation in a reverberant room based on beamforming with a large-aperture microphone array," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016.

[65] J. Benesty, M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*. Springer Verlag, pp. 1023-1029, 2008.

[66] M. Ikram and D. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1041–1044, 2000.

[67] E. Weinstein, K. Steele, A. Agarwal, and J. Glass, "LOUD: A 1020-node microphone array and acoustic," *International Conference on Sound and Vibration (ICSV)*, July, 2007.

[68] A. Levi and H. Silverman, "An alternate approach to adaptive beamforming using SRP-PHAT," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2726–2729, 2010.

[69] H. Kuttruff, "On the audibility of phase distortions in rooms and its significance for sound reproduction and digital simulation in room acoustics," *Acustica*, vol. 74, no. 1, pp. 3–7, June 1991.

[70] J. Sachar, H. Silverman, and W. Patterson III, "Microphone position and gain calibration for a large-aperture microphone array," *IEEE Transactions of Speech and Audio Processing*, vol. 13, no. 2, Jaunary 2005.

[71] T. Kim, H. Attias, S. Lee, and T. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 70–79, Jan 2007.

[72] C. Bishop, *Pattern Recognition and Machine Learning.* Springer, pp. 423-427, 2006.

[73] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[74] C. Fevotte, R. Gribonval, and E. Vincent, "BSS EVAL Toolbox User Guide Revision 2.0," tech. rep., April 2005.

[75] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ)," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 749–752, May 2001.

[76] K. Mardia and P. Jupp, *Directional Statistics.* Wiley, pp. 15, 2000.

[77] M. A. Richards, *Fundamentals of Radar Signal Processing.* McGraw-Hill, 2014. Supplemental Notes: Noncoherent Integration Gain, and its Approximation. Online: http://users.ece.gatech.edu/mrichard/Noncoherent0Approximations.pdf.

[78] K. Hara and R. Chellappa, "Growing regression forests by classification," *European Conference on Computer Vision (ECCV)*, pp. 552–567, July 2014.

[79] A. Chasin, M. Nag, A. Bhoolokam, K. Myny, S. Steudel, S. Schols, J. Genoe, G. Gielen, and P. Heremans, "Gigahertz operation of a-igzo schottky diodes," *IEEE Transactions on Electron Devices*, vol. 60, pp. 3407–3412, Oct. 2013.

[80] C. Coleman, *An Introduction to Radio Frequency Engineering.* Cambridge University Press, pp. 86-89, 2004.

[81] J. Hillenbrand and G. M. Sessler, "Stacked piezoelectret microphones of simple design and high sensitivity," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 13, pp. 973–978, Oct 2006.

[82] A. Mellinger, M. Wegener, W. Wirges, R. R. Mallepally, and R. Gerhard-Multhaupt, "Thermal and temporal stability of ferroelectret films made from cellular polypropylene/air composites," *Ferroelectrics*, vol. 331, pp. 189–199, Oct. 2006.

[83] K. Ramadan, D. Sameoto, and S. Evoy, "A review of piezoelectric polymers as functional materials for electromechanical transducers," *Smart Materials and Structures*, vol. 23, pp. 1–26, Jan. 2014.

[84] W. M. Humphreys, Q. A. Shams, S. S. Graves, B. S. Sealey, S. M. Bartram, and T. Comeaux, "Application of mems microphone array technology to airframe noise measurements," *11th AIAA/CEAS Aeroacoustics Conference*, vol. 3004, pp. 1–20, May 2005.

[85] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining.* Pearson, pp. 487-568, 2005.

[86] M. Woelfel and J. McDonough, *Distant Speech Recognition.* Wiley, 2009.

[87] W. Hartmann, A. Narayanan, E. Fosler-Lussier, and D. Wang, "A direct masking approach to robust asr," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, pp. 1993–2005, Oct. 2013.

[88] Y.-H. Kim, *Springer Handbook of Acoustics.* Springer-Verlag New York, pp. 1115-1137, 2014.

[89] M. Vorlander, *Handbook of Engineering Acoustics.* Springer Verlag, pp. 28, 2013.

[90] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *Audio Engineering Society Convention 108*, vol. 5093, pp. 1–23, Feb. 2000.

[91] H. Jia, J. Sanz-Robinson, S. Wagner, J. C. Sturm, and N. Verma, "Hybrid flexible electronics and cmos ic system for multi-user voice command reconstruction and 3d remote gesture sensing," *Flexible and Printed Electronics Conference and Exhibition (FLEX)*, March 2016.