

## 16 Blind Separation of Acoustic Signals

Scott C. Douglas

Southern Methodist University, Dallas TX, USA

**Abstract.** This chapter presents an overview of criteria and algorithms for the blind separation of linearly mixed acoustic signals. Particular attention is paid to the underlying statistical formulations of various approaches to the convolutive blind signal separation task, and comparisons to other blind inverse problems are made. Several algorithms are described, including a novel algorithm that largely maintains the spectral content of the original mixtures in the extracted source signals. Numerical experiments are also provided to explore the behaviors of the algorithms in real-world blind signal separation tasks.

### 16.1 Introduction

#### 16.1.1 The Cocktail Party Effect

Consider for a moment the last time you had a conversation with a friend at a social gathering. Numerous conversations were occurring around you, and yet you had the capability of focusing your attention on your friend's phrases. This so-called "cocktail party effect" [1] is well-known and taken for granted by almost everyone. What are the system elements at work in this situation?

- *An acoustic sensor array.* In our case, two sensors (our ears) collect acoustic information, although there is nothing precluding the use of more than two sensors in the task from a system design standpoint.
- *A computational processing system.* Most auditory models allow for several layers of signal processing, from lower-level frequency-dependent feature extraction within the inner ear to higher-level pattern recognition within the brain [2].

The current boom in electronic devices, while creating a revolution in human lifestyles, also creates the need for sensing and translating acoustical information into useful electronic forms. Since we use our ears as information sensors in conversations with others, it seems only natural to develop similar listening capabilities in information-processing devices. This quest remains a challenging problem, despite several decades of research in microphone arrays, voice recognition, and natural language processing. Moreover, while speech can be accurately collected using close-to-the-mouth microphones, there are applications in which such placements are impractical or undesirable, such as in audio teleconferencing systems and automobile speakerphones. Clearly,

34. P. Regalia and P. Loubaton, "Rational subspace estimation using adaptive lossless filters," *IEEE Trans. Signal Processing*, vol. 40, pp. 2392-2405, Oct. 1992.
35. S. Douglas, S. Amari, and S.-Y. Kung, "Adaptive paraunitary filter banks for spatio-temporal principal and minor subspace analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-99)*, Phoenix AZ, USA, pp. 1089-1092, Mar. 1999.
36. X. Sun and S. Douglas, "Multichannel blind deconvolution of arbitrary signals: Adaptive algorithms and stability analyses," in *Proc. 34th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove CA, USA, pp. 1412-1416, Oct. 2000.
37. The Bobs, "Boy around the corner," from *Songs For Tomorrow Morning*, [audio recording], Rhino Records, 1988.



## 17 Future Directions for Microphone Arrays

Gary W. Elko

Media Signal Processing Research, Agere Systems, Murray Hill NJ, USA

### 17.1 Introduction

The telephone, invented more than 140 years ago, has changed very little with respect to acoustic design in the last 40 years. The most recent technological improvements were the replacement of the carbon microphone with an electret microphone and the development of the speakerphone for hands-free telephony in the early sixties. It might seem quite surprising that in spite of the rapid changes in communication systems: digital switching, digital wireless, packet-based data communications and the like, that very little has changed to the acoustic front-end in telephony systems. For instance, with the general availability of broadband communication networks, why have we not significantly increased the audio communication bandwidth? Why don't we see stereo or multichannel in telephony and communication systems? Why don't we see more ubiquitous adoption of voice recognition in consumer products? Why can't we tailor the audio response of our communication systems the same way we do with our entertainment systems? Why do our cell phones sound so tinny, distorted, and noisy? Why is it so difficult to understand people who call us from noisy environments or call us in noisy environments? These are good questions when one ponders the explosion of inexpensive digital signal processing integrated circuits combined with the current bandwidth of communication systems and those that are about to be introduced. Clearly front-end acoustic signal processing will play a large role in addressing these questions.

In our increasingly mobile society we find that hands-free operation of communication systems is becoming more and more the norm, whether we talk to other humans or to machines. There are significant challenges in enabling high quality speech communication in difficult acoustic environments. Microphone arrays are an enabling technology for hands-free communication since these systems offer directional gain to improve the signal-to-noise ratio. This chapter will address some new possibilities as well as outstanding research issues that remain.

### 17.2 Hands-Free Communication

The hands-free speakerphone was first released by AT&T's Western Electric in 1954. The speakerphone allowed many users gathered in one location to

Since turbulent length scales are orders of magnitude lower than acoustic propagating wavelengths, there is the possibility of using adaptive frequency-wavenumber filtering to remove high wavenumber turbulent sensitivity from the microphone array. MEMS technology also allows one to easily build complex combinations of pressure, velocity, and acceleration signals. Combinations of these different types of sensors with integral signal processing opens up a whole new area for investigation.

## 17.4 Conclusions

Microphone array beamforming for human-to-human and human-to-machine communication is now a viable and cost effective solution. Microphone arrays will play an essential role in the solution to the difficult problem of hands-free communication. The biggest challenge that remains is that it will be essentially impossible to obtain an SNR that is as high as a microphone mounted close to the desired talker (a telephone handset for instance). Linear combinations of microphones can only offer a maximum gain of  $20 \log N$ , where  $N$  is the number of microphone elements, and typical realizable gains are much less than this limit. Thus one is led to conclude that we need to search for new parametric and nonlinear processing algorithms in hopes of obtaining directional gains that significantly exceed the gains for linear processing. This is the challenge that has to be met if we ever hope to attain handset quality speech in hands-free conditions. One possible solution to this quality gap is to resort to wearable acoustic communication devices that can be thought of as "acoustic jewelry".

## References

1. G. W. Elko, "Superdirectional microphone arrays", in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, eds., Kluwer Academic Publishers, chapter 10, pp. 181-237, 2000.
2. K. D. Kryter, "Validation of articulation index", *J. Acoust. Soc. Am.*, vol. 34, no. 11, pp. 1698-1702, Nov. 1962.
3. W. Kellermann, "A self steering digital microphone array", in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pp. 3581-3584, 1991.
4. G. W. Elko, R. A. Kubli, D. R. Morgan, and J. E. West, *Adjustable filter for differential microphones*, U.S. Patent US05303307, 1994.
5. W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays", in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pp. 219-222, 1997.
6. P. M. Zurek, J. E. Greenberg, and P. M. Peterson, "Sensitivity to design parameters in an adaptive-beamforming hearing aid", in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pp. 1129-1132, 1990.

## 18 Future Directions in Microphone Array Processing

Dirk Van Compernelle

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.

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
- beam pattern, 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

EASI algorithm, 371  
 echo path gain, 293  
 echo return loss enhancement, 274, 283  
 echo shaping, 272  
 end-point detection, 333, 340  
 endfire array, 234, 275  
 entropy, 364  
 ESPRIT, 182  
 extended Kalman filter, 210  
 Eyring's formula, 118  
  
 face tracking, 204, 218  
 farfield, 25  
 farfield assumption, 3, 290  
 fast Newton algorithm, 286  
 feature extraction, 333, 340, 344  
 filter-and-sum beamformer, 5, 159, 171, 204, 290  
 filterbank, 45, 116, 184, 317  
 fixed beamformer, 234, 310  
 focused beamformer, 159  
 frequency-domain beamforming, 5  
 front-to-back ratio, 24, 30  
  
 generalized cross-correlation, 161, 183, 191  
 generalized sidelobe canceler, 35, 40, 88, 113, 127, 236, 291, 310, 338  
 generalized singular value decomposition, 115  
 gradient microphones, 35  
  
 hands-free speech acquisition, 181, 255, 307  
 harmonic nesting, 5, 340  
 hearing aid, 393  
 hearing impairment, 229  
 hidden Markov model, 334  
 high-resolution spectral estimation, 158  
 hypercardioid microphone, 36, 243, 259, 274  
  
 image method, 52, 118, 182, 342  
 in situ calibrated microphone array, 312  
 injected noise, 311  
 instantaneous blind signal separation, 360  
 intelligibility-weighted directivity, 232  
 intelligibility-weighted gain, 232

interacting multiple model, 222  
 interference suppression, 240  
 inverse Kalman filter, 213  
  
 Jacobian matrix, 211  
 joint diagonalization, 115, 367  
  
 Kalman filter, 206  
 Kullback-Leibler divergence measure, 362  
  
 language model, 332, 335  
 leaky adaptive filters, 92  
 leaky LMS, 312  
 linear intersection method, 195  
 linear prediction, 141, 185, 334, 371  
 LMS algorithm, 236  
 location-estimation null-steering algorithm, 243  
 Lombard effect, 332  
 loudspeaker-enclosure-microphone system, 283  
  
 magnitude squared coherence function, 257  
 matched filter, 134, 308, 310, 338  
 matrix inversion lemma, 43  
 maximum likelihood linear regression, 337  
 MDL, 195  
 mean square error, 114, 290  
 mel-scaled cepstral coefficients, 334  
 MEMS, 386  
 microphone channel response, 165  
 microphone directivity, 232, 257, 261, 357  
 microphone orientation, 257  
 microphone placement, 157, 261, 308  
 MIN-NORM, 182  
 minimum mutual information, 363  
 minimum variance spectral estimation, 160  
 MMSE beamformer, 313, 314  
 MMSE filter, 263  
 multi-channel blind deconvolution, 361  
 multi-resolution time-frequency adaptive beamformer, 319  
 MUSIC, 160, 182, 186  
 MVDR beamformer, 25, 39

- narrowband assumption, 3
- natural gradient adaptation, 368
- nearfield, 25, 290, 385
- nearfield superdirective array, 44
- nested array, 290
- NLMS algorithm, 266, 268, 284, 313
- noise canceler, 34, 89
- noise cancellation, 61, 256, 308
- noise subspace, 182, 309
- nonholonomic constraints, 370
- nonlinear optimization, 367
- normalized interference suppression, 322
- normalized kurtosis, 365
- normalized noise suppression, 322
- omnidirectional microphone, 232, 257, 331
- order permutation, 372
- parallel model combination, 337
- phase transform, 162, 167
- phone, 335
- phoneme recognition, 230
- PLP coefficients, 334
- post-filter, 39, 255, 338
- power spectral density matrix, 21
- pseudo-inversion, 368
- QR-decomposition, 122
- recursive state estimation, 205
- reverberation distance, 258
- reverberation time, 118, 164, 173, 283
- RLS algorithm, 285, 313
- robust estimation, 195
- robustness, 88, 121, 336, 377
- room impulse response, 162, 367
- ROOT-MUSIC, 189
- Sabine's equation, 258
- scaled projection algorithm, 29, 242
- scoring problem, 335
- segmental SNR, 269
- sensor calibration, 309
- sensor fusion, 205
- sensor mismatch, 235
- short-time Fourier transform, 317
- short-time modified coherence, 341
- signal bias removal, 337
- signal cancellation, 90, 338
- signal subspace, 182
- signal-to-interference ratio, 374
- signal-to-noise plus interference power ratio, 316
- single-channel speech enhancement, 136, 255
- singular value decomposition, 189, 193
- source activity detection, 267, 294
- spatial aliasing, 4, 26
- spatial clustering, 183
- spatial correlation function, 61
- spatial dithering, 312
- speakerphone, 383
- spectral subtraction, 40, 242, 308
- speech corpora, 333, 341
- speech intelligibility, 230
- speech quality measures, 51
- speech reception threshold, 229
- speech recognition, 281, 292, 307, 331, 385
- speech-to-interference ratio, 230
- spherically isotropic noise field, 26, 65
- state error information vector, 214
- statistical pattern recognition, 333
- steered response power, 158
- steering vector, 169, 184
- stochastic matching, 337
- stochastic region contraction, 159
- subband decomposition, 314
- subspace tracking, 369
- super-resolution, 310
- superdirective array with Wiener post-filter, 44
- superdirectivity, 19, 234, 290, 338
- supergain, 19
- teleconferencing, 281, 300, 355, 393
- time delay estimation, 40, 161, 267, 338
- time-difference of arrival, 158, 168, 183, 204
- time-domain beamforming, 5
- time-invariant beamforming, 290
- time-varying beamforming, 291
- Toeplitz matrix, 117, 191
- training data contamination, 337
- training problem, 335
- transform-domain structure, 287

- uncorrelated noise field, 24
- universal noise subspace, 186
- universal signal subspace, 186
- universal spatial covariance matrix, 185
- unvoiced speech, 137, 140
- varechoic chamber, 342
- videoconferencing, 163, 181, 203
- Viterbi algorithm, 335
- voice activity detector, 312, 343, 347
- voiced speech, 137, 140
- wavelet transform, 141
- white noise gain, 24, 47
- whitening filter, 190, 266
- wideband weighted subspace fitting, 186
- Wiener filter, 39, 41, 62, 114, 263, 291, 308, 338
- Wiener solution, 35
- Wiener-Hopf equation, 42, 284
- word recognition rate, 346
- WSF, 182

# APPENDIX 1003B

DIGITAL SIGNAL PROCESSING

M. Brandstein • D. Ward (Eds.)

# Microphone Arrays

Signal Processing Techniques  
and Applications



Springer

Metas Platform, Inc. - Exhibit 1015

138 of 250

Brandstein • Ward



Microphone Arrays

5986

53

001

PPY 1

LIBRARY OF CONGRESS



0 007 875 135 6

The study and implementation of 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. 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. 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.

ISBN 3-540-41953-5



9 783540 419532

<http://www.springer.de>

Brandenstein, Ward



Microphone Arrays

TK 59

.M53

2001

COPY

Springer

*Berlin*

*Heidelberg*

*New York*

*Barcelona*

*Hongkong*

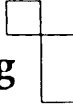
*London*

*Milan*

*Paris*

*Singapore*

*Tokyo*

Engineering  ONLINE LIBRARY

<http://www.springer.de/engine/>

Michael Brandstein · Darren Ward (Eds.)

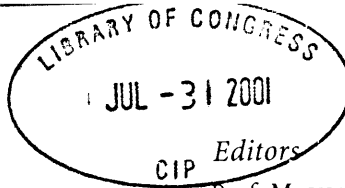
# Microphone Arrays

Signal Processing  
Techniques and Applications

With 149 Figures



Springer



*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

*CIP 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 be 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 individ-

ual 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-

## VIII Preface

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  
London, UK  
January 2001

*Michael Brandstein*  
*Darren Ward*

# Contents

---

## Part I. Speech Enhancement

---

### 1 Constant Directivity Beamforming

|                                                                      |    |
|----------------------------------------------------------------------|----|
| <i>Darren B. Ward, Rodney A. Kennedy, Robert C. Williamson</i> ..... | 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

|                                                       |    |
|-------------------------------------------------------|----|
| <i>Joerg Bitzer, K. Uwe Simmer</i> .....              | 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 |

|                                                                                                 |     |
|-------------------------------------------------------------------------------------------------|-----|
| 2.4.2 Comparison with Gradient Microphones .....                                                | 35  |
| 2.5 Conclusion .....                                                                            | 36  |
| References .....                                                                                | 37  |
| <b>3 Post-Filtering Techniques</b>                                                              |     |
| <i>K. Uwe Simmer, Joerg Bitzer, Claude Marro</i> .....                                          | 39  |
| 3.1 Introduction .....                                                                          | 39  |
| 3.2 Multi-channel Wiener Filtering in Subbands .....                                            | 41  |
| 3.2.1 Derivation of the Optimum Solution .....                                                  | 41  |
| 3.2.2 Factorization of the Wiener Solution .....                                                | 42  |
| 3.2.3 Interpretation .....                                                                      | 45  |
| 3.3 Algorithms for Post-Filter Estimation .....                                                 | 46  |
| 3.3.1 Analysis of Post-Filter Algorithms .....                                                  | 47  |
| 3.3.2 Properties of Post-Filter Algorithms .....                                                | 49  |
| 3.3.3 A New Post-Filter Algorithm .....                                                         | 50  |
| 3.4 Performance Evaluation .....                                                                | 51  |
| 3.4.1 Simulation System .....                                                                   | 52  |
| 3.4.2 Objective Measures .....                                                                  | 52  |
| 3.4.3 Simulation Results .....                                                                  | 54  |
| 3.5 Conclusion .....                                                                            | 57  |
| <b>4 Spatial Coherence Functions for Differential Microphones<br/>in Isotropic Noise Fields</b> |     |
| <i>Gary W. Elko</i> .....                                                                       | 61  |
| 4.1 Introduction .....                                                                          | 61  |
| 4.2 Adaptive Noise Cancellation .....                                                           | 61  |
| 4.3 Spherically Isotropic Coherence .....                                                       | 65  |
| 4.4 Cylindrically Isotropic Fields .....                                                        | 73  |
| 4.5 Conclusions .....                                                                           | 77  |
| References .....                                                                                | 84  |
| <b>5 Robust Adaptive Beamforming</b>                                                            |     |
| <i>Osamu Hoshuyama, Akihiko Sugiyama</i> .....                                                  | 87  |
| 5.1 Introduction .....                                                                          | 87  |
| 5.2 Adaptive Beamformers .....                                                                  | 88  |
| 5.3 Robustness Problem in the GJBF .....                                                        | 90  |
| 5.4 Robust Adaptive Microphone Arrays — Solutions to Steering-<br>Vector Errors .....           | 92  |
| 5.4.1 LAF-LAF Structure .....                                                                   | 92  |
| 5.4.2 CCAF-LAF Structure .....                                                                  | 94  |
| 5.4.3 CCAF-NCAF Structure .....                                                                 | 95  |
| 5.4.4 CCAF-NCAF Structure with an AMC .....                                                     | 97  |
| 5.5 Software Evaluation of a Robust Adaptive Microphone Array .....                             | 99  |
| 5.5.1 Simulated Anechoic Environment .....                                                      | 99  |
| 5.5.2 Reverberant Environment .....                                                             | 101 |

|       |                                                                   |     |
|-------|-------------------------------------------------------------------|-----|
| 5.6   | Hardware Evaluation of a Robust Adaptive Microphone Array . . . . | 104 |
| 5.6.1 | Implementation . . . . .                                          | 104 |
| 5.6.2 | Evaluation in a Real Environment . . . . .                        | 104 |
| 5.7   | Conclusion . . . . .                                              | 106 |
|       | References . . . . .                                              | 106 |

## 6 GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement

|       |                                                           |     |
|-------|-----------------------------------------------------------|-----|
|       | <i>Simon Doclo, Marc Moonen</i> . . . . .                 | 111 |
| 6.1   | Introduction . . . . .                                    | 111 |
| 6.2   | GSVD-Based Optimal Filtering Technique . . . . .          | 113 |
| 6.2.1 | Optimal Filter Theory . . . . .                           | 114 |
| 6.2.2 | General Class of Estimators . . . . .                     | 116 |
| 6.2.3 | Symmetry Properties for Time-Series Filtering . . . . .   | 117 |
| 6.3   | Performance of GSVD-Based Optimal Filtering . . . . .     | 118 |
| 6.3.1 | Simulation Environment . . . . .                          | 118 |
| 6.3.2 | Spatial Directivity Pattern . . . . .                     | 119 |
| 6.3.3 | Noise Reduction Performance . . . . .                     | 121 |
| 6.3.4 | Robustness Issues . . . . .                               | 121 |
| 6.4   | Complexity Reduction . . . . .                            | 122 |
| 6.4.1 | Linear Algebra Techniques for Computing GSVD . . . . .    | 122 |
| 6.4.2 | Recursive and Approximate GSVD-Updating Algorithms . . .  | 123 |
| 6.4.3 | Downsampling Techniques . . . . .                         | 125 |
| 6.4.4 | Simulations . . . . .                                     | 125 |
| 6.4.5 | Computational Complexity . . . . .                        | 126 |
| 6.5   | Combination with ANC Postprocessing Stage . . . . .       | 127 |
| 6.5.1 | Creation of Speech and Noise References . . . . .         | 127 |
| 6.5.2 | Noise Reduction Performance of ANC Postprocessing Stage . | 128 |
| 6.5.3 | Comparison with Standard Beamforming Techniques . . . . . | 129 |
| 6.6   | Conclusion . . . . .                                      | 129 |
|       | References . . . . .                                      | 130 |

## 7 Explicit Speech Modeling for Microphone Array Speech Acquisition

|       |                                                          |     |
|-------|----------------------------------------------------------|-----|
|       | <i>Michael Brandstein, Scott Griebel</i> . . . . .       | 133 |
| 7.1   | Introduction . . . . .                                   | 133 |
| 7.2   | Model-Based Strategies . . . . .                         | 136 |
| 7.2.1 | Example 1: A Frequency-Domain Model-Based Algorithm . .  | 137 |
| 7.2.2 | Example 2: A Time-Domain Model-Based Algorithm . . . . . | 140 |
| 7.3   | Conclusion . . . . .                                     | 148 |
|       | References . . . . .                                     | 151 |

---

## Part II. Source Localization

---

**8 Robust Localization in Reverberant Rooms**

|                                                                            |     |
|----------------------------------------------------------------------------|-----|
| <i>Joseph H. DiBiase, Harvey F. Silverman, Michael S. Brandstein</i> ..... | 157 |
| 8.1 Introduction .....                                                     | 157 |
| 8.2 Source Localization Strategies .....                                   | 158 |
| 8.2.1 Steered-Beamformer-Based Locators .....                              | 159 |
| 8.2.2 High-Resolution Spectral-Estimation-Based Locators .....             | 160 |
| 8.2.3 TDOA-Based Locators .....                                            | 161 |
| 8.3 A Robust Localization Algorithm .....                                  | 164 |
| 8.3.1 The Impulse Response Model .....                                     | 164 |
| 8.3.2 The GCC and PHAT Weighting Function .....                            | 166 |
| 8.3.3 ML TDOA-Based Source Localization .....                              | 167 |
| 8.3.4 SRP-Based Source Localization .....                                  | 169 |
| 8.3.5 The SRP-PHAT Algorithm .....                                         | 170 |
| 8.4 Experimental Comparison .....                                          | 172 |
| References .....                                                           | 178 |

**9 Multi-Source Localization Strategies**

|                                                                                      |     |
|--------------------------------------------------------------------------------------|-----|
| <i>Elio D. Di Claudio, Raffaele Parisi</i> .....                                     | 181 |
| 9.1 Introduction .....                                                               | 181 |
| 9.2 Background .....                                                                 | 184 |
| 9.2.1 Array Signal Model .....                                                       | 184 |
| 9.2.2 Incoherent Approach .....                                                      | 185 |
| 9.2.3 Coherent Signal Subspace Method (CSSM) .....                                   | 185 |
| 9.2.4 Wideband Weighted Subspace Fitting (WB-WSF) .....                              | 186 |
| 9.3 The Issue of Coherent Multipath in Array Processing .....                        | 187 |
| 9.4 Implementation Issues .....                                                      | 188 |
| 9.5 Linear Prediction-ROOT-MUSIC TDOA Estimation .....                               | 189 |
| 9.5.1 Signal Pre-Whitening .....                                                     | 189 |
| 9.5.2 An Approximate Model for Multiple Sources in Reverberant<br>Environments ..... | 191 |
| 9.5.3 Robust TDOA Estimation via ROOT-MUSIC .....                                    | 192 |
| 9.5.4 Estimation of the Number of Relevant Reflections .....                         | 194 |
| 9.5.5 Source Clustering .....                                                        | 195 |
| 9.5.6 Experimental Results .....                                                     | 196 |
| References .....                                                                     | 198 |

**10 Joint Audio-Video Signal Processing for Object Localization and Tracking**

|                                                                       |     |
|-----------------------------------------------------------------------|-----|
| <i>Norbert Strobel, Sascha Spors, Rudolf Rabenstein</i> .....         | 203 |
| 10.1 Introduction .....                                               | 203 |
| 10.2 Recursive State Estimation .....                                 | 205 |
| 10.2.1 Linear Kalman Filter .....                                     | 206 |
| 10.2.2 Extended Kalman Filter due to a Measurement Nonlinearity ..... | 210 |
| 10.2.3 Decentralized Kalman Filter .....                              | 212 |
| 10.3 Implementation .....                                             | 218 |

|                                       |     |
|---------------------------------------|-----|
| 10.3.1 System description .....       | 218 |
| 10.3.2 Results .....                  | 219 |
| 10.4 Discussion and Conclusions ..... | 221 |
| References .....                      | 222 |

---

### Part III. Applications

---

#### 11 Microphone-Array Hearing Aids

|                                                               |     |
|---------------------------------------------------------------|-----|
| <i>Julie E. Greenberg, Patrick M. Zurek</i> .....             | 229 |
| 11.1 Introduction .....                                       | 229 |
| 11.2 Implications for Design and Evaluation .....             | 230 |
| 11.2.1 Assumptions Regarding Sound Sources .....              | 230 |
| 11.2.2 Implementation Issues .....                            | 231 |
| 11.2.3 Assessing Performance .....                            | 232 |
| 11.3 Hearing Aids with Directional Microphones .....          | 233 |
| 11.4 Fixed-Beamforming Hearing Aids .....                     | 234 |
| 11.5 Adaptive-Beamforming Hearing Aids .....                  | 235 |
| 11.5.1 Generalized Sidelobe Canceler with Modifications ..... | 236 |
| 11.5.2 Scaled Projection Algorithm .....                      | 242 |
| 11.5.3 Direction of Arrival Estimation .....                  | 243 |
| 11.5.4 Other Adaptive Approaches and Devices .....            | 243 |
| 11.6 Physiologically-Motivated Algorithms .....               | 244 |
| 11.7 Beamformers with Binaural Outputs .....                  | 245 |
| 11.8 Discussion .....                                         | 246 |
| References .....                                              | 249 |

#### 12 Small Microphone Arrays with Postfilters for Noise and Acoustic Echo Reduction

|                                                                       |     |
|-----------------------------------------------------------------------|-----|
| <i>Rainer Martin</i> .....                                            | 255 |
| 12.1 Introduction .....                                               | 255 |
| 12.2 Coherence of Speech and Noise .....                              | 257 |
| 12.2.1 The Magnitude Squared Coherence .....                          | 257 |
| 12.2.2 The Reverberation Distance .....                               | 258 |
| 12.2.3 Coherence of Noise and Speech in Reverberant Enclosures ..     | 259 |
| 12.3 Analysis of the Wiener Filter with Symmetric Input Signals ..... | 263 |
| 12.3.1 No Near End Speech .....                                       | 265 |
| 12.3.2 High Signal to Noise Ratio .....                               | 265 |
| 12.4 A Noise Reduction Application .....                              | 266 |
| 12.4.1 An Implementation Based on the NLMS Algorithm .....            | 266 |
| 12.4.2 Processing in the 800 – 3600 Hz Band .....                     | 268 |
| 12.4.3 Processing in the 240 – 800 Hz Band .....                      | 269 |
| 12.4.4 Evaluation .....                                               | 269 |
| 12.4.5 Alternative Implementations of the Coherence Based Postfilter  | 271 |
| 12.5 Combined Noise and Acoustic Echo Reduction .....                 | 271 |

|                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------|-----|
| 12.5.1 Experimental Results .....                                                                     | 274 |
| 12.6 Conclusions .....                                                                                | 275 |
| References .....                                                                                      | 276 |
| <b>13 Acoustic Echo Cancellation for Beamforming</b>                                                  |     |
| <b>Microphone Arrays</b>                                                                              |     |
| <i>Walter L. Kellermann</i> .....                                                                     | 281 |
| 13.1 Introduction .....                                                                               | 281 |
| 13.2 Acoustic Echo Cancellation .....                                                                 | 282 |
| 13.2.1 Adaptation algorithms .....                                                                    | 284 |
| 13.2.2 AEC for multi-channel sound reproduction .....                                                 | 287 |
| 13.2.3 AEC for multi-channel acquisition .....                                                        | 287 |
| 13.3 Beamforming .....                                                                                | 288 |
| 13.3.1 General structure .....                                                                        | 288 |
| 13.3.2 Time-invariant beamforming .....                                                               | 290 |
| 13.3.3 Time-varying beamforming .....                                                                 | 291 |
| 13.3.4 Computational complexity .....                                                                 | 292 |
| 13.4 Generic structures for combining AEC with beamforming .....                                      | 292 |
| 13.4.1 Motivation .....                                                                               | 292 |
| 13.4.2 Basic options .....                                                                            | 293 |
| 13.4.3 'AEC first' .....                                                                              | 293 |
| 13.4.4 'Beamforming first' .....                                                                      | 296 |
| 13.5 Integration of AEC into time-varying beamforming .....                                           | 297 |
| 13.5.1 Cascading time-invariant and time-varying beamforming ...                                      | 297 |
| 13.5.2 AEC with GSC-type beamforming structures .....                                                 | 301 |
| 13.6 Combined AEC and beamforming for multi-channel recording and<br>multi-channel reproduction ..... | 302 |
| 13.7 Conclusions .....                                                                                | 303 |
| References .....                                                                                      | 303 |
| <b>14 Optimal and Adaptive Microphone Arrays for Speech In-<br/>put in Automobiles</b>                |     |
| <i>Sven Nordholm, Ingvar Claesson, Nedelko Grbić</i> .....                                            | 307 |
| 14.1 Introduction: Hands-Free Telephony in Cars .....                                                 | 307 |
| 14.2 Optimum and Adaptive Beamforming .....                                                           | 309 |
| 14.2.1 Common Signal Modeling .....                                                                   | 309 |
| 14.2.2 Constrained Minimum Variance Beamforming and the Gen-<br>eralized Sidelobe Canceler .....      | 310 |
| 14.2.3 <i>In Situ</i> Calibrated Microphone Array (ICMA) .....                                        | 312 |
| 14.2.4 Time-Domain Minimum-Mean-Square-Error Solution .....                                           | 313 |
| 14.2.5 Frequency-Domain Minimum-Mean-Square-Error Solution ..                                         | 314 |
| 14.2.6 Optimal Near-Field Signal-to-Noise plus Interference Beam-<br>former .....                     | 316 |
| 14.3 Subband Implementation of the Microphone Array .....                                             | 317 |
| 14.3.1 Description of LS-Subband Beamforming .....                                                    | 318 |

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| 14.4 Multi-Resolution Time-Frequency Adaptive Beamforming .....     | 319 |
| 14.4.1 Memory Saving and Improvements .....                         | 319 |
| 14.5 Evaluation and Examples .....                                  | 320 |
| 14.5.1 Car Environment .....                                        | 320 |
| 14.5.2 Microphone Configurations .....                              | 321 |
| 14.5.3 Performance Measures .....                                   | 321 |
| 14.5.4 Spectral Performance Measures .....                          | 322 |
| 14.5.5 Evaluation on car data .....                                 | 323 |
| 14.5.6 Evaluation Results .....                                     | 323 |
| 14.6 Summary and Conclusions .....                                  | 324 |
| References .....                                                    | 326 |
| <br><b>15 Speech Recognition with Microphone Arrays</b>             |     |
| <i>Maurizio Omologo, Marco Matassoni, Piergiorgio Svaizer</i> ..... | 331 |
| 15.1 Introduction .....                                             | 331 |
| 15.2 State of the Art .....                                         | 332 |
| 15.2.1 Automatic Speech Recognition .....                           | 332 |
| 15.2.2 Robustness in ASR .....                                      | 336 |
| 15.2.3 Microphone Arrays and Related Processing for ASR .....       | 337 |
| 15.2.4 Distant-Talker Speech Recognition .....                      | 339 |
| 15.3 A Microphone Array-Based ASR System .....                      | 342 |
| 15.3.1 System Description .....                                     | 342 |
| 15.3.2 Speech Corpora and Task .....                                | 345 |
| 15.3.3 Experiments and Results .....                                | 346 |
| 15.4 Discussion and Future Trends .....                             | 348 |
| References .....                                                    | 349 |
| <br><b>16 Blind Separation of Acoustic Signals</b>                  |     |
| <i>Scott C. Douglas</i> .....                                       | 355 |
| 16.1 Introduction .....                                             | 355 |
| 16.1.1 The Cocktail Party Effect .....                              | 355 |
| 16.1.2 Chapter Overview .....                                       | 356 |
| 16.2 Blind Signal Separation of Convolutional Mixtures .....        | 357 |
| 16.2.1 Problem Structure .....                                      | 357 |
| 16.2.2 Goal of Convolutional BSS .....                              | 359 |
| 16.2.3 Relationship to Other Problems .....                         | 360 |
| 16.3 Criteria for Blind Signal Separation .....                     | 362 |
| 16.3.1 Overview of BSS Criteria .....                               | 362 |
| 16.3.2 Density Modeling Criteria .....                              | 362 |
| 16.3.3 Contrast Functions .....                                     | 364 |
| 16.3.4 Correlation-Based Criteria .....                             | 366 |
| 16.4 Structures and Algorithms for Blind Signal Separation .....    | 367 |
| 16.4.1 Filter Structures .....                                      | 367 |
| 16.4.2 Density Matching BSS Using Natural Gradient Adaptation ..    | 368 |
| 16.4.3 Contrast-Based BSS Under Prewhitening Constraints .....      | 370 |

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| 16.4.4 Temporal Decorrelation BSS for Nonstationary Sources ..... | 372 |
| 16.5 Numerical Evaluations .....                                  | 373 |
| 16.6 Conclusions and Open Issues .....                            | 375 |
| References .....                                                  | 378 |

---

## Part IV. Open Problems and Future Directions

---

### 17 Future Directions for Microphone Arrays

|                                                        |     |
|--------------------------------------------------------|-----|
| <i>Gary W. Elko</i> .....                              | 383 |
| 17.1 Introduction .....                                | 383 |
| 17.2 Hands-Free Communication .....                    | 383 |
| 17.3 The "Future" of Microphone Array Processing ..... | 385 |
| 17.4 Conclusions .....                                 | 387 |

### 18 Future Directions in Microphone Array Processing

|                                                                                      |     |
|--------------------------------------------------------------------------------------|-----|
| <i>Dirk Van Compernelle</i> .....                                                    | 389 |
| 18.1 Lessons From the Past .....                                                     | 389 |
| 18.2 A Future Focused on Applications .....                                          | 391 |
| 18.2.1 Automotive .....                                                              | 391 |
| 18.2.2 Desktop .....                                                                 | 392 |
| 18.2.3 Hearing Aids .....                                                            | 393 |
| 18.2.4 Teleconferencing .....                                                        | 393 |
| 18.2.5 Very Large Arrays .....                                                       | 393 |
| 18.2.6 The Signal Subspace Approach - An Alternative to Spatial<br>Filtering ? ..... | 393 |
| 18.3 Final Remarks .....                                                             | 394 |

|             |     |
|-------------|-----|
| Index ..... | 395 |
|-------------|-----|

## List of Contributors

**Joerg Bitzer**

Houpert Digital Audio  
Bremen, Germany

**Michael S. Brandstein**

Harvard University  
Cambridge MA, USA

**Ingvar Claesson**

Blekinge Inst. of Technology  
Ronneby, Sweden

**Joseph H. DiBiase**

Brown University  
Providence RI, USA

**Elio D. Di Claudio**

University of Rome "La Sapienza"  
Rome, Italy

**Simon Doclo**

Katholieke Universiteit Leuven  
Leuven, Belgium

**Scott C. Douglas**

Southern Methodist University  
Dallas TX, USA

**Gary W. Elko**

Agere Systems  
Murray Hill NJ, USA

**Nedelko Grbić**

Blekinge Inst. of Technology  
Ronneby, Sweden

**Julie E. Greenberg**

Massachusetts Inst. of Technology  
Cambridge MA, USA

**Scott M. Griebel**

Harvard University  
Cambridge MA, USA

**Osamu Hoshuyama**

NEC Media Research Labs  
Kawasaki, Japan

**Walter L. Kellermann**

University Erlangen-Nuremberg  
Erlangen, Germany

**Rodney A. Kennedy**

The Australian National University  
Canberra, Australia

**Claude Marro**

France Télécom R&D  
Lannion, France

**Rainer Martin**

Aachen University of Technology  
Aachen, Germany

**Marco Matassoni**

Istituto per la Ricerca Scientifica e  
Tecnologica  
Povo, Italy

**Marc Moonen**

Katholieke Universiteit Leuven  
Leuven, Belgium

XVIII List of Contributors

**Sven Nordholm**  
Curtin University of Technology  
Perth, Australia

**Maurizio Omologo**  
Istituto per la Ricerca Scientifica e  
Tecnologica  
Povo, Italy

**Raffaele Parisi**  
University of Rome "La Sapienza"  
Rome, Italy

**Rudolf Rabenstein**  
University Erlangen-Nuremberg  
Erlangen, Germany

**Harvey F. Silverman**  
Brown University  
Providence RI, USA

**K. Uwe Simmer**  
Aureca GmbH  
Bremen, Germany

**Sascha Spors**  
University Erlangen-Nuremberg  
Erlangen, Germany

**Norbert Strobel**  
Siemens Medical Solutions  
Erlangen, Germany

**Akihiko Sugiyama**  
NEC Media Research Labs  
Kawasaki, Japan

**Piergiorgio Svaizer**  
Istituto per la Ricerca Scientifica e  
Tecnologica  
Povo, Italy

**Dirk Van Compernelle**  
Katholieke Universiteit Leuven  
Leuven, Belgium

**Darren B. Ward**  
Imperial College of Science,  
Technology and Medicine  
London, UK

**Robert C. Williamson**  
The Australian National University  
Canberra, Australia

**Patrick M. Zurek**  
Sensimetrics Corporation  
Somerville MA, USA

Part I

## Speech Enhancement

# 1 Constant Directivity Beamforming

Darren B. Ward<sup>1</sup>, Rodney A. Kennedy<sup>2</sup>, and Robert C. Williamson<sup>2</sup>

<sup>1</sup> Imperial College of Science, Technology and Medicine, London, UK

<sup>2</sup> The Australian National University, Canberra, Australia

**Abstract.** Beamforming, or spatial filtering, is one of the simplest methods for discriminating between different signals based on the physical location of the sources. Because speech is a very wideband signal, covering some four octaves, traditional narrowband beamforming techniques are inappropriate for hands-free speech acquisition. One class of broadband beamformers, called constant directivity beamformers, aim to produce a constant spatial response over a broad frequency range. In this chapter we review such beamformers, and discuss implementation issues related to their use in microphone arrays.

## 1.1 Introduction

Beamforming is one of the simplest and most robust means of *spatial filtering*, i.e., discriminating between signals based on the physical locations of the signal sources [1]. In a typical microphone array environment, the desired speech signal originates from a talker's mouth, and is corrupted by interfering signals such as other talkers and room reverberation. Spatial filtering can be useful in such an environment, since the interfering sources generally originate from points in space separate from the desired talker's mouth. By exploiting the spatial dimension of the problem, microphone arrays attempt to obtain a high-quality speech signal without requiring the talker to speak directly into a close-talking microphone.

In most beamforming applications two assumptions simplify the analysis: (i) the signals incident on the array are narrowband (the *narrowband assumption*); and (ii) the signal sources are located far enough away from the array that the wavefronts impinging on the array can be modeled as plane waves (the *farfield assumption*). For many microphone array applications, the farfield assumption is valid. However, the narrowband assumption is never valid, and it is this aspect of the beamforming problem that we focus on in this chapter (see [2] for techniques that also lift the nearfield assumption).

To understand the inherent problem in using a narrowband array for broadband signals, consider a linear array with a fixed number of elements separated by a fixed inter-element distance. The important dimension in measuring array performance is its size in terms of operating wavelength. Thus for high frequency signals (having a small wavelength) a fixed array will appear large and the main beam will be narrow. However, for low frequencies

## 2 Superdirective Microphone Arrays

Joerg Bitzer<sup>1</sup> and K. Uwe Simmer<sup>2</sup>

<sup>1</sup> Houpert Digital Audio, Bremen, Germany

<sup>2</sup> Aureca GmbH, Bremen, Germany

**Abstract.** This chapter gives an overview of so-called superdirective beamformers, which can be derived by applying the minimum variance distortionless response (MVDR) principle to theoretically well-defined noise fields, as for example the diffuse noise field. We show that all relevant performance measures for beamformer designs are functions of the coherence matrix of the noise field. Additionally, we present unconstrained and constrained MVDR-solutions using modified coherence functions. Solutions for different choices of the optimization criterion are given including a new solution to optimize the front-to-back ratio. Finally, we present a comparison of superdirective beamformers to gradient microphones and an alternative generalized sidelobe canceler (GSC) implementation of the superdirective beamformer.

### 2.1 Introduction

What is “super” about a superdirective microphone array? Compared to the standard delay-and-sum beamformer a superdirective array achieves a higher directivity. Therefore, “super”-directivity indicates that summing is not the optimal choice for combining sensor signals, if optimal directivity is desired. The term directivity describes the ability of a beamformer to suppress noise coming from all directions without affecting a desired signal from one principal direction.

A short historical overview in [6] shows that superdirectivity (or supergain) in connection with array processing was first mentioned in the first half of the last century. The solutions provided at that time were of academic interest only, since a lot of practical problems occurred which restricted the use of the theoretical work. The main reasons for failure were the self-noise and the gain and phase errors of the microphones. In order to overcome these problems a first constrained solution was published by Gilbert and Morgan in 1955 [15]. Early applications with slight modifications were seismic and sonar techniques [5]. It was not until the 90's that supergain was connected to microphone applications. Research in hearing aids highlighted the advantages of fixed beamformers over adaptive solutions [17]. Modern designs of superdirective beamformers include nearfield assumptions and the possibility to adapt the constraining to the actual problem.

This chapter is organized as follows: Section 2.2 introduces the measures to judge the different designs. In section 2.3 the optimal design will be derived

### 3 Post-Filtering Techniques

K. Uwe Simmer<sup>1</sup>, Joerg Bitzer<sup>2</sup>, and Claude Marro<sup>3</sup>

<sup>1</sup> Aureca GmbH, Bremen, Germany

<sup>2</sup> Houpert Digital Audio, Bremen, Germany

<sup>3</sup> France Télécom R&D, Lannion, France

**Abstract.** In the context of microphone arrays, the term post-filtering denotes the post-processing of the array output by a single-channel noise suppression filter. A theoretical analysis shows that Wiener post-filtering of the output of an optimum distortionless beamformer provides a minimum mean squared error solution. We examine published methods for post-filter estimation and develop a new algorithm. A simulation system is presented to compare the performance of the discussed algorithms.

#### 3.1 Introduction

What can be gained by additional post-filtering if the Minimum Variance Distortionless Response (MVDR) beamformer already provides the optimum solution for a given sound field?

Assuming that signal and noise are mutually uncorrelated the MVDR beamformer minimizes the noise power (or variance) subject to the constraint of a distortionless look direction response. The solution can be shown to be optimum in the Maximum Likelihood (ML) sense and produces the best possible Signal to Noise Ratio (SNR) for a narrowband input [1]. However, it does not maximize the SNR for a broadband input such as speech. Furthermore, the MVDR beamformer does not provide a broadband Minimum Mean Squared Error (MMSE) solution. The best possible linear filter in the MMSE sense is the multi-channel Wiener filter. As shown below the broadband multi-channel MMSE solution can be factorized into a MVDR beamformer followed by a single-channel Wiener post-filter. The multi-channel Wiener filter generally produces a higher output SNR than the MVDR filter. Therefore, additional post-filtering can significantly improve the SNR, which motivates this chapter.

The squared error minimized by the single-channel Wiener filter is the sum of residual noise and signal distortion components at the output of the filter. As a result, linear distortion of the desired signal cannot be avoided entirely if Wiener filtering is used. Additional Wiener filtering is advantageous in practice, however, because signal distortions can be masked by residual noise and a compromise between signal distortion and noise suppression can be found. Using MVDR beamforming alone often does not provide sufficient noise reduction due to its limited ability to reduce diffuse noise and reverberation.

## 4 Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields

Gary W. Elko

Media Signal Processing Research, Agere Systems, Murray Hill NJ, USA

**Abstract.** The spatial correlation function between directional microphones is useful in the design and analysis of the performance of these microphones in actual acoustic noise fields. These correlation functions are well known for omnidirectional receivers, but not well known for directional receivers. This chapter investigates the spatial correlation functions for  $N$ th-order differential microphones in both spherically and cylindrically isotropic noise fields. The results are used to calculate the amount of achievable cancellation from an adaptive noise cancellation application using combinations of differential microphones to remove unwanted noise from a desired signal. The results are useful in determining signal-to-noise ratio gains from arbitrarily positioned differential microphone elements in microphone array applications.

### 4.1 Introduction

The spatial correlation function is important in the design of optimal beamformers that maximize the signal-to-noise ratio (SNR), source direction finding algorithms, the calculation of actual SNR gain from arrays, and other array signal processing areas. The space-time correlation functions are well known for omnidirectional receivers in two specific environments: spherically and cylindrically isotropic noise fields. One area of large concern that has been a topic of ongoing work has been the design and performance of directional differential microphone systems. One application of these systems is in adaptive noise cancellation schemes. In order to predict the expected performance gains of these adaptive cancellation systems, the spatial correlation functions between directional microphones are required. Results are presented here for the specific cases of general orientation for first-order differential microphones in both spherically and cylindrically isotropic fields. Specific results are given for the general  $N$ th-order cases for differential arrays that have collinear axes.

### 4.2 Adaptive Noise Cancellation

The use of adaptive noise cancellation in communication devices has been under investigation for more than two decades [1], [2]. The early studies predicted SNR gains on the order of 10 dB and higher. However, it was

# 5 Robust Adaptive Beamforming

Osamu Hoshuyama and Akihiko Sugiyama

NEC Media Research Labs, Kawasaki, Japan

**Abstract.** This chapter presents robust adaptive beamforming techniques designed specifically for microphone array applications. The basics of adaptive beamformers are first reviewed with the Griffiths-Jim beamformer (GJBF). Its robustness problems caused by steering vector errors are then discussed with some conventionally proposed robust beamformers. As better solutions to the conventional robust beamformers, GJBFs with an adaptive blocking matrix are presented in the form of a microphone array. Simulation results and real-time evaluation data show that a new robust adaptive microphone array achieves improved robustness against steering vector errors. Good sound quality of the output signal is also confirmed by a subjective evaluation.

## 5.1 Introduction

Beamforming is a technique which extracts the desired signal contaminated by interference based on directivity, i.e. spatial signal selectivity [1]–[5]. This extraction is performed by processing the signals obtained by multiple sensors such as microphones, antennas, and sonar transducers located at different positions in the space. The principle of beamforming has been known for a long time. Because of the vast amount of necessary signal processing, most research and development effort has been focused on geological investigations and sonar, which can afford a higher cost. With the advent of LSI technology, the required amount of signal processing has become relatively small. As a result, a variety of research projects where acoustic beamforming is applied to consumer-oriented applications, have been carried out [6].

Applications of beamforming include microphone arrays for speech enhancement. The goal of speech enhancement is to remove undesirable signals such as noise and reverberation. Among research areas in the field of speech enhancement are teleconferencing [7]–[8], hands-free telephones [9]–[11], hearing aids [12]–[21], speech recognition [22]–[23], intelligibility improvement [24]–[25], and acoustic measurement [26].

Beamforming can be considered as multidimensional signal processing in space and time. Ideal conditions assumed in most theoretical discussions are not always maintained. The target DOA (direction of arrival), which is assumed to be stable, does change with the movement of the speaker. The sensor gains, which are assumed uniform, exhibit significant distribution. As a result, the performance obtained by beamforming may not be as good as

## 6 GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement

Simon Doclo and Marc Moonen

Katholieke Universiteit Leuven, Leuven, Belgium

**Abstract.** In this chapter a class of multi-microphone signal enhancement techniques is described, which are based on a Generalized Singular Value Decomposition (GSVD) and which amount to a specific optimal filtering problem when the so-called desired response signal cannot be observed. When applying this GSVD-based optimal filtering technique to noise reduction in multi-microphone speech recordings, simulations show that this technique has a better noise reduction performance than standard beamforming techniques for all reverberation times and that it is more robust to deviations from the nominal situation, e.g. encountered in uncalibrated microphone arrays. If the GSVD-based procedure is used to create speech as well as noise references, as in a Generalized Sidelobe Canceler (GSC) structure, and an adaptive noise canceler (ANC) is added, the noise reduction performance can still be improved. Because computational complexity may appear as a bottleneck for this algorithm, recursive GSVD-updating and downsampling techniques are needed to make this signal enhancement technique amenable to real-time implementation.

### 6.1 Introduction

In many speech communication applications, like audio-conferencing and hands-free mobile telephony, the recorded speech signals are corrupted by acoustic background noise and far-end echo signals (see Fig. 6.1). This is mainly due to the fact that the speaker is located at a distance from the recording microphones, allowing the microphones to record the noise and echo sources too. Background noise and far-end echo signals cause a signal degradation which can lead to total unintelligibility of the speech and which decreases the performance of speech coding and speech recognition systems. Therefore efficient acoustic noise and echo reduction algorithms are called for.

Several single-microphone techniques for noise reduction in speech have been proposed that are based on a (generalized) singular value decomposition. While early techniques only dealt with white noise [1][2], more advanced techniques also incorporated colored noise [3][4]. However, since these techniques are single-microphone speech enhancement techniques, they only perform a (signal-adaptive) frequency filtering on the noisy speech signal.

When performing multi-microphone speech enhancement, both the frequency and the spatial characteristics of the speech and the noise sources can be exploited. Recently a multi-microphone noise reduction technique for

## 7 Explicit Speech Modeling for Microphone Array Applications

Michael Brandstein and Scott Griebel

Harvard University, Cambridge MA, USA

**Abstract.** In this chapter we address the limitations of current approaches to using microphone arrays for speech acquisition and advocate the development of multi-channel techniques which employ an explicit model of the speech signal. The goal is to combine the advantages of spatial filtering achieved through beamforming with knowledge of the desired time-series attributes. We offer two examples of algorithms which incorporate this principle. The first is a frequency domain approach illustrating the utility of model-based microphone array processing for improved speech spectral analysis. The second is a time domain procedure which highlights the benefits of speech modeling/spatial filtering fusion for the purpose of improving speech quality.

### 7.1 Introduction

A major thrust of microphone array research has focused on improving the spatial filtering capability of the system in its operating environment. The previous chapters have directly addressed a number of the methods which have proven beneficial in this regard. The goal of this chapter is to present an alternative and complementary strategy which emphasizes the incorporation of explicit speech modeling into the microphone array processing. As will be shown, by combining specific knowledge of the desired time-series attributes with the advantages of multi-channel spatial filtering, this approach can significantly improve the quality of the speech signal and its derived quantities obtained in challenging acoustic environments.

Much of this work is motivated from current research in the field of single-channel speech systems where, due to necessity perhaps, improvements have been obtained by focusing on the underlying content of the signal of interest in addition to its environmental degradations. There is a rich history of work addressing the use of single channel methods for speech enhancement. Summaries of these techniques may be found in texts on the subject, such as [1-3]. While capable of improving perceived quality in restrictive environments (additive noise, no multipath, high to moderate signal-to-noise ratio (SNR), single source), these approaches do not perform well in the face of reverberant distortions, competing sources, and severe noise conditions. In recent years, sophisticated speech models have been applied to the enhancement problem. In addition to utilizing the periodic features of the speech, as in the case of comb filtering, these systems exploit the signal's mixture of harmonic and

## 8 Robust Localization in Reverberant Rooms

Joseph H. DiBiase<sup>1</sup>, Harvey F. Silverman<sup>1</sup>, and Michael S. Brandstein<sup>2</sup>

<sup>1</sup> Brown University, Providence RI, USA

<sup>2</sup> Harvard University, Cambridge MA, USA

**Abstract.** Talker localization with microphone arrays has received significant attention lately as a means for the automated tracking of individuals in an enclosure and as a necessary component of any general purpose speech capture system. Several algorithmic approaches are available for speech source localization with multi-channel data. This chapter summarizes the current field and comments on the general merits and shortcomings of each genre. A new localization method is then presented in detail. By utilizing key features of existing methods, this new algorithm is shown to be significantly more robust to acoustical conditions, particularly reverberation effects, than the traditional localization techniques in use today.

### 8.1 Introduction

The primary goal of a speech localization system is accuracy. In general, estimate precision is dependent upon a number of factors. Major issues include (1) the quantity and quality of microphones employed, (2) microphone placement relative to each other and the speech sources to be analyzed, (3) the ambient noise and reverberation levels, and (4) the number of active sources and their spectral content. The performance of localization techniques generally improves with the number of microphones in the array, particularly when adverse acoustic effects are present. This has spawned the research and construction of large array systems (e.g. 512 elements) [1]. However, when acoustic conditions are favorable and the microphones are positioned judiciously, source localization can be performed adequately using a modest number (e.g. 4 elements) of microphones. Performance is clearly affected by the array geometry. The optimal design of the array based on localization criteria is typically dependent on the room layout, speaking scenarios, and the acoustic conditions [2]. In practice, many of these design considerations are very dependent on the specific application conditions, the hardware available, and non-scientific cost criteria. In an effort to make its applicability as general as possible, this chapter will focus primarily on speech localization effectiveness as a function of the acoustic degradations present, namely background noise and reverberations, rather than attempt to address more specific environmental scenarios.

In addition to high accuracy, these location estimates must be updated frequently in order to be useful in practical tracking and beamforming appli-

## 9 Multi-Source Localization Strategies

Elio D. Di Claudio and Raffaele Parisi

INFOCOM Dept., University of Rome "La Sapienza", Rome, Italy

**Abstract.** Localization of multiple acoustic sources is an important task in many practical applications. However, in most cases adopted models are not fully adequate to describe real scenarios. In particular, in the presence of reverberation, the signal model should explicitly take into account both signals radiated by multiple sources and reflections. In alternative to Generalized Cross-Correlation methods, array processing concepts can be effectively applied to multi-source localization in reverberant environments. In this chapter, main features and limitations of wide-band array processing approaches are briefly reviewed. Furthermore, a new integrated localization and classification system, based on a robust frequency-domain Time Delay Of Arrival (TDOA) estimation followed by a spatial clustering of raw location estimates, is presented. The proposed method efficiently incorporates TDOA and array processing concepts in a unified approach. Results on simulated data are supplied.

### 9.1 Introduction

Many practical applications rely on microphone arrays to localize multiple acoustic sources. Direct examples include automated systems for surveillance, videoconferencing and hands-free speech acquisition [1]. Localization data is also used indirectly for a variety of other tasks: dereverberation of speech, fault prediction and analysis in machinery, cuing and tracking of television cameras, speaker verification, etc.

Ideally, localization methods should exploit knowledge of the particular acoustical transfer functions associated with the sources and sensors. However, over reliance on these functions in the form of excessive calibration to a specific environment may result in limited system flexibility and poor adaptation capabilities to different and varying acoustical conditions. This suggests that localization algorithms should be based on a compact and robust parameterization of the transfer functions.

If reverberation is present, a general signal model should explicitly take into account both the signals radiated by multiple sources and their respective multi-path reflections. The effects of reflective surfaces in closed environments can be studied in a rigorous manner by solving the wave equations with proper boundary and initial conditions. This task is generally too expensive and can be very sensitive to numerical errors and changes in assumptions about the environment. A physically acceptable approximation may be derived from

# 10 Joint Audio-Video Signal Processing for Object Localization and Tracking

Norbert Strobel<sup>1</sup>, Sascha Spors<sup>2</sup>, and Rudolf Rabenstein<sup>2</sup>

<sup>1</sup> Siemens Medical Solutions, Erlangen, Germany

<sup>2</sup> University Erlangen-Nuremberg, Erlangen, Germany

**Abstract.** Applications such as videoconferencing, automatic scene analysis, or security surveillance involving acoustic sources can benefit from object localization within a complex scene. Many single-sensor techniques already exist for this purpose. They are, e.g., based on microphone arrays, video cameras, or range sensors. Since all of these sensors have their specific strengths and weaknesses, it is often advantageous to combine information from various sensor modalities to arrive at more robust position estimates.

This chapter presents a joint audio-video signal processing methodology for object localizing and tracking. The approach is based on a decentralized Kalman filter structure modified such that different sensor measurement models can be incorporated. Such a situation is typical for combined audio-video sensing, since different coordinate systems are usually used for the camera system and the microphone array.

At first, the decentralized estimation algorithm is presented. Then a speaker localization example is discussed. Finally, some estimation results are shown.

## 10.1 Introduction

Microphone arrays are becoming increasingly important for applications related to the recording of speech or music for subsequent processing, transmission or storage. In most of these cases, the goal is to extract information by listening. In object localization and tracking, however, things are different. There, the information conveyed by acoustic waves is not related to the pattern of the signal they carry but to the position of the sound emitting source.

Of course, sound waves are not the only physical quantity that can be used for source localization. Other sensors may, for example, measure the range of the source or detect its retinal image. Since all sensors have their specific strengths and weaknesses, there is no single modality for object localization that always outperforms all others. Therefore, it is desirable to integrate the information of various sensor modalities to exploit different physical phenomena.

A practical example for such a multi-sensor system may comprise a microphone array and a video camera, for example. Either sensor can be used to estimate a (single-sensor) source position. Taken together, they can be viewed as a joint audio-video object localization system. Unfortunately, such

# 11 Microphone-Array Hearing Aids

Julie E. Greenberg<sup>1</sup> and Patrick M. Zurek<sup>2</sup>

<sup>1</sup> Massachusetts Institute of Technology, Cambridge MA, USA

<sup>2</sup> Sensimetrics Corporation, Somerville MA, USA

**Abstract.** Microphone-array hearing aids provide a promising solution to the problems encountered by hearing-impaired persons when listening to speech in the presence of background noise. This chapter first discusses implementation issues and performance metrics specific to the hearing-aid application. A review of previous work on microphone-array hearing aids includes systems with directional microphones, fixed beamformers, adaptive beamformers, physiologically-motivated processing, and binaural outputs. Recent simulation results of one promising adaptive beamforming system are presented. The performance of microphone-array hearing aids depends heavily on the acoustic environments in which they are used. Additional information about the level of reverberation, number of interferers, and relative levels of interferers encountered by hearing-aid users in everyday situations is required to quantify the benefit of microphone-array hearing aids and to select the optimal processing strategy.

## 11.1 Introduction

Hearing impairments affect 10–16% of the total adult population and at least one-third of the elderly population [1–5]. Recent surveys in the United States indicate that roughly 20% of the hearing-impaired population (5–6 million people) use hearing aids, but only 60% of users are satisfied with their devices [6].

Users' overall satisfaction with their hearing aids is highly dependent on the situations in which they are used. While 90% of hearing aid users indicated that they were satisfied with their hearing aid for one-on-one listening, 62% were satisfied in small group settings, 45% in workplaces and restaurants and only 25% in large groups [7]. In an analysis of why hearing aid owners do not use their hearing aids, the second most prevalent reason, given by 25% of respondents, was that the hearing aids "did not work" in background noise [8].

These complaints are not surprising given the greater susceptibility of hearing-impaired listeners to interference from background noise. Normal-hearing listeners typically exhibit speech reception thresholds<sup>1</sup> (SRT's) of

<sup>1</sup> The speech reception threshold is the signal-to-interference ratio at which speech is 50% intelligible.

## 12 Small Microphone Arrays with Postfilters for Noise and Acoustic Echo Reduction

Rainer Martin

Institute of Communication Systems and Data Processing  
Aachen University of Technology, Aachen, Germany

**Abstract.** This chapter presents arrays with two microphones and a postfilter for noise reduction and acoustic echo cancellation. The postfilter algorithm exploits the spatial coherence of the microphone signals. In contrast to single-microphone enhancement algorithms, it does not need an explicit noise power spectral density estimate. An analysis of the mean square error reveals that the coherence properties of the microphone signals are of paramount importance for the performance of the postfilter. Coherence measurements of signals in various acoustic environments are presented. The influence of the directivity and orientation of the microphones on the measured coherence is discussed and rules for the design of the acoustic interface are given. Finally, applications of this approach are presented. The two-microphone algorithm is employed to reduce the non-stationary noise in the voice intercom of a computed tomography scanner. It is also combined with echo cancellers to be used in a robust desktop conferencing device.

### 12.1 Introduction

Hands-free operation of voice communication terminals presents a challenging signal processing task. The relatively large distance between speaker and microphones, the feedback of acoustic echoes, and the "anywhere and anytime" paradigm of mobile communications can all contribute to considerably disturbed speech signals. To achieve a reasonable communication quality the hands-free terminal must therefore reduce disturbing environmental acoustic noise as well as acoustic echoes in received speech signals.

Because of their ease of implementation and use, single-microphone speech enhancement systems are favored in many applications. However, multi-microphone systems have considerable advantages over single-microphone systems when the noise is non-stationary or the speech is reverberated. Multi-microphone systems take the spatial correlation of sound fields into account. The spatial correlation can be exploited to dereverberate the desired speech signal and to reduce noise and acoustic echoes. The simplest system which takes advantage of some of these benefits is the two-microphone array.

In contrast to larger arrays, the two-microphone approach relies less on the beamforming gain of the array but more on the noise and echo suppression of a postfilter. The postfilter combines and processes the two microphone signals in order to compute an estimate of the clean speech signal. The array can be easily implemented using widely available stereo A/D converters.

# 13 Acoustic Echo Cancellation for Beamforming Microphone Arrays

Walter L. Kellermann

University Erlangen – Nürnberg, Germany

**Abstract.** Acoustic feedback from loudspeakers to microphones constitutes a major challenge for digital signal processing in interfaces for natural, full-duplex human-machine speech interaction. Two techniques, each one successful on its own, are combined here to jointly achieve maximum echo cancellation in real environments: For one, acoustic echo cancellation (AEC), which has matured for single-microphone signal acquisition, and, secondly, beamforming microphone arrays, which aim at dereverberation of desired local signals and suppression of local interferers, including acoustic echoes. Structural analysis shows that straightforward combinations of the two techniques either multiply the considerable computational cost of AEC by the number of array microphones or sacrifice algorithmic performance if the beamforming is time-varying. Striving for increased computational efficiency without performance loss, the integration of AEC into time-varying beamforming is examined for two broad classes of beamforming structures. Finally, the combination of AEC and beamforming is discussed for multi-channel recording and multi-channel reproduction schemes.

## 13.1 Introduction

For natural human-machine interaction, acoustic interfaces are desirable that support seamless full-duplex communication without requiring the user to wear or hold special devices. For that, the general scenario of Figure 13.1 foresees several loudspeakers for multi-channel sound reproduction and a microphone array for acquisition of desired signals in the local acoustic environment. Acoustic signal processing is employed to support services such as speech transmission, speech recognition, or sound field synthesis offered by communication networks or autonomous interactive systems. Such hands-free acoustic interfaces may be tailored for incorporation into a wide variety of communication terminals, including teleconferencing equipment, mobile phones and computers, car information systems, and home entertainment equipment.

For signal acquisition, microphone arrays allow spatial filtering of arriving signals and, thus, desired signals can be enhanced and interferers can be suppressed. With full-duplex communication, echoes of the loudspeaker signals will join local interferers to corrupt the desired source signals. Beamforming, however, does not exploit the available loudspeaker signals as reference information for suppressing the acoustic echoes. This is accomplished by acoustic

## 14 Optimal and Adaptive Microphone Arrays for Speech Input in Automobiles

Sven Nordholm<sup>1</sup>, Ingvar Claesson<sup>2</sup>, and Nedelko Grbić<sup>2</sup>

<sup>1</sup> Curtin University of Technology, Perth, Australia

<sup>2</sup> Blekinge Institute of Technology, Ronneby, Sweden

**Abstract.** In this chapter, speech enhancement and echo cancellation for hands-free mobile telephony are discussed. A number of microphone array methods have been tested and results from car measurements are given. Traditional methods such as linearly constrained beamforming and generalized sidelobe cancelers are discussed as well as array gain optimization methods. An *in situ* calibrated method which gives an overall improved performance is also presented. Algorithms such as Least-Squares (LS) and Normalized-Least-Mean-Squares (NLMS) are used to find optimal weights. Improved performance using an LS-method is shown, but at the cost of increased numerical complexity limiting its implementation in real-time applications. By introducing subband processing this issue can be avoided. The results show a noise suppression of approximately 18 dB and hands-free loudspeaker suppression of the same order.

### 14.1 Introduction: Hands-Free Telephony in Cars

The increased use of mobile telephones in cars has created a greater demand for hands-free, in-car installations. The advantages of hands-free telephones are safety and convenience. In many countries and regions hand-held telephony in cars is prohibited by legislation. The car manufacturers also prohibit such use since it will interact with other electronic devices in the car such as air bags, navigation equipment, etc. This means that a mobile telephone should be properly installed and an external antenna should be used. However, by installing the microphone far away from the user, a number of disadvantages, such as poor sound quality and acoustic feedback from the far-end side, are introduced. This means that some form of filtering is required in order to obtain sound quality comparable to that of hand-held telephony. This filtering operation must suppress the loudspeaker, as well as background noise and room reverberation, without causing severe speech distortion. A number of potential methods will be presented to address this problem.

For automobile applications, there has also been the desire to replace many hand-controlled functions with voice controls. The signal degradations in this context have many similarities to those encountered in distant-talker speech recognition applications. A study of recognition in car environments was presented in [1,2]. However this topic is beyond the primary goal here and is the specific subject of Chapter 15.

## 15 Speech Recognition with Microphone Arrays

Maurizio Omologo, Marco Matassoni, and Piergiorgio Svaizer

ITC-IRST (Istituto per la Ricerca Scientifica e Tecnologica), Povo (Trento), Italy

**Abstract.** Microphone arrays can be advantageously employed in Automatic Speech Recognition (ASR) systems to allow distant-talking interaction. Their beamforming capabilities are used to enhance the speech message, while attenuating the undesired contribution of environmental noise and reverberation. In the first part of this chapter the state of the art of ASR systems is briefly reviewed, with a particular concern about robustness in distant-talker applications. The objective is the reduction of the mismatch between the real noisy data and the acoustic models used by the recognizer. Beamforming, speech enhancement, feature compensation, and model adaptation are the techniques adopted to this end. The second part of the chapter is dedicated to the description of a microphone-array based speech recognition system developed at ITC-IRST. It includes a linear array beamformer, an acoustic front-end for speech activity detection and feature extraction, a recognition engine based on Hidden Markov Models and the modules for training and adaptation of the acoustic models. Finally the performance of this system on a typical recognition task is reported.

### 15.1 Introduction

During the last decade research on ASR technology has made significant advances. As a result, high performance systems are now available for situations where there is a good match between testing and training conditions [1,2]. However, these same systems tend to suffer from a limited robustness to variability in their operating environment [3-8].

One of the most attractive potential features of ASR technology is the flexibility afforded through hands-free interaction. Not being encumbered by a hand-held or head-mounted microphone may be of considerable utility to the user. Of particular concern is the distant-talker case where the user is beyond the normal acquisition range of the system microphone<sup>1</sup> (e.g. at more than one meter in the case of an omnidirectional microphone). For ASR applications of moderate/high language complexity this represents a very ambitious task.

The development of distant-talker ASR will allow for the expansion of voice activated technology into a number of areas where it has until now

<sup>1</sup> The distinctions among different types of microphones will not be treated here. However, the choices available and their relative characteristics should be specifically addressed in the design of the application.

## 16 Blind Separation of Acoustic Signals

Scott C. Douglas

Southern Methodist University, Dallas TX, USA

**Abstract.** This chapter presents an overview of criteria and algorithms for the blind separation of linearly mixed acoustic signals. Particular attention is paid to the underlying statistical formulations of various approaches to the convolutive blind signal separation task, and comparisons to other blind inverse problems are made. Several algorithms are described, including a novel algorithm that largely maintains the spectral content of the original mixtures in the extracted source signals. Numerical experiments are also provided to explore the behaviors of the algorithms in real-world blind signal separation tasks.

### 16.1 Introduction

#### 16.1.1 The Cocktail Party Effect

Consider for a moment the last time you had a conversation with a friend at a social gathering. Numerous conversations were occurring around you, and yet you had the capability of focusing your attention on your friend's phrases. This so-called "cocktail party effect" [1] is well-known and taken for granted by almost everyone. What are the system elements at work in this situation?

- *An acoustic sensor array.* In our case, two sensors (our ears) collect acoustic information, although there is nothing precluding the use of more than two sensors in the task from a system design standpoint.
- *A computational processing system.* Most auditory models allow for several layers of signal processing, from lower-level frequency-dependent feature extraction within the inner ear to higher-level pattern recognition within the brain [2].

The current boom in electronic devices, while creating a revolution in human lifestyles, also creates the need for sensing and translating acoustical information into useful electronic forms. Since we use our ears as information sensors in conversations with others, it seems only natural to develop similar listening capabilities in information-processing devices. This quest remains a challenging problem, despite several decades of research in microphone arrays, voice recognition, and natural language processing. Moreover, while speech can be accurately collected using close-to-the-mouth microphones, there are applications in which such placements are impractical or undesirable, such as in audio teleconferencing systems and automobile speakerphones. Clearly,

## 17 Future Directions for Microphone Arrays

Gary W. Elko

Media Signal Processing Research, Agere Systems, Murray Hill NJ, USA

### 17.1 Introduction

The telephone, invented more than 140 years ago, has changed very little with respect to acoustic design in the last 40 years. The most recent technological improvements were the replacement of the carbon microphone with an electret microphone and the development of the speakerphone for hands-free telephony in the early sixties. It might seem quite surprising that in spite of the rapid changes in communication systems: digital switching, digital wireless, packet-based data communications and the like, that very little has changed to the acoustic front-end in telephony systems. For instance, with the general availability of broadband communication networks, why have we not significantly increased the audio communication bandwidth? Why don't we see stereo or multichannel in telephony and communication systems? Why don't we see more ubiquitous adoption of voice recognition in consumer products? Why can't we tailor the audio response of our communication systems the same way we do with our entertainment systems? Why do our cell phones sound so tinny, distorted, and noisy? Why is it so difficult to understand people who call us from noisy environments or call us in noisy environments? These are good questions when one ponders the explosion of inexpensive digital signal processing integrated circuits combined with the current bandwidth of communication systems and those that are about to be introduced. Clearly front-end acoustic signal processing will play a large role in addressing these questions.

In our increasingly mobile society we find that hands-free operation of communication systems is becoming more and more the norm, whether we talk to other humans or to machines. There are significant challenges in enabling high quality speech communication in difficult acoustic environments. Microphone arrays are an enabling technology for hands-free communication since these systems offer directional gain to improve the signal-to-noise ratio. This chapter will address some new possibilities as well as outstanding research issues that remain.

### 17.2 Hands-Free Communication

The hands-free speakerphone was first released by AT&T's Western Electric in 1954. The speakerphone allowed many users gathered in one location to

# 18 Future Directions in Microphone Array Processing

Dirk Van Compernelle

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.

# APPENDIX 1003C

# LIBRARY

# CATALOG



 1 of 1
 

BOOK

## Microphone arrays : signal processing techniques and applications

**Full Record**

[MARC Tags](#)

Main title

Microphone arrays : signal processing techniques and applications / Michael Brandstein, Darren Ward (eds.).

Published/Created

New York : Springer, c2001.

[Request this Item](#)

 [LC Find It](#)

### Links



Links

Publisher description <http://www.loc.gov/catdir/enhancements/fy0826/2001031428-d.html>

Table of contents only <http://www.loc.gov/catdir/enhancements/fy0826/2001031428-t.html>

### More Information



LCCN Permalink

<https://lccn.loc.gov/2001031428>

Description

xviii, 398 p. : ill. ; 25 cm.

|                       |                                                                                          |
|-----------------------|------------------------------------------------------------------------------------------|
| ISBN                  | 3540419535 (alk. paper)                                                                  |
| LC classification     | TK5986 .M53 2001                                                                         |
| Related names         | <a href="#">Brandstein, Michael, 1967-</a><br><a href="#">Ward, Darren, 1970-</a>        |
| LC Subjects           | <a href="#">Microphone.</a>                                                              |
| Browse by shelf order | <a href="#">TK5986</a>                                                                   |
| Notes                 | Includes bibliographical references and index.                                           |
| Series                | Digital signal processing<br><a href="#">Digital signal processing (Springer-Verlag)</a> |
| LCCN                  | 2001031428                                                                               |
| Dewey class no.       | 621.382/84                                                                               |
| Type of material      | Book                                                                                     |

Item Availability

>

|             |                                            |
|-------------|--------------------------------------------|
| CALL NUMBER | <a href="#">TK5986 .M53 2001</a><br>Copy 1 |
| Request in  | Jefferson or Adams Building Reading Rooms  |
| Status      | Not Charged                                |

# LIBRARY

# CATALOG



1 of 1



BOOK

## Microphone arrays : signal processing techniques and applications

[Full Record](#)

**MARC Tags**

|     |                                                                                                                                                                                                  |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 000 | 01325cam a22003014a 4500                                                                                                                                                                         |
| 001 | 12382050                                                                                                                                                                                         |
| 005 | 20080312100222.0                                                                                                                                                                                 |
| 008 | 010417s2001 nyua b 001 0 eng                                                                                                                                                                     |
| 906 | —  a 7  b cbc  c orignew  d 1  e ocip  f 20  g y-gencatlg                                                                                                                                        |
| 925 | 0_  a acquire  b 2 shelf copies  x policy default                                                                                                                                                |
| 955 | —  a to ASCD pc17 04-17-01; jg14 04-18-01; jg12 04-19-01; jg07 04-19-01; to Dewey 04-19-01; aa07 04-19-01  a ps16 2001-07-16 bk rec'd, to CIP ver.  f yf13 2001-09-04  g yf13 2001-09-04 to BCCD |
| 010 | —  a 2001031428                                                                                                                                                                                  |
| 020 | —  a 3540419535 (alk. paper)                                                                                                                                                                     |
| 040 | —  a DLC  c DLC  d DLC                                                                                                                                                                           |
| 042 | —  a pcc                                                                                                                                                                                         |
| 050 | 00  a TK5986  b .M53 2001                                                                                                                                                                        |
| 082 | 00  a 621.382/84  2 21                                                                                                                                                                           |
| 245 | 00  a Microphone arrays :  b signal processing techniques and applications /  c Michael Brandstein, Darren Ward (eds.).                                                                          |
| 260 | —  a New York :  b Springer,  c c2001.                                                                                                                                                           |
| 300 | —  a xviii, 398 p. :  b ill. ;  c 25 cm.                                                                                                                                                         |
| 490 | 1_  a Digital signal processing                                                                                                                                                                  |
| 504 | —  a Includes bibliographical references and index.                                                                                                                                              |
| 650 | _0  a Microphone.                                                                                                                                                                                |

|     |    |                                                                                             |
|-----|----|---------------------------------------------------------------------------------------------|
| 700 | 1_ | a Brandstein, Michael,  d 1967-                                                             |
| 700 | 1_ | a Ward, Darren,  d 1970-                                                                    |
| 830 | _0 | a Digital signal processing (Springer-Verlag)                                               |
| 856 | 42 | 3 Publisher description  u http://www.loc.gov/catdir/enhancements/fy0826/2001031428-d.html  |
| 856 | 41 | 3 Table of contents only  u http://www.loc.gov/catdir/enhancements/fy0826/2001031428-t.html |

Request this Item

 LC Find It

## Item Availability



|             |                                           |
|-------------|-------------------------------------------|
| CALL NUMBER | TK5986 .M53 2001<br>Copy 1                |
| Request in  | Jefferson or Adams Building Reading Rooms |
| Status      | Not Charged                               |

# APPENDIX 1003D



SEARCH

Account Log in

ANN ARBOR

FLINT

Keyword

[Advanced](#)

❗ **Search Tip:** Enter one or more keywords. Use quotes to search for a phrase (e.g., solar power; polar bears; "systems of oppression"). See tips about [Basic Keyword Searching](#).

[Everything](#) [Catalog](#) [Articles](#) [Databases](#) [Online Journals](#) [Guides and More](#)[Catalog](#) / Record

Book



# Microphone arrays : signal processing techniques and applications / Michael Brandstein, Darren Ward (eds.).

## Record info:

### Format

Book

### Contributors

[Brandstein, Michael, 1967-](#)[Ward, Darren, 1970-](#)

### Published/Created

New York : Springer, c2001.

### Series (transcribed)

Digital signal processing

### Language

English

### Physical Description



Ask a Librarian



xviii, 398 p. : ill. ; 25 cm.

### Bibliography

Includes bibliographical references and index.

### ISBN

3540419535

### Call Number

TK 5986 .M531 2001

### OCLC Number

46791090

### Subjects (LCSH)

[Microphone.](#)

### Academic Discipline

[Engineering](#) > [Electrical Engineering](#)

The University of Michigan Library aims to describe library materials in a way that respects the people and communities who create, use, and are represented in our collections. Report harmful or offensive language in catalog records, finding aids, or elsewhere in our collections anonymously through our [metadata feedback form](#). More information at [Remediation of Harmful Language](#).

### Available at:

 Offsite Shelving · 1 item

| Action                   | Description | Status     | Call Number        |
|--------------------------|-------------|------------|--------------------|
| <a href="#">Get This</a> |             | ✓ On shelf | TK 5986 .M531 2001 |

### Actions

Select what to do with this record.

 Email    Text    Citation    Export Citation (EndNote)    Export Citation (RIS)    Copy link

Date Last Indexed: 20231208

### MARC Data

LDR                      00000nam a22003134a 4500

 Ask a Librarian ^

|     |     |                                                                                                                           |
|-----|-----|---------------------------------------------------------------------------------------------------------------------------|
| 001 |     | 990041953010106381                                                                                                        |
| 005 |     | 200112100000000.0                                                                                                         |
| 008 |     | 011210s2001 nyua b 001 0 eng                                                                                              |
| 035 |     | a (MiU)004195301MIU01                                                                                                     |
| 010 |     | a 2001031428                                                                                                              |
| 016 | 7   | a 961270640  2 GyFmDB                                                                                                     |
| 020 |     | a 3540419535  q (alk. paper)                                                                                              |
| 035 |     | a (OCoLC)ocm46791090                                                                                                      |
| 035 |     | a (III)iiio40623737                                                                                                       |
| 040 |     | a DLC  c DLC  d C#P  d OHX  d EYM                                                                                         |
| 042 |     | a pcc                                                                                                                     |
| 050 | 0 0 | a TK5986  b .M53 2001                                                                                                     |
| 072 | 7   | a TK  2 lcco                                                                                                              |
| 082 | 0 0 | a 621.382/84  2 21                                                                                                        |
| 245 | 0 0 | a Microphone arrays :<br> b signal processing techniques and applications /<br> c Michael Brandstein, Darren Ward (eds.). |
| 260 |     | a New York :  b Springer,  c c2001.                                                                                       |
| 300 |     | a xviii, 398 p. :  b ill. ;  c 25 cm.                                                                                     |
| 490 | 0   | a Digital signal processing                                                                                               |
| 504 |     | a Includes bibliographical references and index.                                                                          |
| 650 | 0   | a Microphone.<br> 0 <a href="http://id.loc.gov/authorities/subject">http://id.loc.gov/authorities/subject</a>             |



Ask a Librarian



|       |                                                                                                                                                                                |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 700 1 | a Brandstein, Michael,  d 1967-<br> 0 http://id.loc.gov/authorities/names/n2001008996<br> 0 http://viaf.org/viaf/sourceID/LC n2001008996                                       |
| 700 1 | a Ward, Darren,  d 1970-<br> 0 http://id.loc.gov/authorities/names/n2001007946<br> 0 http://viaf.org/viaf/sourceID/LC n2001007946                                              |
| 998   | c mcp rj                                                                                                                                                                       |
| 958   | a MiU                                                                                                                                                                          |
| 959   | a (notis)ULBBH8249                                                                                                                                                             |
| 995   | a 20                                                                                                                                                                           |
| BIB   | u 2021-07-16 23:51:34 US/Eastern<br> c 2021-06-21 18:34:52 US/Eastern  s false                                                                                                 |
| 852 0 | b OFFS  a MiU  c ELLS  h TK 5986 .M531 2001  8 22977996060006381                                                                                                               |
| 974   | 8 22977996060006381  f 1  c ELLS  m BOOK  a 39015053152867<br> e ELLS  7 23977996050006381  p 01<br> r 2001-08-22 06:59:00 US/Eastern  h TK 5986 .M531 2001  d OFFS<br> b OFFS |

[Give feedback about this page](#) 



Ask a Librarian



---

[Home](#)

[Accessibility](#)

[Tips for Using Library Search](#)

[Get research help](#)

[About Library Search](#)

[Make an I.L.L. Request](#)

©2023 Regents of the University of Michigan. For details and exceptions, see the [Copyright Policy](#).



Ask a Librarian



# APPENDIX 1003E

## Appendix 1003E

### Citations to Brandstein

#### Source Document

Brandstein, M., & Ward, D. (Eds.). (2001). *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Science & Business Media

#### Citations: 2001 – 2002 (in chronological order)

Herbordt, W., & Kellermann, W. (2001, October). Efficient frequency-domain realization of robust generalized, sidelobe cancellers. In *2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No. 01TH8564)* (pp. 377-382). IEEE.

Spors, S., Rabenstein, R., & Strobel, N. (2001, November). A Multi-Sensor Object Localization System. In *VMV* (Vol. 1, pp. 19-26).

Cohen, I., & Berdugo, B. (2002, May). Microphone array post-filtering for non-stationary noise suppression. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. I-901). IEEE.

Sachar, J. M., Silverman, H. F., & Patterson, W. R. (2002, May). Position calibration of large-aperture microphone arrays. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 2, pp. II-1797). IEEE.

Yao, K., Chen, J. C., & Hudson, R. E. (2002, May). Maximum-likelihood acoustic source localization: experimental results. In *2002 IEEE international conference on acoustics, speech, and signal processing* (Vol. 3, pp. III-2949). IEEE.

Balan, R., & Rosca, J. (2002, August). Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002* (pp. 209-213). IEEE.

Varma, K., Ikuma, T., & Beex, A. A. (2002, August). Robust TDE-based DOA estimation for compact audio arrays. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002* (pp. 214-218). IEEE.

Cutler, R., Rui, Y., Gupta, A., Cadiz, J. J., Tashev, I., He, L. W., ... & Silverberg, S. (2002, December). Distributed meetings: A meeting capture and broadcasting system. In *Proceedings of the tenth ACM international conference on Multimedia* (pp. 503-512).

# APPENDIX 1006A

# Near-field Adaptive Beamformer for Robust Speech Recognition

Iain A. McCowan, Darren C. Moore, and S. Sridharan

Speech Research Laboratory, RCSAVT, School of EESE, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia

E-mail: [iain@ieee.org](mailto:iain@ieee.org); [moore@idiap.ch](mailto:moore@idiap.ch); [s.sridharan@qut.edu.au](mailto:s.sridharan@qut.edu.au)

---

McCowan, I. A., Moore, D. C., and Sridharan, S., Near-field Adaptive Beamformer for Robust Speech Recognition, *Digital Signal Processing* **12** (2002) 87–106.

This paper investigates a new microphone array processing technique specifically for the purpose of speech enhancement and recognition. The main objective of the proposed technique is to improve the low frequency directivity of a conventional adaptive beamformer, as low frequency performance is critical in speech processing applications. The proposed technique, termed *near-field adaptive beamforming* (NFAB), is implemented using the standard generalized sidelobe canceler (GSC) system structure, where a near-field superdirective (NFSD) beamformer is used as the fixed upper-path beamformer to improve the low frequency performance. In addition, to minimize signal leakage into the adaptive noise canceling path for near-field sources, a compensation unit is introduced prior to the blocking matrix. The advantage of the technique is verified by comparing the directivity patterns with those of conventional filter-sum, NFSD, and GSC systems. In speech enhancement and recognition experiments, the proposed technique outperforms the standard techniques for a near-field source in adverse noise conditions. © 2002 Elsevier Science (USA)

**Key Words:** microphone array; beamforming; near-field; adaptive; superdirectivity; speech recognition.

---

## 1. INTRODUCTION

Currently, much research is being undertaken to improve the robustness of speech recognition systems in real environments. This paper focuses on the use of a microphone array to enhance the noisy input speech signal prior to recognition. While the use of microphone arrays for speech recognition has been studied for some time by a number of researchers, a persistent problem has been the poor low frequency directivity of conventional beamforming techniques with

practical array dimensions. Low frequency performance is critical for speech processing applications, as significant speech energy is located below 1 kHz.

By explicitly maximizing the array gain, superdirective beamforming techniques are able to achieve greater directivity than conventional techniques with closely spaced sensor arrays [1]. This directivity generally comes at the expense of a controlled reduction in the white noise gain of the array. Recent work has demonstrated the suitability of superdirective beamforming for speech enhancement and recognition tasks [2, 3]. By employing a spherical propagation model in its formulation, rather than assuming a far-field model, *near-field superdirectivity* (NFSD) succeeds in achieving high directivity at low frequencies for near-field speech sources in diffuse noise conditions [4]. In previous work, near-field superdirectivity has been shown to lead to good speech recognition performance in high noise conditions for a near-field speaker [5].

Superdirective techniques are typically formulated assuming a diffuse noise field. While this is a good approximation to many practical noise conditions, further noise reduction would result from a more accurate model of the actual noise conditions during operation. Adaptive array processing techniques continually update their parameters based on the statistics of the measured input noise. The *generalized sidelobe canceler* (GSC) [6] presents a structure that can be used to implement a variety of adaptive beamformers. A block diagram of the basic GSC system is shown in Fig. 1. The GSC separates the adaptive beamformer into two main processing paths—a standard fixed beamformer,  $\mathbf{w}$ , with  $L$  constraints on the desired signal response, and an adaptive path, consisting of a blocking matrix,  $\mathbf{B}$ , and a set of adaptive filters,  $\mathbf{a}$ . As the desired signal has been constrained in the upper path, the lower path filters can be updated using an unconstrained adaptive algorithm, such as the least-mean-square (LMS) algorithm.

While the theory of adaptive techniques promises greater signal enhancement, this is not always the case in real situations. A common problem with the GSC system is leakage of the desired signal through the blocking matrix, resulting in signal degradation at the beamformer output. This is particularly problematic for broadband signals, such as speech, and especially for speech recognition applications where signal distortion is critical.

In this paper we propose a system that is suited to speech enhancement in a practical near-field situation, having both the good low frequency performance of near-field superdirectivity and the adaptability of a GSC system, while taking

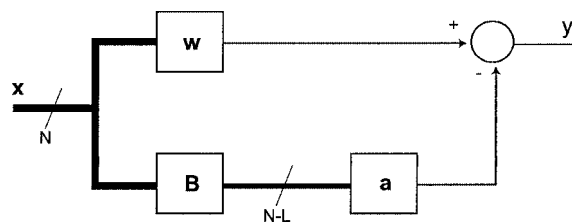


FIG. 1. Generalized sidelobe canceler structure.

care to minimize the problem of signal degradation for near-field sources. We begin by formulating a concise model for near-field sound propagation in Section 2. This model is then used in Section 3 to develop the proposed *near-field adaptive beamforming* (NFAB) technique. To demonstrate the benefit of the technique over existing methods, an experimental evaluation assessing directivity patterns, speech enhancement performance, and speech recognition performance is detailed in Sections 4 and 5.

## 2. NEAR-FIELD SOUND PROPAGATION MODEL

In sensor array applications, a succinct means of characterizing both the array geometry and the location of a signal source is via the *propagation vector*. The propagation vector concisely describes the theoretical propagation of the signal from its source to each sensor in the array. In this section, we develop an expression for the propagation vector of a sound source located in the near-field of a microphone array using a spherical propagation model. This expression is then used in the formulation of the proposed near-field adaptive beamformer in the following sections.

Many microphone array processing techniques assume a planar signal wavefront. This is reasonable for a far-field source, but when the desired source is close to the array a more accurate spherical wavefront model must be employed. For a microphone array of length  $L$ , a source is considered to be in the near-field if  $r < 2L^2/\lambda$ , where  $r$  is the distance to the source and  $\lambda$  is the wavelength.

We define the reference microphone as the origin of a 3-dimensional vector space, as shown in Fig. 2. The position vector for a source in direction  $(\theta_s, \phi_s)$ , at distance  $r_s$  from the reference microphone, is denoted  $\mathbf{p}_s$  and is given by:

$$\mathbf{p}_s = r_s [\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}] \begin{bmatrix} \cos \theta_s \sin \phi_s \\ \sin \theta_s \sin \phi_s \\ \cos \phi_s \end{bmatrix}. \quad (1)$$

The microphone position vectors, denoted as  $\mathbf{p}_i$  ( $i = 1, \dots, N$ ), are similarly defined. The distance from the source to microphone  $i$  is thus

$$d_i = \|\mathbf{p}_s - \mathbf{p}_i\|, \quad (2)$$

where  $\|\cdot\|$  is the Euclidean vector norm.

In such a model, the differences in distance to each sensor can be significant for a near-field source, resulting in phase misalignment across sensors. The difference in propagation time to each microphone with respect to the reference microphone ( $i = 1$ ) is given by

$$\tau_i = \frac{d_i - d_1}{c}, \quad (3)$$

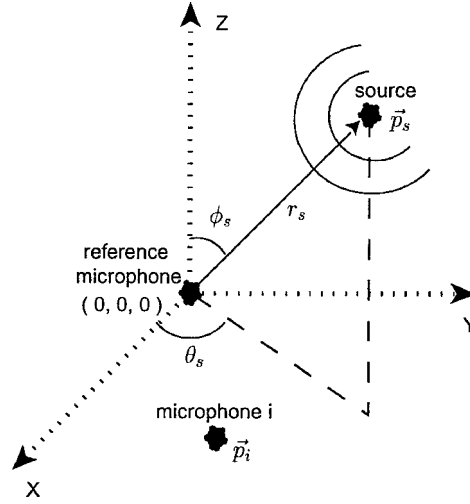


FIG. 2. Near-field propagation model.

where  $c = 340 \text{ ms}^{-1}$  for sound. In addition, the wavefront amplitude decays at a rate proportional to the distance traveled. The resulting amplitude differences across sensors are negligible for far-field sources, but can be significant in the near-field case. The microphone attenuation factors, with respect to the amplitude on the reference microphone, are given by

$$\alpha_i = \frac{d_1}{d_i}. \quad (4)$$

Thus, if  $x_1(f)$  is the desired source at the reference microphone, the signal on the  $i$ th microphone is given by

$$x_i(f) = \alpha_i x_1(f) e^{-j2\pi f \tau_i}. \quad (5)$$

Consequently, we define the near-field propagation vector for a source at distance  $r$  and direction  $(\theta, \phi)$  as

$$\mathbf{d}(f, r, \theta, \phi) = [\alpha_1 e^{-j2\pi f \tau_1} \dots \alpha_i e^{-j2\pi f \tau_i} \dots \alpha_N e^{-j2\pi f \tau_N}]^T. \quad (6)$$

### 3. NEAR-FIELD ADAPTIVE BEAMFORMING

The proposed system structure is shown in Fig. 3. The objective of the proposed technique is to add the benefit of good low frequency directivity to a standard adaptive beamformer, as low frequency performance is critical in speech processing applications. The upper path consists of a fixed near-field superdirective beamformer, while the lower path contains a near-field compensation unit, a blocking matrix and an adaptive noise canceling filter. The principal components of the system are discussed in the following sections.

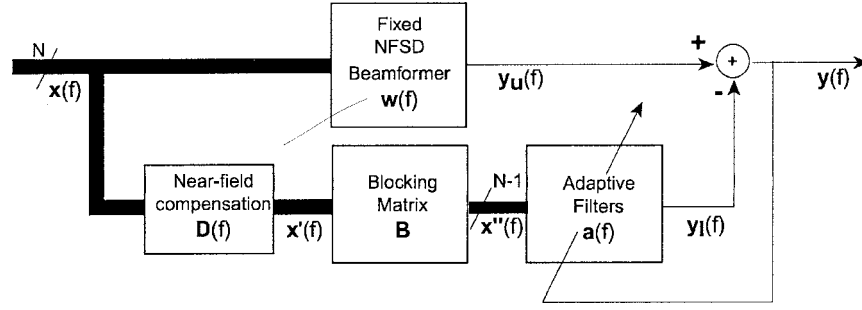


FIG. 3. Near-field adaptive beamformer.

Section 3.1 gives an explanation of the near-field superdirective beamformer. Section 3.2 proposes the inclusion of a near-field compensation unit in the adaptive sidelobe canceling path and examines its effect on reducing signal distortion at the output. Once this near-field compensation has been performed, a standard generalized sidelobe canceling blocking matrix and adaptive filters can be applied to reduce the output noise power, as discussed in Section 3.3.

### 3.1. Near-field Superdirective Beamformer

Superdirective beamforming techniques are based upon the maximization of the array gain, or directivity index. The array gain is defined as the ratio of output signal-to-noise ratio to input signal-to-noise ratio and for the general case can be expressed in matrix notation as [1]

$$G(f) = \frac{\mathbf{w}(f)^H \mathbf{P}(f) \mathbf{w}(f)}{\mathbf{w}(f)^H \mathbf{Q}(f) \mathbf{w}(f)}, \quad (7)$$

where  $\mathbf{w}(f)$  is a column vector of channel gains,

$$\mathbf{w}(f) = [w_1(f) \dots w_i(f) \dots w_N(f)]^T, \quad (8)$$

$(\ )^H$  is the complex conjugate transpose operator, and  $\mathbf{P}(f)$  and  $\mathbf{Q}(f)$  are the cross-spectral density matrices of the signal and noise respectively. In practical speech processing applications the form of the signal and noise cross-spectral density matrices is generally unknown and must be estimated, either from mathematical models (fixed beamformers) or from the statistics of the multichannel inputs (adaptive beamformers). Superdirective beamformers are calculated based on assumed mathematical models for the  $\mathbf{P}(f)$  and  $\mathbf{Q}(f)$  matrices.

When the desired signal is known to emanate from a single source at location  $(r_s, \theta_s, \phi_s)$ , the signal cross-spectral matrix  $\mathbf{P}$  simplifies to the propagation vector of the source, and the array gain can be expressed as

$$G(f) = \frac{|\mathbf{w}(f)^H \mathbf{d}(f, r_s, \theta_s, \phi_s)|^2}{\mathbf{w}(f)^H \mathbf{Q}(f) \mathbf{w}(f)}, \quad (9)$$

where  $\mathbf{d}(f, r, \theta, \phi)$  is the propagation vector for the desired source, as defined in Eq. (6).

A diffuse (spherically isotropic) noise field is often a good approximation for many practical situations, particularly in reverberant closed spaces, such as in a car or an office [7, 8]. For diffuse noise, the noise cross-spectral density matrix  $\mathbf{Q}$  can be formulated as

$$\mathbf{Q}(f) = \frac{1}{4\pi} \int_{\phi} \int_{\theta} \mathbf{d}(f, \theta, \phi) \mathbf{d}(f, \theta, \phi)^H \sin \theta d\theta d\phi, \quad (10)$$

where  $\mathbf{d}(f, \theta, \phi)$  is the propagation vector of a far-field noise source ( $r \gg 2L^2/\lambda$ ) in direction  $(\theta, \phi)$ .

The superdirectivity problem is thus formulated as:

$$\max_{\mathbf{w}(f)} \frac{|\mathbf{w}(f)^H \mathbf{d}(f, r_s, \theta_s, \phi_s)|^2}{\mathbf{w}(f)^H \mathbf{Q}(f) \mathbf{w}(f)}. \quad (11)$$

By using a spherical propagation model to formulate the propagation vector,  $\mathbf{d}$ , the standard superdirective formulation can be optimized for a near-field source [9, 4]. As such, the only difference in the calculation of the standard and near-field superdirective channel filters is the form of the propagation vector,  $\mathbf{d}$ . For a near-field source, the assumption of plane wave (far-field) propagation leads to errors in the array response to the desired signal due to curvature of the direct wavefront. A thorough discussion of the use of a near-field model for superdirective microphone arrays is given by Ryan and Goubran [9].

Cox [10] gives the general superdirective filter solution subject to

1.  $L$  linear constraints,  $\mathbf{C}(f)^H \mathbf{w}(f) = \mathbf{g}(f)$  (explained below); and
2. a constraint on the maximum white noise gain,  $\mathbf{w}(f)^H \mathbf{w}(f) = \delta^{-2}$ , where  $\delta^2$  is the desired white noise gain.

as

$$\mathbf{w}(f) = [\mathbf{Q}(f) + \epsilon \mathbf{I}]^{-1} \mathbf{C}(f) \{ \mathbf{C}(f)^H [\mathbf{Q}(f) + \epsilon \mathbf{I}]^{-1} \mathbf{C}(f) \}^{-1} \mathbf{g}(f), \quad (12)$$

where  $\epsilon$  is a Lagrange multiplier that is iteratively adjusted to satisfy the white noise gain constraint. The white noise gain is the array gain for spatially white (incoherent) noise; that is,  $\mathbf{Q}(f) = \mathbf{I}$ . A constraint on the white noise gain is necessary as an unconstrained superdirective solution will in fact result in significant gain to any incoherent noise, particularly at low frequencies. Cox [10] states that the technique of adding a small amount to each diagonal matrix element prior to inversion is in fact the optimum means of solving this problem. A study of the relationship between the multiplier  $\epsilon$  and the desired white noise gain  $\delta^2$ , shows that the white noise gain increases monotonically with increasing  $\epsilon$ . One possible means of obtaining the desired value of  $\epsilon$  is thus an iterative technique employing a binary search algorithm between a specified minimum and maximum value for  $\epsilon$ . The computational expense of the iterative procedure is not critical, as the beamformer filters depend only on the source

location and array geometry, and thus must only be calculated once for a given configuration.

The constraint matrix,  $\mathbf{C}^H(f)$ , is of order  $L \times N$ , where there are  $L$  linear constraints being applied, and the vector  $\mathbf{g}(f)$  is a length- $L$  column vector of constraining values. The constraints generally include one specifying unity response for the desired signal,  $\mathbf{d}^H(f)\mathbf{w}(f) = 1$ , and where this is the sole constraint the above solution can be simplified by substituting  $\mathbf{C}(f) = \mathbf{d}(f)$  and  $\mathbf{g}(f) = 1$ , giving

$$\mathbf{w}(f) = \frac{[\mathbf{Q}(f) + \epsilon\mathbf{I}]^{-1}\mathbf{d}(f)}{\mathbf{d}(f)^H[\mathbf{Q}(f) + \epsilon\mathbf{I}]^{-1}\mathbf{d}(f)}. \quad (13)$$

Once the optimal filters  $\mathbf{w}(f)$  have been calculated, the near-field superdirective beamformer output is calculated as

$$y_u(f) = \mathbf{w}(f)^H \mathbf{x}(f), \quad (14)$$

where  $\mathbf{x}(f)$  is the  $N$ -channel input column vector

$$\mathbf{x}(f) = [x_1(f) \ \dots \ x_i(f) \ \dots \ x_N(f)]^T. \quad (15)$$

### 3.2. Near-field Compensation Unit

The first element in the adaptive path of standard GSC is the blocking matrix [6]. Its purpose is to block the desired signal from the adaptive noise estimate. To ensure complete blocking, the desired signal must both be time aligned and have equal amplitudes across all channels. If this is the case, cancellation occurs if each row of the blocking matrix sums to zero, and all rows are linearly independent.

For a near-field desired source, to align the desired signal on all channels, a near-field compensation must first be applied to the input channels prior to blocking. To ensure full cancellation we need to compensate for both phase misalignment and amplitude scaling of the desired signal across sensors. We define the diagonal matrix

$$\mathbf{D}(f) = [\text{diag}(\mathbf{d}(f))]^{-1}, \quad (16)$$

where  $\mathbf{d}(f)$  is the near-field propagation vector from Eq. (6). In this paper we define the diagonal operator,  $\text{diag}()$ , to produce a diagonal matrix from a vector parameter. Conversely, if invoked with a matrix parameter, it produces a row vector corresponding to the matrix diagonal. The near-field compensation can be applied as

$$\mathbf{x}'(f) = \mathbf{D}(f)\mathbf{x}(f). \quad (17)$$

Once this near-field compensation has been performed, a standard GSC blocking matrix can be employed to block the desired signal from the adaptive path.

The inclusion of this compensation unit is critical for a near-field desired signal. Without compensation for both phase and amplitude differences between sensors, blocking of the desired signal will not be ensured, leading to signal

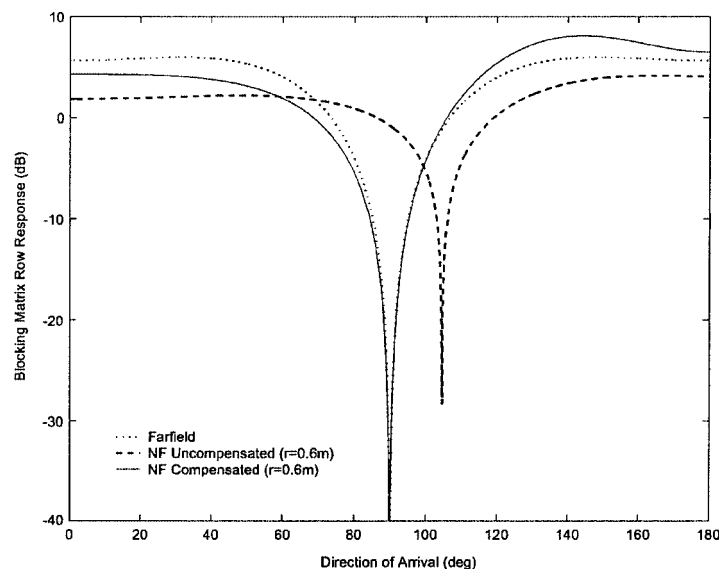


FIG. 4. Comparison of blocking matrix row beam-patterns.

cancellation at the output. The near-field compensation effectively ensures that a true null exists in the beam-pattern of each blocking matrix row in the direction *and* distance corresponding to the desired source. To illustrate, Fig. 4 shows the directivity pattern at 2 kHz for the first row in the blocking matrix using the array shown in Fig. 5, with the desired source directly in front of the center microphone at a distance of 0.6 m. The figure shows the compensated response in the far- and near-fields, as well as the uncompensated near-field response. It is clear that the uncompensated system will allow a high degree of signal leakage into the adaptive path as it blocks noise sources rather than the desired signal.

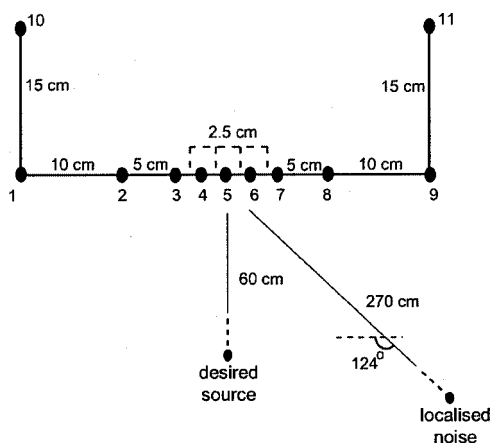


FIG. 5. Experimental configuration.

### 3.3. Blocking Matrix and Adaptive Noise Canceling Filter

The blocking matrix and adaptive noise canceling filters are taken from the standard GSC technique [6]. The order of the blocking matrix is  $N \times (N - L)$ , where there are  $L$  constraints applied in the fixed upper path beamformer. Generally only a unity constraint on the desired signal is specified, and the standard  $N \times (N - 1)$  Griffiths–Jim blocking matrix is used:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \vdots & \vdots & \vdots \\ 0 & -1 & \ddots & 0 & 0 \\ 0 & 0 & \ddots & 1 & 0 \\ \vdots & \vdots & \vdots & -1 & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}. \quad (18)$$

The output of the blocking matrix is calculated as

$$\mathbf{x}''(f) = \mathbf{B}^H \mathbf{x}'(f), \quad (19)$$

where  $\mathbf{x}''(f)$  is an  $(N - 1)$ -length column vector. Defining the  $(N - 1)$ -length adaptive filter column vector as

$$\mathbf{a}(f) = [a_1(f) \ \cdots \ a_i(f) \ \cdots \ a_{N-1}(f)]^T, \quad (20)$$

the output of the lower path is given as

$$y_l(f) = \mathbf{a}(f)^H \mathbf{x}''(f). \quad (21)$$

The NFAB output is then calculated from the upper and lower path outputs as

$$y(f) = y_u(f) - y_l(f) \quad (22)$$

and the adaptive filters are updated using the standard unconstrained LMS algorithm

$$\mathbf{a}_{k+1}(f) = \mathbf{a}_k(f) + \mu \mathbf{x}_k''(f) y_k(f), \quad (23)$$

where  $\mu$  is the adaptation step size and  $k$  denotes the current frame.

### 3.4. Summary of Technique

In summary, the proposed NFAB technique is characterized by the series of equations

$$y_u(f) = \mathbf{w}(f)^H \mathbf{x}(f) \quad (24a)$$

$$\mathbf{x}_k'' = \mathbf{B}^H \mathbf{D}(f) \mathbf{x}(f) \quad (24b)$$

$$y_l(f) = \mathbf{a}(f)^H \mathbf{x}_k''(f) \quad (24c)$$

$$y(f) = y_u(f) - y_l(f) \quad (24d)$$

$$\mathbf{a}_{k+1}(f) = \mathbf{a}_k(f) + \mu \mathbf{x}_k''(f) y_k(f), \quad (24e)$$

where all terms have been defined in the preceding discussion.

#### 4. EXPERIMENTAL CONFIGURATION

For the experimental evaluation in this paper, we used the 11 element array shown in Fig. 5. The array consists of a nine element broadside array, with an additional two microphones situated directly behind the end microphones. The total array is 40 cm wide and 15 cm deep in the horizontal plane. The broadside microphones are arranged according to a standard broadband subarray design, where different subarrays are used for different frequency ranges for the fixed upper path beamformer. The two endfire microphones are included for use by the near-field superdirective beamformer in the low frequency range. The four subarrays are thus

- ( $f < 1$  kHz): microphones 1–11;
- ( $1 \text{ kHz} < f < 2 \text{ kHz}$ ): microphones 1, 2, 5, 8, and 9;
- ( $2 \text{ kHz} < f < 4 \text{ kHz}$ ): microphones 2, 3, 5, 7, and 8; and
- ( $4 \text{ kHz} < f < 8 \text{ kHz}$ ): microphones 3–7.

The array was situated in a computer room, with different sound source locations, as shown in Fig. 5. The two sound sources were

1. the desired speaker situated 60 cm from the center microphone, directly in front of the array; and
2. a localized noise source at an angle of  $124^\circ$  and a distance of 270 cm from the array.

Impulse responses of the acoustic path between each source and microphone were measured from multichannel recordings made in the room with the array using the maximum length sequence technique detailed in Rife and Vanderkooy [11]. As the impulse responses were calculated from real recordings made simultaneously across all input channels, they take into account the real acoustic properties of the room and the array. The multichannel desired speech and localized noise microphone inputs were then generated by convolving the original single-channel speech and noise signals with these impulse responses. In addition, a real multichannel background noise recording of normal operating conditions was made in the room with other workers present. This recording is referred to in the experiments as the ambient noise signal and is approximately diffuse in nature. It consists mainly of computer noise, a variable level of background speech, and noise from an air-conditioning unit. The ambient noise effectively represents a diffuse noise field, while the localized noise represents a coherent noise source. In this paper, we specify the levels of the two different noise sources independently, as the signal to ambient-noise ratio (SANR) and

signal to localized-noise ratio (SLNR). These values are calculated as the average segmental SNR from the speech and noise input, as measured at the center microphone of the array.

In this way, realistic multichannel input signals can be simulated for specified levels of ambient and localized noise. As well as facilitating the generation of different noise conditions, simulating the multichannel inputs using the impulse response method is more practical than making real recordings for speech recognition experiments, as existing single channel speech corpora may be used.

## 5. EXPERIMENTAL RESULTS

This section presents the results of the experimental evaluation. The proposed NFAB technique is compared to a conventional fixed filter-sum beamformer, a fixed near-field superdirective beamformer, and a conventional GSC adaptive beamformer. These beamformers are specified in Table 1.

The techniques are first assessed in terms of the directivity pattern in order to demonstrate the advantage of the proposed NFAB over conventional beamforming techniques, particularly at low frequencies. Following this, the techniques are evaluated for speech enhancement in terms of the improvement in signal to noise ratio and the log area ratio. Finally, the techniques are compared in a hands-free speech recognition task in noisy conditions using the TIDIGITS database [12].

### 5.1. Directivity Analysis

As has been stated, the main objective of the proposed technique is to produce an *adaptive beamformer* that exhibits good *low frequency* performance for *near-field* speech sources. To assess the effectiveness of the proposed technique in achieving this objective, in this section we analyze the horizontal directivity pattern. The directivity of a filter-sum beamformer is expressed in matrix notation as

$$h(f, r, \theta, \phi) = \mathbf{w}_o(f)^H \mathbf{d}(f, r, \theta, \phi), \quad (25)$$

where  $\mathbf{w}_o$  is the length  $N$  channel filter vector

$$\mathbf{w}_o(f) = [w_{o,1}(f) \ \dots \ w_{o,i}(f) \ \dots \ w_{o,N}(f)]^T. \quad (26)$$

**TABLE 1**  
Beamforming Techniques in Evaluation

| Technique | Description                                    | Filters                                                                                   |
|-----------|------------------------------------------------|-------------------------------------------------------------------------------------------|
| FS        | Conventional FS beamformer                     | $\mathbf{w}_o(f) = [\text{diag}(\mathbf{D}(f))]^H$                                        |
| NFSD      | Near-field superdirective beamformer           | $\mathbf{w}_o(f) = \mathbf{w}(f)$                                                         |
| GSC       | GSC system with FS fixed upper path beamformer | $\mathbf{w}_o(f) = [\text{diag}(\mathbf{D}(f))]^H - \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$ |
| NFAB      | Near-field adaptive beamformer                 | $\mathbf{w}_o(f) = \mathbf{w}(f) - \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$                  |

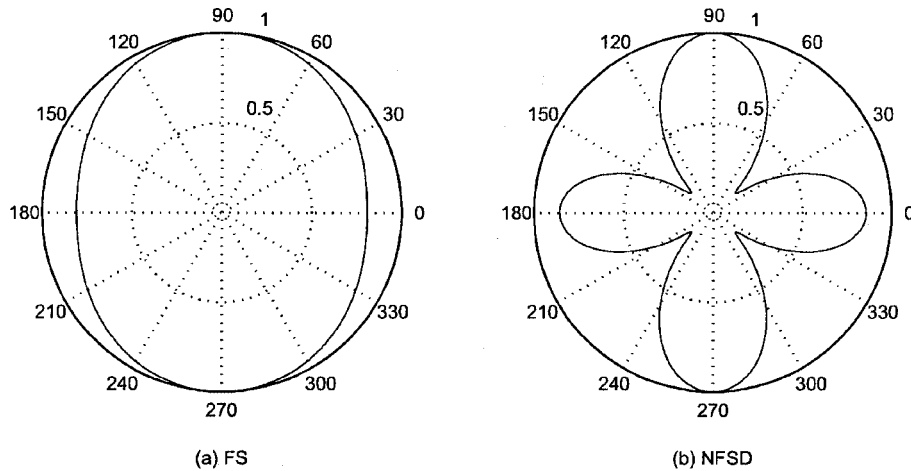


FIG. 6. Upper path directivity pattern at 300 Hz.

**5.1.1. Upper path directivity.** First, we seek to demonstrate the directivity improvement that NFSD achieves at low frequencies compared to a conventional filter-sum (FS) beamformer. For the FS beamformer, a common solution is to choose  $\mathbf{w}_o(f) = [\text{diag}(\mathbf{D}(\mathbf{f}))]^H$ . This effectively ensures that the desired signal is aligned for phase and amplitude across sensors using a spherical propagation model. For NFSD, we use the filter vector  $\mathbf{w}(f)$  described in Section 3.1. Figure 6 shows the near-field directivity pattern at 300 Hz for the FS and NFSD. From these figures, it is clear that the NFSD technique results in greater directional discrimination at low frequencies compared to a conventional beamformer. At higher frequencies ( $f > 1$  kHz), conventional beamformers offer reasonable directivity, and so the FS and NFSD techniques give comparable performance.

**5.1.2. Lower path directivity.** Second, we wish to demonstrate the effect of the noise canceling path. The directivity of the noise canceling filters can be obtained by using the channel filters  $\mathbf{w}_o(f) = \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$ . The blocking matrix and adaptive filters essentially implement a conventional (nonsuperdirective) beamformer that adaptively focuses on the major sources of noise. To examine the directivity of the lower path filters, the beamformer was run on an input speech signal with a white localized noise source (at the location shown in Fig. 5) added at an SLNR of 0 dB and a low level of ambient noise (SANR = 20 dB). The steady-state adaptive filter vector,  $\mathbf{a}(f)$ , was written to file for both the proposed NFAB technique and the conventional GSC beamformer. The near-field directivity patterns of the lower path filters are plotted in Figs. 7 and 8 for 300 and 5000 Hz, respectively. We see that the lower path adaptive filters for both beamformers converge to similar solutions in terms of directivity, producing a main lobe in the direction of the coherent noise source ( $\approx 124^\circ$  from Fig. 5), as well as a null in the location of the desired speaker. As expected, the directivity of the adaptive path is poor at low frequencies, as seen in Fig. 7.

**5.1.3. Overall beamformer directivity.** Finally, we examine the directivity pattern of the overall beamformer for the NFAB and conventional adaptive

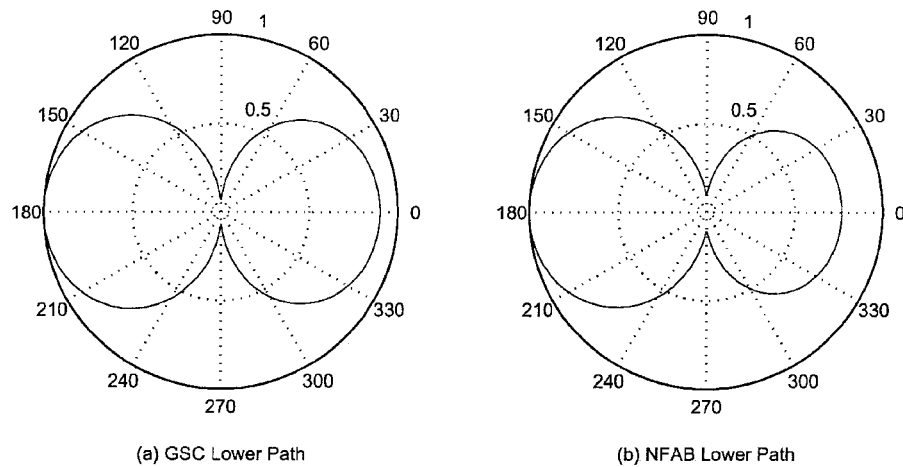


**FIG. 7.** Lower path directivity pattern at 300 Hz.

systems. The near-field directivity patterns at 300 Hz are shown in Fig. 9. We see that the directivity pattern of the NFAB system exhibits a true null in the direction and at the distance of the noise source, while the directivity of the conventional beamformer is too poor to significantly attenuate the noise at this frequency. At frequencies above 1 kHz the directivity performance of both techniques is comparable.

*5.1.4. Summary of beamformer directivity.* In summary we see that, in terms of directivity, the proposed NFAB system:

- outperforms the conventional FS system in terms of low frequency performance and the ability to attenuate coherent noise sources,
- outperforms the NFSD system due to the ability to attenuate coherent noise sources, and

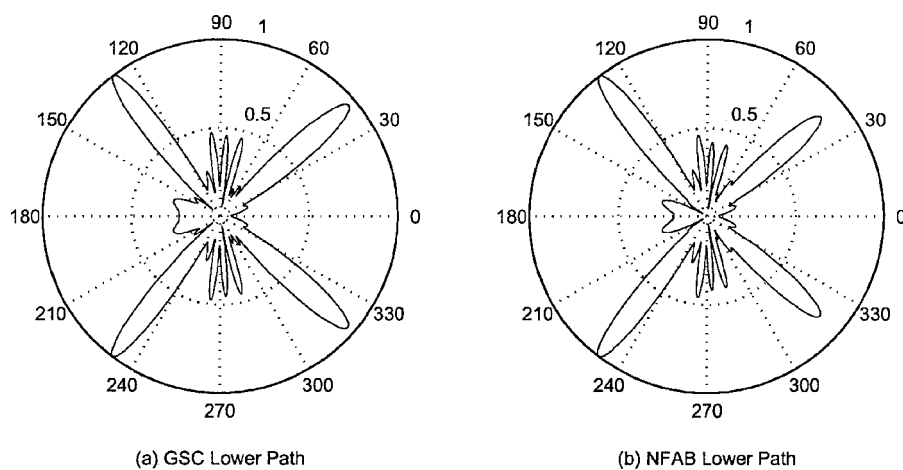


**FIG. 8.** Lower path directivity pattern at 5000 Hz.

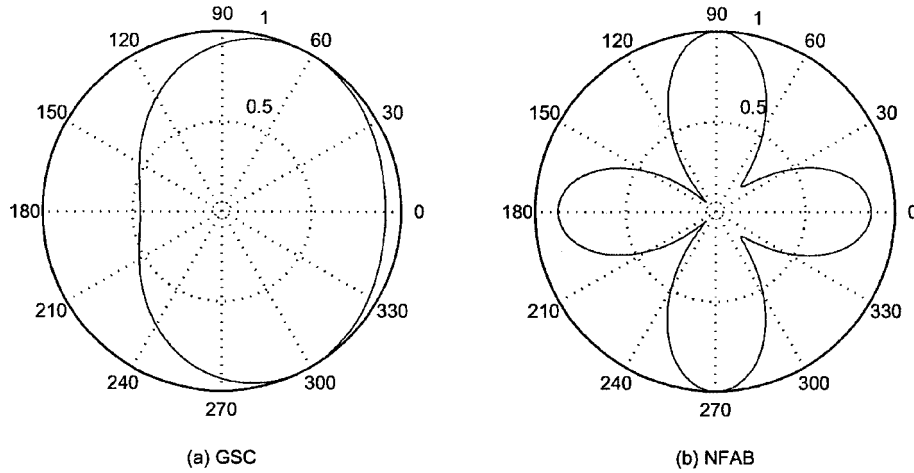


**FIG. 9.** Overall beamformer directivity pattern at 300 Hz.

- outperforms the conventional GSC system in terms of low frequency performance.

In this way, we see that the proposed system succeeds in meeting the stated objectives and should therefore demonstrate improved performance in speech processing applications.

### 5.2. Speech Enhancement Analysis

The signal plots in Fig. 10 give an indication of the level of enhancement achieved by the NFAB technique. For the desired speech signal, we used a segment of speech from the TIDIGITS database corresponding to the digit sequence *one-nine-eight-six*. Ambient noise was added at an SANR level of 10 dB, and a localized white noise signal was added at an SLNR level of 0 dB. The plots indicate that NFAB succeeds in reducing the noise level with negligible distortion to the desired signal.

To better measure the level of enhancement, objective speech measures were used to compare the different techniques. Two measures were used, these being the SNR improvement and the log area ratio distortion measure. The *SNR improvement* is defined as the difference in SNR at the array output and input. As the true SNR cannot be measured, it is estimated as the average segmental signal-plus-noise to noise ratio. While the signal to noise ratio is a useful measure for assessing noise reduction, it does not necessarily give a good indication of how much distortion has been introduced to the desired speech signal. The *log area ratio* (LAR) measure of speech quality is more highly correlated with perceptual intelligibility in humans [13]. The log area ratio measure for a frame of speech is calculated as

$$\text{LAR}(n) = \left| \frac{1}{P} \sum_{i=1}^P \left[ \log \frac{1+r_o(i)}{1-r_o(i)} - \log \frac{1+r_p(i)}{1-r_p(i)} \right] \right|^{1/2}, \quad (27)$$

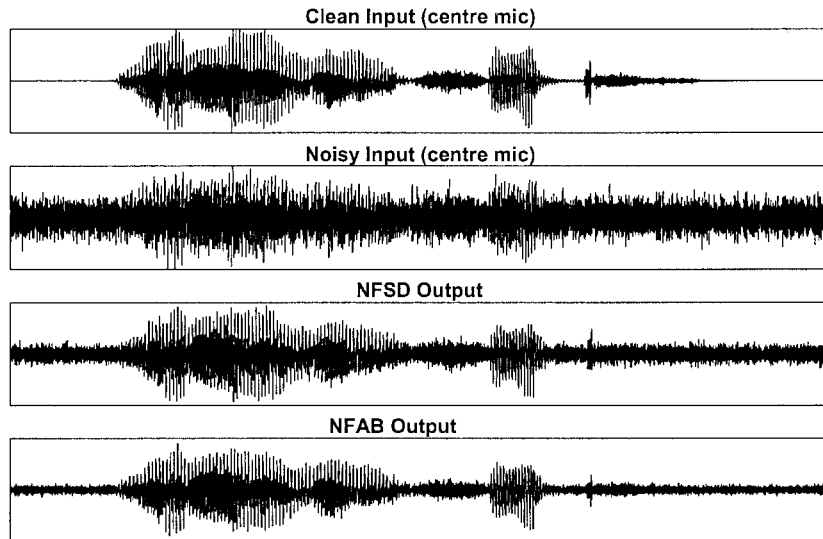


FIG. 10. Sample enhanced signal.

where  $n$  is the frame number, and  $r_o$  and  $r_p$  are the original and processed  $P$ th-order linear predictive coefficients of the  $n$ th frame, respectively. The overall log area ratio distortion measure for the signal is calculated as the average distortion over all input frames.

A set of experiments was conducted in which the localized white noise was replaced with a localized speech-like noise source taken from the NOISEX database [14]. This is essentially a white noise signal that has been shaped with a speech-like spectral envelope and thus represents a more realistic noise scenario than white noise. The signal to localized noise ratio (SLNR) was varied from 20 to 0 dB, with the ambient noise present at a constant SANR level of 10 dB. The output signal to noise ratio improvement and log area ratios are given in Tables 2 and 3 for the different enhancement techniques.<sup>1</sup> The measures have been averaged over 10 randomly chosen speech segments taken

**TABLE 2**  
Signal to Noise Ratio Improvement (SANR = 10 dB)

| Technique | SLNR (dB) |     |     |     |     |
|-----------|-----------|-----|-----|-----|-----|
|           | 0         | 5   | 10  | 15  | 20  |
| FS        | 0.5       | 0.4 | 0.3 | 0.1 | 0.1 |
| NFSD      | 1.4       | 1.6 | 1.1 | 0.5 | 0.2 |
| GSC       | 1.6       | 1.8 | 1.9 | 2.5 | 3.3 |
| NFAB      | 5.5       | 5.9 | 6.4 | 7.5 | 7.9 |

<sup>1</sup> Sample sound files are also available at <http://www.speech.qut.edu.au/pages/people/mccowani>.

**TABLE 3**  
Log Area Ratio: (SANR = 10 dB)

| Technique   | SLNR (dB) |     |     |     |     |
|-------------|-----------|-----|-----|-----|-----|
|             | 0         | 5   | 10  | 15  | 20  |
| Noisy input | 3.6       | 3.3 | 2.9 | 2.6 | 2.5 |
| FS          | 3.1       | 2.7 | 2.5 | 2.4 | 2.3 |
| NFSD        | 2.8       | 2.3 | 2.1 | 2.0 | 1.9 |
| GSC         | 2.6       | 2.1 | 1.8 | 1.7 | 1.6 |
| NFAB        | 2.5       | 1.9 | 1.6 | 1.6 | 1.4 |

from different speakers in the TIDIGITS database. These results are plotted in Fig. 11.

The SNR results show that the proposed NFAB gives considerably greater noise reduction compared to the FS, NFSD, and GSC techniques, providing approximately 6–8 dB of SNR improvement compared to the noisy input signal. Even with a relatively low level of localized noise (high SLNR), the NFAB technique offers significantly greater noise reduction than these other methods. In addition, the proposed technique gives less distortion than the other techniques, as measured by the LAR. As would be expected, the fixed NFSD technique gives slightly less distortion than the adaptive GSC technique.

From these results we see that, in a high level of diffuse and coherent noise with a near-field desired speech source, the proposed NFAB technique succeeds in significantly reducing the noise level, while minimizing the distortion to the speech signal.

### 5.3. Speech Recognition Analysis

The same noise scenario was used for experiments in robust speech recognition. The training and test data for the experiments were taken from the male adult portion of the TIDIGITS database. Tied-state triphone hidden Markov models and standard MFCC parameterization with energy, delta, and acceler-

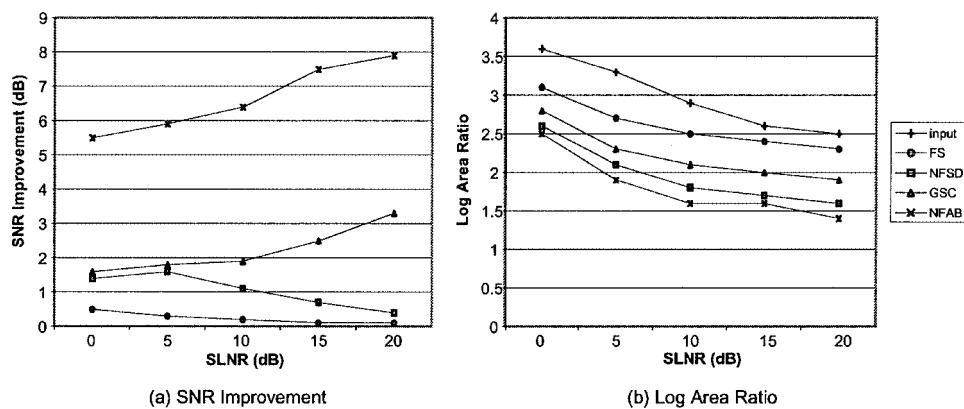


**FIG. 11.** Speech enhancement measures: (a) SNR improvement and (b) LAR.

**TABLE 4**  
Speech Recognition Results: Word Recognition Rates

| Technique   | SLNR (dB) |      |      |      |
|-------------|-----------|------|------|------|
|             | 10        | 5    | 0    | -5   |
| Noisy input | 86.8      | 65.9 | 23.2 | 13.1 |
| FS          | 89.2      | 81.7 | 62.9 | 36.4 |
| NFSD        | 97.7      | 93.2 | 77.2 | 45.4 |
| GSC         | 88.8      | 83.8 | 73.8 | 56.8 |
| NFAB        | 98.2      | 96.7 | 91.1 | 76.7 |

ation coefficients were used. The models were trained with the clean input to the center microphone and then refined using MAP adaptation to better match the noisy environment. The noise segments used in the adaptation process were taken from a separate recording made in the room. The recognition results are given as percentage word recognition rates in Table 4 and shown graphically in Fig. 12.

The results clearly show that NFAB gives excellent robustness to adverse noise conditions in a near-field speech recognition application. The results at low noise levels show that the baseline recognition system is already quite robust to noise, due to the use of MAP adaptation. At more realistic noise levels, however, unenhanced performance is clearly unsatisfactory. For example, at an SLNR of 0 dB and SANR of 10 dB, the word *error* rate for the unprocessed input is 76.8%. While standard GSC and NFSD are able to reduce this to 26.2 and 22.8%, respectively, the proposed NFAB technique succeeds in reducing the error rate to 8.9%. As would be expected, the figure shows that NFAB offers

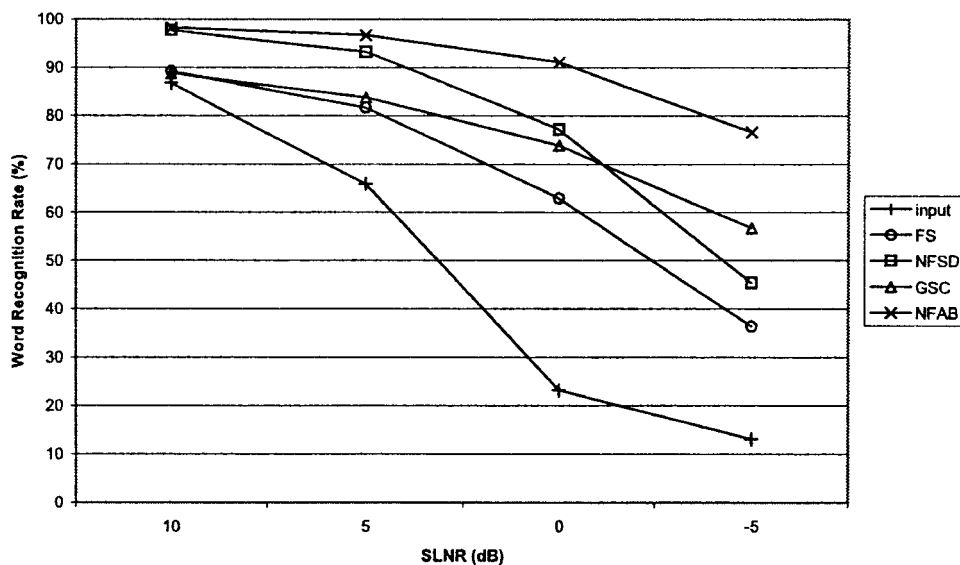


**FIG. 12.** Speech recognition results: Word recognition rates.

similar performance to NFSD when the noise is approximately diffuse (high SLNR) and demonstrates improved ability to attenuate any coherent noise sources (low SLNR) due to the GSC-style adaptive noise canceling path. The recognition performance of NFAB is seen to be similar to NFSD at low levels of coherent noise (high SLNR) and degrades at a rate comparable to GSC with increasing levels of coherent noise.

It is apparent from these results that NFAB is an enhancement technique that is well suited to speech recognition. The experimental results for both speech enhancement and recognition demonstrate that for an adaptive beamformer to be applicable in speech processing applications, it should exhibit good directivity at low frequencies and take care to minimize any signal degradation.

## 6. CONCLUSIONS

A new microphone array processing technique designed specifically for near-field speech processing applications has been proposed, termed near-field adaptive beamforming (NFAB). The technique incorporates a fixed near-field superdirective beamformer into a GSC-style adaptive beamforming structure and as such exhibits the benefits of good low frequency performance and the ability to adaptively attenuate coherent noise signals. Distortion due to the adaptive noise canceling path is minimized by the introduction of a near-field compensation unit.

Two major problems with common conventional microphone array techniques are their poor low frequency performance and the introduction of signal distortion in adaptive techniques. By taking care to address both these issues, the proposed NFAB technique succeeds in significantly outperforming conventional beamforming techniques in terms of objective speech quality measures and speech recognition results in both diffuse and coherent noise.

Speech enhancement results indicate that NFAB succeeds in significantly reducing the output noise power, while also minimizing the distortion to the desired signal. These characteristics make it ideal as an enhancement technique for robust speech recognition. In a high noise configuration, with a signal to localized noise ratio of 0 dB, and a signal to ambient noise ratio of 10 dB, the proposed technique succeeds in increasing the recognition rate from 23.2 to 91.1%. For the same configuration, near-field superdirectivity and conventional GSC only achieve 77.2 and 73.8%, respectively.

In summary, near-field adaptive beamforming has been shown to be a speech enhancement technique that produces a high quality, highly intelligible signal for applications requiring hands-free speech acquisition where the desired speaker is in the array's near-field.

## ACKNOWLEDGMENTS

The near-field superdirective technique was developed at France Telecom R&D [4]. The initial application of these techniques in a speech recognition application was researched by the author during a six month research period at France Telecom R&D in 1999 [5], under the supervision of Yannick Mahieux. The authors specifically wish to acknowledge Claude Marro for his helpful and insightful review of this paper.

## REFERENCES

1. Cox, H., Zeskind, R., and Kooij, T., Practical supergain. *IEEE Trans. Acoustics, Speech Signal Process.* **34** (1986), 393–398.
2. Bitzer, J., Simmer, K. U., and Kammeyer, K., Multi-microphone noise reduction techniques for hands-free speech recognition—a comparative study. In *Robust Methods for Speech Recognition in Adverse Conditions (ROBUST-99)*, Tampere, Finland, May 1999, pp. 171–174.
3. Doerbecker, M., Speech enhancement using small microphone arrays with optimized directivity. In *Proc. Int. Workshop on Acoustic Echo and Noise Control*, September 1997, pp. 100–103.
4. Täger, W., Near field superdirectivity (NFSD). In *Proceedings of ICASSP'98*, 1998, pp. 2045–2048.
5. McCowan, I., Marro, C., and Mauuary, L., Robust speech recognition using near-field superdirective beamforming with post-filtering. In *Proceedings of ICASSP'2000*, 2000, Vol. 3, pp. 1723–1726.
6. Griffiths, L. and Jim, C., An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagation* **30** (1982), 27–34.
7. Meyer, J. and Simmer, K. U., Multi-channel speech enhancement in a car environment using wiener filtering and spectral subtraction. In *Proceedings of ICASSP'97*, 1997, Vol. 2, pp. 1167–1170.
8. Bitzer, J., Simmer, K. U., and Kammeyer, K., Theoretical noise reduction limits of the generalized sidelobe canceller (gac) for speech enhancement. In *Proceedings of ICASSP'99*, 1999, Vol. 5, pp. 2965–2968.
9. Ryan, J. G. and Goubran, R. A., Near-field beamforming for microphone arrays. In *Proceedings of ICASSP'97*, 1997, pp. 363–366.
10. Cox, H., Zeskind, R., and Owen, M., Robust adaptive beamforming. *IEEE Trans. Acoustics, Speech Signal Process.* **35** (1987), 1365–1376.
11. Rife, D. and Vanderkooy, J., Transfer-function measurement with maximum-length sequences. *J. Audio Eng. Soc.* **37** (1989), 419–444.
12. Texas Instruments and NIST. Studio quality speaker-independent connected-digit corpus (TIDIGITS). CD-ROM, February 1991. NIST Speech Discs 4-1, 4-2, and 4-3.
13. Quackenbush, S. R., Barnwell, T. P., and Clements, M. A., *Objective Measures of Speech Quality*. Prentice-Hall, Englewood Cliffs, NJ, 1988.
14. Defence Research Agency Speech Research Unit. NOISEX-92. CD-ROM, June 1992.

---

IAIN A. MCCOWAN received the B.Eng (Hons) and B.InfoTech from the Queensland University of Technology, Brisbane, in 1996. In February 1998 he joined the Research Concentration in Speech, Audio, and Video Technology at the Queensland University of Technology where he is currently completing his Ph.D. His main research interests are in the fields of robust speech recognition and speech enhancement using microphone arrays. He is a student member of the Institute of Electrical and Electronic Engineers.

DARREN C. MOORE received the B.Eng (Hons) and B.InfoTech from the Queensland University of Technology, Brisbane, in 1997. In February 1998 he joined the Research Concentration in Speech, Audio, and Video Technology at the Queensland University of Technology, where he is currently completing a M.Eng. His professional interests lie in the field of speech enhancement using microphone arrays and in the implementation of real-time DSP solutions.

S. SRIDHARAN obtained his B.Sc (electrical engineering) and M.Sc (communication engineering) from the University of Manchester Institute of Science and Technology, United Kingdom and Ph.D (signal processing) from the University of New South Wales, Australia. Dr. Sridharan is senior member of the IEEE, USA and a corporate member of IEE, United Kingdom and IEAust of Australia. He is currently a professor in the School of Electrical and Electronic Systems Engineering of the Queensland University of Technology (QUT) and is also the Head of the Research Concentration in Speech, Audio, and Video Technology at QUT.

# APPENDIX 1006B

# LIBRARY

# CATALOG

PERIODICAL OR NEWSPAPER

## Digital signal processing

Full Record

MARC Tags

Main title

Digital signal processing.

Published/Created

Duluth, MN : Academic Press, c1991-

Request this Item

 LC Find It

### More Information



LCCN Permalink <https://lcn.loc.gov/91650983>

Publication history Vol. 1, no. 1 (Jan. 1991)-

Description v. : ill. ; 28 cm.

Current frequency Bimonthly, <2005->

Former frequency Quarterly, <-2003>

ISSN 1051-2004

Linking ISSN 1051-2004

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LC classification       | TK5102.5 .D4463                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Serial key title        | Digital signal processing (Print)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Abbreviated title       | Digit. signal process. (Print)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| LC Subjects             | <a href="#">Signal processing--Digital techniques--Periodicals.</a><br><a href="#">Signal processing--Digital techniques.</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Other Subjects          | Signal processing--Digital techniques.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Form/Genre              | <a href="#">Periodicals.</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Browse by shelf order   | <a href="#">TK5102.5</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Notes                   | Title from cover.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Reproduction no./Source | Academic Press, Inc., 1 E. 1st St., Duluth, MN 55802                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Additional formats      | Digital signal processing (Online) 1095-4333 (DLC)sn 97006904<br>(OCoLC)36980269                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| LCCN                    | 91650983                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Invalid LCCN            | sn 90002038                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| CODEN                   | DSPREJ                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Dewey class no.         | 621.382/2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Other system no.        | (OCoLC)o cm21889588                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Additional Links        | <a href="http://www.idealibrary.com">http://www.idealibrary.com</a> Logon procedure and access to this title is available via the I.D.E.A.L. (service provider) homepage <a href="http://firstsearch.oclc.org">http://firstsearch.oclc.org</a> Address for accessing the journal using authorization number and password through OCLC FirstSearch Electronic Collections Online. Subscription to online journal required for access to abstracts and full text <a href="http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP">http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP</a> Address for accessing the journal from an authorized IP address through OCLC FirstSearch Electronic Collections Online. Subscription to online journal required for access to abstracts and full text <a href="http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP">http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP</a> |
| Type of material        | Periodical or Newspaper                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

Source of description

Latest issue consulted: Vol. 1, no. 2 (Apr. 1991).

---

**Item Availability**

CALL NUMBER

[TK5102.5 .D4463](#)

Set 1

Request in

Jefferson or Adams Building Reading Rooms

Status

c.1 v. 23, no. 5 2013 Sept In Process 03-31-2015

c.1 v. 72-74 2018 Jan-Mar In Process 07-30-2019

c.1 v. 72-74 2018 Jan-Mar At Bindery - 04-18-2019

c.1 v. 115 2021 Aug In Process 11-14-2023

c.1 v. 121-122 2022 Mar-Apr In Process 11-14-2023

c.1 v. 123 2022 Apr In Process 11-14-2023

c.1 v. 124-125 2022 May-June In Process 11-14-2023

Older receipts

v.1-v.2:no.2 (1991-1992:Apr.), v.3-v.12:no.1 (1993-2002:Jan.), v.13:no.1

(2003:Jan.), v.13:no.4 (2003:Oct.), v.15:no.5 (2005:Sept.), v.16:no.6-

v.17:no.1 (2006:Nov.-2007:Jan.), v.19:no.6 (2009:Dec.), v.20:no.2

(2010:Mar.), v.20:no.6-v.23:no.1 (2010:Dec.-2013:Jan.), v.23:no.3-v.58

(2013:May-2016:Nov.), v.60-v.80 (2018:Sept.), v.82 (2018:Nov.)

v.84-v.101 (2019-2020:June), v.104-v.110 (2020:Sept.-2021:Mar.),

v.112-v.115 (2021:May-2021:Aug.), v.121-v.125 (2022:Mar.-2022:June)

---

CALL NUMBER[TK5102.5 .D4463](#)

Copy 1

Unbound issues

Request in

Newspaper &amp; Current Periodical Reading Room (Madison LM133)

# LIBRARY

# CATALOG



1 of 1



PERIODICAL OR NEWSPAPER

## Digital signal processing

[Full Record](#)

**MARC Tags**

|     |                                                                                                                                                                    |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 000 | 02696cas a2200541 a 4500                                                                                                                                           |
| 001 | 11390503                                                                                                                                                           |
| 005 | 20220221095916.0                                                                                                                                                   |
| 008 | 900620c19919999mnubr p 0 a0eng c                                                                                                                                   |
| 010 | ___  a 91650983  z sn 90002038                                                                                                                                     |
| 022 | 0_  a 1051-2004  l 1051-2004  2 1                                                                                                                                  |
| 030 | ___  a DSPREJ                                                                                                                                                      |
| 035 | ___  a (OCoLC)ocm21889588                                                                                                                                          |
| 037 | ___  b Academic Press, Inc., 1 E. 1st St., Duluth, MN 55802                                                                                                        |
| 040 | ___  a NSD  b eng  c NSD  d DLC  d NST  d CAS  d NST  d NSD  d EYM  d NSD  d OCL  d L<br>YU  d NSD  d OCLCQ  d DLC  d CIT  d U9S  d OCLCQ  d OCLCF  d IUL  d OCLCO |
| 042 | ___  a nsdp  a pcc                                                                                                                                                 |
| 050 | 00  a TK5102.5  b .D4463                                                                                                                                           |
| 082 | 00  a 621.382/2  2 20                                                                                                                                              |
| 210 | 0_  a Digit. signal process.  b (Print)                                                                                                                            |
| 222 | _0  a Digital signal processing  b (Print)                                                                                                                         |
| 245 | 00  a Digital signal processing.                                                                                                                                   |
| 260 | ___  a Duluth, MN :  b Academic Press,  c c1991-                                                                                                                   |
| 300 | ___  a v. :  b ill. ;  c 28 cm.                                                                                                                                    |
| 310 | ___  a Bimonthly,  b <2005->                                                                                                                                       |
| 321 | ___  a Quarterly,  b <-2003>                                                                                                                                       |
| 362 | 0_  a Vol. 1, no. 1 (Jan. 1991)-                                                                                                                                   |
| 500 | ___  a Title from cover.                                                                                                                                           |

|     |    |                                                                                                                                                                                                                                                                                                                                                                           |
|-----|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 588 | —  | a Latest issue consulted: Vol. 1, no. 2 (Apr. 1991).                                                                                                                                                                                                                                                                                                                      |
| 590 | —  | a SERBIB/SERLOC merged record                                                                                                                                                                                                                                                                                                                                             |
| 650 | _0 | a Signal processing  x Digital techniques  v Periodicals.                                                                                                                                                                                                                                                                                                                 |
| 650 | _0 | a Signal processing  x Digital techniques.                                                                                                                                                                                                                                                                                                                                |
| 650 | _6 | a Traitement du signal  x Techniques numériques.                                                                                                                                                                                                                                                                                                                          |
| 650 | _6 | a Traitement du signal  x Techniques numériques  v Périodiques.                                                                                                                                                                                                                                                                                                           |
| 650 | _7 | a Signal processing  x Digital techniques.  2 fast  0 (OCoLC)fst01118285                                                                                                                                                                                                                                                                                                  |
| 655 | _7 | a Periodicals.  2 fast  0 (OCoLC)fst01411641                                                                                                                                                                                                                                                                                                                              |
| 776 | 08 | t Digital signal processing (Online)  x 1095-4333  w (DLC)sn<br>97006904  w (OCoLC)36980269                                                                                                                                                                                                                                                                               |
| 850 | —  | a CPT  a CU-I  a CU-SB  a DLC  a InLP  a MdBj  a MiU  a NN  a NRU-Mus  a NcD                                                                                                                                                                                                                                                                                              |
| 859 | 7_ | u <a href="http://www.idealibrary.com">http://www.idealibrary.com</a>  z Logon procedure and access to this title is available via<br>the I.D.E.A.L. (service provider) homepage  2 http                                                                                                                                                                                  |
| 859 | 7_ | u <a href="http://firstsearch.oclc.org">http://firstsearch.oclc.org</a>  z Address for accessing the journal using authorization<br>number and password through OCLC FirstSearch Electronic Collections Online.<br>Subscription to online journal required for access to abstracts and full text  2 http                                                                  |
| 859 | 7_ | u <a href="http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP">http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP</a>  z Address for<br>accessing the journal from an authorized IP address through OCLC FirstSearch Electronic<br>Collections Online. Subscription to online journal required for access to abstracts and full<br>text  2 http |
| 853 | 20 | 8 1  a v.  b no.  u 4  v r  i (year)  j (month)  w q  x 01                                                                                                                                                                                                                                                                                                                |
| 863 | 41 | 8 1.1  a <11>  b <1>  i <2001>  j <01>                                                                                                                                                                                                                                                                                                                                    |
| 890 | —  | a Digital signal processing.  i 91-650983                                                                                                                                                                                                                                                                                                                                 |
| 906 | —  | a 7  b cbc  c serials  d 3  e ncip  f 19  g n-oclcserc                                                                                                                                                                                                                                                                                                                    |
| 920 | —  | a Keep 1                                                                                                                                                                                                                                                                                                                                                                  |
| 984 | —  | a srvf  d 2001-01-09                                                                                                                                                                                                                                                                                                                                                      |
| 985 | —  | e eserial 200406                                                                                                                                                                                                                                                                                                                                                          |
| 991 | —  | b c-GenColl  h TK5102.5  i .D4463  w SERIALS                                                                                                                                                                                                                                                                                                                              |
| 992 | —  | b SER  w SERLOC                                                                                                                                                                                                                                                                                                                                                           |

[Request this Item](#) [LC Find It](#)

---

## Item Availability



CALL NUMBER

TK5102.5 .D4463

## Set 1

|                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Request in     | Jefferson or Adams Building Reading Rooms                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Status         | c.1 v. 23, no. 5 2013 Sept In Process 03-31-2015<br>c.1 v. 72-74 2018 Jan-Mar In Process 07-30-2019<br>c.1 v. 72-74 2018 Jan-Mar At Bindery - 04-18-2019<br>c.1 v. 115 2021 Aug In Process 11-14-2023<br>c.1 v. 121-122 2022 Mar-Apr In Process 11-14-2023<br>c.1 v. 123 2022 Apr In Process 11-14-2023<br>c.1 v. 124-125 2022 May-June In Process 11-14-2023                                                                                                                               |
| Older receipts | v.1-v.2:no.2 (1991-1992:Apr.), v.3-v.12:no.1 (1993-2002:Jan.), v.13:no.1 (2003:Jan.), v.13:no.4 (2003:Oct.), v.15:no.5 (2005:Sept.), v.16:no.6-v.17:no.1 (2006:Nov.-2007:Jan.), v.19:no.6 (2009:Dec.), v.20:no.2 (2010:Mar.), v.20:no.6-v.23:no.1 (2010:Dec.-2013:Jan.), v.23:no.3-v.58 (2013:May-2016:Nov.), v.60-v.80 (2018:Sept.), v.82 (2018:Nov.) v.84-v.101 (2019-2020:June), v.104-v.110 (2020:Sept.-2021:Mar.), v.112-v.115 (2021:May-2021:Aug.), v.121-v.125 (2022:Mar.-2022:June) |

---

CALL NUMBER [TK5102.5 .D4463](#)  
Copy 1  
Unbound issues

Request in Newspaper & Current Periodical Reading Room (Madison LM133)

# APPENDIX 1006C

Volume 12, Number 1  
**January 2002**

c.  
**Digital  
Signal  
Processing**

*A Review Journal*

**IDEAL<sup>®</sup> First**  
Articles published online first  
<http://www.idealibrary.com>

Editors

**Jim Schroeder**

**Joe Campbell**

ISSN 1051-2004



**ACADEMIC  
PRESS**

Microfilm and microfiche editions also available

1015  
219 of 250

# Digital Signal Processing

---

*A Review Journal*

## Editors

**Jim Schroeder**

*SPRI/CSSIP*

*Adelaide, SA, Australia*

E-mail: [schroeder@cssip.edu.au](mailto:schroeder@cssip.edu.au)

**Joe Campbell**

*M.I.T. Lincoln Laboratory*

*Lexington, Massachusetts*

E-mail: [j.campbell@ieee.org](mailto:j.campbell@ieee.org)

## Editorial Board

**Maurice Bellanger**

*CNAM*

*Paris, France*

**Robert E. Bogner**

*University of Adelaide*

*Adelaide, SA, Australia*

**Johann F. Böhme**

*Ruhr-Universität Bochum*

*Bochum, Germany*

**James A. Cadzow**

*Vanderbilt University*

*Nashville, Tennessee*

**G. Clifford Carter**

*NUWC*

*Newport, Rhode Island*

**A. G. Constantinides**

*Imperial College*

*London, England*

**Petar M. Djuric**

*State University of New York*

*Stony Brook, New York*

**Anthony D. Fagan**

*University College Dublin*

*Dublin, Ireland*

**Sadaoki Furui**

*Tokyo Institute of Technology*

*Tokyo, Japan*

**John E. Hershey**

*General Electric Company*

*Schenectady, New York*

**B. R. Hunt**

*University of Arizona*

*Tucson, Arizona*

**James F. Kaiser**

*Duke University*

*Durham, North Carolina*

**R. Lynn Kirlin**

*University of Victoria*

*Victoria, British Columbia, Canada*

**Ercan Kuruoğlu**

*Istituto di Elaborazione della Informazione*

*Ghezzano, Italy*

**Meemong Lee**

*Jet Propulsion Laboratory*

*Pasadena, California*

**Petre Stoica**

*Uppsala University*

*Uppsala, Sweden*

**Mati Wax**

*Wavion, Ltd.*

*Yotneam, Israel*

**Rao Yarlagadda**

*Oklahoma State University*

*Stillwater, Oklahoma*

## Digital Signal Processing

Volume 12, Number 1, January 2002

© 2002 Elsevier Science (USA)

All Rights Reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the Publisher. *Exceptions:* Explicit permission from Academic Press is not required to reproduce a maximum of two figures or tables from an Academic Press article in another scientific or research publication provided that the material has not been credited to another source and that full credit to the Academic Press article is given. In addition, authors of work contained herein need not obtain permission in the following cases only: (1) to use their original figures or tables in their future works; (2) to make copies of their papers for use in their classroom teaching; and (3) to include their papers as part of their dissertations.

The appearance of the code at the bottom of the first page of an article in this journal indicates the Publisher's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per copy fee through the Copyright Clearance Center, Inc. (222 Rosewood Drive, Danvers, Massachusetts 01923), for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. Copy fees for pre-2002 articles are as shown on the article title pages; if no fee code appears on the title page, the copy fee is the same as those for current articles.

1051-2004/02 \$35.00

MADE IN THE UNITED STATES OF AMERICA

This journal is printed on acid-free paper.



DIGITAL SIGNAL PROCESSING (ISSN 1051-2004)

Published quarterly by Elsevier Science.

Editorial and Production Offices: 525 B Street, Suite 1900, San Diego, CA 92101-4495

Accounting and Circulation Offices: 6277 Sea Harbor Drive, Orlando, FL 32887-4900

2002: Volume 12. Price \$343.00 U.S.A. and Canada; \$374.00 all other countries

All prices include postage and handling

Information concerning personal subscription rates may be obtained by writing to the Publishers. All correspondence, permission requests, and subscription orders should be addressed to the office of the Publishers at 6277 Sea Harbor Drive, Orlando, FL 32887-4900 (telephone: 407-345-2000). Send notices of change of address to the office of the Publishers at least 6 to 8 weeks in advance. Please include both old and new addresses. POSTMASTER: Send changes of address to *Digital Signal Processing*, 6277 Sea Harbor Drive, Orlando, FL 32887-4900.

LIBRARY

FEB 18 2002

Meta Platforms, Inc. Exhibit 0015  
of Science 271 of 250 gy

# Digital Signal Processing

*A Review Journal*

**Volume 12, Number 1, January 2002**

## *Contents*

- 1      Rapid Power-Line Frequency Monitoring  
         **Ronald M. Adelson**
- 12     A Constrained Block Iterative Algorithm for Multiple-Access  
         Interference Suppression  
         **John F. Doherty and Yu-Tsun Hsieh**
- 21     An Adaptive Combined Classifier System for Invariant  
         Face Recognition  
         **G. A. Khuwaja**
- 47     Representation and Recognition of 3D Free-Form Objects  
         **George Mamic and Mohammed Bennamoun**
- 77     A New Algorithm for 16QAM Carrier Phase Estimation Using  
         QPSK Partitioning  
         **Feng Rice, Mark Rice, and Bill Cowley**
- 87     Near-field Adaptive Beamformer for Robust Speech  
         Recognition  
         **Iain A. McCowan, Darren C. Moore, and S. Sridharan**
- 107    Multiplierless Adaptive Filtering  
         **Tamal Bose, Anand Venkatachalam,  
         and Ratchaneekorn Thamvichai**
- 119    Nonlinear Least  $l_p$ -Norm Filters for Nonlinear Autoregressive  
          $\alpha$ -Stable Processes  
         **Ercan E. Kuruoğlu**

The homepage for Digital Signal Processing can be found  
at <http://www.academicpress.com/dsp>. All articles are  
available online at <http://www.idealibrary.com>.



# Near-field Adaptive Beamformer for Robust Speech Recognition

Iain A. McCowan, Darren C. Moore, and S. Sridharan

Speech Research Laboratory, RCSAVT, School of EESE, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia

E-mail: [iain@ieee.org](mailto:iain@ieee.org); [moore@idiap.ch](mailto:moore@idiap.ch); [s.sridharan@qut.edu.au](mailto:s.sridharan@qut.edu.au)

---

McCowan, I. A., Moore, D. C., and Sridharan, S., Near-field Adaptive Beamformer for Robust Speech Recognition, *Digital Signal Processing* 12 (2002) 87–106.

This paper investigates a new microphone array processing technique specifically for the purpose of speech enhancement and recognition. The main objective of the proposed technique is to improve the low frequency directivity of a conventional adaptive beamformer, as low frequency performance is critical in speech processing applications. The proposed technique, termed *near-field adaptive beamforming* (NFAB), is implemented using the standard generalized sidelobe canceler (GSC) system structure, where a near-field superdirective (NFSD) beamformer is used as the fixed upper-path beamformer to improve the low frequency performance. In addition, to minimize signal leakage into the adaptive noise canceling path for near-field sources, a compensation unit is introduced prior to the blocking matrix. The advantage of the technique is verified by comparing the directivity patterns with those of conventional filter-sum, NFSD, and GSC systems. In speech enhancement and recognition experiments, the proposed technique outperforms the standard techniques for a near-field source in adverse noise conditions. © 2002 Elsevier Science (USA)

**Key Words:** microphone array; beamforming; near-field; adaptive; superdirectivity; speech recognition.

---

## 1. INTRODUCTION

Currently, much research is being undertaken to improve the robustness of speech recognition systems in real environments. This paper focuses on the use of a microphone array to enhance the noisy input speech signal prior to recognition. While the use of microphone arrays for speech recognition has been studied for some time by a number of researchers, a persistent problem has been the poor low frequency directivity of conventional beamforming techniques with

practical array dimensions. Low frequency performance is critical for speech processing applications, as significant speech energy is located below 1 kHz.

By explicitly maximizing the array gain, superdirective beamforming techniques are able to achieve greater directivity than conventional techniques with closely spaced sensor arrays [1]. This directivity generally comes at the expense of a controlled reduction in the white noise gain of the array. Recent work has demonstrated the suitability of superdirective beamforming for speech enhancement and recognition tasks [2, 3]. By employing a spherical propagation model in its formulation, rather than assuming a far-field model, *near-field superdirectivity* (NFSD) succeeds in achieving high directivity at low frequencies for near-field speech sources in diffuse noise conditions [4]. In previous work, near-field superdirectivity has been shown to lead to good speech recognition performance in high noise conditions for a near-field speaker [5].

Superdirective techniques are typically formulated assuming a diffuse noise field. While this is a good approximation to many practical noise conditions, further noise reduction would result from a more accurate model of the actual noise conditions during operation. Adaptive array processing techniques continually update their parameters based on the statistics of the measured input noise. The *generalized sidelobe canceler* (GSC) [6] presents a structure that can be used to implement a variety of adaptive beamformers. A block diagram of the basic GSC system is shown in Fig. 1. The GSC separates the adaptive beamformer into two main processing paths—a standard fixed beamformer,  $\mathbf{w}$ , with  $L$  constraints on the desired signal response, and an adaptive path, consisting of a blocking matrix,  $\mathbf{B}$ , and a set of adaptive filters,  $\mathbf{a}$ . As the desired signal has been constrained in the upper path, the lower path filters can be updated using an unconstrained adaptive algorithm, such as the least-mean-square (LMS) algorithm.

While the theory of adaptive techniques promises greater signal enhancement, this is not always the case in real situations. A common problem with the GSC system is leakage of the desired signal through the blocking matrix, resulting in signal degradation at the beamformer output. This is particularly problematic for broadband signals, such as speech, and especially for speech recognition applications where signal distortion is critical.

In this paper we propose a system that is suited to speech enhancement in a practical near-field situation, having both the good low frequency performance of near-field superdirectivity and the adaptability of a GSC system, while taking

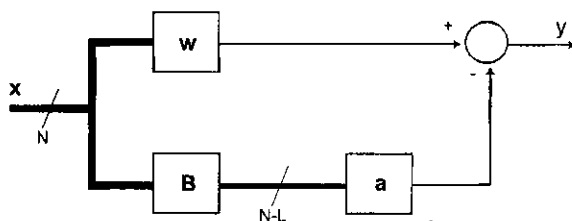


FIG. 1. Generalized sidelobe canceler structure.

care to minimize the problem of signal degradation for near-field sources. We begin by formulating a concise model for near-field sound propagation in Section 2. This model is then used in Section 3 to develop the proposed *near-field adaptive beamforming* (NFAB) technique. To demonstrate the benefit of the technique over existing methods, an experimental evaluation assessing directivity patterns, speech enhancement performance, and speech recognition performance is detailed in Sections 4 and 5.

## 2. NEAR-FIELD SOUND PROPAGATION MODEL

In sensor array applications, a succinct means of characterizing both the array geometry and the location of a signal source is via the *propagation vector*. The propagation vector concisely describes the theoretical propagation of the signal from its source to each sensor in the array. In this section, we develop an expression for the propagation vector of a sound source located in the near-field of a microphone array using a spherical propagation model. This expression is then used in the formulation of the proposed near-field adaptive beamformer in the following sections.

Many microphone array processing techniques assume a planar signal wavefront. This is reasonable for a far-field source, but when the desired source is close to the array a more accurate spherical wavefront model must be employed. For a microphone array of length  $L$ , a source is considered to be in the near-field if  $r < 2L^2/\lambda$ , where  $r$  is the distance to the source and  $\lambda$  is the wavelength.

We define the reference microphone as the origin of a 3-dimensional vector space, as shown in Fig. 2. The position vector for a source in direction  $(\theta_s, \phi_s)$ , at distance  $r_s$  from the reference microphone, is denoted  $\mathbf{p}_s$  and is given by:

$$\mathbf{p}_s = r_s [\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}] \begin{bmatrix} \cos \theta_s \sin \phi_s \\ \sin \theta_s \sin \phi_s \\ \cos \phi_s \end{bmatrix}. \quad (1)$$

The microphone position vectors, denoted as  $\mathbf{p}_i$  ( $i = 1, \dots, N$ ), are similarly defined. The distance from the source to microphone  $i$  is thus

$$d_i = \|\mathbf{p}_s - \mathbf{p}_i\|, \quad (2)$$

where  $\|\cdot\|$  is the Euclidean vector norm.

In such a model, the differences in distance to each sensor can be significant for a near-field source, resulting in phase misalignment across sensors. The difference in propagation time to each microphone with respect to the reference microphone ( $i = 1$ ) is given by

$$\tau_i = \frac{d_i - d_1}{c}, \quad (3)$$

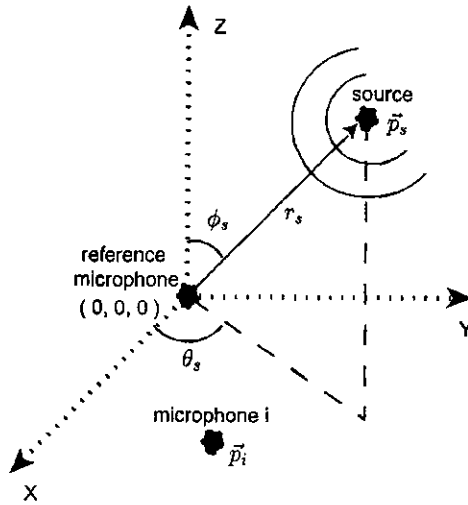


FIG. 2. Near-field propagation model.

where  $c = 340 \text{ ms}^{-1}$  for sound. In addition, the wavefront amplitude decays at a rate proportional to the distance traveled. The resulting amplitude differences across sensors are negligible for far-field sources, but can be significant in the near-field case. The microphone attenuation factors, with respect to the amplitude on the reference microphone, are given by

$$\alpha_i = \frac{d_1}{d_i}. \quad (4)$$

Thus, if  $x_1(f)$  is the desired source at the reference microphone, the signal on the  $i$ th microphone is given by

$$x_i(f) = \alpha_i x_1(f) e^{-j2\pi f \tau_i}. \quad (5)$$

Consequently, we define the near-field propagation vector for a source at distance  $r$  and direction  $(\theta, \phi)$  as

$$\mathbf{d}(f, r, \theta, \phi) = [\alpha_1 e^{-j2\pi f \tau_1} \dots \alpha_i e^{-j2\pi f \tau_i} \dots \alpha_N e^{-j2\pi f \tau_N}]^T. \quad (6)$$

### 3. NEAR-FIELD ADAPTIVE BEAMFORMING

The proposed system structure is shown in Fig. 3. The objective of the proposed technique is to add the benefit of good low frequency directivity to a standard adaptive beamformer, as low frequency performance is critical in speech processing applications. The upper path consists of a fixed near-field superdirective beamformer, while the lower path contains a near-field compensation unit, a blocking matrix and an adaptive noise canceling filter. The principal components of the system are discussed in the following sections.

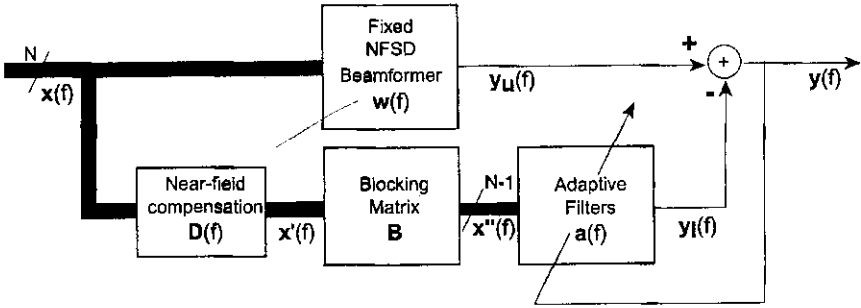


FIG. 3. Near-field adaptive beamformer.

Section 3.1 gives an explanation of the near-field superdirective beamformer. Section 3.2 proposes the inclusion of a near-field compensation unit in the adaptive sidelobe canceling path and examines its effect on reducing signal distortion at the output. Once this near-field compensation has been performed, a standard generalized sidelobe canceling blocking matrix and adaptive filters can be applied to reduce the output noise power, as discussed in Section 3.3.

### 3.1. Near-field Superdirective Beamformer

Superdirective beamforming techniques are based upon the maximization of the array gain, or directivity index. The array gain is defined as the ratio of output signal-to-noise ratio to input signal-to-noise ratio and for the general case can be expressed in matrix notation as [1]

$$G(f) = \frac{\mathbf{w}(f)^H \mathbf{P}(f) \mathbf{w}(f)}{\mathbf{w}(f)^H \mathbf{Q}(f) \mathbf{w}(f)}, \quad (7)$$

where  $\mathbf{w}(f)$  is a column vector of channel gains,

$$\mathbf{w}(f) = [w_1(f) \dots w_i(f) \dots w_N(f)]^T, \quad (8)$$

$(\ )^H$  is the complex conjugate transpose operator, and  $\mathbf{P}(f)$  and  $\mathbf{Q}(f)$  are the cross-spectral density matrices of the signal and noise respectively. In practical speech processing applications the form of the signal and noise cross-spectral density matrices is generally unknown and must be estimated, either from mathematical models (fixed beamformers) or from the statistics of the multichannel inputs (adaptive beamformers). Superdirective beamformers are calculated based on assumed mathematical models for the  $\mathbf{P}(f)$  and  $\mathbf{Q}(f)$  matrices.

When the desired signal is known to emanate from a single source at location  $(r_s, \theta_s, \phi_s)$ , the signal cross-spectral matrix  $\mathbf{P}$  simplifies to the propagation vector of the source, and the array gain can be expressed as

$$G(f) = \frac{|\mathbf{w}(f)^H \mathbf{d}(f, r_s, \theta_s, \phi_s)|^2}{\mathbf{w}(f)^H \mathbf{Q}(f) \mathbf{w}(f)}, \quad (9)$$

where  $\mathbf{d}(f, r, \theta, \phi)$  is the propagation vector for the desired source, as defined in Eq. (6).

A diffuse (spherically isotropic) noise field is often a good approximation for many practical situations, particularly in reverberant closed spaces, such as in a car or an office [7, 8]. For diffuse noise, the noise cross-spectral density matrix  $\mathbf{Q}$  can be formulated as

$$\mathbf{Q}(f) = \frac{1}{4\pi} \int_{\phi} \int_{\theta} \mathbf{d}(f, \theta, \phi) \mathbf{d}(f, \theta, \phi)^H \sin \theta d\theta d\phi, \quad (10)$$

where  $\mathbf{d}(f, \theta, \phi)$  is the propagation vector of a far-field noise source ( $r \gg 2L^2/\lambda$ ) in direction  $(\theta, \phi)$ .

The superdirectivity problem is thus formulated as:

$$\max_{\mathbf{w}(f)} \frac{|\mathbf{w}(f)^H \mathbf{d}(f, r_s, \theta_s, \phi_s)|^2}{\mathbf{w}(f)^H \mathbf{Q}(f) \mathbf{w}(f)}. \quad (11)$$

By using a spherical propagation model to formulate the propagation vector,  $\mathbf{d}$ , the standard superdirective formulation can be optimized for a near-field source [9, 4]. As such, the only difference in the calculation of the standard and near-field superdirective channel filters is the form of the propagation vector,  $\mathbf{d}$ . For a near-field source, the assumption of plane wave (far-field) propagation leads to errors in the array response to the desired signal due to curvature of the direct wavefront. A thorough discussion of the use of a near-field model for superdirective microphone arrays is given by Ryan and Goubran [9].

Cox [10] gives the general superdirective filter solution subject to

1.  $L$  linear constraints,  $\mathbf{C}(f)^H \mathbf{w}(f) = \mathbf{g}(f)$  (explained below); and
2. a constraint on the maximum white noise gain,  $\mathbf{w}(f)^H \mathbf{w}(f) = \delta^{-2}$ , where  $\delta^2$  is the desired white noise gain.

as

$$\mathbf{w}(f) = [\mathbf{Q}(f) + \epsilon \mathbf{I}]^{-1} \mathbf{C}(f) \{ \mathbf{C}(f)^H [\mathbf{Q}(f) + \epsilon \mathbf{I}]^{-1} \mathbf{C}(f) \}^{-1} \mathbf{g}(f), \quad (12)$$

where  $\epsilon$  is a Lagrange multiplier that is iteratively adjusted to satisfy the white noise gain constraint. The white noise gain is the array gain for spatially white (incoherent) noise; that is,  $\mathbf{Q}(f) = \mathbf{I}$ . A constraint on the white noise gain is necessary as an unconstrained superdirective solution will in fact result in significant gain to any incoherent noise, particularly at low frequencies. Cox [10] states that the technique of adding a small amount to each diagonal matrix element prior to inversion is in fact the optimum means of solving this problem. A study of the relationship between the multiplier  $\epsilon$  and the desired white noise gain  $\delta^2$ , shows that the white noise gain increases monotonically with increasing  $\epsilon$ . One possible means of obtaining the desired value of  $\epsilon$  is thus an iterative technique employing a binary search algorithm between a specified minimum and maximum value for  $\epsilon$ . The computational expense of the iterative procedure is not critical, as the beamformer filters depend only on the source

location and array geometry, and thus must only be calculated once for a given configuration.

The constraint matrix,  $\mathbf{C}^H(f)$ , is of order  $L \times N$ , where there are  $L$  linear constraints being applied, and the vector  $\mathbf{g}(f)$  is a length- $L$  column vector of constraining values. The constraints generally include one specifying unity response for the desired signal,  $\mathbf{d}^H(f)\mathbf{w}(f) = 1$ , and where this is the sole constraint the above solution can be simplified by substituting  $\mathbf{C}(f) = \mathbf{d}(f)$  and  $\mathbf{g}(f) = 1$ , giving

$$\mathbf{w}(f) = \frac{[\mathbf{Q}(f) + \epsilon \mathbf{I}]^{-1} \mathbf{d}(f)}{\mathbf{d}(f)^H [\mathbf{Q}(f) + \epsilon \mathbf{I}]^{-1} \mathbf{d}(f)}. \quad (13)$$

Once the optimal filters  $\mathbf{w}(f)$  have been calculated, the near-field superdirective beamformer output is calculated as

$$y_u(f) = \mathbf{w}(f)^H \mathbf{x}(f), \quad (14)$$

where  $\mathbf{x}(f)$  is the  $N$ -channel input column vector

$$\mathbf{x}(f) = [x_1(f) \ \dots \ x_i(f) \ \dots \ x_N(f)]^T. \quad (15)$$

### 3.2. Near-field Compensation Unit

The first element in the adaptive path of standard GSC is the blocking matrix [6]. Its purpose is to block the desired signal from the adaptive noise estimate. To ensure complete blocking, the desired signal must both be time aligned and have equal amplitudes across all channels. If this is the case, cancellation occurs if each row of the blocking matrix sums to zero, and all rows are linearly independent.

For a near-field desired source, to align the desired signal on all channels, a near-field compensation must first be applied to the input channels prior to blocking. To ensure full cancellation we need to compensate for both phase misalignment and amplitude scaling of the desired signal across sensors. We define the diagonal matrix

$$\mathbf{D}(f) = [\text{diag}(\mathbf{d}(f))]^{-1}, \quad (16)$$

where  $\mathbf{d}(f)$  is the near-field propagation vector from Eq. (6). In this paper we define the diagonal operator,  $\text{diag}(\cdot)$ , to produce a diagonal matrix from a vector parameter. Conversely, if invoked with a matrix parameter, it produces a row vector corresponding to the matrix diagonal. The near-field compensation can be applied as

$$\mathbf{x}'(f) = \mathbf{D}(f)\mathbf{x}(f). \quad (17)$$

Once this near-field compensation has been performed, a standard GSC blocking matrix can be employed to block the desired signal from the adaptive path.

The inclusion of this compensation unit is critical for a near-field desired signal. Without compensation for both phase and amplitude differences between sensors, blocking of the desired signal will not be ensured, leading to signal

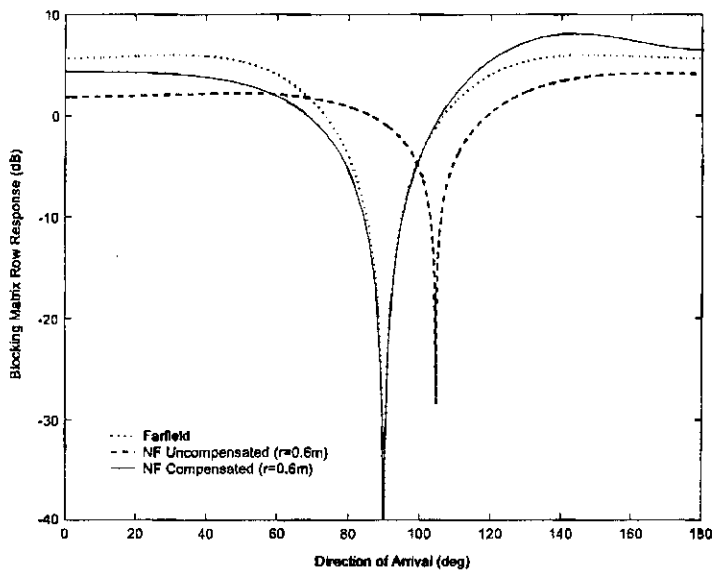
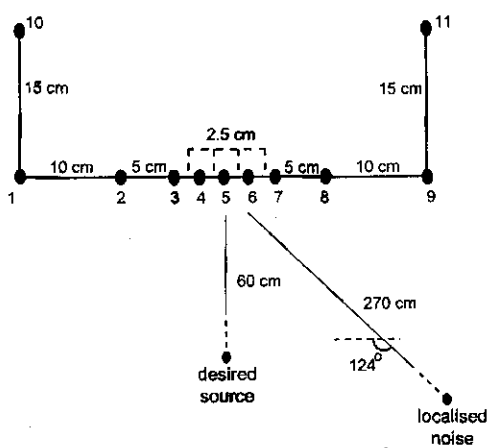


FIG. 4. Comparison of blocking matrix row beam-patterns.

cancellation at the output. The near-field compensation effectively ensures that a true null exists in the beam-pattern of each blocking matrix row in the direction *and* distance corresponding to the desired source. To illustrate, Fig. 4 shows the directivity pattern at 2 kHz for the first row in the blocking matrix using the array shown in Fig. 5, with the desired source directly in front of the center microphone at a distance of 0.6 m. The figure shows the compensated response in the far- and near-fields, as well as the uncompensated near-field response. It is clear that the uncompensated system will allow a high degree of signal leakage into the adaptive path as it blocks noise sources rather than the desired signal.



Meta Platforms, Inc. - Exhibit 1015  
FIG. 5. Experimental configuration. 230 of 250

### 3.3. Blocking Matrix and Adaptive Noise Canceling Filter

The blocking matrix and adaptive noise canceling filters are taken from the standard GSC technique [6]. The order of the blocking matrix is  $N \times (N - L)$ , where there are  $L$  constraints applied in the fixed upper path beamformer. Generally only a unity constraint on the desired signal is specified, and the standard  $N \times (N - 1)$  Griffiths–Jim blocking matrix is used:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \vdots & \vdots & \vdots \\ 0 & -1 & \ddots & 0 & 0 \\ 0 & 0 & \ddots & 1 & 0 \\ \vdots & \vdots & \vdots & -1 & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}. \quad (18)$$

The output of the blocking matrix is calculated as

$$\mathbf{x}''(f) = \mathbf{B}^H \mathbf{x}'(f), \quad (19)$$

where  $\mathbf{x}''(f)$  is an  $(N - 1)$ -length column vector. Defining the  $(N - 1)$ -length adaptive filter column vector as

$$\mathbf{a}(f) = [a_1(f) \ \dots \ a_i(f) \ \dots \ a_{N-1}(f)]^T, \quad (20)$$

the output of the lower path is given as

$$y_l(f) = \mathbf{a}(f)^H \mathbf{x}''(f). \quad (21)$$

The NFAB output is then calculated from the upper and lower path outputs as

$$y(f) = y_u(f) - y_l(f) \quad (22)$$

and the adaptive filters are updated using the standard unconstrained LMS algorithm

$$\mathbf{a}_{k+1}(f) = \mathbf{a}_k(f) + \mu \mathbf{x}_k''(f) y_k(f), \quad (23)$$

where  $\mu$  is the adaptation step size and  $k$  denotes the current frame.

### 3.4. Summary of Technique

In summary, the proposed NFAB technique is characterized by the series of equations

$$y_u(f) = \mathbf{w}(f)^H \mathbf{x}(f) \quad (24a)$$

$$\mathbf{x}_k'' = \mathbf{B}^H \mathbf{D}(f) \mathbf{x}(f) \quad (24b)$$

$$y_l(f) = \mathbf{a}(f)^H \mathbf{x}_k''(f) \quad (24c)$$

$$y(f) = y_u(f) - y_l(f) \quad (24d)$$

$$\mathbf{a}_{k+1}(f) = \mathbf{a}_k(f) + \mu \mathbf{x}_k''(f) y_k(f), \quad (24e)$$

where all terms have been defined in the preceding discussion.

#### 4. EXPERIMENTAL CONFIGURATION

For the experimental evaluation in this paper, we used the 11 element array shown in Fig. 5. The array consists of a nine element broadside array, with an additional two microphones situated directly behind the end microphones. The total array is 40 cm wide and 15 cm deep in the horizontal plane. The broadside microphones are arranged according to a standard broadband subarray design, where different subarrays are used for different frequency ranges for the fixed upper path beamformer. The two endfire microphones are included for use by the near-field superdirective beamformer in the low frequency range. The four subarrays are thus

- ( $f < 1$  kHz): microphones 1–11;
- ( $1 \text{ kHz} < f < 2$  kHz): microphones 1, 2, 5, 8, and 9;
- ( $2 \text{ kHz} < f < 4$  kHz): microphones 2, 3, 5, 7, and 8; and
- ( $4 \text{ kHz} < f < 8$  kHz): microphones 3–7.

The array was situated in a computer room, with different sound source locations, as shown in Fig. 5. The two sound sources were

1. the desired speaker situated 60 cm from the center microphone, directly in front of the array; and
2. a localized noise source at an angle of  $124^\circ$  and a distance of 270 cm from the array.

Impulse responses of the acoustic path between each source and microphone were measured from multichannel recordings made in the room with the array using the maximum length sequence technique detailed in Rife and Vanderkooy [11]. As the impulse responses were calculated from real recordings made simultaneously across all input channels, they take into account the real acoustic properties of the room and the array. The multichannel desired speech and localized noise microphone inputs were then generated by convolving the original single-channel speech and noise signals with these impulse responses. In addition, a real multichannel background noise recording of normal operating conditions was made in the room with other workers present. This recording is referred to in the experiments as the ambient noise signal and is approximately diffuse in nature. It consists mainly of computer noise, a variable level of background speech, and noise from an air-conditioning unit. The ambient noise effectively represents a diffuse noise field, while the localized noise represents a coherent noise source. In this paper, we specify the levels of the two different noise sources independently, as the signal to ambient-noise ratio (SANR) and

signal to localized-noise ratio (SLNR). These values are calculated as the average segmental SNR from the speech and noise input, as measured at the center microphone of the array.

In this way, realistic multichannel input signals can be simulated for specified levels of ambient and localized noise. As well as facilitating the generation of different noise conditions, simulating the multichannel inputs using the impulse response method is more practical than making real recordings for speech recognition experiments, as existing single channel speech corpora may be used.

5. EXPERIMENTAL RESULTS

This section presents the results of the experimental evaluation. The proposed NFAB technique is compared to a conventional fixed filter-sum beamformer, a fixed near-field superdirective beamformer, and a conventional GSC adaptive beamformer. These beamformers are specified in Table 1.

The techniques are first assessed in terms of the directivity pattern in order to demonstrate the advantage of the proposed NFAB over conventional beamforming techniques, particularly at low frequencies. Following this, the techniques are evaluated for speech enhancement in terms of the improvement in signal to noise ratio and the log area ratio. Finally, the techniques are compared in a hands-free speech recognition task in noisy conditions using the TIDIGITS database [12].

5.1. Directivity Analysis

As has been stated, the main objective of the proposed technique is to produce an *adaptive beamformer* that exhibits good *low frequency* performance for *near-field* speech sources. To assess the effectiveness of the proposed technique in achieving this objective, in this section we analyze the horizontal directivity pattern. The directivity of a filter-sum beamformer is expressed in matrix notation as

$$h(f, r, \theta, \phi) = \mathbf{w}_o(f)^H \mathbf{d}(f, r, \theta, \phi), \tag{25}$$

where  $\mathbf{w}_o$  is the length  $N$  channel filter vector

$$\mathbf{w}_o(f) = [w_{o,1}(f) \ \dots \ w_{o,i}(f) \ \dots \ w_{o,N}(f)]^T. \tag{26}$$

TABLE 1  
Beamforming Techniques in Evaluation

| Technique | Description                                    | Filters                                                                                   |
|-----------|------------------------------------------------|-------------------------------------------------------------------------------------------|
| FS        | Conventional FS beamformer                     | $\mathbf{w}_o(f) = [\text{diag}(\mathbf{D}(f))]^H$                                        |
| NFSD      | Near-field superdirective beamformer           | $\mathbf{w}_o(f) = \mathbf{w}(f)$                                                         |
| GSC       | GSC system with FS fixed upper path beamformer | $\mathbf{w}_o(f) = [\text{diag}(\mathbf{D}(f))]^H - \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$ |
| NFAB      | Near-field adaptive beamformer                 | $\mathbf{w}_o(f) = \mathbf{w}(f) - \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$                  |

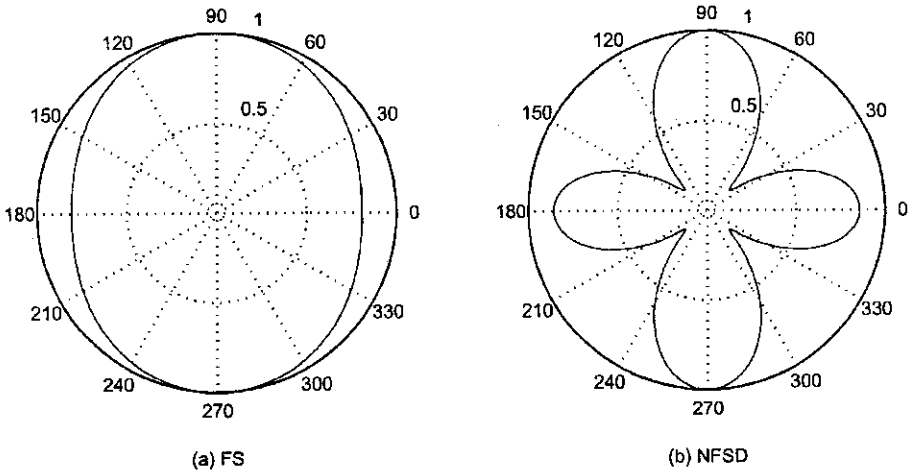


FIG. 6. Upper path directivity pattern at 300 Hz.

**5.1.1. Upper path directivity.** First, we seek to demonstrate the directivity improvement that NFSD achieves at low frequencies compared to a conventional filter-sum (FS) beamformer. For the FS beamformer, a common solution is to choose  $\mathbf{w}_o(f) = [\text{diag}(\mathbf{D}(\mathbf{f}))]^H$ . This effectively ensures that the desired signal is aligned for phase and amplitude across sensors using a spherical propagation model. For NFSD, we use the filter vector  $\mathbf{w}(f)$  described in Section 3.1. Figure 6 shows the near-field directivity pattern at 300 Hz for the FS and NFSD. From these figures, it is clear that the NFSD technique results in greater directional discrimination at low frequencies compared to a conventional beamformer. At higher frequencies ( $f > 1$  kHz), conventional beamformers offer reasonable directivity, and so the FS and NFSD techniques give comparable performance.

**5.1.2. Lower path directivity.** Second, we wish to demonstrate the effect of the noise canceling path. The directivity of the noise canceling filters can be obtained by using the channel filters  $\mathbf{w}_o(f) = \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$ . The blocking matrix and adaptive filters essentially implement a conventional (nonsuperdirective) beamformer that adaptively focuses on the major sources of noise. To examine the directivity of the lower path filters, the beamformer was run on an input speech signal with a white localized noise source (at the location shown in Fig. 5) added at an SLNR of 0 dB and a low level of ambient noise (SANR = 20 dB). The steady-state adaptive filter vector,  $\mathbf{a}(f)$ , was written to file for both the proposed NFAB technique and the conventional GSC beamformer. The near-field directivity patterns of the lower path filters are plotted in Figs. 7 and 8 for 300 and 5000 Hz, respectively. We see that the lower path adaptive filters for both beamformers converge to similar solutions in terms of directivity, producing a main lobe in the direction of the coherent noise source ( $\approx 124^\circ$  from Fig. 5), as well as a null in the location of the desired speaker. As expected, the directivity of the adaptive path is poor at low frequencies, as seen in Fig. 7.

**5.1.3. Overall beamformer directivity.** Finally, we examine the directivity pattern of the overall beamformer for the NFAB and conventional adaptive

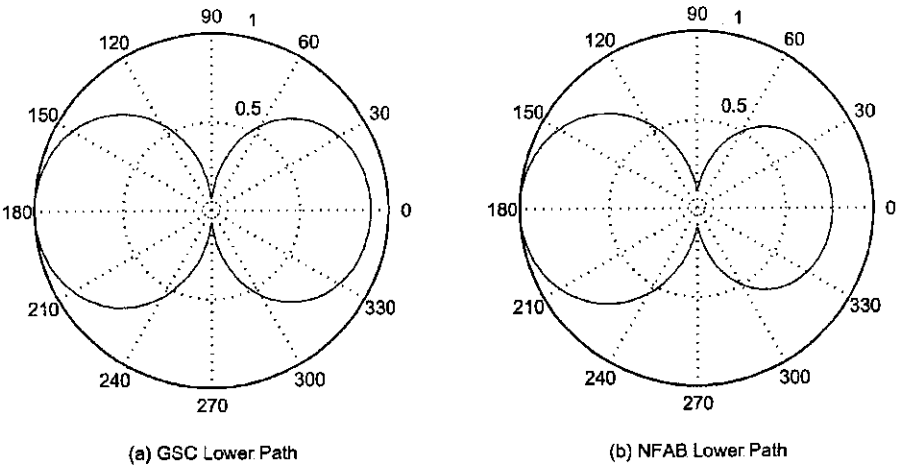


FIG. 7. Lower path directivity pattern at 300 Hz.

systems. The near-field directivity patterns at 300 Hz are shown in Fig. 9. We see that the directivity pattern of the NFAB system exhibits a true null in the direction and at the distance of the noise source, while the directivity of the conventional beamformer is too poor to significantly attenuate the noise at this frequency. At frequencies above 1 kHz the directivity performance of both techniques is comparable.

5.1.4. Summary of beamformer directivity. In summary we see that, in terms of directivity, the proposed NFAB system:

- outperforms the conventional FS system in terms of low frequency performance and the ability to attenuate coherent noise sources,
- outperforms the NFSD system due to the ability to attenuate coherent noise sources, and

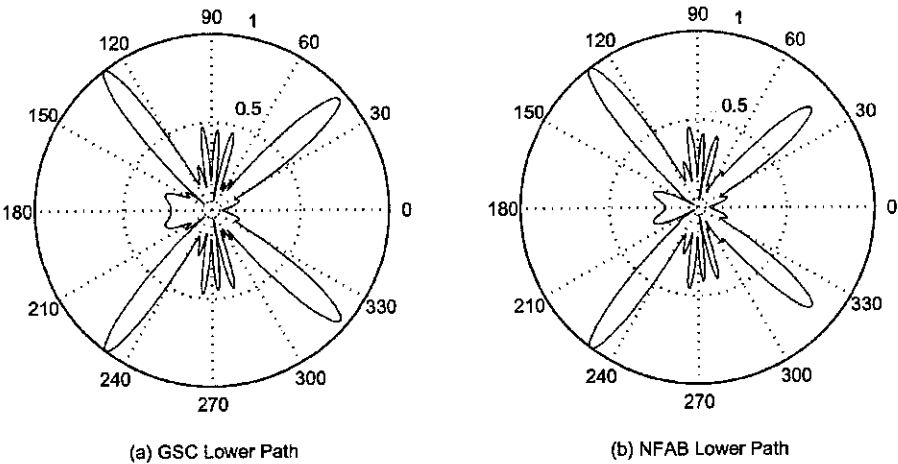


FIG. 8. Lower path directivity pattern at 300 Hz.

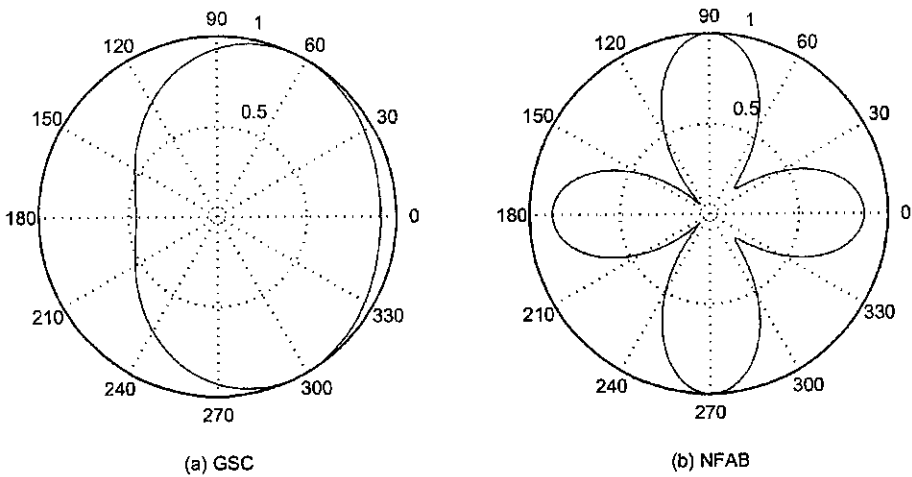


FIG. 9. Overall beamformer directivity pattern at 300 Hz.

• outperforms the conventional GSC system in terms of low frequency performance.

In this way, we see that the proposed system succeeds in meeting the stated objectives and should therefore demonstrate improved performance in speech processing applications.

5.2. Speech Enhancement Analysis

The signal plots in Fig. 10 give an indication of the level of enhancement achieved by the NFAB technique. For the desired speech signal, we used a segment of speech from the TIDIGITS database corresponding to the digit sequence *one-nine-eight-six*. Ambient noise was added at an SANR level of 10 dB, and a localized white noise signal was added at an SLNR level of 0 dB. The plots indicate that NFAB succeeds in reducing the noise level with negligible distortion to the desired signal.

To better measure the level of enhancement, objective speech measures were used to compare the different techniques. Two measures were used, these being the SNR improvement and the log area ratio distortion measure. The *SNR improvement* is defined as the difference in SNR at the array output and input. As the true SNR cannot be measured, it is estimated as the average segmental signal-plus-noise to noise ratio. While the signal to noise ratio is a useful measure for assessing noise reduction, it does not necessarily give a good indication of how much distortion has been introduced to the desired speech signal. The *log area ratio* (LAR) measure of speech quality is more highly correlated with perceptual intelligibility in humans [13]. The log area ratio measure for a frame of speech is calculated as

$$LAR(n) = \left| \frac{1}{P} \sum_{i=1}^P \left[ \log \frac{1 + r_o(i)}{1 - r_o(i)} - \log \frac{1 + r_p(i)}{1 - r_p(i)} \right] \right|^{1/2}$$

Meta Platforms, Inc. - Exhibit 101(27)  
236 of 250

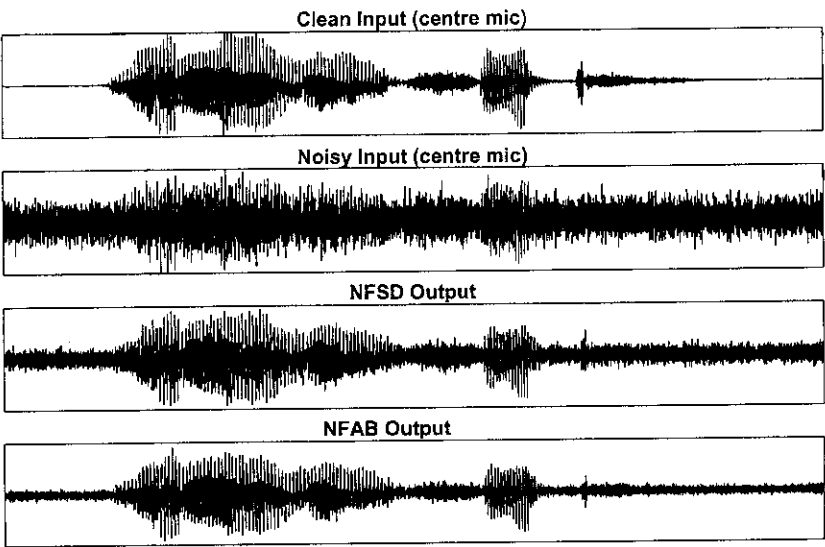


FIG. 10. Sample enhanced signal.

where  $n$  is the frame number, and  $r_o$  and  $r_p$  are the original and processed  $P$ th-order linear predictive coefficients of the  $n$ th frame, respectively. The overall log area ratio distortion measure for the signal is calculated as the average distortion over all input frames.

A set of experiments was conducted in which the localized white noise was replaced with a localized speech-like noise source taken from the NOISEX database [14]. This is essentially a white noise signal that has been shaped with a speech-like spectral envelope and thus represents a more realistic noise scenario than white noise. The signal to localized noise ratio (SLNR) was varied from 20 to 0 dB, with the ambient noise present at a constant SANR level of 10 dB. The output signal to noise ratio improvement and log area ratios are given in Tables 2 and 3 for the different enhancement techniques.<sup>1</sup> The measures have been averaged over 10 randomly chosen speech segments taken

TABLE 2  
Signal to Noise Ratio Improvement (SANR = 10 dB)

| Technique | SLNR (dB) |     |     |     |     |
|-----------|-----------|-----|-----|-----|-----|
|           | 0         | 5   | 10  | 15  | 20  |
| FS        | 0.5       | 0.4 | 0.3 | 0.1 | 0.1 |
| NFSD      | 1.4       | 1.6 | 1.1 | 0.5 | 0.2 |
| GSC       | 1.6       | 1.8 | 1.9 | 2.5 | 3.3 |
| NFAB      | 5.5       | 5.9 | 6.4 | 7.5 | 7.9 |

<sup>1</sup> Sample sound files are also available at <http://www.speech.qut.edu.au/pages/people/mccowan/>.  
237 of 250

**TABLE 3**  
Log Area Ratio: (SANR = 10 dB)

| Technique   | SLNR (dB) |     |     |     |     |
|-------------|-----------|-----|-----|-----|-----|
|             | 0         | 5   | 10  | 15  | 20  |
| Noisy input | 3.6       | 3.3 | 2.9 | 2.6 | 2.5 |
| FS          | 3.1       | 2.7 | 2.5 | 2.4 | 2.3 |
| NFSD        | 2.8       | 2.3 | 2.1 | 2.0 | 1.9 |
| GSC         | 2.6       | 2.1 | 1.8 | 1.7 | 1.6 |
| NFAB        | 2.5       | 1.9 | 1.6 | 1.6 | 1.4 |

from different speakers in the TIDIGITS database. These results are plotted in Fig. 11.

The SNR results show that the proposed NFAB gives considerably greater noise reduction compared to the FS, NFSD, and GSC techniques, providing approximately 6–8 dB of SNR improvement compared to the noisy input signal. Even with a relatively low level of localized noise (high SLNR), the NFAB technique offers significantly greater noise reduction than these other methods. In addition, the proposed technique gives less distortion than the other techniques, as measured by the LAR. As would be expected, the fixed NFSD technique gives slightly less distortion than the adaptive GSC technique.

From these results we see that, in a high level of diffuse and coherent noise with a near-field desired speech source, the proposed NFAB technique succeeds in significantly reducing the noise level, while minimizing the distortion to the speech signal.

5.3. Speech Recognition Analysis

The same noise scenario was used for experiments in robust speech recognition. The training and test data for the experiments were taken from the male adult portion of the TIDIGITS database. Tied-state triphone hidden Markov models and standard MFCC parameterization with energy, delta, and accel-

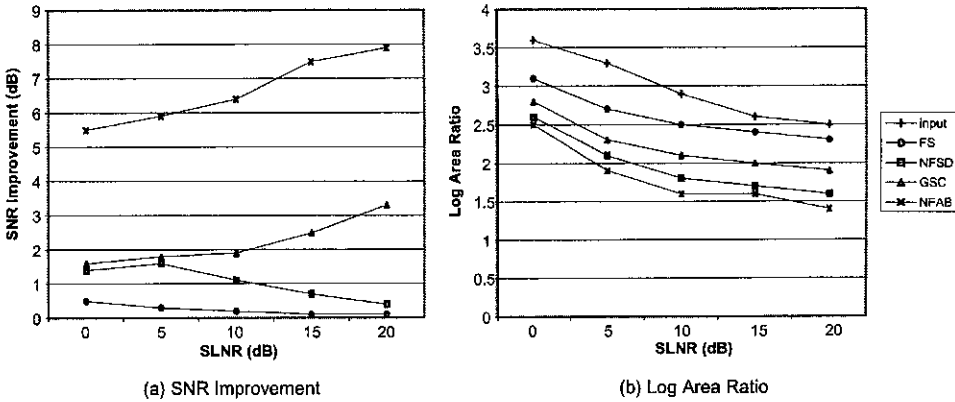


**FIG. 11.** Speech enhancement measures (SNR Improvement and Log Area Ratio) versus SLNR (dB).  
Meta Platforms Inc and Exhibit 1015  
238 of 250

TABLE 4  
Speech Recognition Results: Word Recognition Rates

| Technique   | SLNR (dB) |      |      |      |
|-------------|-----------|------|------|------|
|             | 10        | 5    | 0    | -5   |
| Noisy input | 86.8      | 65.9 | 23.2 | 13.1 |
| FS          | 89.2      | 81.7 | 62.9 | 36.4 |
| NFSD        | 97.7      | 93.2 | 77.2 | 45.4 |
| GSC         | 88.8      | 83.8 | 73.8 | 56.8 |
| NFAB        | 98.2      | 96.7 | 91.1 | 76.7 |

ation coefficients were used. The models were trained with the clean input to the center microphone and then refined using MAP adaptation to better match the noisy environment. The noise segments used in the adaptation process were taken from a separate recording made in the room. The recognition results are given as percentage word recognition rates in Table 4 and shown graphically in Fig. 12.

The results clearly show that NFAB gives excellent robustness to adverse noise conditions in a near-field speech recognition application. The results at low noise levels show that the baseline recognition system is already quite robust to noise, due to the use of MAP adaptation. At more realistic noise levels, however, unenhanced performance is clearly unsatisfactory. For example, at an SLNR of 0 dB and SANR of 10 dB, the word *error* rate for the unprocessed input is 76.8%. While standard GSC and NFSD are able to reduce this to 26.2 and 22.8%, respectively, the proposed NFAB technique succeeds in reducing the error rate to 8.9%. As would be expected, the figure shows that NFAB offers

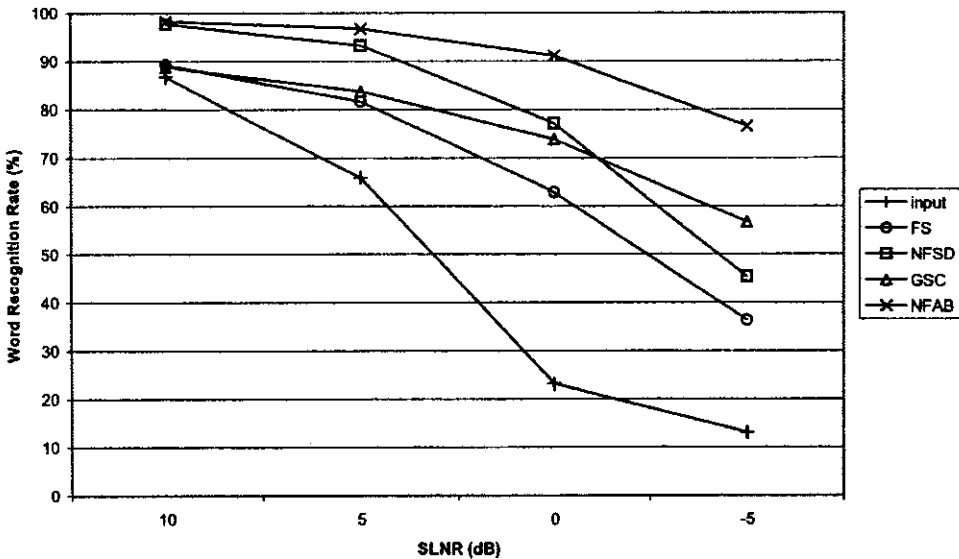


FIG. 12. Speech recognition results: Word recognition rates. Exhibit 1015  
Meta Platforms, Inc. 239 of 250

similar performance to NFSD when the noise is approximately diffuse (high SLNR) and demonstrates improved ability to attenuate any coherent noise sources (low SLNR) due to the GSC-style adaptive noise canceling path. The recognition performance of NFAB is seen to be similar to NFSD at low levels of coherent noise (high SLNR) and degrades at a rate comparable to GSC with increasing levels of coherent noise.

It is apparent from these results that NFAB is an enhancement technique that is well suited to speech recognition. The experimental results for both speech enhancement and recognition demonstrate that for an adaptive beamformer to be applicable in speech processing applications, it should exhibit good directivity at low frequencies and take care to minimize any signal degradation.

## 6. CONCLUSIONS

A new microphone array processing technique designed specifically for near-field speech processing applications has been proposed, termed near-field adaptive beamforming (NFAB). The technique incorporates a fixed near-field superdirective beamformer into a GSC-style adaptive beamforming structure and as such exhibits the benefits of good low frequency performance and the ability to adaptively attenuate coherent noise signals. Distortion due to the adaptive noise canceling path is minimized by the introduction of a near-field compensation unit.

Two major problems with common conventional microphone array techniques are their poor low frequency performance and the introduction of signal distortion in adaptive techniques. By taking care to address both these issues, the proposed NFAB technique succeeds in significantly outperforming conventional beamforming techniques in terms of objective speech quality measures and speech recognition results in both diffuse and coherent noise.

Speech enhancement results indicate that NFAB succeeds in significantly reducing the output noise power, while also minimizing the distortion to the desired signal. These characteristics make it ideal as an enhancement technique for robust speech recognition. In a high noise configuration, with a signal to localized noise ratio of 0 dB, and a signal to ambient noise ratio of 10 dB, the proposed technique succeeds in increasing the recognition rate from 23.2 to 91.1%. For the same configuration, near-field superdirectivity and conventional GSC only achieve 77.2 and 73.8%, respectively.

In summary, near-field adaptive beamforming has been shown to be a speech enhancement technique that produces a high quality, highly intelligible signal for applications requiring hands-free speech acquisition where the desired speaker is in the array's near-field.

## ACKNOWLEDGMENTS

The near-field superdirective technique was developed at France Telecom R&D [4]. The initial application of these techniques in a speech recognition application was researched by the author during a six month research period at France Telecom R&D in 1999 [5], under the supervision of Yannick Mahieux. The authors specifically wish to acknowledge Claude Marro for his helpful and insightful review of this paper.

## REFERENCES

1. Cox, H., Zeskind, R., and Kooij, T., Practical supergain. *IEEE Trans. Acoustics, Speech Signal Process.* **34** (1986), 393–398.
2. Bitzer, J., Simmer, K. U., and Kammeyer, K., Multi-microphone noise reduction techniques for hands-free speech recognition—a comparative study. In *Robust Methods for Speech Recognition in Adverse Conditions (ROBUST-99)*, Tampere, Finland, May 1999, pp. 171–174.
3. Doerbecker, M., Speech enhancement using small microphone arrays with optimized directivity. In *Proc. Int. Workshop on Acoustic Echo and Noise Control*, September 1997, pp. 100–103.
4. Täger, W., Near field superdirectivity (NFSD). In *Proceedings of ICASSP'98*, 1998, pp. 2045–2048.
5. McCowan, I., Marro, C., and Mauuary, L., Robust speech recognition using near-field superdirective beamforming with post-filtering. In *Proceedings of ICASSP'2000*, 2000, Vol. 3, pp. 1723–1726.
6. Griffiths, L. and Jim, C., An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagation* **30** (1982), 27–34.
7. Meyer, J. and Simmer, K. U., Multi-channel speech enhancement in a car environment using wiener filtering and spectral subtraction. In *Proceedings of ICASSP'97*, 1997, Vol. 2, pp. 1167–1170.
8. Bitzer, J., Simmer, K. U., and Kammeyer, K., Theoretical noise reduction limits of the generalized sidelobe canceller (gac) for speech enhancement. In *Proceedings of ICASSP'99*, 1999, Vol. 5, pp. 2965–2968.
9. Ryan, J. G. and Goubran, R. A., Near-field beamforming for microphone arrays. In *Proceedings of ICASSP'97*, 1997, pp. 363–366.
10. Cox, H., Zeskind, R., and Owen, M., Robust adaptive beamforming. *IEEE Trans. Acoustics, Speech Signal Process.* **35** (1987), 1365–1376.
11. Rife, D. and Vanderkooy, J., Transfer-function measurement with maximum-length sequences. *J. Audio Eng. Soc.* **37** (1989), 419–444.
12. Texas Instruments and NIST. Studio quality speaker-independent connected-digit corpus (TDIGITS). CD-ROM, February 1991. NIST Speech Discs 4-1, 4-2, and 4-3.
13. Quackenbush, S. R., Barnwell, T. P., and Clements, M. A., *Objective Measures of Speech Quality*. Prentice-Hall, Englewood Cliffs, NJ, 1988.
14. Defence Research Agency Speech Research Unit. NOISEX-92. CD-ROM, June 1992.

---

IAIN A. MCCOWAN received the B.Eng (Hons) and B.InfoTech from the Queensland University of Technology, Brisbane, in 1996. In February 1998 he joined the Research Concentration in Speech, Audio, and Video Technology at the Queensland University of Technology where he is currently completing his Ph.D. His main research interests are in the fields of robust speech recognition and speech enhancement using microphone arrays. He is a student member of the Institute of Electrical and Electronic Engineers.

DARREN C. MOORE received the B.Eng (Hons) and B.InfoTech from the Queensland University of Technology, Brisbane, in 1997. In February 1998 he joined the Research Concentration in Speech, Audio, and Video Technology at the Queensland University of Technology, where he is currently completing a M.Eng. His professional interests lie in the field of speech enhancement using microphone arrays and in the implementation of real-time DSP solutions.

S. SRIDHARAN obtained his B.Sc (electrical engineering) and M.Sc (communication engineering) from the University of Manchester Institute of Science and Technology, United Kingdom and Ph.D (signal processing) from the University of New South Wales, Australia. Dr. Sridharan is senior member of the IEEE, USA and a corporate member of IEE, United Kingdom and IEAust of Australia. He is currently a professor in the School of Electrical and Electronic Systems Engineering of the Queensland University of Technology (QUT) and is also the Head of the Research Concentration in Speech, Audio, and Video Technology at QUT.

# APPENDIX 1006D



JOURNAL

## Digital signal processing.

Available at Parks Library Storage Building (TK5102.5 D4463) >

Online access (Coverage: 1991-2019) >

View Journal Contents

TOP

SEARCH INSIDE

VIEW ONLINE

GET IT

TAGS

DETAILS

SEND TO

LINKS

VIRTUAL BROWSE

Send to



QR



ENDNOTE



MENDELEY



CITATION



EXPORT BIBTEX



EXPORT RIS



EXPORT TO  
EXCEL



PERMALINK



PRINT



E-MAIL

[https://quicksearch.lib.iastate.edu/permalink/01IASU\\_INST/174tg9m/alma990008738310102756](https://quicksearch.lib.iastate.edu/permalink/01IASU_INST/174tg9m/alma990008738310102756)



COPY THE PERMALINK TO CLIPBOARD

See something that doesn't look right? [Report a Problem](#)

- TOP
- SEARCH INSIDE
- VIEW ONLINE
- GET IT
- TAGS
- DETAILS
- SEND TO
- LINKS
- VIRTUAL BROWSE

Search inside

Search for articles within this journal

Article title or keyword

View Online

Availability

|                                                                                   |              |                   |
|-----------------------------------------------------------------------------------|--------------|-------------------|
| Elsevier ScienceDirect Journals                                                   | SHOW LICENSE | <a href="#">↗</a> |
| Available from 01/01/1991 volume: 1 issue: 1 until 12/31/2019 volume: 95.         |              |                   |
| Elsevier ScienceDirect Journals - Owned                                           | SHOW LICENSE | <a href="#">↗</a> |
| Available from 2006 until 2019.                                                   |              |                   |
| Elsevier SD Backfile Computer                                                     | SHOW LICENSE | <a href="#">↗</a> |
| Available from 01/01/1991 volume: 1 issue: 1 until 10/31/1994 volume: 4 issue: 4. |              |                   |
| Elsevier SD Backfile Computer Science Supplement                                  | SHOW LICENSE | <a href="#">↗</a> |
| Available from 01/01/1991 volume: 1 issue: 1 until 10/31/1994 volume: 4 issue: 4. |              |                   |
| Elsevier SD Backfile Engineering and Technology                                   | SHOW LICENSE | <a href="#">↗</a> |
| Available from 01/01/1991 volume: 1 issue: 1 until 10/31/1994 volume: 4 issue: 4. |              |                   |

- TOP
- SEARCH INSIDE
- VIEW ONLINE
- GET IT
- TAGS
- DETAILS
- SEND TO
- LINKS
- VIRTUAL BROWSE

Please sign in to see request options and to place requests.  Sign in to request, save searches, and [more](#)

 [BACK TO LOCATIONS](#)

LOCATION ITEMS

Parks Library

Available, Storage Building ; TK5102.5 D4463

Summary holdings: v. 4 (1994)-v. 15 (2005)

 LOCATE  

|                                                          |                                                                                       |
|----------------------------------------------------------|---------------------------------------------------------------------------------------|
| Item in place (0 requests) v. 15,no.1-2,2005<br>Loanable |    |
| Item in place (0 requests) v. 14,2004<br>Loanable        |    |
| Item in place (0 requests) v. 13,2003<br>Loanable        |    |
| Item in place (0 requests) v. 12,2002<br>Loanable        |   |
| Item in place (0 requests) v. 11,2001<br>Loanable        |  |

[SHOW MORE ITEMS](#)

Title [Digital signal processing.](#)

Subjects [Signal processing](#) -- [Digital techniques](#) >

Genre [Periodicals.](#) >

|                |            |                                                                                 |
|----------------|------------|---------------------------------------------------------------------------------|
|                | Details    |                                                                                 |
| TOP            | Title      | Digital signal processing.                                                      |
| SEARCH INSIDE  | Subjects   | Signal processing -- Digital techniques >                                       |
| VIEW ONLINE    | Genre      | Periodicals. ><br>Periodicals >                                                 |
| GET IT         | Publisher  | Duluth, MN : Academic Press<br>Vol. 1, no. 1 (Jan. 1991)-                       |
| TAGS           | Notes      | Title from cover.<br>Imprint varies.                                            |
| DETAILS        | Language   | English                                                                         |
| SEND TO        | Format     | volumes : illustrations ; 28 cm                                                 |
| LINKS          | Frequency  | Bimonthly (varies)                                                              |
| VIRTUAL BROWSE | Source     | Library Catalog                                                                 |
|                | Identifier | LC : 91650983<br>ISSN : 1051-2004<br>OCLC : (OCoLC)21889588<br>ISSN : 1095-4333 |
|                | Record ID  | 990008738310102756                                                              |

Tags

 ADD YOUR OWN TAGS

Virtual Browse

  
Digital signal processing : a

  
Detection of abrupt

  
Digital signal processing

  
Digital signal processing :

  
Digital signal processing

  
Digital signal processing --

  
Digit

```

leader 01465nas a2200433 a 4500
001 9900008738310102756
005 20210426142726.0
008 900620c19919999mnubr1p 0 a0eng d
010 ##$a91650983 $zsn 90002038
012 ##$a3 $i9111 $k1 $l1
020 ##$c$110.00
022 0#$a1051-2004
035 ##$a(aleph)000873831ISU01
035 9#$a873831
035 ##$a(0CoLC)21889588
035 ##$a(BIBL)990008738310102756
037 ##$bAcademic Press, Inc., 1 E. 1st St., Duluth, MN 55802
040 ##$aNSD $beng $cNSD $dDLC $dIWA
042 ##$ansdp $alc
050 00$aTK5102.5 D4463
210 0#$aDigit. signal process.
222 #0$aDigital signal processing
245 00$aDigital signal processing.
260 ##$aDuluth, MN : $bAcademic Press, $cc1991-.
300 ##$avolumes : $billustrations ; $c28 cm
310 ##$aBimonthly (varies)
336 ##$atext $btxt $2rdacontent
337 ##$aunmediated $bn $2rdamedia
338 ##$avolume $bnc $2rdacarrier
362 0#$aVol. 1, no. 1 (Jan. 1991)-
500 ##$aTitle from cover.
500 ##$aImprint varies.
530 ##$aAlso issued online.
650 #0$aSignal processing $xDigital techniques $vPeriodicals.
655 #7$aPeriodicals. $2lcgft $0https://id.loc.gov/authorities/genreForms/gf2014026139
776 0#$aDigital signal processing (Online) $c(Online) $d[Amsterdam?] : Elsevier $tDigital signal processing $x1095-4333
890 ##$aDigital signal processing.
911 ##$aMARCIVE 20230126
999 ##$a873831

```

# APPENDIX 1006E

**Appendix 1006E**  
Citations to McCowan

**Source Document:**

McCowan, I. A., Moore, D. C., & Sridharan, S. (2002, January). Near-field adaptive beamformer for robust speech recognition. *Digital Signal Processing*, vol. 12, no. 1, pp. 87-106.

**Citations: 2004 – April 2007** (in chronological order)

Li, W., Itakura, F., & Takeda, K. (2004, October). Optimizing regression for in-car speech recognition using multiple distributed microphones. In *Eighth International Conference on Spoken Language Processing*.

Moir, T. J. (2004). Analysis of a non-minimum phase acoustic beamformer, Res. Lett. Inf. Math. Sci., vol. 6, pp. 59-69.  
<http://iims.massey.ac.nz/research/letters/>

Li, W., Takeda, K., & Itakura, F. (2005, March). Adaptive log-spectral regression for in-car speech recognition using multiple distributed microphones. *IEEE Signal Processing Letters*, 12(4), 340-343.

Li, W., Shinde, T., Fujimura, H., Miyajima, C., Nishino, T., Itou, K., ... & Itakura, F. (2005, March). Multiple regression of log spectra for in-car speech recognition using multiple distributed microphones. *IEICE transactions on information and systems*, 88(3), 384-390.

Musha, T., & Kikuchi, T. (2005, October). Adaptive signal processing for reducing nearby generated flow noise. *Applied Acoustics*, 66(10), 1206-1217.

Ser, W., Chen, H., & Yu, Z. L. (2007, March). Self-calibration-based robust near-field adaptive beamforming for microphone arrays. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 54(3), 267-271.

Mustiere, F., Bouchard, M., & Bolic, M. (2007, April). Quality assessment of speech enhanced using particle filters. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* (Vol. 3, pp. III-1197). IEEE.