Springer
Berlin
Heidelberg
New York
Barcelona
Hongkong
London
Milan
Paris
Singapore
Tokyo

Engineering ONLINE LIBRARY
http://www.springer.de/engine/

Michael Brandstein · Darren Ward (Eds.)

Microphone Arrays

Signal Processing Techniques and Applications

Springer

Jawbone Exhibit No. 2003, IPR2024-00350
Page 001

Series Editors

Prof. Dr.-Ing. Arild Lacroix
Johann-Wolfgang-Goethe-Universität
Institut für angewandte Physik
Robert-Mayer-Str. 2-4
D-60325 Frankfurt

Prof. Dr.-Ing.
Anastasios Venetsanopoulos
University of Toronto
Dept. of Electrical and Computer Engineering
10 King's College Road
M5S 3G4 Toronto, Ontario
Canada

Editors

Prof. Michael Brandstein
Harvard University,
Div. of Eng. and Applied Sciences
33 Oxford Street
MA 02138 Cambridge
USA
e-mail: msb@hrl.harvard.edu

Dr. Darren Ward
Imperial College, Dept. of Electrical Engineering
Exhibition Road
SW7 2AZ London
GB
e-mail: d.ward@ic.ac.uk

ISBN 3-540-41953-5 Springer-Verlag Berlin Heidelberg New York

Cip data applied for

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in other ways, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution act under German Copyright Law.

Springer-Verlag is a company in the BertelsmannSpringer publishing group
http://www.springer.de

© Springer-Verlag Berlin Heidelberg New York 2001
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready copy by authors
Cover-Design: de'blik, Berlin
SPIN: 10836055 62/3020 5 4 3 2 1 0 Printed on acid-free paper

Preface

The study and implementation of microphone arrays originated over 20 years ago. Thanks to the research and experimental developments pursued to the present day, the field has matured to the point that array-based technology now has immediate applicability to a number of current systems and a vast potential for the improvement of existing products and the creation of future devices.

In putting this book together, our goal was to provide, for the first time, a single complete reference on microphone arrays. We invited the top researchers in the field to contribute articles addressing their specific topic(s) of study. The reception we received from our colleagues was quite enthusiastic and very encouraging. There was the general consensus that a work of this kind was well overdue. The results provided in this collection cover the current state of the art in microphone array research, development, and technological application.

This text is organized into four sections which roughly follow the major areas of microphone array research today. Parts I and II are primarily theoretical in nature and emphasize the use of microphone arrays for speech enhancement and source localization, respectively. Part III presents a number of specific applications of array-based technology. Part IV addresses some open questions and explores the future of the field.

Part I concerns the problem of enhancing the speech signal acquired by an array of microphones. For a variety of applications, including human-computer interaction and hands-free telephony, the goal is to allow users to roam unfettered in diverse environments while still providing a high quality speech signal and robustness against background noise, interfering sources, and reverberation effects. The use of microphone arrays gives one the opportunity to exploit the fact that the source of the desired speech signal and the noise sources are physically separated in space. Conventional array processing techniques, typically developed for applications such as radar and sonar, were initially applied to the hands-free speech acquisition problem. However, the environment in which microphone arrays is used is significantly different from that of conventional array applications. Firstly, the desired speech signal has an extremely wide bandwidth relative to its center frequency, meaning that conventional narrowband techniques are not suitable. Secondly, there

is significant multipath interference caused by room reverberation. Finally, the speech source and noise signals may located close to the array, meaning that the conventional far-field assumption is typically not valid. These differences (amongst others) have meant that new array techniques have had to be formulated for microphone array applications. Chapter 1 describes the design of an array whose spatial response does not change appreciably over a wide bandwidth. Such a design ensures that the spatial filtering performed by the array is uniform across the entire bandwidth of the speech signal. The main problem with many array designs is that a very large physical array is required to obtain reasonable spatial resolution, especially at low frequencies. This problem is addressed in Chapter 2, which reviews so-called superdirective arrays. These arrays are designed to achieve spatial directivity that is significantly higher than a standard delay-and-sum beamformer. Chapter 3 describes the use of a single-channel noise suppression filter on the output of a microphone array. The design of such a post-filter typically requires information about the correlation of the noise between different microphones. The spatial correlation functions for various directional microphones are investigated in Chapter 4, which also describes the use of these functions in adaptive noise cancellation applications. Chapter 5 reviews adaptive techniques for microphone arrays, focusing on algorithms that are robust and perform well in real environments. Chapter 6 presents optimal spatial filtering algorithms based on the generalized singular-value decomposition. These techniques require a large number of computations, so the chapter presents techniques to reduce the computational complexity and thereby permit real-time implementation. Chapter 7 advocates a new approach that combines explicit modeling of the speech signal (a technique which is well-known in single-channel speech enhancement applications) with the spatial filtering afforded by multi-channel array processing.

Part II is devoted to the source localization problem. The ability to locate and track one or more speech sources is an essential requirement of microphone array systems. For speech enhancement applications, an accurate fix on the primary talker, as well as knowledge of any interfering talkers or coherent noise sources, is necessary to effectively steer the array, enhancing a given source while simultaneously attenuating those deemed undesirable. Location data may be used as a guide for discriminating individual speakers in a multi-source scenario. With this information available, it would then be possible to automatically focus upon and follow a given source on an extended basis. Of particular interest lately, is the application of the speaker location estimates for aiming a camera or series of cameras in a video-conferencing system. In this regard, the automated localization information eliminates the need for a human or number of human camera operators. Several existing commercial products apply microphone-array technology in small-room environments to steer a robotic camera and frame active talkers. Chapter 8 summarizes the various approaches which have been explored to accurately locate an individual in a practical acoustic environment. The emphasis is on precision in the face of adverse conditions, with an appropriate method presented in detail. Chapter 9 extends the problem to the case of multiple active sources. While again considering realistic environments, the issue is complicated by the presence of several talkers. Chapter 10 further generalizes the source localization scenario to include knowledge derived from non-acoustic sensor modalities. In this case both audio and video signals are effectively combined to track the motion of a talker.

Part III of this text details some specific applications of microphone array technology available today. Microphone arrays have been deployed for a variety of practical applications thus far and their utility and presence in our daily lives is increasing rapidly. At one extreme are large aperture arrays with tens to hundreds of elements designed for large rooms, distant talkers, and adverse acoustic conditions. Examples include the two-dimensional, harmonic array installed in the main auditorium of Bell Laboratories, Murray Hill and the 512-element Huge Microphone Array (HMA) developed at Brown University. While these systems provide tremendous functionality in the environments for which they are intended, small arrays consisting of just a handful (usually 2 to 8) of microphones and encompassing only a few centimeters of space have become far more common and affordable. These systems are intended for sound capture in close-talking, low to moderate noise conditions (such as an individual dictating at a workstation or using a hands-free telephone in an automobile) and have exhibited a degree of effectiveness, especially when compared to their single microphone counterparts. The technology has developed to the point that microphone arrays are now available in off-the-shelf consumer electronic devices available for under $150. Because of their growing popularity and feasibility we have chosen to focus primarily on the issues associated with small-aperture devices. Chapter 11 addresses the incorporation of multiple microphones into hearing aid devices. The ability of beamforming methods to reduce background noise and interference has been shown to dramatically improve the speech understanding of the hearing impaired and to increase their overall satisfaction with the device. Chapter 12 focuses on the case of a simple two-element array combined with postfiltering to achieve noise and echo reduction. The performance of this configuration is analyzed under realistic acoustic conditions and its utility is demonstrated for desktop conferencing and intercom applications. Chapter 13 is concerned with the problem of acoustic feedback inherent in full-duplex communications involving loudspeakers and microphones. Existing single-channel echo cancellation methods are integrated within a beamforming context to achieve enhanced echo suppression. These results are applied to single- and multi-channel conferencing scenarios. Chapter 14 explores the use of microphone arrays for sound capture in automobiles. The issues of noise, interference, and echo cancellation specifically within the car environment are addressed and a particularly effective approach is detailed. Chapter 15 discusses the applica-

Jawbone Exhibit No. 2003, IPR2024-00350
Page 003

tion of microphone arrays to improve the performance of speech recognition systems in adverse conditions. Strategies for effectively coupling the acoustic signal enhancements afforded through beamforming with existing speech recognition techniques are presented. A specific adaptation of a recognizer to function with an array is presented. Finally, Chapter 16 presents an overview of the problem of separating blind mixtures of acoustic signals recorded at a microphone array. This represents a very new application for microphone arrays, and is a technique that is fundamentally different to the spatial filtering approaches detailed in earlier chapters.

In the final section of the book, Part IV presents expert summaries of current open problems in the field, as well as personal views of what the future of microphone array processing might hold. These summaries, presented in Chapters 17 and 18, describe both academically-oriented research problems, as well as industry-focused areas where microphone array research may be headed.

The individual chapters that we selected for the book were designed to be tutorial in nature with a specific emphasis on recent important results. We hope the result is a text that will be of utility to a large audience, from the student or practicing engineer just approaching the field to the advanced researcher with multi-channel signal processing experience.

Cambridge MA, USA — *Michael Brandstein*
London, UK — *Darren Ward*
January 2001

Contents

Part I. Speech Enhancement

1 Constant Directivity Beamforming
Darren B. Ward, Rodney A. Kennedy, Robert C. Williamson 3
1.1 Introduction 3
1.2 Problem Formulation 6
1.3 Theoretical Solution 7
1.3.1 Continuous sensor 7
1.3.2 Beam-shaping function 8
1.4 Practical Implementation 9
1.4.1 Dimension-reducing parameterization 9
1.4.2 Reference beam-shaping filter 11
1.4.3 Sensor placement 12
1.4.4 Summary of implementation 12
1.5 Examples 13
1.6 Conclusions 16
References 16

2 Superdirective Microphone Arrays
Joerg Bitzer, K. Uwe Simmer 19
2.1 Introduction 19
2.2 Evaluation of Beamformers 20
2.2.1 Array-Gain 21
2.2.2 Beampattern 22
2.2.3 Directivity 23
2.2.4 Front-to-Back Ratio 24
2.2.5 White Noise Gain 24
2.3 Design of Superdirective Beamformers 24
2.3.1 Delay-and-Sum Beamformer 26
2.3.2 Design for spherical isotropic noise 26
2.3.3 Design for Cylindrical Isotropic Noise 30
2.3.4 Design for an Optimal Front-to-Back Ratio 30
2.3.5 Design for Measured Noise Fields 32
2.4 Extensions and Details 33
2.4.1 Alternative Form 33

18 Future Directions in Microphone Array Processing

Dirk Van Compernolle

Katholieke Universiteit Leuven, Leuven, Belgium

18.1 Lessons From the Past

Antenna array processing has had long-standing impact on phased array radars, sonars and radio astronomy for several decades. The gigantic antenna arrays that were constructed for deep space observation must stand out as some of the most impressive engineering achievements of any discipline. Success in these related fields of signal processing have without any doubt stimulated interest in microphone array processing. And these successes did not only generate interest, they did much more—they created high expectations. Another interest generating stimulus came from a very different field, i.e. the one of anatomy and physiology. Nature has endowed virtually all species with two ears. One good reason, of course, is that there is always a second as backup when one of the two fails. But at the same time we all know that our sense for orientation is helped considerably by the use of two ears instead of one and that it helps us understand each other in the midst of a noisy crowd.

After 20 years of active research, however, we cannot claim that microphone array processing has had the success many of us hoped for, and many will wonder when the great breakthrough in microphone array processing will finally come, if ever. Nevertheless, progress in computer technology has helped us in a big way. In the early days only analog schemes of limited signal processing complexity were possible. This was followed by early years of high cost DSP computations, where computational cost seemed to impede widespread use of the technology. Today we have affordable DSPs that allow us to implement all but the most complex schemes cheaply in digital signal processing technology in real-time. But this in itself was not enough. Apart from breaking through the computing bottleneck, our understanding of the problems at hand has significantly progressed, as witnessed in this book. Most of the results presented are from recent years and give new insight into both the potential and the limitations of microphone array processing. However, too often the same problems that were considered too hard ten or twenty years ago are still set apart for 'future research'. Admitted weaknesses to proposed solutions are similar to the ones that we have been struggling with for a long time. Generally speaking we may say that many proposed solutions add to our understanding but lack robustness in order to make a bright future for themselves.

So should we not ask ourselves if there is a fundamental issue with microphone array processing? And my answer is 'nothing is fundamentally wrong'. Microphone array processing has only proven to be quite a bit harder than other previously successful array processing applications. We have known the problems from the beginning, but have underestimated the impact of some of them in real-life situations.

The basic problems fall into a small number of categories: (i) the speech signal is broadband; (ii) in many practical situations the desired source is in a reverberant space in the near or mid field, is moving, and cannot be assumed to be a point source; and (iii) the speech signal changes rapidly, it is intermittent, and shares many characteristics with the competing and interfering signals.

It is very difficult to tackle all these issues at once. It is especially difficult to come up with tractable mathematical models for this complex environment. The result of this complex situation is that a lot of research effort has gone into, and continues to go into the search for optimal beamforming strategies that rely on extra assumptions and constraints. Sadly enough, this all too often leads to solutions that lack robustness when evaluated in a variety of real-life situations. It may be that a far field assumption is required, it may be that less reverberation would be sufficient, or it may be that a perfect predictive speech detector will bring the breakthrough. Surely these mathematical developments are relevant and give us a better understanding of broadband beamforming in general. Simultaneously we should admit to ourselves that robustness has been, and still is today, one of the main issues.

The drive to achieve (mathematically) optimal solutions is a natural underpinning of our science and engineering nature. But is microphone array processing not too complex to be solved with optimal approaches? Should we not expect real breakthroughs to come from so-called robust solutions that are clearly sub-optimal for any given circumstance, but applicable in a relatively wide range of situations? Also is it not obvious that there will not be a single solution, but that we need quite different solutions depending on the target application(s)? These observations go hand in hand with one of the major problems that has faced microphone arrays since their debut: size and cost. A large size always seemed a must from the requirement of uniform broadband beamforming. Some of the first microphone arrays, especially the one constructed in the auditorium at Bell Labs, were magically impressive by their shear size and number of microphones. They were great fun as a research project. Also they resulted in functional solutions. At the same time the price of such systems seems exorbitant. Later on we saw many arrays on the order of, say, 1 meter. Any such design is still only applicable to a very limited number of applications such as conference rooms. In the majority of potential applications such a bulky design has no place. No industry has screamed more for tiny and low cost solutions than the hearing aid industry. Here spacing of a few cm are the maximum and processing power is an order of magnitude less than in desktop applications. In all these situations we should not be surprised about the small size and limited number of sensors (two) in human hearing. It is far from optimal, but it works.

18.2 A Future Focused on Applications

If we ask ourselves what will the future bring for microphone array processing, we must envisage a range of widely differing solutions of different sizes and costs.

In the sequel I analyze the potential of the most important market segments and go looking for killer applications. If commercialization has not yet started, the question is of course what hampered commercial introduction and when, if ever, will we see usage of microphone arrays in each of these application fields.

18.2.1 Automotive

If any 'killer application' exists for microphone arrays, then it should be voice input in the car. It has all the right ingredients. Mobile telephony and speech recognition scream for hands-free voice input in a noisy environment. Signal-to-noise ratios obtained by single microphones are just not sufficient. Thus microphone arrays seem the logical solution. There are extra features that should help. The speakers inside a car are not mobile and their position is reasonably constant from session to session. Ultimately, a market potential of tens of millions of units per year should be commercially convincing. All of this should be sufficient for successful uptake of microphone array technology, but is it?

Today, penetration of microphone arrays in cars is minimal, except for a few top brands that are not all that noisy by themselves, and therefore have the least need for it. The major concern of car equipment manufacturers and car manufacturers alike is cost: multiple microphones, multiple wires, extra DSP power required, etc. Every cent in every component counts when putting a car together and microphone arrays have been judged as too expensive. Also, at least for the foreseeable future one should envisage that most microphones are mounted into existing cars, further complicating the story for arrays.

Given this large cost concern I do not believe that large arrays spanning the entire car will ever be viable. On the contrary, the car is an ideal environment for a microphone array that behaves like a traditional directional microphone but with a slightly steerable beam. Such an embedded array can be mounted as any regular microphone by not so specialized technicians. I do believe that the development of better microphones for usage inside the car will be a point of focus for microphone developers in the coming years.

Jawbone Exhibit No. 2003, IPR2024-00350
Page 006

18.2.2 Desktop

Cost has been the stumbling block as well for desktop microphone arrays in conjunction with speech recognition in the PC environment. People just do not like it if you tell them that the accuracy of a $50 speech recognition software package will drastically improve if they buy a $150 microphone array to go with it. It is still unclear if we will ever overcome the cost hurdle in this case. It may just be a question of whether large enough volumes will ever be reached such that current prices can be lowered drastically.

Microphone arrays for the desktop have just started to appear on the market. The reviews so far are ambiguous. In *quiet* environments they work as well as any headset worn. So if you do not want to be physically hooked up to your computer, this is the way to go the reviews say. At the same time the reviews will warn you that the existing commercial array microphones do not work well in considerable noise, and that one should not move around. Current reviews unanimously advise a wireless headset if one needs to move around a lot.

It seems therefore that current commercial implementations only solve a small part of the problem. All of the designs rely primarily on fixed beamforming, most often with limited directionality adjustment. On top of this, some additional noise suppression may be used. The 'speech seeking' part seems to be insufficient in all of the produced arrays. Also the quality and speed of tracking is substandard. It just shows how great the robustness issue really is when bringing microphone array technology to consumer products.

All in all there is reason for optimism, however. Desktop arrays are very pragmatic in their designs. These microphones are built for applications that use a PC screen or monitor, and they sit perfectly well on top of a monitor or attach to the front of it. Overall size is limited to about 20 cm, all computing is done inside the array, and the array connects to other equipment just as any other microphone would do. We have come a very long way to bring prices down enough such that a single enclosure with multiple elements, A/D converters and a DSP can be made at prices competing with traditional high end microphones. And let us not forget that these are first generation devices and that volumes are still very small.

Given some more time, I believe that there is hope that microphone arrays will capture a part of this market. Who knows, 5 years from now microphone arrays may be standard equipment on laptop and desktop computers. There is also a chicken and egg situation here. A wider usage of speech recognition would put more pressure on hardware manufactures to include higher end microphones, including arrays. On the other hand, one of the main hurdles in improving performance and subsequent acceptance of speech recognition is the low quality audio input on most systems today.

18.2.3 Hearing Aids

Hearing aids form a market by themselves. Restrictions on size and computational power are an order of magnitude more stringent than in other areas, leading to substantially different designs. Array sizes of 5 to 20 cm have been used in experiments with hearing aids, but have overall been met with disapproval. Nevertheless, here we have also seen the introduction of a range of new multi-microphone based products in the last couple of years. Many of these products do not use classical arrays, but a combination of microphones with different characteristics, used as inputs to a noise suppression stage. Perhaps even more obvious than in the automotive or desktop case, the evolution is towards an adaptive speech seeking and noise suppressing microphone. The distinction here between microphone technology and array technology is not entirely clear (but that does not really matter).

18.2.4 Teleconferencing

Teleconferencing was for some time seem as one of the potential killer applications. But I think that this is no longer true. On the one hand, the expansion of the teleconferencing market seems to have come down to slow growth and we see nothing of the explosion that some had hoped for. Therefore, the hope for a massive market does not seem justified. Acoustic echo cancellation is the crucial issue and it can not be solved by array processing. When using arrays, as with any multi-microphone input, the problem becomes significantly worse. Special microphone designs, including radial arrays, have been constructed and will continue to play a role in this market. Large wall mounted microphone arrays, however, are unlikely to find their way into teleconferencing rooms in any big way.

18.2.5 Very Large Arrays

Teleconferencing was one of the potential markets for large arrays. Another one is the virtual microphone in large auditoria. However, this can not be considered a booming market either. Design and manufacturing of these arrays is costly and a large degree of optimization may be required from site to site, making the picture even worse. Hence large microphone arrays are doomed to remain a niche market. They will certainly survive in high profile demonstration projects, and as a research topic they will carry on for many years to come. Another (quite niche) market for very large arrays exists in the acoustic monitoring industry.

18.2.6 The Signal Subspace Approach - An Alternative to Spatial Filtering ?

Finally, we should ask ourselves the question if we should not look for alternative solutions to plain spatial filtering. We may think in two directions:

Jawbone Exhibit No. 2003, IPR2024-00350
Page 007

blind signal separation and signal subspace approaches. These techniques do not require sensitive geometric information about the array layout but work with any configuration.

These techniques should result in higher configuration robustness. But at the same time they are computationally very demanding and, while making fewer assumptions about the layout, they make in general more assumptions about the signals. Practical implementations have not appeared so far, but demonstration results are often impressive. So we should keep an eye open for these techniques. It is unlikely we will find them in products in the coming years, but in later generation array processing techniques, they may become the standard way to go.

18.3 Final Remarks

The near-term trend is in one direction: small arrays with few microphones and a high degree of robustness that behave as speech seeking, directional, and noise canceling microphones. Depending on the target application designs may vary from less than a 1 cm in diameter for the hearing aid market, over 5 cm for the car, to a maximum of 20 cm for desktop. After all, human hearing does very well with two ears spaced about 20 cm apart. These designs will not reach maximal noise suppression in any theoretical sense. Their goal is clear: a few dB gain in signal-to-noise ratio across the board at a cost which is only marginally above that of other microphones. A market of several million units for such medium cost devices is realistic and therefore economically viable. Economic potential for large arrays is much more limited and will therefore remain a niche market.

Index

acoustic echo cancellation, 272, 281, 308, 386
acoustic model, 335, 337
adaptation control, 286
adaptive beamformer, 88, 163, 235, 309, 338
affine projection algorithm, 285
AIC, 195
array calibration, 310
array gain, 21, 46, 293
array geometry, 340
articulation index theory, 232
artificial neural network, 334
automotive microphone array, 309, 391
autoregressive modeling, 160, 191

Bark spectral distortion, 148
Baum-Welch algorithm, 335, 345
beamforming, 3, 87, 288
beampattern, 6, 22
blind signal separation, 325, 355
blocking matrix, 34, 88, 291

cardioid microphone, 69, 257
causal separation model, 358
cepstral mean subtraction, 337
cepstral prefiltering, 162
close-talking microphone, 345, 355, 385
clustering, 195
cochlear implant, 230
cocktail party effect, 355
coefficient-constrained adaptive filters, 94
coherence, 22, 63, 143, 161, 244
coherence matrix, 22, 45, 47
coherence of car noise, 259
coherence of office noise, 259
coherence of speech, 259
coherence-based post-filter, 271
coherent signal subspace method, 183
computed tomography scanner, 256, 266
constant directivity beamformer, 4
constrained minimum variance beamformer, 310
constrained optimization, 309
contrast functions, 364
correlation matrix, 114, 160, 314
correlation-based criteria, 366
cross-correlation, 40, 166
cross-correlation matrix, 114
cross-power spectrum, 189, 309
cross-power spectrum phase, 190, 338
cumulant-based separation criteria, 366
cylindrically isotropic noise field, 30, 73

data association problem, 222
decentralized Kalman filter, 212
decorrelation system, 371
delay-and-sum beamformer, 26, 159, 169, 234, 290, 338
density modeling criteria, 362
desktop microphone array, 392
differential microphone, 67
diffuse noise field, 23, 309, 338
diffuse sound field, 49, 257
dipole microphone, 69
direction of arrival, 162, 182, 243, 290, 302
directional gain, 384
directivity index, 23, 44, 293
distributed multi-microphone, 348
disturbed harmonics model, 189
double-talk, 65
dual excitation speech model, 137
dynamic time warping, 334

Jawbone Exhibit No. 2003, IPR2024-00350
Page 008