UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

AMAZON.COM, INC.,
Petitioner,

v.

JAWBONE INNOVATIONS, LLC,
Patent Owner.

IPR2023-00251
U.S. Patent No. 11,122,357

DECLARATION OF CAROL S. PETERSON REGARDING *BRANDSTEIN*, *GANNOT*, *GRIFFITHS-JIM*, and *MCCOWAN*

I, Carol S. Peterson, state and declare as follows:

I. Introduction

1. I am currently a Research Librarian with the law firm of Knobbe, Martens, Olson & Bear, LLP, located at 2040 Main Street, 14th Floor, Irvine, CA 92614.

2. I am over eighteen years of age. I am competent to make this Declaration, and I make this Declaration based on my own personal knowledge as well as my knowledge of library science practices.

3. I earned a Master of Library Science from the University of California, Los Angeles in 1979 and a Bachelors in English from the University of California, Davis, in 1977. I have worked as a research librarian at Knobbe, Martens, Olson & Bear for approximately 40 years.

4. I understand that a petition for *inter partes* review of U.S. Patent No. 11,122,357 will be filed concurrently with this Declaration.

5. I understand that Exhibit 1003 of the concurrently filed petition is a copy excerpts from MICROPHONE ARRAYS: SIGNAL PROCESSING TECHNIQUES AND APPLICATIONS (Michael Brandstein & Darren Ward eds., Springer 2001) ("*Brandstein*").

6. Appendix A to this Declaration is a true and correct copy of the title page, bibliographic and publication information page, preface, and excerpts from

Brandstein that was obtained from the University of California, Irvine Library. I have reviewed Exhibit 1003 of the concurrently filed petition and Appendix A, and, based on this review, it is my opinion that Exhibit 1003 is a true and correct copy of *Brandstein*.

7. Exhibit 1004 of the concurrently filed petition is a true and accurate copy of Sharon Gannot et al., *Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech*, vol. 49, no. 8 IEEE TRANSACTIONS ON SIGNAL PROCESSING, 1614 (Aug. 2001) ("*Gannot*") that was obtained from the University of California, Los Angeles Library.

8. Exhibit 1005 of the concurrently filed petition is a true and accurate copy of Lloyd Griffiths & Charles Jim, *An Alternative Approach to Linearly Constrained Adaptive Beamforming*, vol. AP-30, no. 1 IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, 27 (Jan. 1982) ("*Griffiths-Jim*") that was obtained from the University of California, Los Angeles Library.

9. I understand that Exhibit 1006 of the concurrently filed petition is copy of Iain A. McCowan et al., *Near-Field Adaptive Beamformer for Robust Speech Recognition*, vol. 12, no. 1 DIGITAL SIGNAL PROCESSING, 87 (Jan. 2002) ("*McCowan*").

10. Appendix R to this Declaration is a true and accurate copy of excerpts of vol. 12, no. 1 of Digital Signal Processing including *McCowan* that was obtained

from the Stanford University Library. I have reviewed Exhibit 1006 and Appendix R, and, based on this review, it is my opinion that Exhibit 1006 is a true and correct copy of *McCowan*.

II. Standard Library Practices for Making Materials Publicly Available

11. I have personal knowledge of standard library practices for making materials available to the public.

12. I have knowledge of and experience with the Machine-Readable Cataloging (MARC) system, an industry-wide standard that libraries use to catalog materials. Since at least 1980, MARC has been the international standard for cataloging bibliographic data in libraries' systems. Each MARC record contains fields that provide specific information about how cataloged items are held and made available to the public. MARC records can be accessed through many libraries' electronic cataloging systems.

III. MARC Records

13. The MARC record system uses a specific three-digit numeric code (from 001-999) to identify each field in a catalog record. For example:

a. field tag 008 provides bibliographic information and includes the six-digit date of when the item was cataloged in the "**YYMMDD**" format (*Date entered on file*), the four-digit year of the item's publication date;

b. field tag 022 provides the International Standard Serial Number (ISSN), which is a unique identification number assigned to serial publications;

c. field tag 245 identifies the full title statement for the work;

d. field tags 260 and 264 provide publication information that often includes the place of publication, name of publisher, publication date, and copyright date of the publication;

e. field tag 300 identifies the physical description, often including holdings information, for the work;

f. field tag 310 provides the current publication frequency of a serial;

g. field tag 362 indicates the numbering used for chronological cataloging of individual issues of a serial including information on when the serial began, which affects how issues of the serial are checked in, processed, and added to the library's main collection; and

h. field tags 9XX provide local holdings information for the resource.

14. More information about the MARC record system can be found on the Library of Congress website at: https://www.loc.gov/marc/bibliographic/bd20x24x.html.

15. After an item is received by a library, the standard practice for the library is to process, catalog, date stamp, and then shelve the item. The public may then access the item by searching the catalog and either requesting a print or electronic copy of the item. If a resource is available online or electronically, standard practice is to provide a link in the library's online catalog through which the resource can be downloaded. Standard practice is to make the item available to the public within a few days or weeks of cataloging it.

IV. Serial Publications

16. A serial publication, often known as a "journal," is a resource that is periodically published in successive parts ("issues") and has no predetermined conclusion. Each issue is usually chronologically numbered and dated. The presence of enumeration, years of coverage, and/or other chronological information also indicates a serial publication.

17. There are differences between cataloging finite resources (e.g., books) and continuing resources (e.g., serial publications). For serial publications, the MARC record provides information about the serial as a whole, including the first or earliest available issue. It also provides information about the volumes and issues

held by a library, including the dates a serial publication's issues were received by the library and the date they were made available to the public.

18. Serial publications contain unique identifying characteristics that are slightly different from non-serial publications such as textbooks. The issue date for a print serial publication, for example, generally appears on the front or back cover, the masthead page, title page (if any), table of contents page(s), or on the pages of the individual articles contained in the issue. More information regarding the unique aspects of cataloging serials can be found at:

https://www.loc.gov/aba/pcc/conser/scctppt/Basic-2014/Basic-Trainee-Mannual.pdf.

V. Brandstein

19. As detailed below, I have reviewed the University of California, Irvine Library and Library of Congress catalogs and MARC records for *Brandstein*.

A. University of California, Irvine Library Records for *Brandstein*

20. Appendix B to this Declaration is a true and correct copy of the University of California, Irvine Library public catalog record for *Brandstein*, which was downloaded from:

https://uci.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma991027958079704701&context=L&vid=01CDL_IRV_INST:UCI&lang=en.

21. Appendix B provides the public with availability and holdings information for accessing *Brandstein*, including the location, scope of holdings, and call number for locating *Brandstein*. Appendix B provides the following information for the public to access *Brandstein*:

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

Publication Date: 2001;

Publisher: New York : Springer;

ISBNs: 3540419535, 9783540419532, 3642075479, 9783642075476; and

Call Number: TK5986 .M53 2001.

Appendix B also indicates that the University of California, Irvine's copy of *Brandstein* that is held in the Science Library and is currently out on loan until December 12, 2022. Appendix B shows that three other University of California schools have copies of *Brandstein* available.

22. Appendix C to this Declaration is a true and accurate copy of the University of California, Irvine Library MARC record for its copy of *Brandstein*, which was downloaded from https://uci.primo.exlibrisgroup.com/discovery/sourceRecord?vid=01CDL_IRV_INST:UCI&docId=alma991027958079704701&recordOwner=01UCS_NETWORK.

23. The University of California, Irvine Library MARC record (Appendix C) for *Brandstein* indicates that:

a. The data elements of field tag 008 (010417s2001 nyua b 001 0 eng) provide the "date entered on file" for the record, which includes the six-digit date entry "010417," indicating that the MARC record for *Brandstein* was created on April 17, 2001, with the character position "s" followed by "2001" indicating that the item has a single known date of publication in 2001. This denotes that *Brandstein* was first cataloged by the University of California, Irvine Library on April 17, 2001. Based on standard library practices, *Brandstein* would have been processed, cataloged, shelved, and made available to the public, on an ongoing, continuing basis, beginning a few days or weeks after April 17, 2001.

b. Field tag 020 (four separate entries) provide the ISBNs for *Brandstein* (3540419535, 9783540419532, 3642075479, and 9783642075476).

c. Field tag 245 denotes the title statement of the work as Microphone arrays: signal processing techniques and

applications, and that Michael Brandstein and Darren Ward were editors.

d. Field tag 260 indicates that the work was first published or distributed in New York in 2001 and that the publisher was "Springer."

B. Library of Congress Records for *Brandstein*

24. Appendix D to this Declaration is a true and correct copy of the Library of Congress public catalog record for *Brandstein*, which was downloaded from: https://catalog.loc.gov/vwebv/holdingsInfo?&recCount=25&recPointer=0&bibId=12382050.

25. Appendix D provides the public with availability and holdings information for accessing *Brandstein*, including the location, scope of holdings, and call number for locating *Brandstein*. Appendix D provides the following information for the public to access *Brandstein*:

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

Published/Created: New York : Springer, c2001;

ISBN: 3540419535; and

Call Number: TK5986 .M53 2001.

Appendix D also indicates that *Brandstein* is available by request in the Jefferson or Adams Building Reading Rooms.

26. Appendix E to this Declaration is a true and accurate copy of the Library of Congress MARC record for its copy of *Brandstein*, which was downloaded from: https://catalog.loc.gov/vwebv/staffView?bibId=12382050.

27. The Library of Congress MARC record (Appendix E) for *Brandstein* indicates that:

a. The data elements of field tag 008 (010417s2001 nyua b 001 0 eng) provide the "date entered on file" for the record, which includes the six-digit date entry "010417," indicating that the MARC record for *Brandstein* was created on April 17, 2001, with the character position "s" followed by "2001" indicating that the item has a single known date of publication in 2001. This denotes that *Brandstein* was first cataloged by the Library of Congress on April 17, 2001. Based on standard library practices, *Brandstein* would have been processed, cataloged, shelved, and made available to the public, on an ongoing, continuing basis, beginning a few days or weeks after April 17, 2001.

b. Field tag 020 provides *Brandstein's* ISBN (3540419535).

c. Field tag 245 denotes the title statement of the work as Microphone arrays: signal processing techniques and applications, and that Michael Brandstein and Darren Ward were editors.

d. Field tag 260 indicates that the work was first published or distributed in New York in 2001 and that the publisher was "Springer.".

C. The Library Catalog and MARC Records are Consistent with Publication Information in *Brandstein*

28. Appendix A indicates that the work is titled "Microphone Arrays: Signal Processing Techniques and Applications," and that it was edited by Michael Brandstein and Darren Ward. Appendix A also provides the work's ISBN (3-540-41953-5) and indicates that it published in 2001.

29. The library catalog information and MARC records for *Brandstein*, discussed above, are consistent with and confirm the bibliographic and publication information provided in Appendix A.

30. Based on the evidence discussed above and my understanding of standard library practices, it is my opinion that *Brandstein* would have been publicly available no later than a few days or weeks after April 17, 2001.

VI. Gannot

31. As detailed below, I have reviewed the University of California, Los Angeles and Library of Congress catalogs and MARC records for "IEEE Transactions on Signal Processing," the serial publication containing *Gannot*.

A. University of California, Los Angeles Library Records for *Gannot*

32. Appendix F to this Declaration is a true and correct copy of the University of California, Los Angeles Library catalog record for IEEE Transactions on Signal Processing, which was downloaded from: https://search.library.ucla.edu/discovery/fulldisplay?docid=alma99246023606533&context=L&vid=01UCS_LAL:UCLA&lang=en.

33. Appendix F provides the public with availability and holdings information for accessing IEEE Transactions on Signal Processing, including the location, scope of holdings, and call number for locating issues of IEEE Transactions on Signal Processing. Appendix F provides the following information for the public to access IEEE Transactions on Signal Processing:

Title: IEEE transactions signal processing : a publication of the IEEE Signal Processing Society;

Publication Information: New York, NY : Institute of Electrical and Electronics Engineers Vol. 39, no. 1 (Jan. 1991);

Type: Journal;

Frequency: Monthly; and

Identifier: "ISSN: 1053-587X" and "ISSN: 1941-0476."

Appendix F also indicates that IEEE Transactions on Signal Processing is available online, and that the library may be contacted to request copies of Vol. 49, nos. 1-5 and 7-12 corresponding with call number TK5981 .l59t in the "Boelter Hall Stacks."

34. Appendix G to this Declaration is a true and accurate copy of the University of California, Los Angeles Library MARC record for its copy of IEEE Transactions on Signal Processing, which was downloaded from: https://search.library.ucla.edu/discovery/sourceRecord?vid=01UCS_LAL:UCLA&docId=alma99246023606533&recordOwner=01UCS_NETWORK.

35. The University of California, Los Angeles MARC record (Appendix G) provides the following information for IEEE Transactions on Signal Processing:

a. The data elements of field tag 008 (901025c19919999nyumr p 0 a0eng c) provide the "date entered on file" for the record, which includes the six-digit date entry "901025" that indicates that the MARC record for the serial publication was created by the University of California, Los Angeles Library no later than October 25, 1990. The character "c" followed by "19919999" indicates that the serial publication is a continuing resource that was first published in 1991 and continues to publish.

b. Field tag 022 indicates that the serial publication's ISSN is 1053-587X.

c. Field tag 245 identifies the title of the serial publication as IEEE transactions on signal processing : publication of the IEEE Signal Processing Society.

d. Field tag 260 indicates that the serial publication was first published or distributed in New York by the Institute of Electrical and Electronics Engineers, and provides publication and copyright dates of "c1991-," indicating the serial publication is still published.

e. Field tag 310 indicates that the serial publication publishes monthly.

f. Field tag 362 indicates that the serial publication began with Volume 39, no. 1, which published in January 1991, and does not contain an ending designation.

B. Library of Congress Records for *Gannot*

36. Appendix H to this Declaration is a true and correct copy of the Library of Congress record for the serial publication IEEE Transactions on Signal Processing, which was downloaded from:

https://catalog.loc.gov/vwebv/holdingsInfo?recCount=25&recPointer=1&bibId=11385724.

37. Appendix H provides the public with availability and holdings information for accessing IEEE Transactions on Signal Processing, including the location, scope of holdings, and call number for locating issues of IEEE Transactions on Signal Processing. Appendix H provides the following information for the public to access IEEE Transactions on Signal Processing:

Main Title: IEEE transactions signal processing : a publication of the IEEE Signal Processing Society;

Published/Created: New York, N.Y: Institute of Electrical and Electronics Engineers, c1991-;

Publication History: Vol. 39, no. 1 (Jan. 1991)-;

Current Frequency: Monthly; and

ISSN: 1053-587X.

Appendix H also indicates that a copy with call number TK5981 .I2 is available by request in the Jefferson or Adams Building Reading Rooms.

38. Appendix I to this Declaration is a true and accurate copy of the Library of Congress MARC record for its copy of the serial publication IEEE Transactions on Signal Processing, which was downloaded from: https://catalog.loc.gov/vwebv/staffView?bibId=11385724.

39. As explained below, the Library of Congress MARC record (Appendix I) provides the following information for the serial publication IEEE Transactions on Signal Processing:

a. The data elements of field tag 008 (901025c19919999nyumr p 0 a0eng c) provide the "date entered on file" for the record, which includes the six-digit date entry "901025" that indicates that the MARC record for the serial publication was created by the Library of Congress no later than October 25, 1990. The character "c" followed by "19919999" indicates that the serial publication is a continuing resource that was first published in 1991 and continues to publish.

b. Field tag 022 indicates that the serial publication's ISSN is 1053-587X.

c. Field tag 245 denotes the title of the serial publication as IEEE transactions signal processing : a publication of the IEEE Signal Processing Society.

d. Field tag 260 indicates that the serial publication was first published or distributed in New York, NY by the Institute of Electrical and Electronics Engineers, and provides publication

and copyright dates of "c1991-" indicating the serial continues to publish.

e. Field tag 310 indicates that the serial publication publishes issues monthly.

f. Field tag 362 indicates that the serial publication began with Volume 39, no. 1, which published in January 1991, and does not contain an ending designation.

C. The Library Catalog and MARC Records are Consistent with Publication Information in *Gannot*

40. Exhibit 1004 of the concurrently filed petition indicates that the article titled "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech" authored by Sharon Gannot et al. was published in Vol. 49, no. 8 of IEEE Transactions on Signal Processing in August 2001. Exhibit 1004 also indicates that vol. 49, no. 8 of IEEE Transactions on Signal Processing has an ISSN of 1053-587X and that it was received and date stamped by the University of California, Los Angeles on August 18, 2001, and is assigned call number TK5981 l59t.

41. The library holdings information and MARC records for IEEE Transactions on Signal Processing, which contains *Gannot*, discussed above, are consistent with and confirm the bibliographic and publication information in Exhibit 1004 of the concurrently filed petition.

42. Based on this evidence and my knowledge of standard library practices, it is my opinion that *Gannot* would have been publicly available and publicly accessible no later than a few days or weeks after August 18, 2001.

VII. Griffiths-Jim

43. As detailed below, I have reviewed the University of California, Los Angeles Library and Library of Congress catalogs and MARC records for "IEEE Transactions on Antennas and Propagation," the serial publication containing *Griffiths-Jim*.

A. University of California, Los Angeles Library Records for *Griffiths-Jim*

44. Appendix J to this Declaration is a true and correct copy of the University of California, Los Angeles Library record for IEEE Transactions on Antennas and Propagation, which was downloaded from:

https://search.library.ucla.edu/discovery/fulldisplay?docid=alma99245553606533&context=L&vid=01UCS_LAL:UCLA&lang=en.

45. Appendix J provides the public with availability and holdings information for accessing IEEE Transactions on Antennas and Propagation, including the location, scope of holdings, and call number for locating issues of the serial publication. Appendix J provides the following information for the public to access IEEE Transactions on Antennas and Propagation:

Title: IEEE Transactions on antennas and propagation;

Publication Information: New York, Institute of Electrical and Electronics Engineers v. AP-11- Jan. 1963-;

Creation Date: 1963;

Type: Journal;

Frequency: Monthly, <Jan. 1984->; and

ISSNs: 0018-926X and 1558-2221.

Appendix J also indicates that IEEE Transactions on Signal Processing is available online, and that the University of California, Los Angeles Library holds copies of volumes 11-50 (1963-2002) corresponding to call number TK7872.A6 l59ta in the "Boelter Hall Stacks."

46. Appendix K to this Declaration is a true and accurate copy of the University of California, Los Angeles Library MARC record for its copy of IEEE Transactions on Antennas and Propagation, which was downloaded from: https://search.library.ucla.edu/discovery/sourceRecord?vid=01UCS_LAL:UCLA&docId=alma99245553606533&recordOwner=01UCS_NETWORK.

47. The University of California, Los Angeles MARC record (Appendix K) provides the following information for IEEE Transactions on Antennas and Propagation:

a. The data elements of field tag 008 (751101c19639999nyumr p 0 a0eng d) provide the "date entered on file" for the record,

which includes the six-digit date entry "751101" that indicates that the MARC record for the serial publication was created by the University of California, Los Angeles Library no later than November 1, 1975. The character "c" followed by "19639999" indicates that the serial publication is a continuing resource that was first published in 1963 and continues to publish.

b. Field tag 022 indicates that the serial publication's ISSN is 0018-926X.

c. Field tag 245 denotes the title of the serial publication as IEEE transactions antennas and propagation.

d. Field tag 260 indicates that the serial publication was first published or distributed in New York by the Institute of Electrical and Electronics Engineers.

e. Field tag 310 indicates that the serial publication publishes monthly and has published monthly since January 1984.

f. Field tag 321 indicates that the serial publication formerly published bimonthly.

g. Field tag 362 indicates that the serial publication began with Volume AP-11 in January 1963 and does not contain an ending designation.

B. Library of Congress Records for *Griffiths-Jim*

48. Appendix L to this Declaration is a true and correct copy of the Library of Congress record for IEEE Transactions on Antennas and Propagation, which was downloaded from: https://catalog.loc.gov/vwebv/holdingsInfo?&bibId=11150579.

49. Appendix L provides the public with availability and holdings information for accessing IEEE Transactions on Antennas and Propagation, including the location, scope of holdings, and call number for locating issues of IEEE Transactions on Antennas and Propagation. Appendix L provides the following information for the public to access IEEE Transactions on Antennas and Propagation:

Main Title: IEEE transactions on antennas and propagation;

Published/Created: [New York], [Institute of Electrical and Electronics Engineers];

Publication History: v. AP-11- Jan. 1963-;

Current Frequency: Monthly, <Jan. 1984->;

Former Frequency: Bimonthly;

ISSN: 0018-926X.

Appendix L also indicates that the Library of Congress's copy of Volumes AP2-AP50, no. 3 of IEEE Transactions on Antennas and Propagation (call number TK7800 .I2) can be requested in the Jefferson or Adams building Reading Rooms.

50. Appendix M to this Declaration is a true and accurate copy of the Library of Congress MARC record for its copy of IEEE Transactions on Antennas and Propagation, which was downloaded from: https://catalog.loc.gov/vwebv/staffView?bibId=11150579.

51. As explained below, the Library of Congress MARC record (Appendix M) provides the following information for the serial publication IEEE Transactions on Antennas and Propagation:

a. The data elements of field tag 008 (751101c19639999nyumr p 0 a0eng d) provide the "date entered on file" for the record, which includes the six-digit date entry "751101" that indicates that the MARC record for the serial publication was created by the Library of Congress no later than November 1, 1975. The character "c" followed by "19639999" indicates that the serial publication is a continuing resource that was first published in 1963 and continues to publish.

b. Field tag 022 indicates that the serial publication's ISSN is 0018-926X.

c. Field tag 245 identifies the title of the serial publication as IEEE antennas and propagation.

d. Field tag 260 indicates that the serial publication was first published or distributed in New York by the Institute of Electrical and Electronics Engineers.

e. Field tag 310 indicates that the serial publication publishes issues monthly and has published monthly since January 1984.

f. Field tag 321 indicates that that the serial publication formerly published bimonthly.

g. Field tag 362 indicates that the serial publication began with Volume AP-11, which published in January 1963, and does not contain an ending designation.

C. The Library Catalog Records and MARC Records are Consistent with Publication Information in *Griffiths-Jim*

52. Exhibit 1005 of the concurrently filed petition indicates that the article titled "An Alternative Approach to Linearly Constrained Adaptive Beamforming" by L. J. Griffiths and C. W. Jim was published on page 27 of Volume AP-30, no. 1 of IEEE Transactions on Antennas and Propagation, which has an ISSN of 0018-926X. Exhibit 1005 further indicates that Volume AP-30, no. 1 of IEEE Transactions on Antennas and Propagation was published in January 1982. Exhibit 1005 bears a date stamp of January 20, 1982, indicating that it was received and

catalogued by the University of California, Los Angeles Library no later than January 20, 1982.

53. The MARC records for IEEE Transactions on Antennas and Propagation, which contains *Griffiths-Jim*, discussed above, are consistent with and confirm the bibliographic and publication information in Exhibit 1005.

54. Based on the evidence discussed above and my understanding of standard library practices, it is my opinion that *Griffiths-Jim* would have been publicly available no later than a few days or weeks after January 20, 1982.

VIII. McCowan

55. As detailed below, I have reviewed the Stanford University Library and Library of Congress catalogs and MARC records for "Digital Signal Processing," the serial publication containing *McCowan*.

A. Stanford University Library Records for *McCowan*

56. Appendix N to this Declaration is a true and correct copy of the University of the Stanford University Library public catalog record for Digital Signal Processing, which was downloaded from:
https://searchworks.stanford.edu/view/3471549.

57. Appendix N provides the public with availability and holdings information for accessing Digital Signal Processing, including the location, scope of

holdings, and call number for locating the serial publication. Appendix N provides the following information for the public to access Digital Signal Processing:

Key Title: Digital signal processing;

Imprint: Duluth, MN : Academic Press, c1991-;

Beginning Date: 1991;

Frequency: Quarterly;

Vol/Date Range: Vol. 1, no. 1 (Jan. 1991)-; and

ISSN: 1015-2004.

Appendix N also shows that the Stanford Library holds copies of volumes 7 through 19 of Digital Signal Processing with call numbers "TK5102.5 .D4463 V.7" through "TK50102.5 .D4463 V.19 2009" off campus at "SAL3." Appendix N indicates that the public does not have free access to SAL3 but that copies can be requested through library staff and are available.

58. Appendix O to this Declaration is a true and accurate copy of the Stanford University Library MARC record for its copy of Digital Signal Processing, which was downloaded from:

https://searchworks.stanford.edu/view/3471549.refworks_marc_txt.

59. The Stanford University Library MARC record (Appendix O) for the serial publication containing *McCowan* indicates that:

a. The data elements of field tag 008 (900620c19919999mnuqr1p 0 a0eng d) provide the "date entered on file" for the record, which includes the six-digit date entry "900620" that indicates that the MARC record for the serial publication was created by the Stanford University Library no later than June 20, 1990. The character "c" followed by "19919999" indicates that the serial publication is a continuing resource that was first published in 1991 continues to publish.

b. Field tag 022 indicates that the serial publication's ISSN is 1051-2004.

c. Field tag 245 denotes the title of the serial publication as Digital signal processing.

d. Field tag 260 indicates that the serial publication was first published or distributed in Duluth, MN by Academic Press, and provides the publication and copyright dates "c1991-" indicating that the serial publication is still published.

e. Field tag 310 indicates that the serial publication publishes quarterly.

f. Field tag 362 indicates that the serial publication began with Vol. 1, no. 1 in 1991, and does not contain an ending designation.

B. Library of Congress Records for *McCowan*

60. Appendix P to this Declaration is a true and correct copy of the Library of Congress public catalog record for Digital Signal Processing, which was downloaded from: https://catalog.loc.gov/vwebv/holdingsInfo?bibId=11390503.

61. Appendix P provides the public with availability and holdings information for accessing Digital Signal Processing, including the location, scope of holdings, and call number for locating Digital Signal Processing. Appendix P provides the following information for the public to access Digital Signal Processing:

Main Title: Digital signal processing;

Published/Created: Duluth, MN : Academic Press, c1991-;

Publication History: Vol. 1, no. 1 (Jan. 1991)-;

Current Frequency: Bimonthly, <2005->;

Former Frequency: Quarterly, <-2003>; and

ISSN: 1051-2004.

Appendix P also shows that the Library of Congress's copy of volume 3 through volume 12, no. 1 is assigned call number TK5102.5 .D4463 and is available by request in the Jefferson or Adams Building Reading Rooms.

62. Appendix Q to this Declaration is a true and accurate copy of the Library of Congress MARC record for its copy of the serial publication Digital

Signal Processing, which was downloaded from:

https://catalog.loc.gov/vwebv/staffView?bibId=11390503.

63. The Library of Congress MARC record (Appendix Q) for Digital Signal Processing indicates that:

a. The data elements of field tag 008 (900620c19919999mnubr p 0 a0eng c) provide the "date entered on file" for the record, which includes the six-digit date entry "900620" that indicates that the MARC record for the serial publication was created by the Library of Congress no later than June 20, 1990. The character "c" followed by "19919999" indicates that the serial publication is a continuing resource that was first published in 1991 continues to publish.

b. Field tag 022 indicates that the serial publication's ISSN is 1051-2004.

c. Field tag 245 denotes the title of the serial publication as Digital signal processing.

d. Field tag 260 indicates that the serial publication was first published or distributed in Duluth, MN by Academic Press, and provides the publication and copyright dates "c1991-" indicating that the serial publication is still published.

e. Field tag 310 indicates that the serial publication publishes bimonthly and has published bimonthly since 2005.

f. Field tag 321 indicates that the serial publication published quarterly through 2003.

g. Field tag 362 indicates that the serial publication began with Vol. 1, no. 1 in 1991, and does not contain an ending designation.

C. The Library Catalog Records and MARC Records are Consistent with Publication Information in *McCowan*

64. Appendix R indicates that the article titled "Near-field Adaptive Beamformer for Robust Speech Recognition" by Iain A. McCowan et al. was published on page 87 of vol. 12, no. 1 of "Digital Signal Processing" in January 2002. Exhibit 1006 also contains a date stamp of February 19, 2002, indicating that it was received and catalogued by the Library of Congress no later than February 19, 2002.

65. The library catalog and MARC records for Digital Signal Processing, which contains *McCowan*, discussed above, are consistent with and confirm the bibliographic and publication information Appendix R and Exhibit 1006.

66. Based on the evidence discussed above and my understanding of standard library practices, it is my opinion that *McCowan* would have been publicly available no later than a few days or weeks after February 19, 2002.

67. I understand that in IPR Case No. 2022-01124, Shauna Wiest similarly submitted a declaration concluding that *McCowan* was publicly available within a few days or weeks of February 19, 2002. This supports my opinion that *McCowan* would have been publicly available within a few days or weeks of February 19, 2002.

IX. Conclusion

I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Executed on November 18, 2022, in Los Alamitos, California.

Carol S. Peterson

Carol S. Peterson

EXHIBIT A

Michael Brandstein · Darren Ward (Eds.)

Microphone Arrays

Signal Processing Techniques and Applications

With 149 Figures

Springer

B3851517

Series Editors

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

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

Editors

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

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

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

Cip data applied for

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

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

© Springer-Verlag Berlin Heidelberg New York 2001

Printed in Germany

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

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

Preface

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

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

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

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

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

Part II is devoted to the source localization problem. The ability to locate and track one or more speech sources is an essential requirement of microphone array systems. For speech enhancement applications, an accurate fix on the primary talker, as well as knowledge of any interfering talkers or coherent noise sources, is necessary to effectively steer the array, enhancing a given source while simultaneously attenuating those deemed undesirable. Location data may be used as a guide for discriminating individual speakers in a multi-source scenario. With this information available, it would then be possible to automatically focus upon and follow a given source on an extended basis. Of particular interest lately, is the application of the speaker location estimates for aiming a camera or series of cameras in a video-conferencing system. In this regard, the automated localization information eliminates the need for a human or number of human camera operators. Several existing commercial products apply microphone-array technology in small-room environments to steer a robotic camera and frame active talkers. Chapter 8 summarizes the various approaches which have been explored to accurately locate an 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 multichannel 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-

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

1 Constant Directivity Beamforming

Darren B. Ward[1], Rodney A. Kennedy[2], and Robert C. Williamson[2]

[1] Imperial College of Science, Technology and Medicine, London, UK
[2] 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

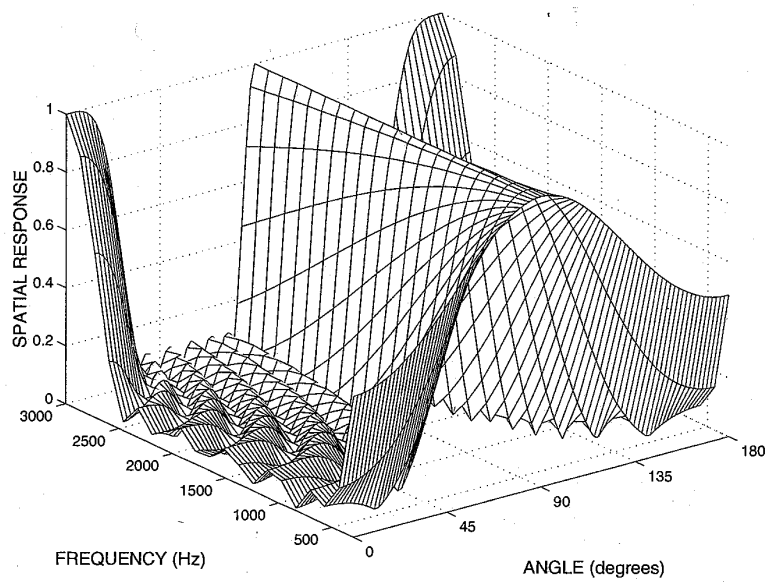


Fig. 1.1. Response of a narrowband array operated over a wide bandwidth.

(large wavelength) the same physical array appears small and the main beam will widen.

This is illustrated in Fig. 1.1 which shows the beampattern of an array designed for 1.5 kHz, but operated over a frequency range of 300 Hz to 3 kHz. If an interfering signal is present at, say, 60°, then ideally it should be attenuated completely by the array. However, because the beam is wider at low frequencies than at high frequencies, the interfering signal will be low-pass filtered rather than uniformly attenuated over its entire band. This "spectral tilt" results in a disturbing speech output if used for speech acquisition, and thus, such a narrowband array is unacceptable for speech applications. Another drawback of this narrowband design is that spatial aliasing is evident at high frequencies.[1]

To overcome this problem, one must use a beamformer that is designed specifically for broadband applications. In this chapter we focus on a specific class of broadband beamformers, called *constant directivity beamformers* (CDB), designed such that the spatial response is the same over a wide frequency band. The response of a typical CDB is shown in Fig. 1.6 on page 15.

There have been several techniques proposed to design a CDB. Most techniques are based on the idea that at different frequencies, a different array should be used that has total size and inter-sensor spacing appropriate for that particular frequency. An example of this idea is the use of harmonically-

[1] Spatial aliasing comes about if a sensor spacing wider than half a wavelength is used. It is analogous to temporal aliasing in discrete-time signal processing.

nested subarrays, e.g., [3–5]. In this case, the array is composed of a set of nested equally-spaced arrays, with each subarray being designed as a narrowband array. The outputs of the various subarrays are then combined by appropriate bandpass filtering. The idea of harmonic nesting is to reduce the beampattern variation to that which occurs within a single octave. This approach can be improved by using a set of subarray filters to interpolate to frequencies between the subarray design frequencies [6].

A novel approach to CDB design was proposed by Smith in [7]. Noting that, for a given array, the beamwidth narrows at high frequencies, Smith's idea was to form several beams and to steer each individual beam in such a way that the width of the overall multi-beam was kept constant. Thus, as the individual beams narrow at higher frequencies, they are progressively "fanned" outwards in an attempt to keep the overall beamwidth constant. Unless a very large number of beams are formed, at high frequencies this fanning will result in notches in the main beam where the progressively narrower beams no longer overlap. This approach was applied to the design of microphone arrays in [8].

The first approach to CDB design that attempted to keep a constant beampattern over the entire spatial region (not just for the main beam) was presented by Doles and Benedict [9]. Using the asymptotic theory of unequally-spaced arrays [10,11], they derived relationships between beampattern characteristics and functional requirements on sensor spacings and weightings. This results in a filter-and-sum array, with the sensor filters creating a space-tapered array: at each frequency the non-zero filter responses identify a subarray having total length and spacing appropriate for that frequency. Although this design technique results in a beampattern that is frequency-invariant over a specified frequency band, it is not a general design technique, since it is based on a specific array geometry and beampattern shape. Other recent techniques for CDB design include [12] (based on a two-dimensional Fourier transform property [13] which exists for equally-spaced arrays) and [14] (based on a beam space implementation).

Prompted by the work of Doles and Benedict, we derived in [15] a very general design method for CDB's, suitable for three-dimensional array geometries. In this chapter we outline this technique, and discuss implementation issues specific to microphone array applications.

Time-domain versus frequency-domain beamforming

There are two general methods of beamforming for broadband signals: time-domain beamforming and frequency-domain beamforming. In time-domain beamforming an FIR filter is used on each sensor, and the filter outputs summed to form the beamformer output. For an array with M sensors, each feeding a L tap filter, there are ML free parameters. In frequency-domain beamforming the signal received by each sensor is separated into narrowband frequency bins (either through bandpass filtering or data segmentation

and discrete Fourier transform), and the data in each frequency bin is processed separately using narrowband techniques. For an array with M sensors, with L frequency bins within the band of interest, there are again ML free parameters. As with most beamformers, the method that we describe in this chapter can be formulated in either domain. A time-domain formulation has previously been given in [16], and hence, we restrict our attention to frequency-domain processing here.

1.2 Problem Formulation

Consider a linear array of $M = 2N+1$ sensors located at $p_n, n = -N, \ldots, N$. Assume that the data received at the nth sensor is separated into narrowband frequency bins, each of width Δf. Let the center frequency of the ith bin be f_i, and denote the frequencies within the bin as

$$F_i = [f_i - \Delta f/2, f_i + \Delta f/2).$$

The array data received in the ith bin at time k, is given by the M-vector:

$$\boldsymbol{x}_i(k) = \boldsymbol{a}(\theta, f_i)s_i(k) + \boldsymbol{v}_i(k).$$

The desired source signal is represented by $s_i(k)$, and the M-vector $\boldsymbol{v}_i(k)$ represents the interfering noise (consisting of reverberation and other unwanted noise sources). The array vector $\boldsymbol{a}(\theta, f)$ represents the propagation of the signal source to the array, and its nth element is given by

$$a_n(\theta, f) = e^{-j2\pi f c^{-1} p_n \cos\theta},$$

where c is the speed of wave propagation, and θ is the direction to the desired source (measured relative to the array axis). To simplify notation we will drop the explicit dependence on k in the sequel.

The beamformer output is formed by applying a weight vector to the received array data, giving

$$y_i = \boldsymbol{w}_i^H \boldsymbol{x}_i, \tag{1.1}$$

where ^{H} denotes Hermitian transpose, and $\boldsymbol{w}_i$ is the M-vector of array weights to apply to the ith frequency bin.[2]

The spatial response of the beamformer is given by

$$b(\theta, f) = \boldsymbol{w}_i^H \boldsymbol{a}(\theta, f), \quad f \in F_i, \tag{1.2}$$

which defines the transfer function between a source at location $\theta \in [-\pi, \pi)$ and the beamformer output. Also of interest is the *beampattern*, defined as the squared magnitude of the spatial response.

[2] Note that it is a notational convention to use $\boldsymbol{w}^H$ rather than $\boldsymbol{w}^T$ [1].

The problem of designing a CDB can now be formulated as finding the array weights in each frequency bin such that the resulting spatial response remains constant over all frequency bins of interest.

One simple (but not very illuminating) approach to solving this problem is to perform a least-squares optimization in each frequency bin, i.e.,

$$\min_{\boldsymbol{w}_i} \int_{2\pi} |b_{\text{FI}}(\theta) - \boldsymbol{w}_i^H \boldsymbol{a}(\theta, f_i)|^2 \; d\theta, \tag{1.3}$$

where $b_{\text{FI}}(\theta)$ is the desired frequency-invariant response. Thus, in each frequency bin there are M free parameters to optimize. Although this is a standard least-squares optimization problem and the required array weights are easily found, the solution provides very little insight into the problem. Specifically, there is no suggestion of any inherent structure in the CDB, and many important questions are left unanswered, such as how many sensors are required, and what range of frequencies can be used.

In an attempt to provide some insight into the problem of designing a CDB, we take an alternative theoretical approach in the following section, and then relate these theoretical results back to the problem of finding the required filter coefficients. As we will see, there is in fact a very strong implicit structure in the CDB, and exploiting this structure enables us to reduce the number of design parameters and find efficient implementations.

1.3 Theoretical Solution

It is well known that the important dimension in determining the array response is the physical array size, measured in wavelengths. Thus, to obtain the same beampattern at different frequencies requires that the array size remains constant in terms of wavelength. Specifically, consider a linear array with N elements located at $p_n, n = 1, \ldots, N$, and assume the array weights are chosen to produce a desired beampattern $b(\theta)$ at a frequency f_1. Then, at a frequency f_2, the same beampattern $b(\theta)$ will be produced if the same array weights are used in an array with elements located at $p_n(f_1/f_2), n = 1, \ldots, N$. In other words, the size of the array must scale directly with frequency to obtain the same beampattern.[3] To obtain the same beampattern over a continuous range of frequencies would theoretically require a continuum of sensors.

1.3.1 Continuous sensor

Motivated by this interpretation, we consider the response of a theoretical continuous sensor. Assume that a signal $x(p, f)$ is received at a point p on

[3] This is precisely the idea used in the harmonically-nested subarray technique.

the sensor at frequency f, and a weight $w(p, f)$ is applied to the sensor at this point and frequency. The output of the sensor is

$$y(f) = \int w(p, f)\, x(p, f)\, dp,$$

and the spatial response for a source at angle θ is

$$b(\theta, f) = \int w(p, f)\, e^{-j2\pi f c^{-1} p \cos\theta}\, dp. \tag{1.4}$$

We assume that the aperture has finite support in p, and thus, the integration has infinite limits.

Let $u = c^{-1}\cos\theta$. The response of the continuous sensor can now be written

$$b_u(u, f) = \int w(p, f)\, e^{-j2\pi f p u}\, dp.$$

Let the sensor weighting function be given by

$$w(p, f) = fB(pf), \tag{1.5}$$

where $B(\cdot)$ is an arbitrary, absolutely-integrable, finite-support function. Substitution gives

$$b_u(u, f) = \int fB(pf)\, e^{-j2\pi f p u}\, dp. \tag{1.6}$$

With the change of variable $\zeta = pf$, and noting that $d\zeta = f dp$, it is easily seen that the resulting spatial response is now independent of frequency, i.e.,

$$b_u(u, f) = \int B(\zeta)\, e^{-j2\pi \zeta u}\, d\zeta = b_{\text{FI}}(u). \tag{1.7}$$

This is an important result, since it states that if the weighting function is given by (1.5), then the resulting spatial response will be independent of frequency. In other words, (1.5) defines the weighting function for a CDB. It was shown in [15], that not only does (1.5) provide a sufficient condition, but it is in fact the necessary condition for a frequency-invariant spatial response.

1.3.2 Beam-shaping function

Equation (1.7) defines a Fourier transform relationship between $B(\cdot)$ and $b_{\text{FI}}(\cdot)$. To achieve some desired spatial response, the required function $B(\zeta)$ is thus easily found by taking the inverse Fourier transform of $b(u)$. We will refer to $B(\cdot)$ as the *beam-shaping* (BS) function, since it has a fundamental role in determining the spatial response.

Because of its symmetry with respect to space and frequency, the BS function can be interpreted as either a filter response at a certain point, i.e., $H_p(f) = B(pf)$, or equivalently, as an aperture weighting function at a certain frequency, i.e., $A_f(p) = B(pf)$.

We will assume that the BS function is Hermitian symmetric, i.e., $B(-\zeta) = B^*(\zeta)$. This implies that the resulting spatial response is real-valued.

1.4 Practical Implementation

Whilst we have shown theoretically that it is possible to produce a beampattern that is exactly frequency-invariant using a continuous sensor, in practise we must attempt to approximate such a response using a finite array of discrete sensors. The problem of approximating a continuous aperture by a discrete array has been considered in [17]. One simple but effective technique is to approximate the integral in (1.6) using a Riemann sum—this is the approach we take here. In particular, we use trapezoidal integration to approximate the integral (1.6) by a summation of the form:

$$\hat{b}_{\mathrm{FI}}(u) = \sum_{n=-N}^{N} f\, B(p_n f)\, e^{-j2\pi f p_n u}\, \Delta_n \tag{1.8}$$

where p_n is the location of the nth discrete sensor, and $\hat{b}_{\mathrm{FI}}$ denotes an approximation of b_{FI}. We assume that the array is Hermitian symmetric about the origin, so that $B(-pf) = B(pf)^*$, and $p_{-n} = -p_n$. Although the technique is suitable for an arbitrary array geometry, a symmetric geometry simplifies implementation, and ensures that the position of the array phase center does not vary with frequency. The length of the nth subinterval is

$$\Delta_n = \frac{p_{n+1} - p_{n-1}}{2}, \tag{1.9}$$

which we refer to as the *spatial weighting term.*

Relating (1.8) to the response of a general array (1.2), we find that for a CDB the weight on the nth sensor in the ith frequency bin is

$$w_{i,n} = f_i\, \Delta_n\, B(p_n f_i), \tag{1.10}$$

where, recall, p_n is the location of the sensor, and f_i is the center frequency of the bin.

1.4.1 Dimension-reducing parameterization

Define the *reference beam-shaping filter response* as

$$H(f) = B(p_{\mathrm{ref}} f), \tag{1.11}$$

where p_{ref} is some reference location (to be defined later). Also define the *beam-shaping filter response* of the nth sensor as

$$H_n(f) = B(p_n f), \quad n = -N, \ldots, N.$$

It immediately follows that the BS filters satisfy the following dilation property:

$$H_n(f) = H(\gamma_n f), \tag{1.12}$$

where

$$\gamma_n = \frac{p_n}{p_{\text{ref}}}$$

is the dilation factor for the nth sensor. This is an extremely important property, since it shows that the filter responses on all sensors can be derived from the single filter response, $H(f)$, and enables the following efficient implementation of the CDB.

Let the reference BS filter response be given by its standard FIR filter representation:

$$H(f) = \sum_l h[l]\, e^{-j2\pi f/f_s l},$$

where f_s is the sampling frequency, and $h[l]$ is a L-vector of *beam-shaping coefficients*. From (1.12), the nth BS filter response is given by

$$\begin{aligned} H_n(f) &= \sum_l h[l]\, e^{-j2\pi f/f_s \gamma_n l} \\ &= \boldsymbol{h}^H \boldsymbol{d}_n(f), \end{aligned} \tag{1.13}$$

where $\boldsymbol{d}_n(f)$ is the L-dimensional *dilation vector* for the nth sensor. From (1.10), we see that the weight to use on the nth sensor in the ith bin is

$$w_{i,n} = \boldsymbol{h}^H \boldsymbol{t}_{i,n}, \tag{1.14}$$

where

$$\boldsymbol{t}_{i,n} = f_i \Delta_n \boldsymbol{d}_n(f_i) \tag{1.15}$$

is a L-dimensional *transformation vector*.

Equation (1.14) demonstrates the efficient parameterization afforded by this particular formulation of the CDB problem. Whereas the naive least-squares approach (1.3) requires an optimization of M parameters $\boldsymbol{w}_i$ in each frequency bin, we find that it is really only necessary to choose L frequency-independent BS parameters $\boldsymbol{h}$. Changing the beampattern shape only requires modification of these BS coefficients, and the implicit structure imposed by the transformation vectors ensures that the resulting response has constant directivity over the design band.

1.4.2 Reference beam-shaping filter

The underlying principle of the CDB is that the size and shape of the active array aperture should scale directly with frequency. This frequency scaling operation is performed by the BS filters. In deciding the coefficients of the reference BS filter, and the location of the reference point p_{ref}, we must consider this scaling property in more detail.

Let the chosen aperture size be Q wavelengths. Assuming the array is symmetric about the origin, this means that at any wavelength λ, sensors further from the origin than $Q\lambda/2$ should be inactive. In other words, the nth sensor should have a low-pass characteristic with a cutoff frequency of

$$f_n = \frac{Qc}{2|p_n|}. \tag{1.16}$$

From (1.13), note that $\gamma_n > 1$ results in compression in the frequency domain, whereas $\gamma_n < 1$ results in frequency expansion. Since the discrete-time frequency response $H(f)$ is periodic, it follows that frequency compression may cause aliasing; this is extremely undesirable. Aliasing can be avoided in one of two ways. First, choosing $p_{\text{ref}} = \max|p_n|$ ensures that $\gamma_n \leq 1, \forall n$, thus avoiding aliasing altogether—however, this requires additional constraints on the reference BS coefficients to impose the low-pass property (1.16). Alternatively, for sensors having $\gamma_n > 1$, the weights $w_{i,n}$ are set to zero for frequency bins $f_i > f_n$—the reference BS weights are now potentially unconstrained. Of these two approaches, the second is preferable, since it removes any constraints on the BS coefficients. Moreover, the requirement that the sensor weights within certain bins are always zero does not complicate implementation.

Assume that the frequency response of the reference BS filter is non-zero for all frequencies up to $f_s/2$, the Nyquist frequency; this is the most general case of $H(f)$. From (1.16), it follows that a sensor with non-zero frequency response up to $f_s/2$ would be positioned at $|p_n| = Qc/f_s$. Thus, for the most general case of $H(f)$ the reference location is chosen as

$$p_{\text{ref}} = \frac{Qc}{f_s}. \tag{1.17}$$

The reference BS coefficients can be found by using the Fourier transform relationship defined by (1.7). Specifically, the BS function $B(\zeta)$ is found by taking the Fourier transform of the desired frequency-invariant spatial response $b_{\text{FI}}(u)$. Setting $f = \zeta/p_{\text{ref}}$, $B(\zeta)$ now defines the frequency response of the reference BS filter. The BS coefficient vector $\boldsymbol{h}$ is found using any standard FIR filter design technique. In practise, low-order implementations of the reference BS filter are generally to be preferred; this point is demonstrated in the following section.

1.4.3 Sensor placement

The most common geometry for array processing applications is typically an equally-spaced array, usually with a spacing of one half-wavelength at the highest frequency of operation. Although such a geometry is valid for a CDB, less sensors are required if a logarithmically spaced array is used. In choosing an appropriate sensor geometry, the most important consideration is to ensure that at any frequency spatial aliasing is avoided.

The idea is to start with an equally-spaced array that is used at the highest frequency, and then progressively add more sensors with wider spacings as frequency decreases (and the wavelength increases). At any frequency f, the total active aperture size should be Qc/f, and the largest spacing within the active array should be $c/(2f)$. These requirements are met (using the least number of sensors) with the following symmetric array geometry:

$$p_n = n\,\frac{c}{2f_U}, \quad 0 \le n \le \frac{Q}{2} \tag{1.18a}$$

$$p_{n+1} = \frac{Q}{Q-1}\,p_n, \quad n > \frac{Q}{2},\ p_n < \frac{(Q-1)c}{2f_L} \tag{1.18b}$$

$$p_{-n} = -p_n. \tag{1.18c}$$

Note that a harmonically-nested subarray geometry is only produced if $Q = 2$.

1.4.4 Summary of implementation

1. Choose a set of L reference BS coefficients, $\boldsymbol{h}$.
2. Position the sensors according to (1.18a)–(1.18c).
3. In the ith frequency bin, the weight on the nth sensor is

$$w_{i,n} = \boldsymbol{h}^H \boldsymbol{t}_{i,n},$$

where

$$\boldsymbol{t}_{i,n} = \begin{cases} f_i \Delta_n \boldsymbol{d}_n(f_i), & f_i < f_n \\ \boldsymbol{0}, & \text{otherwise,} \end{cases}$$

$$f_n = \frac{Qc}{2|p_n|}$$

$$\Delta_n = \frac{p_{n+1} - p_{n-1}}{2}$$

$$\boldsymbol{d}_n(f_i) = \left[e^{j2\pi f/f_s \gamma_n (L-1)/2}, \ldots, e^{-j2\pi f/f_s \gamma_n (L-1)/2}\right]$$

$$\gamma_n = \frac{|p_n|}{p_{\text{ref}}}$$

$$p_{\text{ref}} = \frac{Qc}{f_s}$$

1.5 Examples

We now show an example of the CDB design technique. The design was for a bandwidth of 300–3000 Hz (i.e., the same bandwidth as used in Fig. 1.1), with an aperture size of $Q = 4$ wavelengths. Using an FFT size of 128 resulted in 44 bins within the design band, with each bin having a width of 62.5 Hz. The sensors were positioned according to (1.18a)–(1.18c), resulting in the $M = 25$ sensor array geometry shown in Fig. 1.2. For frequencies of 1000 Hz and 2000 Hz, the active sensors are also indicated in this figure.

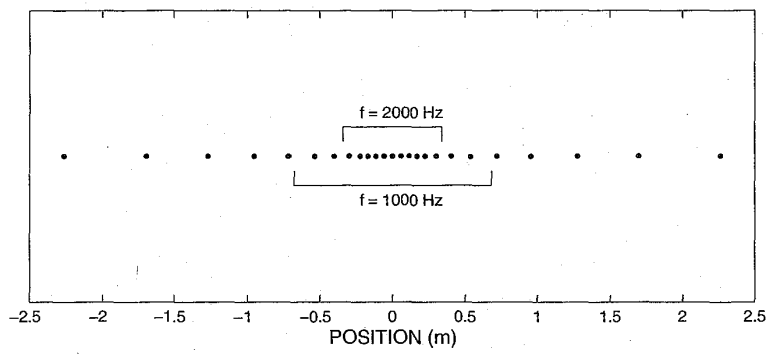


Fig. 1.2. Array geometry used for example CDB.

Assume we wish to design a standard sinc-like response (as produced by a uniformly weighted array). In this case it is known that the aperture function should be uniform. Thus, the BS function $B(\cdot)$ should ideally be a brick-wall low-pass filter. Assume we design the BS vector $\boldsymbol{h}$ to approximate an ideal low-pass filter using $L = 101$ filter coefficients. This results in the BS frequency responses shown in Fig. 1.3; for each sensor in the array, the weight required at each frequency is plotted. Note that these responses are all dilations of a single response, and that each has a low-pass characteristic.

Using these BS coefficients, the resulting spatial response of the CDB is shown in Fig. 1.4. Although the variation is not as great as for the narrow-band design in Fig. 1.1, the spatial response in Fig. 1.4 is far from frequency invariant. Why is this? The answer lies in the fact that the BS frequency response has a very sharp cutoff. Consider a single sensor. At low frequencies the sensor is always on. As frequency increases, there will come a point where the sensor will suddenly turn off, and at this frequency the aperture abruptly changes size. This abrupt change in the active aperture causes the alp-like appearance of the spatial response in Fig. 1.4.

Now, returning to the problem of designing the BS coefficients for the desired uniform spatial response, assume we design the BS vector $\boldsymbol{h}$ to ap-

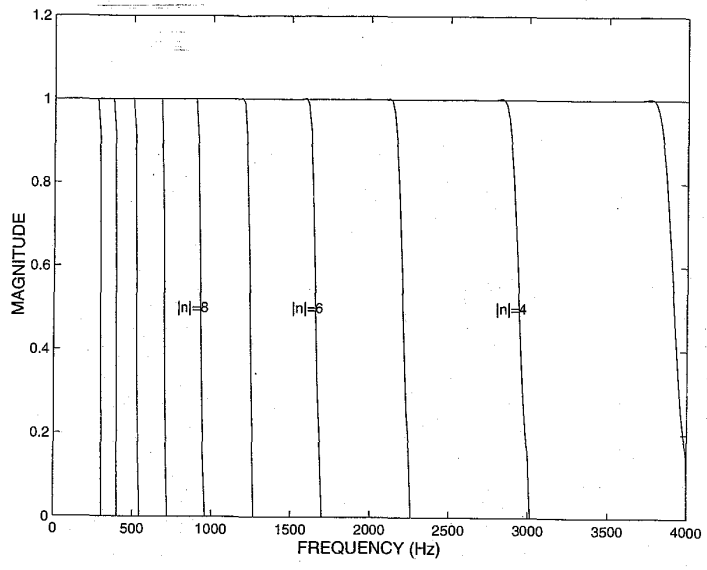


Fig. 1.3. Frequency responses of the weights on each sensor.

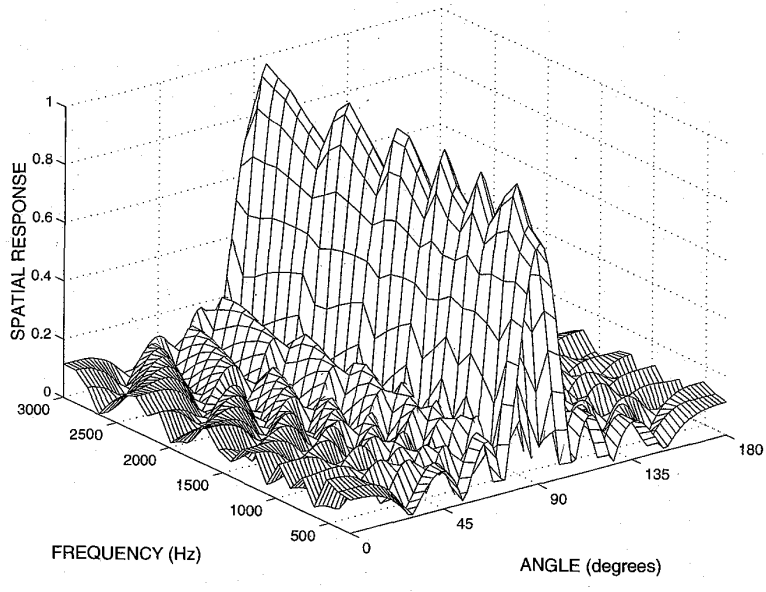


Fig. 1.4. Spatial response of example CDB.

proximate an ideal low-pass filter using only $L = 21$ filter coefficients. This results in the BS frequency responses shown in Fig. 1.5. In comparing this figure with Fig. 1.4, notice that the frequency responses exhibit a more gradual cutoff.

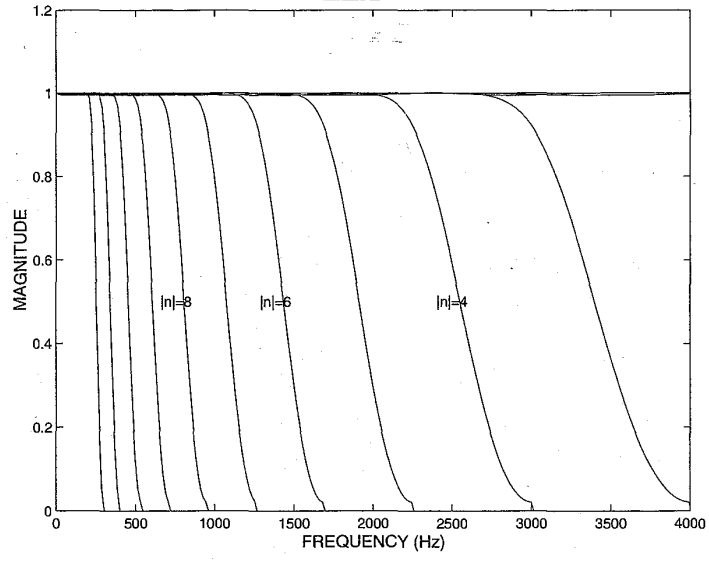


Fig. 1.5. Frequency responses of the weights on each sensor.

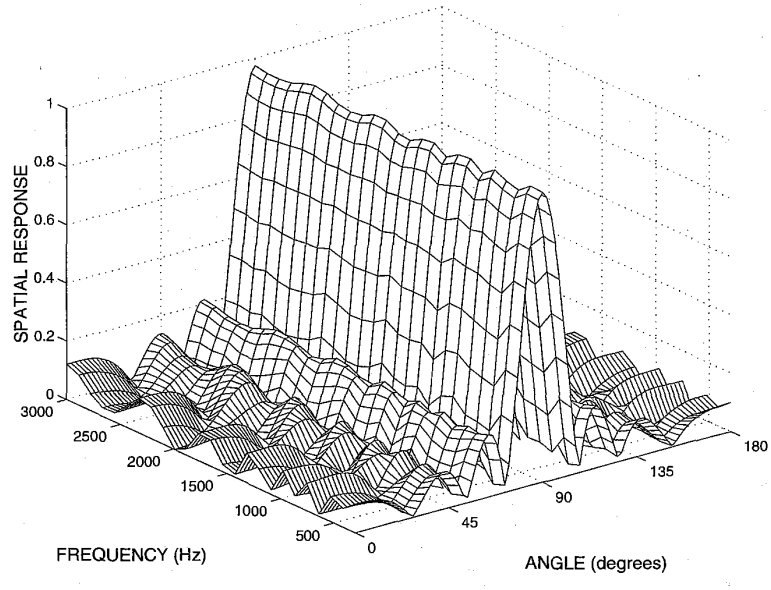


Fig. 1.6. Spatial response of example CDB.

This
s fig-
ıdual

Using these 21 BS coefficients, the resulting spatial response of the CDB is shown in Fig. 1.6. In this case the spatial response shows very little variation with frequency. This demonstrates that one should take careful consideration

of how well the underlying function can be approximated by the discrete array when choosing the required BS function.

1.6 Conclusions

Constant-directivity beamforming is a useful technique for spatial filtering in broadband signal environments in which the desired signal and the interference signals cover approximately the same bandwidth. In this chapter we have developed a technique for designing a CDB, and shown that there is an efficient parameterization and underlying structure exhibited by a CDB. The greatest drawback of a CDB in microphone array applications is that the size of the array is related to the lowest frequency of operation. Thus, producing an array that has a frequency-invariant spatial response down to, say, 300 Hz may require an array that is several meters long. In all but the largest rooms this is impractical. However, a constant spatial response can be readily achieved for mid and high frequencies (above say 1000 Hz) using an array with a total size of less than a meter. For the lower frequencies, other methods (such as the superdirective techniques described in the following chapter) are probably more appropriate.

References

1. B.D. Van Veen and K.M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
2. T.D. Abhayapala, R.A. Kennedy, and R.C. Williamson, "Nearfield broadband array design using a radially invariant modal expansion," *J. Acoust. Soc. Amer.*, vol. 107, no. 1, pp. 392–403, Jan. 2000.
3. J.L. Flanagan, D.A. Berkeley, G.W. Elko, J.E. West, and M.M. Sondhi, "Autodirective microphone systems," *Acustica*, vol. 73, pp. 58–71, 1991.
4. W. Kellermann, "A self-steering digital microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-91)*, 1991, vol. 5, pp. 3581–3584.
5. F. Khalil, J.P. Jullien, and A. Gilloire, "Microphone array for sound pickup in teleconference systems," *J. Audio Eng. Soc.*, vol. 42, no. 9, pp. 691–700, Sept. 1994.
6. J. Lardies, "Acoustic ring array with constant beamwidth over a very wide frequency range," *Acoust. Letters*, vol. 13, no. 5, pp. 77–81, 1989.
7. R. Smith, "Constant beamwidth receiving arrays for broad band sonar systems," *Acustica*, vol. 23, pp. 21–26, 1970.
8. M.M. Goodwin and G.W. Elko, "Constant beamwidth beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-93)*, 1993, vol. 1, pp. 169–172.
9. J.H. Doles III and F.D. Benedict, "Broad-band array design using the asymptotic theory of unequally spaced arrays," *IEEE Trans. Antennas Propagat.*, vol. 36, no. 1, pp. 27–33, Jan. 1988.

10. A. Ishimaru, "Theory of unequally-spaced arrays," *IRE Trans. Antennas Propagat.*, vol. AP-10, pp. 691–702, Nov. 1962.
11. A. Ishimaru and Y.S. Chen, "Thinning and broadbanding antenna arrays by unequal spacings," *IEEE Trans. Antennas Propagat.*, vol. AP-13, pp. 34–42, Jan. 1965.
12. T. Chou, "Frequency-independent beamformer with low response error," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-95)*, Detroit, USA, May 1995, pp. 2995–2998.
13. S. Haykin and J. Kesler, "Relation between the radiation pattern of an array and the two-dimensional discrete Fourier transform," *IEEE Trans. Antennas Propagat.*, vol. AP-23, no. 3, pp. 419–420, May 1975.
14. J.S. Marciano Jr. and T.B. Vu, "Reduced complexity beam space broadband frequency invariant beamforming," *Electronics Letters*, vol. 36, no. 7, pp. 682–683, Mar. 2000.
15. D.B. Ward, R.A. Kennedy, and R.C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Amer.*, vol. 97, no. 2, pp. 1023–1034, Feb. 1995.
16. D.B. Ward, R.A. Kennedy, and R.C. Williamson, "FIR filter design for frequency-invariant beamformers," *IEEE Signal Processing Lett.*, vol. 3, no. 3, pp. 69–71, Mar. 1996.
17. C. Winter, "Using continuous apertures discretely," *IEEE Trans. Antennas Propagat.*, vol. AP-25, pp. 695–700, Sept. 1977.

2 Superdirective Microphone Arrays

Joerg Bitzer[1] and K. Uwe Simmer[2]

[1] Houpert Digital Audio, Bremen, Germany
[2] 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

with respect to the given problems. Further extensions and special details are given in section 2.4. Concluding remarks close this chapter.

2.2 Evaluation of Beamformers

In order to get a better understanding of the features of the different designs of optimal beamformers, we first need to derive the measures to analyze their performance.

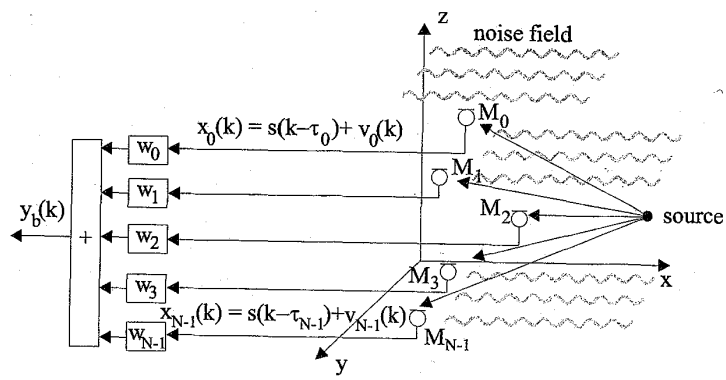


Fig. 2.1. Signal model consisting of noise field and desired source signal

The signal model is shown in Fig. 2.1. We assume that one sample of the discrete input sequence $x(k)$ at each sensor n consists of a delayed and attenuated version of the desired signal $a_i s(k-\tau_n)$ and a noise component $v_n(k)$ with arbitrary spatial statistics.

$$\begin{pmatrix} x_0(k) \\ x_1(k) \\ \vdots \\ x_{N-1}(k) \end{pmatrix} = \begin{pmatrix} a_0 s(k-\tau_0) \\ a_1 s(k-\tau_1) \\ \vdots \\ a_{N-1} s(k-\tau_{N-1}) \end{pmatrix} + \begin{pmatrix} v_0(k) \\ v_1(k) \\ \vdots \\ v_{N-1}(k) \end{pmatrix}$$

$$\boldsymbol{x}(k) = \boldsymbol{a}s(k-\boldsymbol{\tau}) + \boldsymbol{v}(k) \; . \tag{2.1}$$

Since all relevant quantities and designs depend on the frequency, the following examinations are carried out in the frequency domain without any loss of generality. The Fourier-transform leads to

$$\boldsymbol{X}(e^{j\Omega}) = S(e^{j\Omega})\boldsymbol{d} + \boldsymbol{V}(e^{j\Omega}) \; , \tag{2.2}$$

where $\boldsymbol{d}$ is the representation of the delays and the attenuation in the frequency domain which depends on the actual geometry of the array and the

direction of the source signal.

$$\boldsymbol{d}^T = [a_0 \exp(-j\Omega\tau_0), a_1 \exp(-j\Omega\tau_1), \cdots, a_{N-1} \exp(-j\Omega\tau_{N-1})] \ . \quad (2.3)$$

Finally, the output signal

$$Y_b(e^{j\Omega}) = \sum_{n=0}^{N-1} W_n{}^*(e^{j\Omega}) X_n(e^{j\Omega}) = \boldsymbol{W}^H \boldsymbol{X} \ , \quad (2.4)$$

where $W_n(e^{j\Omega})$ denotes the frequency-domain coefficients of the beamformer of sensor n at the frequency Ω and the operator ^{H} denotes a conjugated transposition (Hermitian operator). The inverse Fourier-transform results in the discrete-time output signal $y_b(k)$.

2.2.1 Array-Gain

The array-gain (AG) is the measure which shows the improvement of the signal-to-noise ratio (SNR) between one sensor and the output of the whole array [1]. Therefore,

$$G = \frac{SNR_{\text{Array}}}{SNR_{\text{Sensor}}} \ . \quad (2.5)$$

Assuming stationary signals, the SNR of one sensor is given by the ratio of the power spectral densities (PSD) of the signal Φ_{SS} and the average noise $\Phi_{V_a V_a}$.

The SNR at the output can be computed by deriving the PSD of the output signal

$$\Phi_{Y_b Y_b} = \boldsymbol{W}^H \boldsymbol{\Phi}_{\boldsymbol{XX}} \boldsymbol{W} \ , \quad (2.6)$$

where

$$\boldsymbol{\Phi}_{\boldsymbol{XX}} = \begin{pmatrix} \Phi_{X_0X_0} & \Phi_{X_0X_1} & \cdots & \Phi_{X_0X_{N-1}} \\ \Phi_{X_1X_0} & \Phi_{X_1X_1} & \cdots & \Phi_{X_1X_{N-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{X_{N-1}X_0} & \Phi_{X_{N-1}X_1} & \cdots & \Phi_{X_{N-1}X_{N-1}} \end{pmatrix} \quad (2.7)$$

is a power spectral density matrix of the array input signals. When the desired signal is present only, the output is

$$\Phi_{Y_b Y_b}\Big|_{\text{Signal}} = \Phi_{SS} \left| \boldsymbol{W}^H \boldsymbol{d} \right|^2 \ , \quad (2.8)$$

[1] The dependence on Ω is omitted for the sake of brevity and readability.

and for the noise-only case the output is

$$\Phi_{Y_bY_b}\Big|_{\text{Noise}} = \Phi_{V_aV_a}\boldsymbol{W}^H\boldsymbol{\Phi}_{\boldsymbol{VV}}\boldsymbol{W}\,, \tag{2.9}$$

where $\boldsymbol{\Phi}_{\boldsymbol{VV}}$ is a normalized cross power spectral density matrix of the noise[2]. Therefore,

$$G = \frac{\left|\boldsymbol{W}^H\boldsymbol{d}\right|^2}{\boldsymbol{W}^H\boldsymbol{\Phi}_{\boldsymbol{VV}}\boldsymbol{W}}\,. \tag{2.10}$$

Assuming a homogeneous noise field (2.10) can be expressed in terms of the coherence matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{VV}} = \begin{pmatrix} 1 & \Gamma_{V_0V_1} & \Gamma_{V_0V_2} & \cdots & \Gamma_{V_0V_{N-1}} \\ \Gamma_{V_1V_0} & 1 & \Gamma_{V_1V_2} & \cdots & \Gamma_{V_1V_{N-1}} \\ \vdots & \vdots & \ddots & \vdots & \\ \Gamma_{V_{N-1}V_0} & \Gamma_{V_{N-1}V_1} & \Gamma_{V_{N-1}V_2} & \cdots & 1 \end{pmatrix}, \tag{2.11}$$

where

$$\Gamma_{V_nV_m}(e^{j\Omega}) = \frac{\Phi_{V_nV_m}(e^{j\Omega})}{\sqrt{\Phi_{V_nV_n}(e^{j\Omega})\Phi_{V_mV_m}(e^{j\Omega})}} \tag{2.12}$$

is the coherence function [4].

Thus,

$$G = \frac{\left|\boldsymbol{W}^H\boldsymbol{d}\right|^2}{\boldsymbol{W}^H\boldsymbol{\Gamma}_{\boldsymbol{VV}}\boldsymbol{W}}\,. \tag{2.13}$$

This representation allows an easier examination of beamformers for different noise fields, since many theoretically defined noise fields can be expressed by their coherence function.

2.2.2 Beampattern

One way to evaluate beamformers is to compute the response of the array to a wavefront coming from a specific frequency and a specific angle, depending on azimuth φ and elevation θ in a spherical coordinate system. Computing this response over all angles and frequencies leads to the spatial-temporal transfer function

$$|H(\varphi,\theta)|^2\Big|_{\text{dB}} = -10\log_{10}\left(\frac{\left|\boldsymbol{W}^H\boldsymbol{d}\right|^2}{\boldsymbol{W}^H\boldsymbol{\Gamma}_{\boldsymbol{VV}}\Big|_{\text{Wavefront}}\boldsymbol{W}}\right) \tag{2.14}$$

[2] The normalization factor is set to force the trace of the matrix to equal N.

called the farfield beampattern, which is usually displayed on a logarithmic scale. It can be computed by using (2.13) and the knowledge of the coherence function of a single wavefront with frequency Ω and an angle of arrival φ, θ. Additionally, f_s denotes the sampling frequency, $c = 340$ m/s the speed of sound, and $\boldsymbol{l}_{n\,m}$ the distances between the sensors in the Cartesian coordinate system

$$\Gamma_{V_n V_m}\Big|_{\text{Wavefront}} = \exp(j\Omega\tau_{n\,m}) \ , \tag{2.15}$$

where

$$\tau_{n\,m} = \frac{f_s}{c}\left(l_{x,n\,m}\sin(\theta)\cos(\varphi) + l_{y,n\,m}\sin(\theta)\sin(\varphi) + l_{z,n\,m}\cos(\theta)\right) \ . \tag{2.16}$$

Since the beampattern depends on three variables, it is not possible to display it in a single plot. Fortunately, line arrays aligned to the z-axis have a rotational symmetry and, therefore, the beampattern is independent of φ. Examples of beampatterns for line arrays will be shown in section 2.3.

2.2.3 Directivity

A common quantity to evaluate beamformers is the directivity factor, or its logarithmic equivalent the directivity index (DI) which describes the ability of the array to suppress a diffuse noise field. Therefore, we can compute the directivity factor by using (2.13) and inserting the coherence function of a diffuse noise field:

$$\begin{aligned}\Gamma_{V_n V_m}(e^{j\Omega})\Big|_{\text{Diffuse}} &= \frac{\sin\left(\Omega f_s l_{n\,m}/c\right)}{\Omega f_s l_{n\,m}/c} \\ &= \operatorname{sinc}\left\{\frac{\Omega f_s l_{n\,m}}{c}\right\}\end{aligned} \tag{2.17}$$

where $\operatorname{sinc}(x) = \sin(x)/x$. Thus, the DI is

$$\mathrm{DI}(e^{j\Omega}) = 10\log_{10}\left(\frac{\left|\boldsymbol{W}^H\boldsymbol{d}\right|^2}{\boldsymbol{W}^H\boldsymbol{\Gamma}_{\boldsymbol{VV}}\Big|_{\text{Diffuse}}\boldsymbol{W}}\right) \ . \tag{2.18}$$

Another formal definition uses the transfer function (2.14) and describes the ratio of the transfer function of the look-direction θ_0, φ_0 of the array to the spatial integration over all directions of incoming signals.

$$\mathrm{DI}(e^{j\Omega}) = 10\log_{10}\left(\frac{\left|H(e^{j\Omega},\varphi_0,\theta_0)\right|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\left|H(e^{j\Omega},\varphi,\theta)\right|^2\sin(\theta)d\varphi d\theta}\right) \tag{2.19}$$

2.2.4 Front-to-Back Ratio

In many applications no principal look-direction exists, as for example in video-conferences or the recording of orchestras. Therefore, the DI is not the best quantity to describe the behavior of the array. In such applications a front-to-back ratio (FBR) is a better choice, since in most cases all desired sources are in front of the array and all unwanted disturbances are behind the array [19], [11]. The formal description utilizes the beampattern again:

$$\mathrm{FBR}(e^{j\Omega}) = \frac{\displaystyle\int_{\theta_0-\pi/2}^{\theta_0+\pi/2}\int_{\varphi_0-\pi/2}^{\varphi_0+\pi/2} \left|H(e^{j\Omega},\varphi,\theta)\right|^2 \sin(\theta)d\varphi d\theta}{\displaystyle\int_{\theta_0+\pi/2}^{\theta_0+3\pi/2}\int_{\varphi_0+\pi/2}^{\varphi_0+3\pi/2} \left|H(e^{j\Omega},\varphi,\theta)\right|^2 \sin(\theta)d\varphi d\theta} \tag{2.20}$$

2.2.5 White Noise Gain

This last quantity shows the ability of the array to suppress spatially uncorrelated noise, which can be caused by self-noise of the sensors. Inserting the coherence matrix for this noise field

$$\boldsymbol{\Gamma}_{\boldsymbol{VV}}\Big|_{\mathrm{uncorr}} = \boldsymbol{I} \tag{2.21}$$

into (2.13) results in the white noise gain:

$$\mathrm{WNG}(e^{j\Omega}) = \frac{\left|\boldsymbol{W}^H\boldsymbol{d}\right|^2}{\boldsymbol{W}^H\boldsymbol{W}} \,. \tag{2.22}$$

On a logarithmic scale positive values represent an attenuation of uncorrelated noise, whereas negative values show an amplification.

2.3 Design of Superdirective Beamformers

In order to design optimal beamformers, we have to minimize the power of the output signal $y_b(k)$ of the array. The output PSD is given by (2.6) and is a function of the input signal and the coefficients we want to determine. In order to avoid the trivial solution $W_n = 0$, the minimization is constrained to give an undistorted signal response in the desired look direction, i.e.,

$$\boldsymbol{W}^H\boldsymbol{d} = 1 \,. \tag{2.23}$$

Therefore, the following constrained minimization problem has to be solved:

$$\min_{\boldsymbol{W}} \boldsymbol{W}^H\boldsymbol{\Phi}_{\boldsymbol{XX}}\boldsymbol{W} \quad \text{subject to} \quad \boldsymbol{W}^H\boldsymbol{d} = 1 \,. \tag{2.24}$$

Since we are only interested in the optimal suppression of the noise, and we assume a perfect correspondence between the direction of the desired signal and the look-direction of the array, only the noise PSD-matrix $\boldsymbol{\Phi}_{\boldsymbol{VV}}$ is used.

The well-known solution for (2.24) is called the Minimum Variance Distortionless Response (MVDR) beamformer [6]. It is given by

$$\boldsymbol{W} = \frac{\boldsymbol{\Phi}_{\boldsymbol{VV}}{}^{-1}\boldsymbol{d}}{\boldsymbol{d}^H\boldsymbol{\Phi}_{\boldsymbol{VV}}{}^{-1}\boldsymbol{d}}\,, \tag{2.25}$$

and can be derived by using the Lagrange-multiplier [13] or gradient computation [20], [9]. Assuming a homogeneous noise field the solution is a function of the coherence matrix:

$$\boldsymbol{W} = \frac{\boldsymbol{\Gamma}_{\boldsymbol{VV}}{}^{-1}\boldsymbol{d}}{\boldsymbol{d}^H\boldsymbol{\Gamma}_{\boldsymbol{VV}}{}^{-1}\boldsymbol{d}}\,. \tag{2.26}$$

Equations (2.25) or (2.26) can be interpreted as a spatial decorrelation process followed by a matched filter for the desired signal. The normalization in the denominator leads to unity signal response for the look direction.

The design procedure reduces to the choice of theoretically well-defined noise-fields in order to get optimal designs for different applications. Furthermore, different models for the desired signal can be included, leading to farfield and nearfield designs.

Examples for desired signal models are:

- Standard farfield model for linear arrays with equidistant sensors:

$$\begin{aligned}\boldsymbol{d}^T = [&1, \exp(-j\Omega f_s c^{-1} l\cos(\theta_0)), \exp(-j\Omega f_s c^{-1} 2l\cos(\theta_0)),\\ &\cdots, \exp(-j\Omega f_s c^{-1}(N-1)l\cos(\theta_0))]\end{aligned} \tag{2.27}$$

  where l is the inter-sensor spacing.
- Nearfield design, including attenuation of the desired signal [14], [22]

$$\boldsymbol{d}^T = [a_0\exp(-j\omega\tau_0), a_1\exp(-j\omega\tau_1), \cdots, a_{N-1}\exp(-j\omega\tau_{N-1})]\,, \tag{2.28}$$

$$a_i = \frac{||\boldsymbol{q}-\boldsymbol{p}_{ref}||}{||\boldsymbol{q}-\boldsymbol{p}_i||}\,, \tag{2.29}$$

$$\tau_i = \frac{||\boldsymbol{q}-\boldsymbol{p}_{ref}|| - ||\boldsymbol{q}-\boldsymbol{p}_i||}{c}\,, \tag{2.30}$$

  where $||\boldsymbol{q}-\boldsymbol{p}_{ref}||$ and $||\boldsymbol{q}-\boldsymbol{p}_i||$ denote the distance between the vector location of the source q and a reference sensor p_{ref}, or the sensor p_i, respectively.

More elaborate examples for exact nearfield designs can be found in [18], [23]

2.3.1 Delay-and-Sum Beamformer

Although this chapter is called superdirective microphone arrays the well-known Delay-and-Sum Beamformer (DSB) is included for comparison purposes. It is an 'optimal' beamformer for optimizing the WNG. We can derive the coefficients from (2.26) by inserting the coherence matrix for spatial uncorrelated noise $\boldsymbol{\Gamma} = \boldsymbol{I}$. Thus,

$$\boldsymbol{W} = \frac{1}{N}\boldsymbol{d}\,. \tag{2.31}$$

The WNG is optimal in this case and reaches N. All other standard shading schemes like the Dolph-Chebycheff window [10] worsen the performance subject to WNG.

2.3.2 Design for spherical isotropic noise

In order to optimize the directivity factor, which depends on the noise-field of a spherical isotropic noise field (diffuse), we have to solve (2.26) by using the coherence matrix of the diffuse noise field, given by (2.17). The resulting coefficients represent the classic superdirective beamformer (SDB)[3].

Figure 2.2 shows the beampattern of a DSB and a superdirective beamformer, both using five linear equispaced microphones ($l = 5$ cm) in endfire steering direction ($\theta_0 = \pi$). The x-axis represents the incoming spatial angle ($[0 \cdots 2\pi]$) and the y-axes represents the frequency of the signal in kHz. The sampling-frequency was set to 8 kHz to cover the telephone bandwidth. The grey-scaled image represents the attenuation of the incoming signals in dB.

The look-direction is unmodified at all frequencies due to the linear constraint. Additionally, an unmodified region at higher frequencies can be seen caused by spatial aliasing, since our choice of the parameter does not fulfill the spatial sampling theorem, which is given by

$$l < \frac{\lambda}{2}\,, \tag{2.32}$$

where λ denotes the wavelength. The upper sampling frequency should therefore be restricted to $f_s = 6.8$ kHz, or the distance should not exceed $l = 4.25$ cm. However, in order to show some effects we will keep these parameters in all experiments.

Furthermore, the DSB is unable to suppress low frequency noise sources coming from any direction. In contrast, the superdirective beamformer attenuates very well sources coming from directions other than the look-direction

[3] In this chapter the term superdirective beamformer is used for the beamformer which optimizes the directivity factor, independent of the frequency or the ratio of the wavelength to the distance between the sensor elements. In the classic definition this is often restricted to the case where the wavelength is large with respect to the distance between sensors.

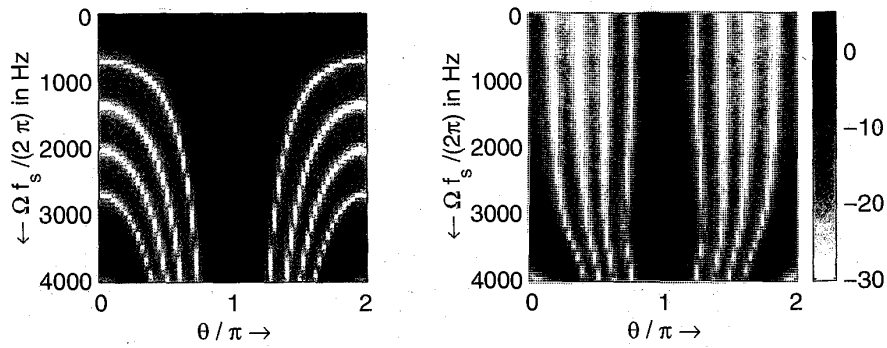


Fig. 2.2. Left: beampattern of a delay-and-sum beamformer. Right: beampattern of an optimal array for isotropic noise (superdirective beamformer). ($l = 5$ cm, $N = 5$, endfire steering direction)

over the whole frequency range. However, at higher frequencies the superdirective beamformer degrades to the DSB, since supergain can only be achieved if the signal wavelength is larger than two times the microphone distance.

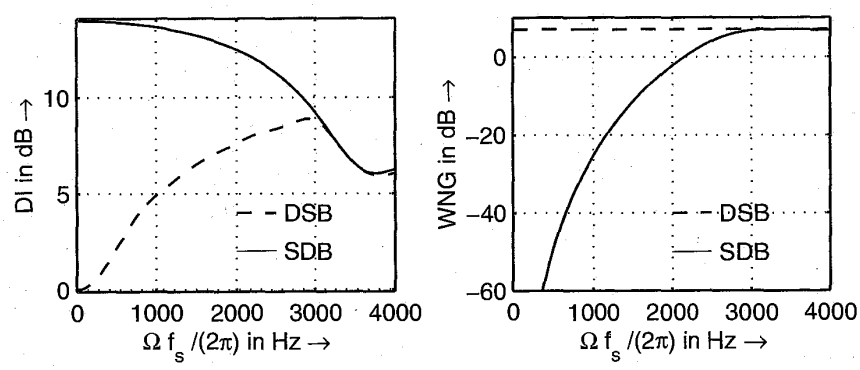


Fig. 2.3. Left: Directivity index (DI) for delay-and-sum beamformer and superdirective beamformer. Right: White noise gain (WNG) for delay-and-sum beamformer and superdirective beamformer. ($l = 5$ cm, $N = 5$, endfire steering direction)

Figure 2.3 shows the DI on the left side and the WNG on the right side for the same parameters as in the previous figure. The directivity index reaches zero at low frequencies for the DSB (as expected by analyzing the beampattern) and N^2 for the superdirective beamformer. The proof for this limit in the endfire steering case can be found in [11]. At higher frequencies the directivity for both designs is nearly the same and it is given by N, since the $\mathrm{sinc}\{\cdot\}$ function tends to zero, and the noise field is uncorrelated in this case.

If we now take a closer look at the WNG, we can see why this design is not suitable in real-world applications. Whereas the DSB suppresses uncorrelated noise equally at all frequencies, the SDB boosts uncorrelated noise at lower frequencies.

In order to give a deeper insight into how supergain works, we will compute the coefficients for an array of only two microphones. The distance is again 5 cm, and endfire steering is used.

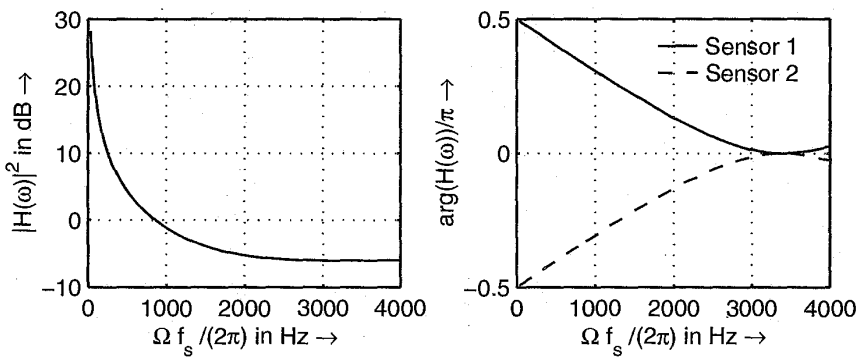


Fig. 2.4. Coefficients of a two channel SDB, left: Magnitude, right: Phase ($l = 5$ cm, $N = 2$, endfire steering direction)

In Fig. 2.4 the squared magnitude and the phase of the two coefficient vectors are shown. First of all, the coefficients are conjugate complex. Secondly, the filters force the phase between the noise components at each sensor to be π. Therefore, the correlated part of the noise will be compensated. Hence, the desired signal is also correlated, and therefore it is reduced as well. To fulfill the constraint of an undisturbed desired signal, the coefficients have to boost the input signals to compensate this behavior, which can be seen on the left side of Fig. 2.4. Therefore, uncorrelated noise will be amplified. At higher frequencies the correlation between the noise components vanishes and the beamformer degrades to the DSB. The magnitude of the coefficients reaches 1/2.

In order to overcome the problem of self-noise amplification in superdirective designs, Gilbert and Morgan have proposed a method for solving (2.24) under a WNG constraint [15]. The method uses a small added scalar μ to the main diagonal of the normalized PSD or coherence matrix:

$$\boldsymbol{W}_c = \frac{(\boldsymbol{\Gamma}_{\boldsymbol{VV}} + \mu \boldsymbol{I})^{-1}\boldsymbol{d}}{\boldsymbol{d}^H(\boldsymbol{\Gamma}_{\boldsymbol{VV}} + \mu \boldsymbol{I}^{-1}\boldsymbol{d}}. \tag{2.33}$$

We prefer a mathematically equivalent form, which preserves the interpretation as a coherence matrix with elements smaller than one. Instead of adding

the scalar to the main diagonal, we divide each non-diagonal element by $1+\mu$. Therefore, μ can be interpreted as the ratio of the sensor noise σ^2 to the ambient noise power Φ_{VV}. For the diffuse noise field the non-diagonal elements are given by

$$\Gamma_{V_n V_m} = \frac{\operatorname{sinc}\left\{\dfrac{\Omega f_s l_{nm}}{c}\right\}}{1+\dfrac{\sigma^2}{\Phi_{VV}}} \quad \forall \quad n \neq m\,. \tag{2.34}$$

The factor μ can vary from zero to infinity, which results in the unconstrained SDB or the DSB respectively. The WNG changes as a monotonic function between the two limits [15]. Typical values for μ are in the range between $-10\,\mathrm{dB}$ to $-30\,\mathrm{dB}$. Unfortunately, there is no simple relation between μ and the resulting WNG. By using a frequency variant μ the WNG can be restricted at all frequencies, but not through direct computation.

There are two different iterative design schemes. The first one was published by Doerbecker [9]. It is a straightforward implementation of a trial-and-error strategy. Another iterative design method uses the scaled projection algorithm developed by Cox *et al.* for adaptive arrays [6]. Instead of the estimated PSD-matrix, the theoretically defined coherence or PSD-matrix is inserted in the scaled projection algorithm. This solution was presented in [17]. Both algorithms result in similar coefficients and can be implemented easily.

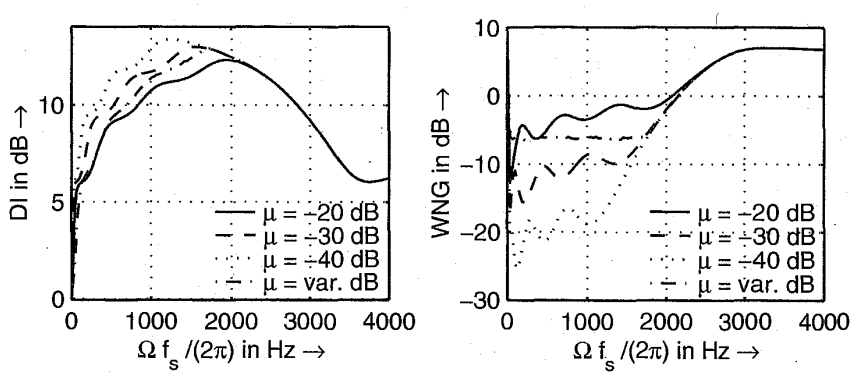


Fig. 2.5. Left: Directivity index (DI) for different constrained designs. Right: White noise gain (WNG) for different constrained designs. ($l = 5$ cm, $N = 5$, endfire steering direction)

Figure 2.5 depicts the effects for three fixed and one variable μ as constraining parameters. For the variable μ, the WNG constraint was set to $-6\,\mathrm{dB}$. The constrained design facilitates a good compromise between DI

and WNG. A careful design can optimize such arrays for a wide range of applications.

2.3.3 Design for Cylindrical Isotropic Noise

In some applications a spherical isotropic noise field is not the best choice or the best approximation of a given noise-field. Another well-defined noise-field can be used, if we reduce the three dimensions to two dimensions. We get a noise-field which is defined by infinite noise sources of a circle with an infinite radius. This kind of noise can arise if a lot of people speak in large rooms where the ceiling and the floor are damped well, or in the free-field (cocktail-party noise) [4]. The coherence between two sensors is given by [7]

$$\Gamma_{X_n X_m}(\omega) = J_0\left(\frac{\omega l_{n\,m}}{c}\right), \tag{2.35}$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. This leads to the solution of [8] as an improved design for speech enhancement for a hearing-aid application. In order to constrain the coefficients, a similar technique as in (2.34) has to be carried out.

In comparison to the design for a diffuse noise-field the differences are not large, but at lower frequencies a better suppression of noise sources behind the look direction can be observed. Elko [11] has shown that the directivity factor is less and its limit is $2N-1$, in contrast to N^2 in the unconstrained case ($\mu = 0$). A design example will be given in the next section.

2.3.4 Design for an Optimal Front-to-Back Ratio

A last data-independent design tries to optimize the front-to-back ratio. In many applications the look direction of the desired signal cannot be predetermined, but in most cases the desired signal is in front of the array and all disturbances are at the rear, e.g. when recording an orchestra or in video-conferences.

Our suggestion for a different design strategy is not to use an isotropic noise field, but to restrict the assumed infinite noise sources to the back half of a circle or a sphere.

The resulting noise-field between two sensors separated by the distance l can be described by an integration over an infinite number of uncorrelated noise sources. The resulting function in the two-dimensional case is:

$$f(e^{j\Omega}, \theta_0) = \frac{1}{\pi}\int_{\theta_0+\pi/2}^{\theta_0+3\pi/2} \exp\left(j\Omega f_s c^{-1} l\cos(\theta)\right)\, d\theta. \tag{2.36}$$

[4] The origin of this cylindrical isotropic noise-field is the sonar application in shallow water.

Using numerical integration methods, inserting the resulting complex values in the coherence matrix and solving (2.26), results in a new design which suppresses noise sources from the rear very well.

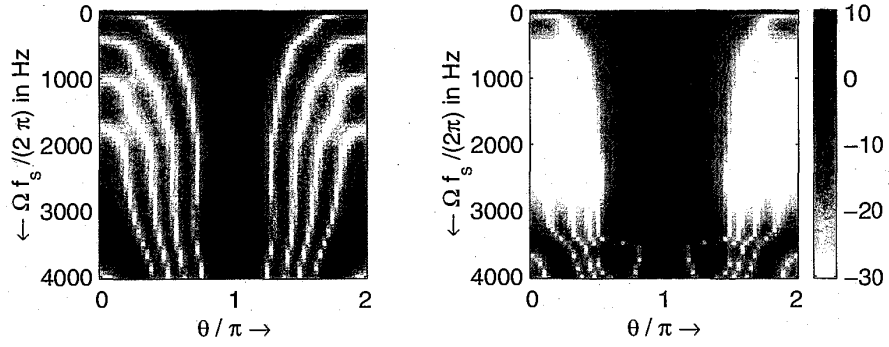


Fig. 2.6. Left: beampattern of a constrained superdirective beamformer. Right: beampattern of a constrained beamformer, designed with (2.36). ($l = 5$ cm, $N = 5$, $\mu = 0.01$, endfire steering direction)

Figure 2.6 shows beampatterns of two constrained beamformers ($\mu = 0.01$). The left side is computed with optimized coefficients for a diffuse noise-field, and the right side uses coefficients designed with the help of (2.36). At lower frequencies the constraining parameter is dominant and therefore, both designs do not perform well. From 300 Hz to 2800 Hz the new design suppresses all signals coming from the rear at the cost of a wider main lobe; this is sometimes an advantage, for example if the source is not exactly in endfire position.

At higher frequencies, especially if spatial aliasing occurs, the new design boosts signals coming from directions near the look direction, which can cause some unnatural coloring of the signal and the remaining noise. Therefore, special care has to be taken when choosing the parameters of the new design scheme.

In order to show the advantages of the new schemes, Fig. 2.7 depicts the DI and the FBR measure for the three different designs. At lower frequencies the small advantage of the cylindrical optimal design against the spherical design for the FBR can be seen, but the differences are very small over the whole frequency range. On the other hand, the behavior of the new design is completely different. Measuring the DI leads to much smaller values, but the FBR is very high, especially in the mid-frequency range.

Interestingly, we can transform between the optimal design for cylindrical isotropic noise and the new design by introducing a new variable which

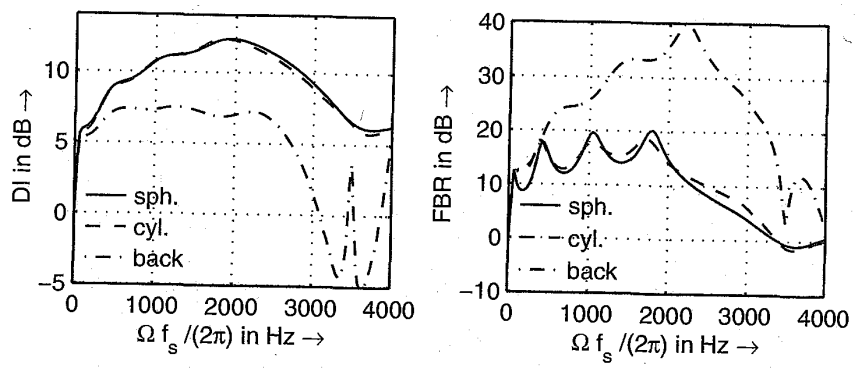


Fig. 2.7. Left: Directivity index (DI) for three optimal designs. Right: Front-to-back ratio (FBR) for three optimal designs. ($l = 5$ cm, $N = 5$, $\mu = 0.01$, endfire steering direction)

adjusts the limits of the integral, i.e.,

$$f(e^{j\Omega}, \theta_0, \delta) = \frac{1}{2(\pi - \delta)} \int_{\theta_0+\delta}^{\theta_0-\delta+2\pi} \exp\left(j\Omega f_s c^{-1} l \cos(\theta)\right) \, d\theta \quad 0 \leq \delta \leq \pi \tag{2.37}$$

Setting $\delta = 0$ corresponds to the isotropic noise case, and $\delta = \pi/2$ results in (2.36).

2.3.5 Design for Measured Noise Fields

So far, only data-independent designs have been considered. If *a priori* knowledge is available, however, it should be used to improve the performance. For example, this information could be a prescribed direction (θ = angle) of an incoming noise source. Assuming the noise source is in the far field of the microphone array, the complex coherence function between two sensors is given by

$$Re\{\Gamma_{X_n X_m}(\omega)\} = \cos\left(\frac{\Omega f_s \cos(\theta) l_{n\,m}}{c}\right) \tag{2.38}$$

$$Im\{\Gamma_{X_n X_m}(\omega)\} = -\sin\left(\frac{\Omega f_s \cos(\theta) l_{n\,m}}{c}\right) \quad . \tag{2.39}$$

Inserting the complete coherence matrix in (2.26) forms a null in that direction over the whole frequency range. In order to restrict the WNG a constrained design is necessary.

Furthermore, if we assume stationarity we can measure the actual noise-field and solve the design equation which results in the MVDR solution. Adaptive algorithms like the constrained projection by Cox [6], or the original

algorithm by Frost [13], will converge exactly to the same solution under the assumption of stationary noise and an infinitely small step-size.

2.4 Extensions and Details

After describing the main form of the MVDR beamformer and typical data-independent designs, we will compare them to their analogue counterparts, the gradient microphones. Furthermore, an alternative implementation structure will be given which can reduce the computational complexity and open superdirective designs for future extensions.

2.4.1 Alternative Form

Assuming a time-aligned input signal, the optimal weights are defined differently, since the look-direction vector $\boldsymbol{d}$ is replaced by the column-vector

$$\mathbf{1} = [\underbrace{1, 1, \cdots, 1}_{N}]^T$$

containing only ones, and the PSD-matrix or the coherence matrix contain the statistical information after time alignment (see Fig. 2.8). This gives

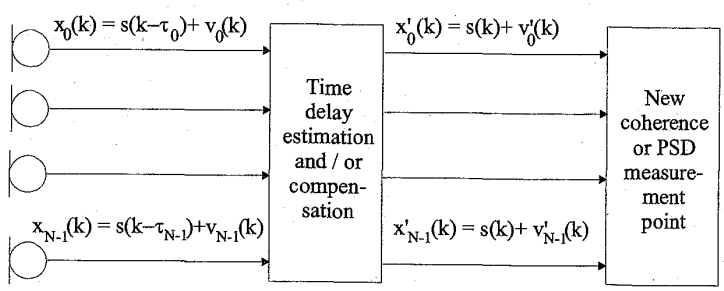


Fig. 2.8. Signal model after time delay compensation

$$\boldsymbol{W}\Big|_{\text{ta}} = \frac{\mathbf{1}^T(\boldsymbol{\Gamma}'_{\boldsymbol{VV}} + \mu \boldsymbol{I})^{-1}}{\mathbf{1}^T(\boldsymbol{\Gamma}'_{\boldsymbol{VV}} + \mu \boldsymbol{I})^{-1}\mathbf{1}}. \tag{2.40}$$

This solution of the constrained minimization problem can be decomposed into two orthogonal parts, following the ideas of Griffith and Jim [16]. One part represents the constraints only and the other part represents the unconstrained coefficients to minimize the output power of the noise.

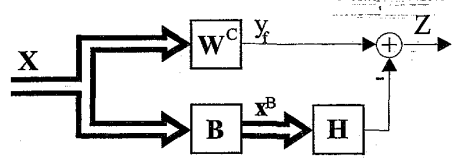


Fig. 2.9. Schematic description of the decomposition of the optimal weight vector into two orthogonal parts

The decomposed structure is depicted in Fig. 2.9. The multi-channel time-aligned input signal $\boldsymbol{X}$ is multiplied by $\boldsymbol{W}^C$ to fulfill the constraints. Furthermore, the input signal is projected onto the noise-only subspace[5] by a blocking matrix $\boldsymbol{B}$. The resulting vector $\boldsymbol{X}^B$ is multiplied by the optimal vector $\boldsymbol{H}$ and then subtracted from the output of the upper part of the structure to get the noise-reduced output signal Z. Several authors have shown the equivalence between this structure and the standard beamformer [16], [3], [12], if

$$\boldsymbol{W}^C = \frac{1}{N}\mathbf{1}\,,$$

which represents a delay-and-sum beamformer. Additionally, $\boldsymbol{B}$ has to fulfill the following properties:

- The size of the matrix is $(N-1) \times N$
- The sum of all values in one row is zero
- The matrix has to be of rank N-1.

An example for $N = 4$ is given by

$$\boldsymbol{B} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{2.41}$$

Another well-known example is the original Griffith-Jim matrix which subtracts two adjacent channels only:

$$\boldsymbol{B} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & -1 \end{pmatrix}.$$

The last step to achieve a solution equivalent to (2.25) is the computation of the optimal filter $\boldsymbol{H}$. A closer look at Fig. 2.9 shows that Y_f, $\boldsymbol{X}^B$ and Z describe exactly the problem of a multiple input noise canceler, described by

[5] Which means that the desired signal is spatially filtered out (blocked).

Widrow and Stearns [24]. Therefore, this structure is called the generalized sidelobe canceler (GSC), if an adaptive implementation is used. The non-adaptive multi-channel Wiener solution of this problem can be found in [21]

$$\boldsymbol{H} = \boldsymbol{\Phi}_{\boldsymbol{X}^B\boldsymbol{X}^B}{}^{-1}\boldsymbol{\Phi}_{\boldsymbol{X}^B Y_f} \ , \tag{2.42}$$

where $\boldsymbol{\Phi}_{\boldsymbol{X}^B\boldsymbol{X}^B}$ denotes the PSD-matrix of all signals after the matrix $\boldsymbol{B}$, and $\boldsymbol{\Phi}_{\boldsymbol{X}^B Y_f}$ is the cross-PSD vector between the fixed beamformer output and the output signals $\boldsymbol{X}^B$. Additionally, the coefficient vector can be computed as a function of the input PSD-matrix:

$$\boldsymbol{H} = (\boldsymbol{B}\boldsymbol{\Phi}_{\boldsymbol{V}\boldsymbol{V}'}\boldsymbol{B}^H)^{-1}\boldsymbol{B}\boldsymbol{\Phi}_{\boldsymbol{V}\boldsymbol{V}'}\boldsymbol{W}^C \ . \tag{2.43}$$

If we now assume a homogeneous noise field, the PSD-matrix can be replaced by the coherence matrix of the delay-compensated noise field to compute the optimal coefficients:

$$\boldsymbol{H} = (\boldsymbol{B}\boldsymbol{\Gamma}_{\boldsymbol{V}\boldsymbol{V}'}\boldsymbol{B}^H)^{-1}\boldsymbol{B}\boldsymbol{\Gamma}_{\boldsymbol{V}\boldsymbol{V}'}\boldsymbol{W}^C. \tag{2.44}$$

Therefore, all designs presented in section 2.3 can be implemented by using the GSC-structure. However, why should we do that? First of all, the new structure needs one filter less than the direct implementation. Using the first blocking matrix (2.41) further reduces the number of filters [1]. Secondly, a DSB output is available which can be used for future extensions. Thirdly, the new structure allows us to combine superdirective beamformers with adaptive post-filters for further noise reduction [2], and the new structure gives a deeper insight into MVDR-beamforming. For example, we can see that optimal beamforming is an averaging process combined with noise compensation.

2.4.2 Comparison with Gradient Microphones

Other devices with superdirectional characteristics are optimized gradient microphones [11]. In Fig. 2.10 a typical structure of a first order gradient microphone and its technical equivalent (composed of two omni-directional microphones) is shown.

The acoustic delay between the two open parts of the microphone can be realized by placing the diaphragm not exactly in the middle, or by using a material with a slower speed of sound.

The output of such systems is given by

$$E(\omega,\theta) = P_0\big(1 - \exp(-j\omega[\tau + c^{-1}l\cos(\theta)])\big) \ , \tag{2.45}$$

where τ is the acoustic delay and P_0 denotes the amplitude of the source signal. If we now assume a small spacing with respect to the wavelength, an approximate solution can be derived:

$$E(\omega,\theta) \approx P_0\omega(\tau + c^{-1}l\cos(\theta)) \ . \tag{2.46}$$

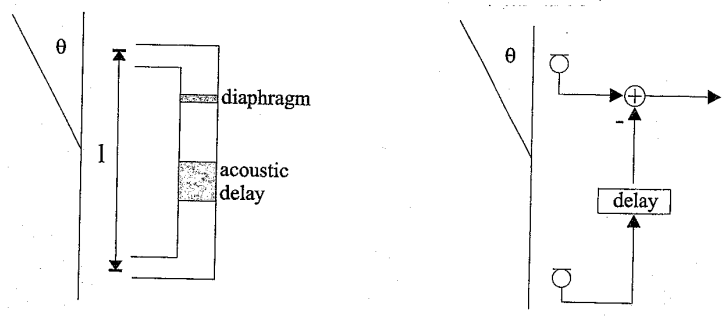


Fig. 2.10. Schematic description of a first order gradient microphone

A proper choice of τ leads to the different superdirective designs, called cardioid, supercardioid and hypercardioid. For example, the beampattern for a hypercardioid first order gradient microphone shows its zeros at $\approx \pm 109°$. This type of microphone is designed to optimize the directivity factor and therefore, it represents the analogue equivalent of a two-sensor superdirective array. For a deeper insight and a complete review of higher order gradient microphones see [11].

At lower frequencies the two systems react more or less equally. The advantages of the analogue system are the smaller size of the device, and that no analogue-to-digital conversion is necessary. The advantages of the digital array technique are its flexibility, the easy scaling for many microphones, and the possible extensions with post-filters or other adaptive techniques.

At higher frequencies, if the assumption of small spacing is not valid anymore, the differences become visible. Through careful manufacturing these frequencies are much higher than the covered bandwidth. However, at some high frequencies the analogue microphone cancels the desired signal completely. On the other hand the array system reacts like a DSB at these frequencies, and no cancellation occurs.

2.5 Conclusion

Designing a so-called superdirective array or an optimal array for theoretically well-defined noise fields can be reduced to solving a single equation. Even nearfield assumptions and measured noise fields can be easily included. We have shown that the spatial characteristic, described by the coherence function, plays a key role in designing arrays. Most of the evaluation tools like the beampattern or the directivity index are directly connected to the coherence function. Beamformer designs with optimized directivity or higher front-to-back ratio also use the coherence.

One of the new aspects included in this chapter was a new noise model to improve the front-to-back ratio. Furthermore, we emphasized the close

relationship between superdirective arrays and adaptive beamformers and their well-known implementation as a generalized sidelobe canceler.

References

1. J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "An alternative implementation of the superdirective beamformer", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio Acoust.*, pp. 7–10, New Paltz, NY, USA, Oct 1999.
2. J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multi-microphone noise reduction by post-filter and superdirective beamformer", in *Proc. Int. Workshop Acoust. Echo and Noise Control*, pp. 100–103, Pocono Manor, USA, Sep 1999.
3. K. M. Buckley, "Broad-band beamforming and the generalized sidelobe canceller", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 34, pp. 1322–1323, Oct 1986.
4. G. C. Carter, *Coherence and Time Delay Estimation*, IEEE Press, 1993.
5. H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 34, pp. 393–398, Jun 1986.
6. H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming", *IEEE Trans. Acoust. Speech Signal Processing*, vol. 35, pp. 1365–1375, Oct 1987.
7. B. F. Cron and C.H. Sherman, "Spatial-correlation functions for various noise models", *J. Acoust. Soc. Amer.*, vol. 34, pp. 1732–1736, Nov 1962.
8. M. Doerbecker, "Speech enhancement using small microphone arrays with optimized directivity", In *Proc. Int. Workshop Acoust. Echo and Noise Control*, pp. 100–103, London, UK, Sep 1997.
9. M. Doerbecker, *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen.* PhD thesis, Dept. of Telecommunications, University of TH Aachen, Verlag der Augustinus Buchhandlung, Aachen, Germany, Aug 1998.
10. C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beamwidth and sidelobe level", *Proc. IRE*, pp. 335–348, Jun 1946.
11. G. W. Elko, "Superdirectional microphone arrays", in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, eds, ch. 10, pp. 181–235, Kluwer Academic Press, 2000.
12. M.H. Er and A. Cantoni, "Transformation of linearly constrained broadband processors to unconstrained partitioned form", *IEE Proc. Pt. H*, vol. 133, pp. 209–212, June 1986.
13. O. L. Frost, "An algorithm for linearly constrained adaptive array processing", *Proc. IEEE*, vol. 60, pp. 926–935, Aug 1972.
14. J.G. Ryan and R. A. Goubran, "Optimal nearfield response for microphone arrays", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio Acoust.*, New Paltz, NY, USA, Oct 1997.
15. E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random variations", *Bell Syst. Tech. J.*, pp. 637–663, May 1955.
16. L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. Antennas Propagat.*, vol. 30, pp. 27–34, 1982.

17. J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array-processing techniques", *J. Acoust. Soc. Amer.*, vol. 99, pp. 3138–3148, May 1996.
18. R. A. Kennedy, T. Abhayapala, D. B. Ward, and R. C. Williamson, "Nearfield broadband frequency invariant beamforming", in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP-96)*, pp. 905–908, April 1996.
19. R. N. Marshall and W. R. Harry, "A new microphone providing uniform directivity over an extended frequency range", *J. Acoust. Soc. Amer.*, vol. 12, pp. 481–497, 1941.
20. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, John Wiley and Sons, New York, 1980.
21. S. Nordholm, I. Claesson, and P. Eriksson, "The broad-band Wiener solution for Griffith-Jim beamformers", *IEEE Trans. Signal Processing*, vol. 40, pp. 474–478, Feb 1992.
22. W. Taeger, "Near field superdirectivity (NFSD)", in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP-98)*, Seattle, WA, USA, 1998.
23. D. B. Ward and G. W. Elko, "Mixed nearfield/farfield beamforming: A new technique for speech acqusition in a reverberant environment", in *Proc. IEEE Workshop Applicat. Signal Processing to Audio Acoust.*, New Paltz, NY, USA, Oct 1997.
24. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, 1985.

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

is expected. Therefore, robustness against steering-vector errors caused by these array imperfections are becoming more and more important.

This chapter presents robust adaptive beamforming with the emphasis on microphone arrays as its application. In Section 2, the basics of adaptive beamformers are reviewed with the Griffiths-Jim beamformer (GJBF). Section 3 discusses robustness problems in the GJBF. Robust adaptive microphone arrays as solutions to the robustness problem are presented in Section 4. Finally in Section 5 evaluations of a robust adaptive microphone array are presented with simulation results and real-time evaluation data.

5.2 Adaptive Beamformers

A beamformer which adaptively forms its directivity pattern is called an adaptive beamformer. It simultaneously performs beam steering and null steering. In most acoustic beamformers, however, only null steering is performed with an assumption that the target DOA is known *a priori*. Due to adaptive processing, deep nulls can be developed even when errors in the propagation model exist. As a result, adaptive beamformers naturally exhibit higher interference suppression capability than its fixed counterpart. Among various adaptive beamformers, the Griffiths-Jim beamformer (GJBF) [27], or the generalized sidelobe canceler, is most widely known.

Figure 5.1 depicts the structure of the GJBF. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM). The FBF is designed to form a beam in the look direction so that the target signal is passed and all other signals are attenuated. On the contrary, the BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed through.

The simplest structure for the BM is a delay-and-subtract beamformer which was described in the previous section. Assuming a look direction perpendicular to the array surface, no delay element is necessary. Thus, a set of subtracters which take the difference between the signals at the adjacent microphones can be used as a BM. This structure is actually the one shown in Fig. 5.1. The BM was named after its function, which is to block the target signal.

The MC is composed of multiple adaptive filters each of which is driven by a BM output, $z_n(k)$ (n=0, 1, $\cdots$, $N-2$). The BM outputs, $z_n(k)$, contain all the signal components except that in the look direction. Based on these signals, the adaptive filters generate replicas of components correlated with the interferences. All the replicas are subtracted from a delayed output signal, $b(k-L_1)$,[1] of the fixed beamformer which has an enhanced target signal component. As a result, in the subtracter output $y(k)$, the target signal is

[1] The L_1-sample delay is introduced to compensate for the signal processing delay in the BM and the MC.

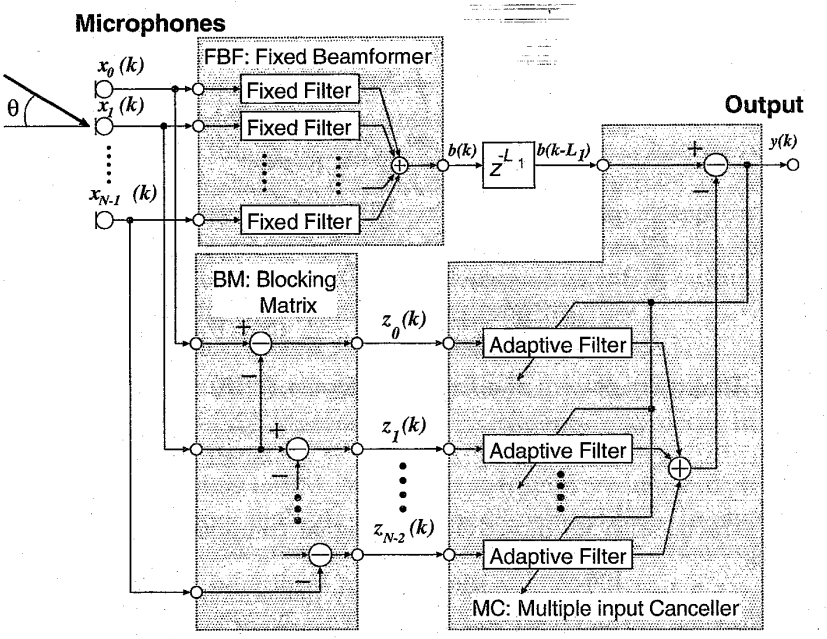


Fig. 5.1. Griffiths-Jim beamformer. It comprises a fixed beamformer (FBF), a multiple-input canceler (MC), and a blocking matrix (BM).

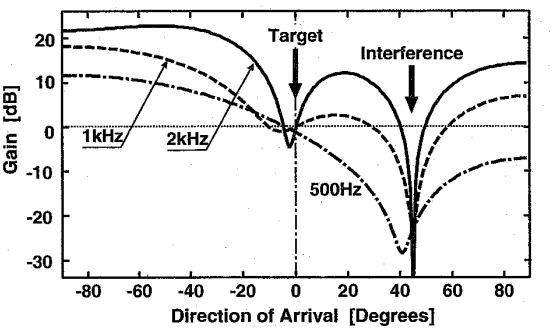


Fig. 5.2. Example directivity pattern of the Griffiths-Jim beamformer.

enhanced and undesirable signals such as ambient noise and interferences are suppressed.

The GJBF can be considered as an adaptive noise canceler with multiple reference signals, each of which is preprocessed by the BM. In an adaptive noise canceler, the auxiliary microphone is located close to the noise source to obtain a best possible noise reference. On the other hand, the BM in the

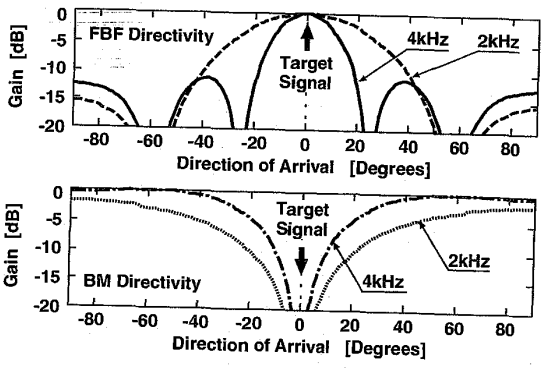


Fig. 5.3. Directivity pattern of a fixed beamformer (FBF) and a blocking matrix (BM).

GJBF extracts, with its directivity, the signal components correlated with the noise.

Figure 5.2 depicts an example directivity pattern obtained by the GJBF. In the direction of the target signal, almost constant gains close to 0 dB are obtained over a wide range of frequencies. On the contrary, in the direction of the interference, a deep null is formed. Although the directivity has frequency dependency, target signal extraction and interference suppression are simultaneously achieved.

With the same microphone array, adaptive beamformers generally achieve better interference suppression than fixed beamformers. This is because nulls are sharper than beams. The effect is demonstrated in Fig. 5.3, where directivity patterns of the FBF and the BM are illustrated. The null of the BM and the main lobe (beam) of the FBF are located in the target direction. It is also clear from the figure that they are orthogonal to each other. The BM in Fig. 5.1 has a simple delay-and-sum structure, however, a filter-and-sum beamformer [28,29] may also be employed.

5.3 Robustness Problem in the GJBF

The GJBF suffers from target-signal cancellation due to steering-vector errors, which is caused by an undesirable phase difference between $x_n(k)$ and $x_{n+1}(k)$ for the target. A phase error leads to target signal leakage into the BM output signal. As a result, blocking of the target becomes incomplete, which results in target signal cancellation at the microphone array output.

Steering-vector errors are inevitable because the propagation model does not always reflect the nonstationary physical environment. The steering vector is sensitive to errors in the microphone positions, those in the microphone characteristics, and those in the assumed target DOA (which is also known

as the look direction). For teleconferencing and hands-free communication in the car, the error in the assumed target DOA is the dominant factor.

A variety of techniques to reduce target-signal cancellation have been proposed mainly in the field of antennas and radars. The beamformers with these techniques are called robust beamformers. Typical approaches are reduction of the target-signal leakage in the BM outputs and restraint of coefficient growth in the MC. The former can be considered as a direct approach which reduces the target leakage in the BM output. The latter takes the form of an indirect approach. Even if there is target leakage in the BM output used as the MC input, the MC tries to minimize its influence.

Techniques to reduce target-signal leakage include:

- Target Tracking: The look direction is steered to the continuously estimated DOA [30]–[32]. Mistracking to interference may occur in the absence of a target signal.
- Multiple Constraints in BM: Multiple constraints are imposed on the BM so that signals from multiple DOAs are eliminated [33]. To compensate for the loss of the degrees of freedom for interference reduction with a large DOA error, additional microphones are needed.
- Constrained Gradient for Look-Direction Sensitivity: Gradient of the sensitivity at the look direction is constrained for a smaller variance of the sensitivity [34,35]. For a large error, loss in the degrees of freedom is inevitable.
- Improved Spatial Filter: A carefully designed spatial filter is used to eliminate the target signal [28]. Such a spatial filter also loses degrees of freedom.

Techniques that attempt to restraint excess coefficient growth include:

- Noise Injection: Artificially-generated noise is added to the error signal used to update the adaptive filters in the MC. This noise causes errors in the adaptive filter coefficients, preventing tap coefficients from growing excessively [36]. A higher noise level is needed to allow a larger look-direction error, resulting in less interference suppression.
- Norm Constraint: The coefficient norm of the adaptive filters in the MC is constrained by an inequality to suppress the growth of the tap coefficients [37]. In spite of its simplicity, interference reduction is degraded when the constraint is designed to allow a large error.
- Leaky Adaptive Algorithm: A leaky coefficient adaptation algorithm such as leaky LMS is used for the adaptive filters in the MC [28]. A large leakage is needed to allow a large look-direction error, leading to degraded interference-reduction.
- Adaptation Mode Control: Coefficient adaptation in the MC is controlled so that adaptation is carried out only when there is no target signal [38]. If there is no target signal when coefficients are adapted in the MC, the target leakage, if any, will have no effect on the performance of the beamformer.

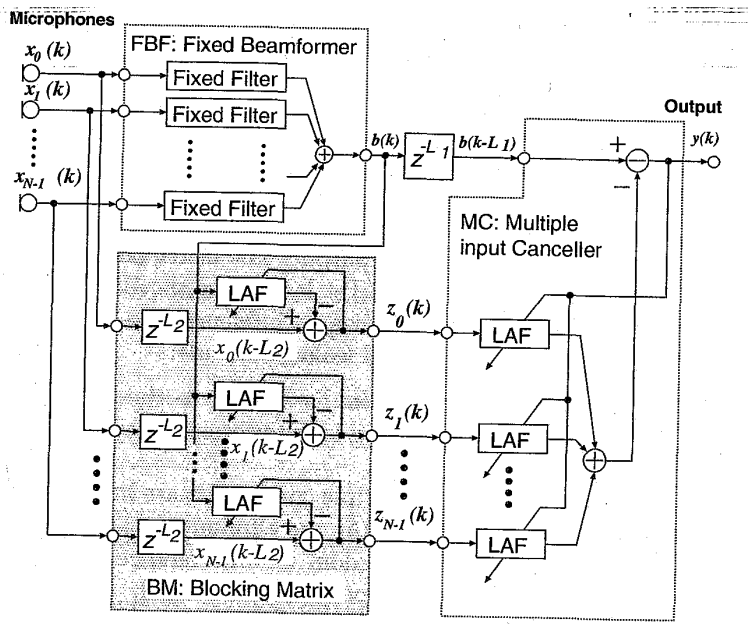


Fig. 5.4. GJBF with a LAF-LAF Structure.

These methods have been developed for a small look-direction error, typically less than 10 degrees. In the case of microphone arrays, the variance of the target DOA is typically much larger than in antennas and radar applications. No single conventional technique for robustness is sufficient for microphone arrays with a larger phase errors.

5.4 Robust Adaptive Microphone Arrays — Solutions to Steering-Vector Errors

5.4.1 LAF-LAF Structure

A target-tracking method with leaky adaptive filters (LAF) in the BM was proposed as a solution to target signal cancellation in [39]. It is combined with leaky adaptive filters in the MC [28], thereby called a LAF-LAF structure. Figure 5.4 depicts its block diagram. The leaky adaptive filters in the BM alleviate the influence of phase error, which results in the robustness. This structure can pick up a target signal with little distortion when the error between the actual and the assumed DOAs is not small. It does not need matrix products, and provides easy implementation. The nth output

$z_n(k)(n = 0, 1, \ldots, N-1)$ of the BM can be obtained as follows:

$$z_n(k) = x_n(k - L_2) - \mathbf{h}_n^T(k)\mathbf{b}(k), \tag{5.1}$$

$$\mathbf{h}_n(k) \triangleq [h_{n,0}(k), h_{n,1}(k), \ldots, h_{n,M_1-1}(k)]^T, \tag{5.2}$$

$$\mathbf{b}(k) \triangleq [b(k), b(k-1), \ldots, b(k-M_1+1)]^T, \tag{5.3}$$

where $[\cdot]^T$ denotes vector transpose and $x_n(k)$ is the nth microphone signal. L_2 is the number of delay samples for causality, $\mathbf{h}_n(k)$ is the coefficient vector of the nth LAF, and $\mathbf{b}(k)$ is the signal vector consisting of delayed signals of $b(k)$ (which is the FBF output). Each LAF is assumed to have M_1 taps. The adaptation by the normalized LMS (NLMS) algorithm [40] is described as follows:

$$\mathbf{h}_n(k+1) = \mathbf{h}_n(k) - \delta \cdot \mathbf{h}_n(k) + \alpha \frac{z_n(k)}{\mathbf{b}(k)^T\mathbf{b}(k)}\mathbf{b}(k), \tag{5.4}$$

where α is the step size for the adaptation algorithm, and $\delta, 0 \leq \delta \leq 1$, is the leakage constant.

LAFs are also used in the MC for enhancing the robustness obtained in the BM. The LAFs prevent undesirable target-signal cancellation caused by the remaining correlation with the target signal in $z_n(k)$. Tap coefficient vectors $\mathbf{w}_n(k)$ of the MC have M_2 taps and are updated by an equation similar to (5.4), where $\mathbf{h}_n$, $\mathbf{b}$, and $z_n(k)$ are replaced with $\mathbf{w}_n$, $\mathbf{z}_n$, and $y(k)$, respectively. The leakage constant δ and the step size α are replaced with γ and β respectively, and may take different values from those in (5.4).

With the LAFs in the BM, the LAF-LAF structure adaptively controls the look direction, which is fixed in the GJBF. Due to robustness by the adaptive control of the look direction, the LAF-LAF structure does not lose degrees of freedom for interference reduction. Thus, no additional microphones are required compared to the conventional robust beamformers. Target signal leakage in the BM is sufficiently small to use a minimum leakage constant γ in the MC even for a large look-direction error. Such a value of γ leads to a higher interference-reduction performance in the MC. The output of the LAFs are summed and subtracted from an L_1 sample delayed version of the FBF output to generate the microphone array output $y(k)$.

The width of the allowable DOA for the target is determined by the leaky constants and the step sizes in both the BM and the MC. Generally, smaller values of these parameters make the allowable target DOA wider. The allowable DOA width for the target is not a simple function of the parameters, however, and is not easy to prescribe. It is reported [39] that the interference is attenuated by more than 18 dB when it is designed, through simulations, to allow 20 degree directional error. Tracking may not be sufficiently precise for a large tracking range. Thus, there is a trade-off between the degree of target-signal cancellation and the amount of interference suppression.

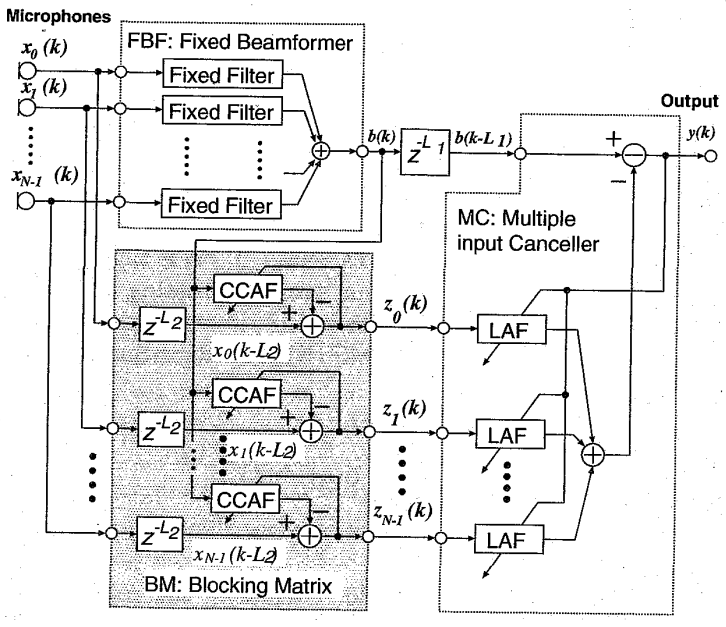


Fig. 5.5. GJBF with a CCAF-LAF Structure.

5.4.2 CCAF-LAF Structure

A more effective solution is to use coefficient-constrained adaptive filters (CCAFs) in the BM [41,42]. When combined with leaky adaptive filters in the MC as depicted in Fig. 5.5, the result is called a CCAF-LAF structure. CCAFs behave like adaptive noise cancelers. The input signal to each CCAF is the output of the FBF, and the output of the CCAF is subtracted from the delayed microphone signal. The CCAF coefficient vectors $\mathbf{h}_n(k)$ are adapted with constraints. Adaptation by the NLMS algorithm is described as follows:

$$\mathbf{h}'_n(k+1) = \mathbf{h}_n(k) + \alpha \frac{z_n(k)}{\mathbf{b}(k)^T \mathbf{b}(k)} \mathbf{b}(k), \tag{5.5}$$

$$\mathbf{h}_n(k+1) = \begin{cases} \boldsymbol{\phi}_n, & \text{for } \mathbf{h}'_n(k+1) > \boldsymbol{\phi}_n \\ \boldsymbol{\psi}_n, & \text{for } \mathbf{h}'_n(k+1) < \boldsymbol{\psi}_n \\ \mathbf{h}'_n(k+1), & \text{otherwise.} \end{cases} \tag{5.6}$$

$$\boldsymbol{\phi}_n \triangleq [\phi_{n,0}, \phi_{n,1}, \cdots, \phi_{n,M_1-1}]^T, \tag{5.7}$$

$$\boldsymbol{\psi}_n \triangleq [\psi_{n,0}, \psi_{n,1}, \cdots, \psi_{n,M_1-1}]^T, \tag{5.8}$$

where each CCAF is assumed to have M_1 taps and $\mathbf{h}'_n(k+1)$ is a temporal coefficient vector for limiting functions. $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$ are the upper and lower

bounds for coefficients. In the output signal $z_n(k)$, the components correlated with $\mathbf{b}(k)$ are cancelled by the CCAFs.

Each coefficient of the CCAFs is constrained based on the fact that filter coefficients for target-signal minimization vary significantly with the target DOA. An example of filter-coefficient variation is illustrated in Fig. 5.6. By the design of the constrained regions of the CCAF coefficients, the maximum allowable look-direction error can be specified. For example, when the CCAF coefficients are constrained in the hatched region in Fig. 5.6, up to 20° error in look direction could be allowed. Only the signal that arrives from a DOA in the limited DOA region is minimized at the outputs of the BM and remains at the output of the MC. If no interference exists in the region, which is common with microphone arrays, no mistracking occurs. For details on the design of upper and lower bounds, refer to [42].

Figure 5.7 illustrates a qualitative comparison between the LAF and the CCAF with respect to look-direction error and coefficient error from the optimum for signal blocking. Both the CCAF and the LAF give error characteristics approximating the ideal nonlinearity for target tracking. However, the coefficient error of the CCAF is a better approximation to the ideal nonlinearity than that of the LAF as shown by Fig. 5.7. The coefficient error of the CCAF becomes effective only when the look-direction error exceeds the threshold, otherwise it has no effect. On the other hand, the coefficient error of the LAF varies continuously with the look-direction error. Therefore, the CCAF leads to precise target tracking, which results in sharper spatial selectivity and less target-signal cancellation.

5.4.3 CCAF-NCAF Structure

It is possible to combine the BM with CCAFs [42] and the MC with norm-constrained adaptive filters (NCAFs) [37]. This is a CCAF-NCAF structure [43]. NCAFs subtract from $b(k-L_1)$ the components correlated with $z_n(k)$ $(n = 0, ..., N-1)$. Let M_2 be the number of taps in each NCAF, and let $\mathbf{w}_n(k)$ and $\mathbf{z}_n(k)$ be the coefficient vector and the signal vector of the nth NCAF, respectively. The signal processing in the MC is described by

$$y(k) = b(k-L_1) - \sum_{n=0}^{N-1} \mathbf{w}_n^T(k)\mathbf{z}_n(k), \tag{5.9}$$

where

$$\mathbf{w}_n(k) \triangleq [w_{n,0}(k), w_{n,1}(k), \ldots, w_{n,M_2-1}(k)]^T, \tag{5.10}$$

$$\mathbf{z}_n(k) \triangleq [z_n(k), z_n(k-1), \ldots, z_n(k-M_2+1)]^T. \tag{5.11}$$

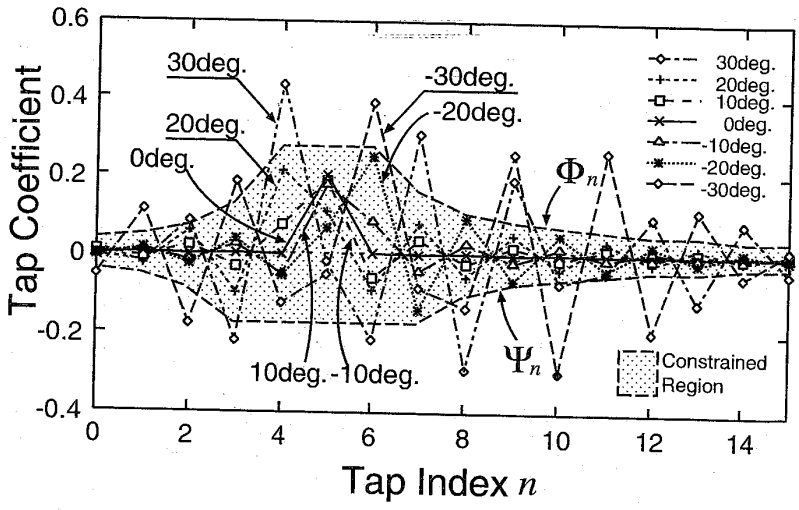


Fig. 5.6. An example of CCAF coefficients to minimize signals from different DOAs and their constraints. When the CCAF coefficients are constrained in the hatched region, up to 20° error in look direction could be allowed.

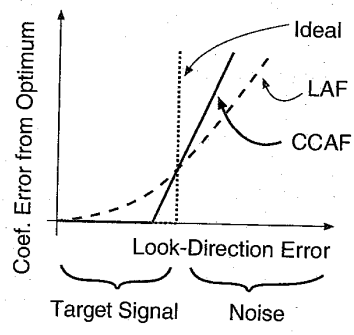


Fig. 5.7. Comparison of selectivity in LAF and CCAF.

Coefficients of the NCAFs are updated by an adaptive algorithm with a norm constraint. Adaptation with the NLMS algorithm is described as follows:

$$\mathbf{w}'_n = \mathbf{w}_n(k) + \beta \frac{y(k)}{\mathbf{z}_j(k)^T \mathbf{z}_j(k)} \mathbf{z}_n(k), \tag{5.12}$$

$$\Omega = \mathbf{w}'^T_n \mathbf{w}'_n, \tag{5.13}$$

$$\mathbf{w}_n(k+1) = \begin{cases} \sqrt{\frac{K}{\Omega}}\, \mathbf{w}'_n & \text{for } \Omega > K \\ \mathbf{w}'_n & \text{otherwise,} \end{cases} \tag{5.14}$$

where β and $\mathbf{w}'_n$ are a step size and a temporal vector for the constraint, respectively. Ω and K are the total squared-norm of $\mathbf{w}_n(k)$ and a threshold. If Ω exceeds K, $\mathbf{w}_n(k+1)$ are restrained by scaling. The norm constraint by scaling restrains excess growth of tap coefficients. The restraint inhibits the undesirable target cancellation when the target signal leaks into the NCAF inputs. If the outputs of the BM have no target signal, the MC cancels only the interference signals. In this ideal case, a norm constraint in the MC is not needed. However, complete rejection of the target signal is almost impossible in the BM, because actual environments have reflection and reverberation. To completely cancel the target signal in a reverberant environment, more than 1,000 taps are needed for each CCAF in the BM. Such a large number of taps leads to slow convergence, large misadjustment, and increased computation. Even with a high-speed processor and a fast convergence algorithm, misadjustment with the adaptive filters is inevitable. Adaptation with a low signal-to-interference ratio (SIR) causes additional misadjustment by the interference, which leads to leakage of the target signal at the BM outputs. Therefore, to avoid the target signal cancellation by leakage, a restraint with the MC such as the NCAF is essential. Because the CCAF-NCAF structure loses no degrees of freedom for interference reduction in the BM, it is robust to large look-direction errors with a small number of microphones.

5.4.4 CCAF-NCAF Structure with an AMC

Adaptations in the BM and in the MC should be performed alternately. This is because the relationship between the desired signal and the noise for the adaptation algorithm in the BM is contrary to that in the MC. For the adaptation algorithm in the BM, the target signal is the desired signal and the noise is the undesired signal. In the MC, however, the noise is the desired signal and the target signal is the undesired signal.

In the robust adaptive beamformers discussed so far, it was implicitly assumed that adaptive filters in the BM are adapted only when the target is active and those in the MC are adapted only when the target is inactive. In a real environment, however, the situation is not so simple, since incorrect adaptation of the BM may cause incomplete target blocking. As a result, the MC directivity may have a null in the direction of the target signal, resulting in target-signal cancellation. Combined with target tracking by the BM, adapting coefficients only when the target signal is absent is an effective strategy for adding robustness to adaptive beamforming [38]–[45]. In order to discriminate active and inactive periods of the target, an adaptation mode controller (AMC) is necessary.

The CCAF-NCAF structure with an AMC [46] depicted in Fig. 5.8 uses a mixed approach of the BM with CCAFs, the MC with NCAFs [37], and an AMC. A BM consisting of CCAFs provides a wider null for the target with sharper edges than leaky adaptive filters. An MC comprising NCAFs reduces

undesirable target-signal cancellation when the MC inputs have some leakage from the target signal.

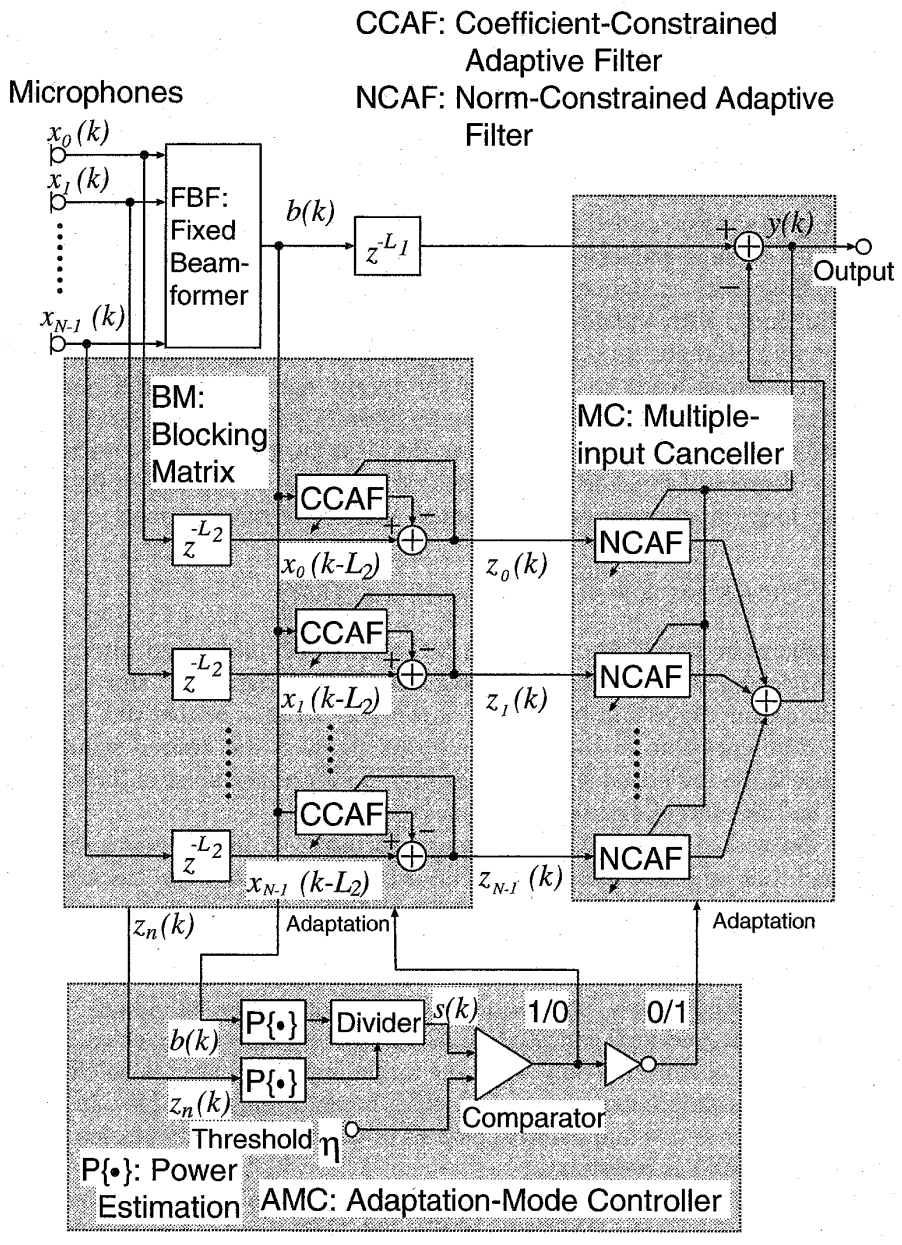


Fig. 5.8. CCAF-NCAF structure with an AMC.

The AMC controls adaptation of the BM and the MC by target-signal detection based on an estimate of the SIR [46]. The SIR is estimated as a power ratio of the output signal $b(k)$ of the FBF, to the output signal $z_n(k)$ of the BM. The main component in the FBF output is the target signal and that in the BM output is the noise. Therefore, the power ratio $s(k)$ can be considered as a direct estimate of the SIR. When the ratio is larger than a threshold η, the adaptation of the BM is performed. Otherwise, the MC is adapted.

5.5 Software Evaluation of a Robust Adaptive Microphone Array

The GJBF with CCAF-NCAF structure combined with an AMC (GJBF-CNA) was evaluated in a computer-simulated anechoic environment and in a real environment with reverberation. In the former environment, it was compared with conventional beamformers in terms of sensitivity pattern. In the latter environment, it was evaluated objectively by SIR and subjectively by mean opinion score (MOS).

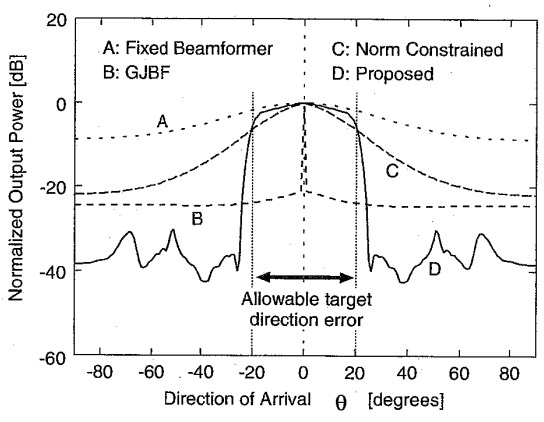


Fig. 5.9. Normalized output power after convergence as a function of DOA.

5.5.1 Simulated Anechoic Environment

A four-channel equi-spaced broadside array was used for these simulations. The spacing between microphones was 4.1 cm.The sampling rate was 8 kHz. The FBF used was a simple beamformer whose output is given by

$$b(k) = \frac{1}{N}\sum_{n=0}^{N-1} x_n(k). \tag{5.15}$$

The first simulation investigated sensitivity (after convergence) as a function of the single-sided DOA. Band-limited (0.3–3.7 kHz) Gaussian signals were used, and the assumed target direction was 0°. The maximum allowable target-direction error was 20°, unless otherwise stated. The number of coefficients for all the CCAFs and all the NCAFs was 16. The parameters were L_1=10, L_2=5, K=10.0, α=0.1, and β=0.2. The constraints of the CCAF were set based on the arrangement of the simulated array and maximum allowable target-direction errors. Total output powers after convergence, normalized by the power of the assumed target direction, are plotted in Fig. 5.9.

The plots are of the FBF (FBF), simple GJBF [27] (GJBF), norm constrained method [37] (Norm Constrained), and the GJBF-CNA (Proposed). The solid line D shows that the GJBF-CNA achieves both robustness against 20° target-direction error and high interference-reduction performance (which is 30 dB at $\theta = \pm 30°$). Similar results for a colored signal instead of the band-limited Gaussian signal have been obtained [43]. The directivity pattern of the GJBF-CNA is slightly degraded for a colored signal. However, the degradation by the norm-constrained method is more serious. This fact shows that the GJBF-CNA exhibits robustness to the power spectrum of input signal.

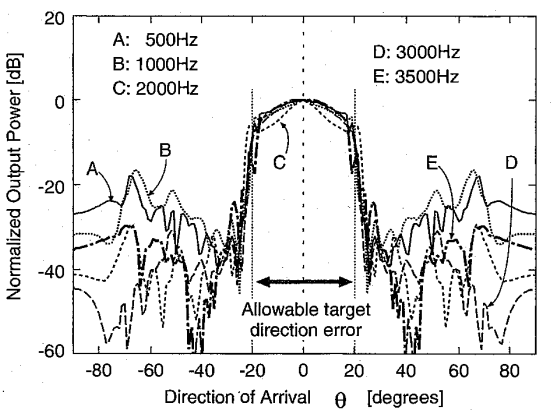


Fig. 5.10. Sensitivities after convergence as a function of DOA at different frequencies.

Frequency dependency of the directivity pattern is shown in Fig. 5.10. In this figure, sensitivities to the frequency component of the target signal are plotted. Frequency dependency of the GJBF-CNA is small, and thus, the GJBF-CNA is suitable for broadband applications such as microphone arrays. The widths of the high-sensitivity regions are almost the same as the allowable target-direction error ($-20° < \theta < 20°$) and the sensitivity in the region is constant.

In the second simulation, sensitivities for different SIRs were investigated. The simulation was performed with amplitude control that was similar to a realistic scenario. A target signal source generated a band-limited white Gaussian signal for the first 50,000 iterations and then stopped. This is a simple simulation of burst characteristics like speech. Another bandlimited white Gaussian signal, which imitates an interference like airconditioner noise, existed throughout the simulation. The SIR is defined as a power ratio of the two signals. The target signal source was placed about 10° off the assumed target DOA and the DOA of the interfering signal source was scanned.

Figure 5.11 shows normalized output power after convergence as a function of interference DOA. Lines G and H have a sharp peak at $\theta = 10°$, which indicates that the target-signal at the output of the BM is sufficiently minimized for the overall robustness. Therefore, when SIR is higher than about 10dB (which is lower than a typical SIR value expected in teleconference) the interference is suppressed even if it arrives from a direction in the allowable target DOA region. When the interference comes from outside the allowable target DOA region, even an SIR of 0 dB causes almost no problem in the GJBF-CNA.

Finally, Fig. 5.12 shows the total output powers for various coefficient constraints with the CCAFs. The signal was bandlimited white Gaussian noise. The allowable target-direction errors are approximately 4, 6, 9, 12, 16, and 20 degrees. These lines demonstrate that the allowable target-direction error can be specified by the user.

5.5.2 Reverberant Environment

Simulations with real sound data captured in a reverberant environment were also performed. The data were recorded with a broad-side linear array. Four omni-directional microphones without calibration were mounted on a universal printed circuit board with an equal spacing of 4.1 cm. The signal of each microphone was bandlimited between 0.3 and 3.4 kHz and sampled at 8 kHz. The number of taps was 16 for both the CCAFs and NCAFs.

Figure 5.13 illustrates the arrangement for sound-data acquisition. The target source was located in front of the array at a distance of 2.0 m. A white noise source was placed about $\theta = 45°$ off the target DOA at a distance of 2.0 m. The reverberation time of the room was about 0.3 second, which is common with actual small offices. All the parameters except the step-sizes were the same as those in the previous subsection. The target source was an English male speech signal.

Objective Evaluation

Output powers for the FBF, the GJBF [27] (GJBF), and the norm-constrained method [37] (Norm Constrained) after convergence are shown in Fig. 5.14. The step-size α for the CCAFs was 0.02 and β for the NCAFs was 0.004.

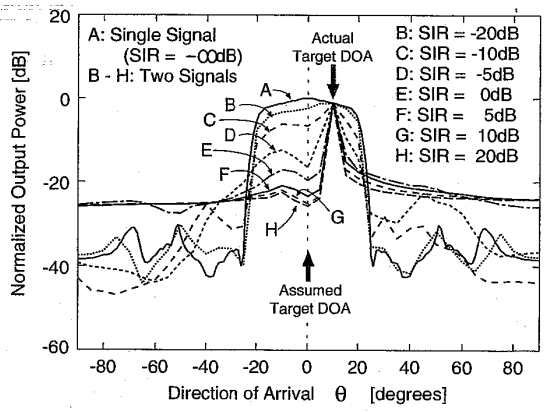


Fig. 5.11. Normalized output power after convergence as a function of DOA with different SIRs.

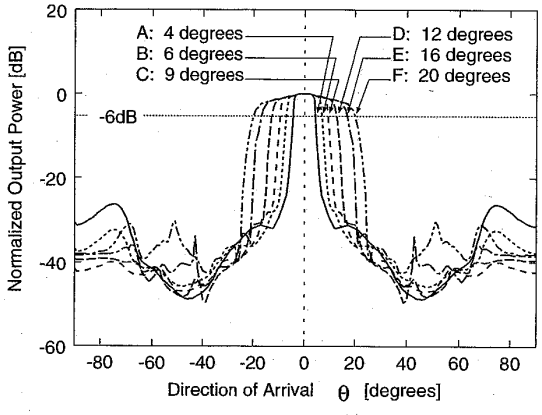


Fig. 5.12. Normalized output power after convergence for different allowable target directions.

These step-sizes were selected so that breathing noise and cancellation of the target signal are sufficiently small subjectively. All other parameters were selected based on the microphone arrangement. If there is any difference between trajectory A and any of B, C, D, E, or F when the voice is active (sample index from 1,720,000 to 1,740,000), the target signal corresponding to the trajectory is partially cancelled. The FBF (B) causes almost no target-signal cancellation. With the GJBF (C), cancellation of the target signal is serious. With the the norm-constrained method (D), and the GJBF-CNA (E), the cancellation of target signal was 2dB, which is subjectively small.

The output powers during voice absence (after sample index 1,760,000) indicate the interference-reduction ratio (IRR). The IRR of the FBF is 3dB,

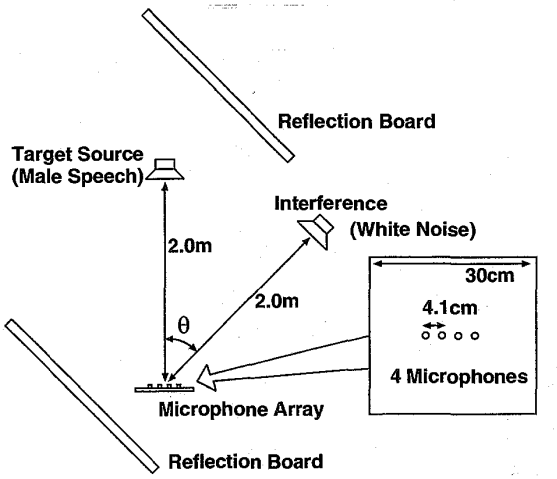


Fig. 5.13. Experimental set-up.

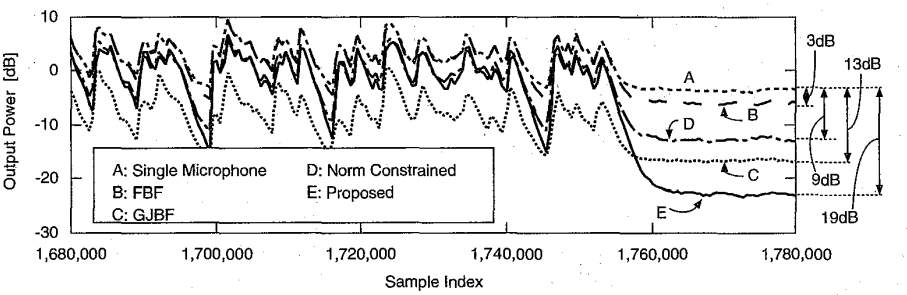


Fig. 5.14. Output Powers for a male speech signal and white noise.

and that of the norm-constrained method is 9dB. On the other hand, with the GJBF-CNA (F), the IRR is as much as 19dB.

Subjective Evaluation

MOS evaluation by 10 nonprofessional subjects was performed based on [47]. As anchors, the signal recorded by a single microphone was used for grade 1 and the original male speech without interference for grade 5. Subjects were instructed that target-signal cancellation should obtain a low score.

Evaluation results are shown in Fig. 5.15. The thick horizontal line on each bar and the number on it represent the score obtained by the corresponding method. The vertical hatched box on each bar indicates ± one standard deviation. The FBF obtained 1.7 points because the number of microphones is so small that its IRR is low. The GJBF reduced the interfer-

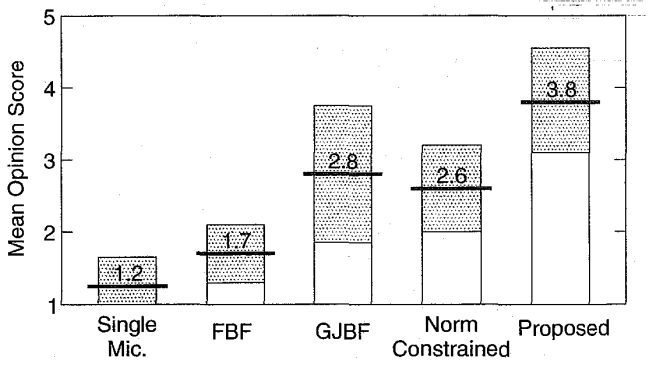


Fig. 5.15. Mean opinion score results.

ence considerably with serious target signal cancellation, thus, it was scored 2.8 points. The norm-constrained method was scored 2.6 points for its 9dB interference-reduction capability. The GJBF-CNA obtained 3.8 points, which is the highest of all the beamformers.

5.6 Hardware Evaluation of a Robust Adaptive Microphone Array

5.6.1 Implementation

The GJBF-CNA was implemented on a portable and flexible DSP system shown in Fig. 5.16 [48,49]. The system comprises a microphone array and a compact touch-panel personal computer which includes a floating point DSP, the ADSP-21062 [50]. The DSP contains a dual on-chip 2-Mbit SRAM and allows 32-bit IEEE floating-point computation. The sampling rate was software-programmed at 8 kHz.

The DSP board has a PCI (Peripheral Component Interconnect) interface, therefore, it can be connected to the PCI bus of any personal computer. A graphical interface has been developed to facilitate ease-of-use and monitoring of the implemented GJBF-CNA. It provides interactive parameter selection and displays the input and the output signals powers as well as the filter coefficients. This graphical display is useful for demonstrating the behavior of the GJBF-CNA and its performance. The system is shown in Fig. 5.16

5.6.2 Evaluation in a Real Environment

The GJBF-CNA in Fig. 5.16 was evaluated using the same linear microphone array as in the previous section. The selected step sizes were 0.02 for the ABM and 0.005 for the MC. The threshold $\eta = 0.65$ was used for the AMC. All other parameters were the same as those in the previous section.

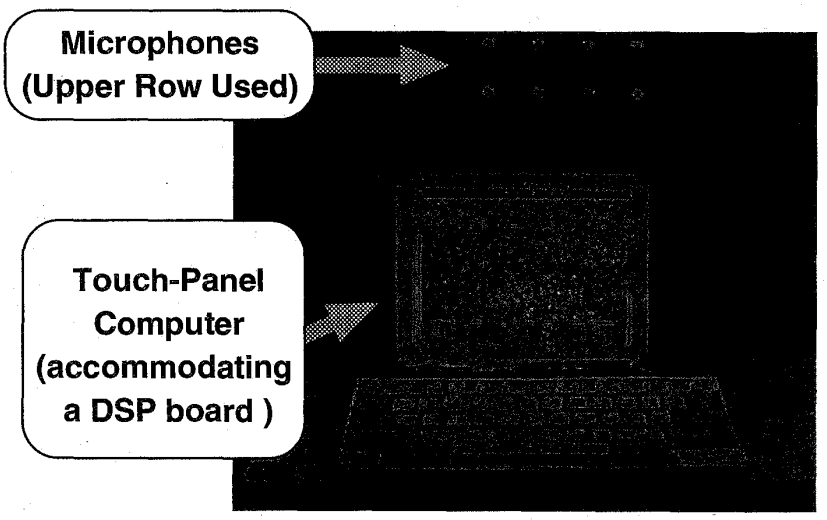


Fig. 5.16. Real-time DSP system.

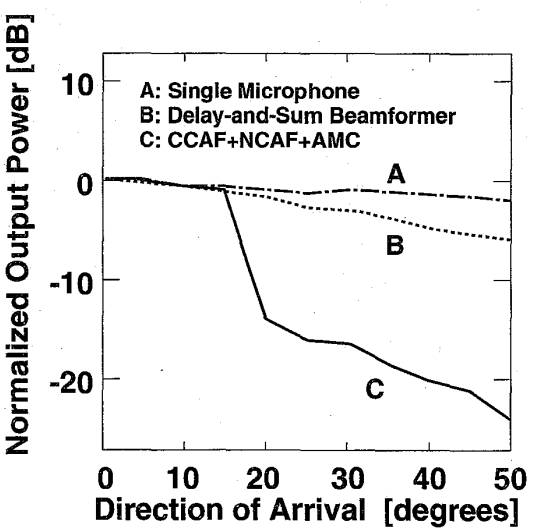


Fig. 5.17. Directivity patterns (i.e., the output powers normalized by the power at the center) measured in 5-degree intervals.

Directivity

Directivity for a single signal-source was measured. A white-noise source was scanned in two directions from 0° to 50° at a distance of 2.0 m from the array. Output powers of the system were measured in 5-degree intervals, and compared with those of a single microphone and an FBF (delay-and-sum beamformer). Figure 5.17 shows the output powers normalized by the power at the center. The figure indicates that the GJBF-CNA can suppress the interference at $\theta = 30°$ by as much as 15 dB when the allowable target DOA is set to ±20 degrees.

Noise Reduction

Noise reduction capability was evaluated in the same room as that for directivity evaluation. There were several computers with noisy fans. In addition, two noise-generating loudspeakers were located on both sides of the array. Stereo music or white noise was used as the noise signal.

In the beginning, breathing noise due to adaptation was observed at almost every utterance. It disappeared in a second and caused almost no problem for conversation. Although the degree of noise reduction depends on the loudspeaker positions, it was typically 8 to 1 dB. These results confirm that the GJBF-CNA is a promising technique for voice communications.

5.7 Conclusion

An overview of robust adaptive beamforming techniques have been presented in this chapter, with an emphasis on systems that are robust to steering-vector errors. It has been shown that the GJBF with the CCAF-NCAF structure and an AMC (GJBF-CNA) is effective in a real environment. Integrated systems with a microphone array, a noise canceler, and an echo canceler will play a key role in future acoustic noise and echo control devices.

References

1. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, New York: Wiley, 1980.
2. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, New York: Prentice-Hall, 1985.
3. S. Haykin ed., *Array Signal Processing*, Englewood Cliffs: Prentice-Hall, 1985.
4. B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, Apr. 1988.
5. D. H. Johnson and D. E. Dudgeon, *Array Signal Processing – Concepts and Techniques*, Englewood Cliffs: Prentice-Hall, 1993.
6. Special Session on Microphone Array Signal Processing, in *Proc. IEEE ICASSP'97*, vol. I, pp. 211–254, Apr. 1997.

7. J. L. Flanagan, D. A. Berkley, G. W. Elko, and M. M. Sondhi, "Autodirective microphone systems," *Acustica*, vol.73, pp. 58–71, Feb. 1991.
8. P. L. Chu, "Superdirective microphone array for a set-top video conferencing system," in *Proc. IEEE ICASSP'97*, vol. I, pp. 235–238, Apr. 1997.
9. I. Claessen, S. Nordholm, B. A. Bengtsson, and P. Eriksson, "A multi-DSP implementation of a broad-band adaptive beamformer for use in a hands-free mobile radio telephone," *IEEE Trans. Vehicular Tech.*, vol. 40, no. 1, pp. 194–202, Feb. 1991.
10. Y. Grenier, "A microphone array for car environments," *Speech Communicat.*, vol. 12, no. 1, pp. 25–39, Mar. 1993.
11. M. Dahl, I. Claesson, and S. Nordebo, "Simultaneous echo cancellation and car noise suppression employing a microphone array," in *Proc. IEEE ICASSP'97*, vol. I, pp. 239–242, Apr. 1997.
12. P. M. Peterson, "Using linearly-constrained adaptive beamforming to reduce interferrence in hearing aids from competing talkers in reverberant rooms," in *Proc. IEEE ICASSP'87*, 5.7.1, pp. 2364–2367, Apr. 1987.
13. P. M. Zurek, J. E. Greenberg, and P. M. Peterson, "Sensitivity to design parameters in an adaptive-beamforming hearing aid,' in *Proc. IEEE ICASSP'90*, A1.10, pp. 1129–1132, Apr. 1990.
14. W. Soede, A. J. Berkhout, and F. Bilson, "Development of a directional hearing instrument based on array technology," *J. Acoust. Soc. Amer.*, vol. 94, no. 2, pt.1, pp. 785–798, Aug. 1993.
15. W. Soede, F. Bilson, and A. J. Berkhout, "Assignment of a directional microphone array for hearing-impared listeners," *J. Acoust. Soc. Amer.*, vol. 94, no. 2, pt.1, pp. 799–808, Aug. 1993.
16. J. M. Kates, "Superdirective arrays for hearing aids," *J. Acoust. Soc. Amer.*, vol.94, no. 4, pp. 1930–1933, Oct. 1993.
17. M. W. Hoffman, T. D. Trine, K. M. Buckley, and D. J. Tasell, "Robust adaptive microphone array processing for hearing aids: Realistic speech enhancement," *J. Acoust. Soc. Amer.*, vol. 96, pp. 759–770, Aug. 1994.
18. F. Asano, Y. Suzuki, and T. Sone, "Weighted RLS adaptive beamforming with initial directivity," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 5, pp. 424–428, Sep. 1995.
19. J. M. Kates, "A comparison of hearing-aid array-processing techniques," *J. Acoust. Soc. Amer.*, vol. 99, no. 5, pp. 3138–3148, May 1996.
20. E. D. McKinney, and V. E. DeBrunner, "A two-microphone adaptive broadband array for hearing aids," in *Proc. IEEE ICASSP'96*, pp. 933–936, May 1996.
21. A. Wang, K. Yao, R. E. Hudson, D. Korompis, S. F. Soli, and S. Gao, "A high performance microphone array system for hearing aid applications," in *Proc. IEEE ICASSP'96*, pp. 3197–3200, May 1996.
22. K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," in *Proc. IEEE ICASSP'97*, vol. I, pp. 215–218, Apr. 1997.
23. M. Omologo, M. Matassoni, P. Svaizer, and D. Giuliani, "Microphone array based speech recognition with different talker-array positions," in *Proc. IEEE ICASSP'97*, vol. I, pp. 227–230, Apr. 1997.
24. T. Nishi, "Relation between objective criteria and subjective factors in a sound field, determined by multivariance analysis," *Acustica*, 76, pp. 153–162, 1992.

25. T. Nishi amd T. Inoue, 'Development of a multi-beam array microphone for multi-channel pickup of sound fields," *Acustica*, 76, pp. 163–172, 1992.
26. W. Täger and Y. Mahieux, "Reverberant sound field analysis using a microphone array," in *Proc. IEEE ICASSP'97*, vol. I, pp. 383–386, Apr. 1997.
27. L. J. Griffiths and C. W. Jim, "An alternative approach to linear constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.
28. I. Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beamforming," *IEEE Trans. Antennas Propagat.*, pp. 1093–1096, Sep. 1992.
29. S. Fischer and K. U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments," *Speech Communication*, vol. 20, pp. 215–227, Apr. 1996.
30. S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Processing* , vol. 5, no. 5, pp. 425–437, Sep. 1997.
31. M. H. Er and B. C. Ng, "A new approach to robust beamforming in the presence of steering vector errors," *IEEE Trans. Signal Processing*, vol. 42, no. 7, pp. 1826–1829, Jul. 1994.
32. G. L. Fudge and D. A. Linebarger, "A calibrated generalized sidelobe canceller for wideband beamforming," *IEEE Trans. Signal Processing*, pp. 2871–2875, Oct. 1994.
33. B. Widrow and M. McCool, "A comparison of adaptive algorithms based on the methods of steepest descent and random search," *IEEE Trans. Antennas Propagat.*, pp. 615–637, Sep. 1976.
34. M. H. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust. Speech Signal Processing*, pp. 1378–1393, Dec. 1983.
35. M. H. Er and A. Cantoni, "An unconstrained partitioned realization for derivative constrained broad-band antenna array processors," *IEEE Trans. Acoust. Speech Signal Processing*, pp. 1376–1379, Dec. 1986
36. N. K. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. Antennas Propagat.*, pp.996–1012, Aug. 1986.
37. H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust. Speech Signal Processing*, pp.1365–1376, Oct. 1987.
38. J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1662–1676, Mar. 1992.
39. O. Hoshuyama and A. Sugiyama, "A robust generalized sidelobe canceller with a blocking matrix using leaky adaptive filters," *Trans. IEICE*, vol.J79-A, no.9, pp.1516–1524, Sep. 1996 (in Japanese). (English version available in *Electron. Communicat. Japan*, vol.80, no.8, pp.56–65, Aug. 1997.)
40. G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*, Englewood Cliffs: Prentice-Hall, 1984.
41. O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," in *Proc. IEEE ICASSP'96*, pp. 925–928, May 1996.
42. O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer with a blocking matrix using coefficient constrained adaptive filters," *Trans. IEICE*, vol.E82-A, no.4, pp.640–647, Apr. 1999.

43. O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
44. O. Hoshuyama, B. Begasse, A. Sugiyama, and A. Hirano, "A robust adaptive microphone array control based on output signals of different beamformers," in *Proc. IEICE 12th DSP Symp.*, A3-4, Nov. 1997.
45. O. Hoshuyama, B. Begasse, and A. Sugiyama, "A new adaptation-mode control based on cross correlation for a robust adaptive microphone array," *IEICE Trans. Fundamentals*, vol. E84-A, Feb. 2001 (to appear).
46. O. Hoshuyama, B. Begasse, A. Sugiyama, and A. Hirano, "A real-time robust adaptive microphone array controlled by an SNR estimate," in *Proc. IEEE ICASSP'98*, pp. 3605–3678, May 1998.
47. H. R. Silbiger, "Audio Subjective Test Methods for Low Bit Rate Codec Evaluations," ISO/IEC JTC1/SC29/WG11/N0981, Jul. 1995.
48. O. Hoshuyama and A. Sugiyama, "An adaptive microphone array with good sound quality using auxiliary fixed beamformers and its DSP implementation," in *Proc. IEEE ICASSP'99*, pp. 949–952, Mar. 1999.
49. O. Hoshuyama and A. Sugiyama: "Realtime adaptive microphone array on a single DSP system," in *Proc. IWAENC'99*, pp. 92–95, Sep. 1999.
50. Analog Devices, *ADSP-2106x SHARC User's Manual*, Mar. 1995.

8 Robust Localization in Reverberant Rooms

Joseph H. DiBiase[1], Harvey F. Silverman[1], and Michael S. Brandstein[2]

[1] Brown University, Providence RI, USA
[2] 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-

cations. Consider the problem of beamforming to a moving speech source. It has been shown that for sources in close proximity to the microphones, the array aiming location must be accurate to within a few centimeters to prevent high-frequency rolloff in the received signal [3] and to allow for effective channel equalization [4]. A practical beamformer must therefore be capable of including a continuous and accurate location procedure within its algorithm. This requirement necessitates the use of a location estimator capable of fine resolution at a high update rate. Additionally, any such estimator would have to be computationally non-demanding and possess a short processing latency to make it practical for real-time systems.

These factors place tight constraints on the microphone data requirements. While the computation time required by the algorithm largely determines the latency of the locator, it is the data requirements that define theoretical limits. The work in [5], for example, focuses on reducing the size of the data segments necessary for accurate source localization in realistic room environments.

The goal of this chapter is to detail the issues associated with the problem of speech source localization in reverberant and noisy rooms and to present an effective methodology for its solution. While the focus will be the single-source scenario, the techniques described, in many cases, are applicable to situations where several individuals are conversing. The more general problem of simultaneous, multi-talker localization is addressed further in Chapter 9. The following section contains a summary of the existing genres for speech source localization using microphone arrays and highlights their relative merits. It is followed in Section 8.3 by the development of a speech source localization algorithm designed specifically for reverberant enclosures which combines two of these general approaches. Section 8.4 then offers some experimental results and conclusions.

8.2 Source Localization Strategies

Existing source localization procedures may be loosely divided into three general categories: those based upon maximizing the steered response power (SRP) of a beamformer, techniques adopting high-resolution spectral estimation concepts, and approaches employing time-difference of arrival (TDOA) information. These broad classifications are delineated by their application environment and method of estimation. The first refers to any situation where the location estimate is derived directly from a filtered, weighted, and summed version of the signal data received at the sensors. The second will be used to term any localization scheme relying upon an application of the signal correlation matrix. The last category includes procedures which calculate source locations from a set of delay estimates measured across various combinations of microphones.

8.2.1 Steered-Beamformer-Based Locators

The first categorization applies to passive arrays for which the system input is an acoustic signal produced by the source. The optimal Maximum Likelihood (ML) location estimator in this situation amounts to a focused beamformer which steers the array to various locations and searches for a peak in output power. Termed *focalization*, derivations of the optimality of the procedure and variations thereof are presented in [6–8]. Theoretical and practical variance bounds obtained via focalization are detailed in [6,7,9] and the steered-beamformer approach has been extended to the case of multiple-signal sources in [10].

The simplest type of steered response is obtained using the output of a delay-and-sum beamformer. This is what is most often referred to as a conventional beamformer. Delay-and-sum beamformers apply time shifts to the array signals to compensate for the propagation delays in the arrival of the source signal at each microphone. These signals are time-aligned and summed together to form a single output signal. More sophisticated beamformers apply filters to the array signals as well as this time alignment. The derivation of the filters in these filter-and-sum beamformers is what distinguishes one method from another.

Beamforming has been used extensively in speech-array applications for voice capture. However, due to the efficiency and satisfactory performance of other methods, it has rarely been applied to the talker localization problem. The physical realization of the ML estimator requires the solution of a nonlinear optimization problem. The use of standard iterative optimization methods, such as steepest descent and Newton-Raphson, for this process was addressed by [10]. A shortcoming of each of these approaches is that the objective function to be minimized does not have a strong global peak and frequently contains several local maxima. As a result, this genre of efficient search methods is often inaccurate and extremely sensitive to the initial search location. In [11] an optimization method appropriate for a multimodal objective function, Stochastic Region Contraction (SRC), was applied specifically to the talker localization problem. While improving the robustness of the location estimate, the resulting search method involved an order of magnitude more evaluations of the objective function in comparison to the less robust search techniques. Overall, the computational requirements of the focalization-based ML estimator, namely the complexity of the objective function itself as well as the relative inefficiency of an appropriate optimization procedure, prohibit its use in the majority of practical, real-time source locators.

Furthermore, the steered response of a conventional beamformer is highly dependent on the spectral content of the source signal. Many optimal derivations are based on *a priori* knowledge of the spectral content of the background noise, as well as the source signal [7,8]. In the presence of significant reverberation, the noise and source signals are highly correlated, making ac-

curate estimation of the noise infeasible. Furthermore, in nearly all array-applications, little or nothing is known about the source signal. Hence, such optimal estimators are not very practical in realistic speech-array environments.

The practical shortcomings of applying correlation-based localization estimation techniques without a great deal of intelligent pruning is typified by the system produced in [12]. In this work a sub-optimal version of the ML steered-beamformer estimator was adapted for the talker-location problem. A source localization algorithm based on multi-rate interpolation of the sum of cross-correlations of many microphone pairs was implemented in conjunction with a real-time beamformer. However, because of the computational requirements of the procedure, it was not possible to obtain the accuracy and update rate required for effective beamforming in real-time given the hardware available.

8.2.2 High-Resolution Spectral-Estimation-Based Locators

This second categorization of location estimation techniques includes the modern beamforming methods adapted from the field of high-resolution spectral analysis: autoregressive (AR) modeling, minimum variance (MV) spectral estimation, and the variety of eigenanalysis-based techniques (of which the popular MUSIC algorithm is an example). Detailed summaries of these approaches may be found in [13,14]. While these approaches have successfully found their way into a variety of array processing applications, they all possess certain restrictions that have been found to limit their effectiveness with the speech-source localization problem addressed here.

Each of these high-resolution processes is based upon the spatiospectral correlation matrix derived from the signals received at the sensors. When exact knowledge of this matrix is unknown (which is most always the case), it must be estimated from the observed data. This is done via ensemble averaging of the signals over an interval in which the sources and noise are assumed to be statistically stationary and their estimation parameters (location in this case) are assumed to be fixed. For speech sources, fulfilling these conditions while allowing sufficient averaging can be very problematic in practice.

With regard to the localization problem at hand, these methods were developed in the context of far-field plane waves projecting onto a linear array. While the MV and MUSIC algorithms have been shown to be extendible to the case of general array geometries and near-field sources [15], the AR model and certain eigenanalysis approaches are limited to the far-field, uniform linear array situation.

With regard to the issue of computational expense, a search of the location space is required in each of these scenarios. While the computational complexity at each iteration is not as demanding as the case of the steered-beamformer, the objective space typically consists of sharp peaks. This property precludes the use of iteratively efficient optimization methods. The sit-

uation is compounded if a more complex source model is adopted (incorporating source orientation or head radiator effects, for instance) in an effort to improve algorithm performance. Additionally, it should be noted that these high-resolution methods are all designed for narrowband signals. They can be extended to wideband signals, including speech, either through simple serial application of the narrowband methods or more sophisticated generalizations of these approaches, such as [16–18]. Either of these routes extends the computational requirements considerably.

These algorithms tend to be significantly less robust to source and sensor modeling errors than conventional beamforming methods [19,20]. The incorporated models typically assume ideal source radiators, uniform sensor channel characteristics, and exact knowledge of the sensor positions. Such conditions are impossible to obtain in real-world environments. While the sensitivity of these high-resolution methods to the modeling assumptions may be reduced, it is at the cost of performance. Additionally, signal coherence, such as that created by the reverberation conditions of primary concern here, is detrimental to algorithmic performance, particularly that of the eigenanalysis approaches. This situation may be improved via signal processing resources, but again at the cost of decreased resolution[21]. Primarily for these reasons, localization methods based upon these high-resolution strategies will not considered further in this work. However, this should not exclude their judicious use in other speech localization contexts, particularly multi-source scenarios.

8.2.3 TDOA-Based Locators

With this third localization strategy, a two-step procedure is adopted. Time delay estimation (TDE) of the speech signals relative to pairs of spatially separated microphones is performed. This data along with knowledge of the microphone positions are then used to generate hyperbolic curves which are then intersected in some optimal sense to arrive at a source location estimate. A number of variations on this principle have been developed, [22–28] are examples. They differ considerably in the method of derivation, the extent of their applicability (2-D vs. 3-D, near source vs. distant source, etc.), and their means of solution. Primarily because of their computational practicality and reasonable performance under amicable conditions, the bulk of passive talker localization systems in use today are TDOA-based.

Accurate and robust TDE is the key to the effectiveness of localizers within this genre. The two major sources of signal degradation which complicate this estimation problem are background noise and channel multi-path due to room reverberations. The noise-alone case has been addressed at length and is well understood. Assuming uncorrelated, stationary Gaussian signal and noise sources with known statistics and no multi-path, the ML time-delay estimate is derived from a SNR-weighted version of the Generalized Cross-Correlation (GCC) function [29]. An ML-type weighting appropriate for non-stationary speech sources was presented in [30] and applied successfully to

speech source localization in low-multipath environments [31]. However, once room reverberations rise above minimal levels, these methods begin to exhibit dramatic performance degradations and become unreliable [32,33]. A basic approach to dealing with multi-path channel distortions in this context has been to make the GCC function more robust by deemphasizing the frequency-dependent weightings. The Phase Transform (PHAT) [29] is one extreme of this procedure which has received considerable attention recently as the basis of speech source localization systems [34–36]. By placing equal emphasis on each component of the cross-spectrum phase, the resulting peak in the GCC-PHAT function corresponds to the dominant delay in the reverberated signal. While effective at reducing some of the degradations due to multi-path, the Phase Transform accentuates components of the spectrum with poor SNR and has the potential to provide poor results, particularly under low reverberation, high noise conditions.

Other approaches for TDE of talkers in adverse environments are available. A procedure which utilizes a speech specific criterion in the design of the GCC weighting function is presented in [37]. Cepstral prefiltering [38] has been used to deconvolve the effects of reverberation prior to applying GCC. However, deconvolution requires long data segments since the duration of a typical small-room impulse response is 200-400 ms. It is also very sensitive to the high variability and non-stationarity of speech signals. In fact, the experiments performed in [38] avoided the use of speech as input altogether. Instead, colored Gaussian noise was used as the source signal. While identification of room impulse responses is extremely problematic when the source signal is unknown, the method proposed in [24], which is based on eigenvalue decomposition, efficiently detects the direct paths of the two impulse responses. This method is effective with speech as input, but requires 250 ms of microphone data to converge. A short-time TDE method, which is more complex than GCC, is presented in [33]. It involves the minimization of a weighted least-squares function of the phase data. It was shown to outperform both GCC-ML and GCC-PHAT in reverberant conditions. However, this improvement comes at the cost of a complicated searching algorithm. The marginal improvement over GCC-PHAT may not justify this added cost in computational complexity. Reverberation effects can also be overcome to some degree by classifying TDE's acquired over time and associating them with the direction of arrival (DOA) of the sound waves [39]. This approach, however, is not suitable for short-time TDE. Under extreme acoustic conditions, a large percentage of the TDE's are anomalous, and it takes a considerable period (1-2 s in [39]) to acquire enough estimates for a statistically meaningful classification.

Among the methods summarized above, those that rely on long data segments generally outperform those that do not. This result may be attributed to the ensemble averaging performed under these conditions to improve the quality of the underlying signal statistics. However, the dynamic environ-

ments of many speech array applications require high update rates, which limit the duration of the data segments used for analysis. For example, the automatic camera steering video-conferencing system detailed in [34] utilizes a TDOA-based method with GCC-PHAT TDE applied at update rates of 200-300 ms. With such long data segments, reliable estimates are produced, even in moderately adverse acoustic conditions. However, applications such as adaptive beamforming and the tracking of multiple talkers using a TDOA-based localizer require an appreciably higher estimate rate; source positions must be acquired from independent data segments as short as 20-30 ms. Over such limited durations, the lack of ensemble averaging has a severe impact on the performance of the TDE.

Given a set of TDOA figures with known error statistics, the second step of obtaining the ML location estimate necessitates solving a set of nonlinear equations. The calculation of this result is considerably less computationally expensive than that required for estimators belonging to the two previously discussed genres. There is an extensive class of sub-optimal, closed-form location estimators. designed to approximate the exact solution to the nonlinear problem. These techniques are computationally undemanding and, in many cases, suffer little detriment in performance relative to their more compute-intensive counterparts. [22,25–28,40,41] are typical of these methods. Regardless of the solution method employed, this third class of location estimation techniques possesses a significant computational advantage over the steered-beamformer or high-resolution spectral-estimation based approaches.

TDOA-based locators do present several disadvantages when used as the basis of a general localization scheme. Their primary limitation is the inability to accommodate multi-source scenarios. These algorithms assume a single-source model. While TDOA-based methods with short analysis intervals may be used to track several individuals in a conversational situation [31,42], the presence of multiple simultaneous talkers, excessive ambient noise, or moderate to high reverberation levels in the acoustic field typically results in poor TDOA figures and subsequently, unreliable location fixes. A TDOA-based locator operating in such an environment would require a means for evaluating the validity and accuracy of the delay and location estimates. These shortcomings may be overcome to some degree through judicious use of appropriate detection methods at each stage in the process [31].

While practical, the application of TDOA-based localization procedures is of limited utility in realistic, acoustic environments. Steered-Beamformer strategies are computationally more intensive, but tend to possess a robustness advantage and require a shorter analysis interval. The two-stage process requiring time-delay estimation prior to the actual location evaluation is suboptimal. The intermediate signal parameterization accomplished by the TDOA estimation procedure represents a significant data reduction at the expense of a decrease in theoretical localization performance. However, in

real situations the performance advantage inherent in the optimal steered-beamformer estimator is lessened because of incomplete knowledge of the signal and noise spectral content as well as unrealistic stationarity assumptions.

With these relative advantages and shortcomings in mind, a new localization method, which combines the best features of the steered-beamformer with those of the Phase Transform weighting of the GCC, was introduced in [5]. The goal was to exploit the inherent robustness and short-time analysis characteristics of the steered response power approach with the insensitivity to signal conditions afforded by the Phase Transform. This new algorithm, termed SRP-PHAT, will be detailed in the following section and will be shown to produce highly reliable location estimates in rooms with reverberation times up to 200 ms, using independent 25 ms data segments.

8.3 A Robust Localization Algorithm

Before describing the SRP-PHAT algorithm, it will be necessary to develop further a number of topics addressed in the prior section. Specifically, the following subsections will provide details of the impulse response model, the GCC and its PHAT implementation, ML TDOA-based localization, and the computation of the SRP. These items will then be tied together in the final subsection to motivate and define the SRP-PHAT algorithm.

8.3.1 The Impulse Response Model

It will be assumed that sound waves propagate as predicted by the linear wave equation [43]. With this assumption, the acoustic paths between sound sources and microphones can be modeled as linear systems [44]. This is clearly advantageous to the analysis and modeling of the signals produced by the microphones of an array. Such linear models are valid under the realistic conditions encountered in small-room speech-array environments and are regularly exploited by array-processing techniques [13].

In the presence of sound-reflecting surfaces, the sound waves produced by a single source propagate along multiple acoustic paths. This gives rise to the familiar effects of reverberation; sounds reflect off objects and produce echoes. The walls of most rooms are reflective enough to create significant reverberation. While it is not always noticeable to the occupants, even mild reverberation can severely impact the performance of speech-array systems. Hence, multi-path propagation must be incorporated into the signal-processing model.

The wave field at a particular location inside a reverberant room may be considered to be linearly related to the source signal, $s(t)$. Let the 3-element vectors, $\boldsymbol{p}_n$ and $\boldsymbol{q}_s$, define the Cartesian coordinates of the nth microphone

and the source, respectively. The received signal at the nth microphone may now be expressed as

$$x_n(t) = s(t) \star h_n(\boldsymbol{q}_s, t) + v_n(t) \tag{8.1}$$

The overall impulse response, $h_n(\boldsymbol{q}_s, t)$, is the result of cascading two filters: the room impulse response and the microphone channel response. The former characterizes all acoustic paths between the source and microphone locations, including the direct path. It is a function of $\boldsymbol{p}_n$ as well as the source location, $\boldsymbol{q}_s$, and is highly dependent on these parameters. In general, the room impulse response is affected by environmental conditions, such as temperature and humidity. It also varies with the movement of furniture and individuals inside the room. While such variations are significant, it is reasonable to assume that these factors remain constant over short periods. Hence, a room impulse response may be considered time-invariant for short periods when the source and microphone are spatially fixed. The microphone channel response accounts for the electrical, mechanical and acoustical properties of the microphone system. In general, the microphone's directivity pattern makes its response a function of its orientation as well as its spatial placement relative to the source. The additive term, $v_n(t)$, is the result of channel noise in the microphone system and any propagating ambient noise such as that due to fans or other mechanical equipment. The propagating noise is usually more significant than the channel noise and tends to dominate this term. Generally, $v_n(t)$ is assumed to be uncorrelated with $s(t)$.

Figure 8.1 illustrates a close-up view of the response that was measured in a typical conference room. The direct-path component and some of the strong reflected components are highlighted in this plot. The peaks corresponding to the reflected sound waves are comparable in size to the direct-path peak. These peaks, which occur within 20 ms of the direct-path, are responsible for many of the erroneous results produced by short-time TDE's, which operate on data blocks as small as 25 ms. The large secondary peaks in the room response are highly correlated with the false peaks in the GCC function [5].

The purpose of TDE is to evaluate the temporal disparity between the direct-path components in the two received microphone signals. To this end, it will be useful to rewrite the impulse response specifically in terms of its direct-path component. Equation 8.1 is modified to:

$$x_n(t) = \frac{1}{r_n} s(t - \tau_n) \star g_n(\boldsymbol{q}_s, t) + v_n(t) \tag{8.2}$$

where r_n is the source-microphone separation distance, τ_n is the direct path time delay, and $g_n(\boldsymbol{q}_s, t)$ is the modified impulse response which encompasses the original response minus the direct path component. The microphone signal model is now expressed explicitly in terms of the parameter of interest, namely the time delay, τ_n.

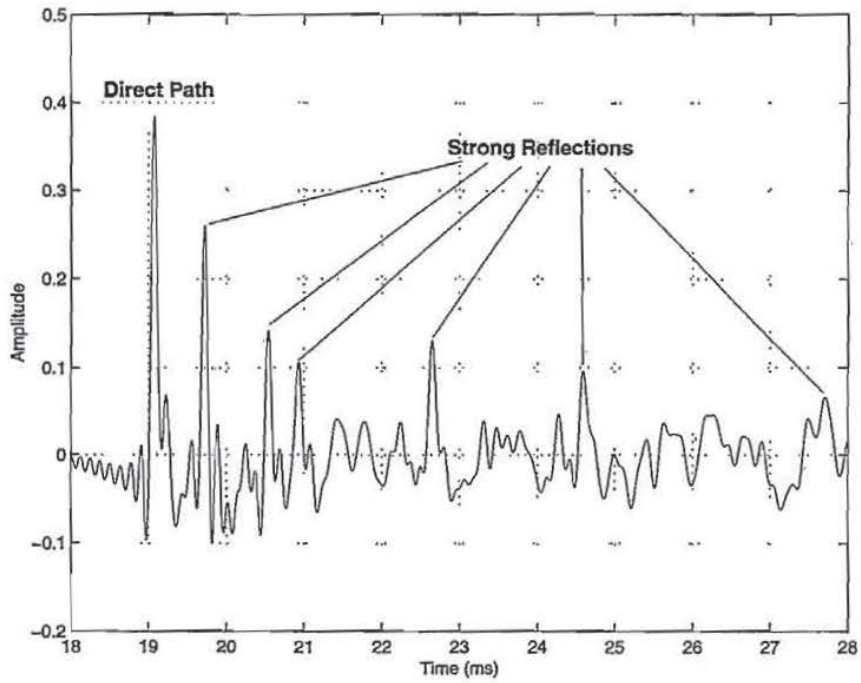


Fig. 8.1. A close-up of a 10-millisecond segment of a room impulse response measured in a typical conference room. The direct-path component and some strong reflected components are highlighted.

8.3.2 The GCC and PHAT Weighting Function

For a pair of microphones, $n = 1, 2$, their associated TDOA, τ_{12}, is defined as

$$\tau_{12} \equiv \tau_2 - \tau_1. \tag{8.3}$$

Applying this definition to their associated received microphone signal models yields

$$\begin{aligned} x_1(t) &= \frac{1}{\tau_1} s(t - \tau_1) \star g_1(\boldsymbol{q}_s, t) + v_1(t) \\ x_2(t) &= \frac{1}{\tau_2} s(t - \tau_1 - \tau_{12}) \star g_2(\boldsymbol{q}_s, t) + v_2(t). \end{aligned} \tag{8.4}$$

If the modified impulse responses for the microphone pair are similar, then (8.4) shows that a scaled version of $s(t - \tau_1)$ is present in the signal from microphone 1 and a time-shifted (and scaled) version of $s(t - \tau_1)$ is present in the signal from microphone 2. The cross-correlation of the two signals should show a peak at the time lag where the shifted versions of $s(t)$ align, corresponding to the TDOA, τ_{12}. The cross correlation of signals and is defined as:

$$c_{12}(\tau) = \int_{-\infty}^{+\infty} x_1(t) x_2(t + \tau) dt \tag{8.5}$$

The GCC function, $R_{12}(\tau)$, is defined as the cross correlation of two filtered versions of $x_1(t)$ and $x_2(t)$ [29]. With the Fourier transforms of these filters denoted by $G_1(\omega)$ and $G_2(\omega)$, respectively, the GCC function can be expressed in terms of the Fourier transforms of the microphone signals

$$R_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \big(G_1(\omega)X_1(\omega)\big)\big(G_2(\omega)X_2(\omega)\big)^* e^{j\omega\tau} d\omega \tag{8.6}$$

Rearranging the order of the signals and filters and defining the frequency dependent weighting function, $\Psi_{12} \equiv G_1(\omega)G_2(\omega)^*$, the GCC function can be expressed as

$$R_{12}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Psi_{12}(\omega)X_1(\omega)X_2(\omega)^* e^{j\omega\tau} d\omega \tag{8.7}$$

Ideally, $R_{12}(\tau)$ will exhibit an explicit global maximum at the lag value which corresponds to the relative delay. The TDOA estimate is calculated from

$$\hat{\tau}_{12} = \underset{\tau \in D}{\operatorname{argmax}}\, R_{12}(\tau). \tag{8.8}$$

The range of potential TDOA values is restricted to a finite interval, D, which is determined by the physical separation between the microphones. In general, $R_{12}(\tau)$ will have multiple local maxima which may obscure the true TDOA peak and subsequently, produce an incorrect estimate. The amplitudes and corresponding time lags of these erroneous maxima depend on a number of factors, typically ambient noise levels and reverberation conditions.

The goal of the weighting function, Ψ_{12}, is to emphasize the GCC value at the true TDOA value over the undesired local extrema. A number of such functions have been investigated. As previously stated, for realsitic acoustical conditions the PHAT weighting [29] defined by

$$\Psi_{12}(\omega) \equiv \frac{1}{|X_1(\omega)X_2^*(\omega)|} \tag{8.9}$$

has been found to perform considerably better than its counterparts designed to be statistically optimal under specific non-reverberant, noise conditions. The PHAT weighting whitens the microphone signals to equally emphasize all frequencies. The utility of this strategy and its extension to steered-beamforming form the basis of the SRP-PHAT algorithm that follows.

8.3.3 ML TDOA-Based Source Localization

Consider the ith pair of microphones with spatial coordinates denoted by the 3-element vectors, $\boldsymbol{p}_{i1}$ and $\boldsymbol{p}_{i2}$, respectively. For a signal source with known

spatial location, $\boldsymbol{q}_s$, the true TDOA relative to the ith sensor pair will be denoted by $T(\{\boldsymbol{p}_{i1}, \boldsymbol{p}_{i2}\}, \boldsymbol{q}_s)$, and is calculated from the expression

$$T(\{\boldsymbol{p}_{i1}, \boldsymbol{p}_{i2}\}, \boldsymbol{q}_s) = \frac{|\boldsymbol{q}_s - \boldsymbol{p}_{i2}| - |\boldsymbol{q}_s - \boldsymbol{p}_{i1}|}{c} \tag{8.10}$$

where c is the speed of sound in air. The estimate of this true TDOA, the result of a TDE procedure involving the signals received at the two microphones, will be given by $\hat{\tau}_i$. In practice, the TDOA estimate is a corrupted version of the true TDOA and in general, $\hat{\tau}_i \neq T(\{\boldsymbol{p}_{i1}, \boldsymbol{p}_{i2}\}, \boldsymbol{q}_s)$.

For a single microphone pair and its TDOA estimate, the locus of potential source locations in 3-space which satisfy (8.10) corresponds to one-half of a hyperboloid of two sheets. This hyperboloid is centered about the midpoint of the microphones and has $\boldsymbol{p}_{i2} - \boldsymbol{p}_{i1}$ as its axis of symmetry.

For sources with a large source-range to microphone-separation ratio, the hyperboloid may be well-approximated by a cone with a constant direction angle relative to the axis of symmetry. The corresponding estimated direction angle, $\hat{\theta}_i$, for the microphone pair is given by:

$$\hat{\theta}_i = \cos^{-1}\left(\frac{c \cdot \hat{\tau}_i}{|\mathbf{m}_{i1} - \mathbf{m}_{i2}|}\right) \tag{8.11}$$

In this manner each microphone pair and TDOA estimate combination may be associated with a single parameter which specifies the angle of the cone relative to the sensor pair axis. For a given source and TDOA estimate, $\hat{\theta}_i$ is referred to as the DOA relative to the ith pair of microphones.

Given a set of M TDOA estimates derived from the signals received at multiple pairs of microphones, the problem remains as how to best estimate the true source location, $\boldsymbol{q}_s$. Ideally, the estimate will be an element of the intersection of all the potential source loci. In practice, however, for more than two pairs of sensors this intersection is, in general, the empty set. This disparity is due in part to imprecision in the knowledge of system parameters (TDOA estimate and sensor location measurement errors) and in part to unrealistic modeling assumptions (point source radiator, ideal medium, ideal sensor characteristics, etc.). With no ideal solution available, the source location must be estimated as the point in space which best *fits* the sensor-TDOA data or more specifically, minimizes an error criterion that is a function of the given data and a hypothesized source location. If the time-delay estimates at each microphone pair are assumed to be independently corrupted by zero-mean additive white Gaussian noise of equal variance then the ML location estimate can be shown to be the position which minimizes the least squares error criterion

$$E(\boldsymbol{q}) = \sum_{i=1}^{M} (\hat{\tau}_i - T(\{\boldsymbol{p}_{i1}, \boldsymbol{p}_{i2}\}, \boldsymbol{q}))^2. \tag{8.12}$$

The location estimate is then found from

$$\hat{q}_s = \underset{q}{\operatorname{argmin}}\, E(q). \tag{8.13}$$

The criterion in (8.12) will be referred to as the LS-TDOA error. As stated earlier, the evaluation of $\hat{q}_s$ in this manner involves the optimization of a non-linear function and necessitates the use of search methods. Closed-form approximations to this method were given earlier.

8.3.4 SRP-Based Source Localization

The microphone signal model in (8.2) shows that for an array of N microphones in the reception region of a source, a delayed, filtered, and noise corrupted version of the source signal, $s(t)$, is present in each of the received microphone signals. The delay-and-sum beamformer time aligns and sums together the $x_n(t)$, in an effort to preserve unmodified the signal from a given spatial location while attenuating to some degree the noise and convolutional components. It is defined as simply as

$$y(t, q_s) = \sum_{n=1}^{N} x_n(t + \Delta_n) \tag{8.14}$$

where Δ_n are the *steering delays* appropriate for focusing the array to the source spatial location, q_s, and compensating for the direct path propagation delay associated with the desired signal at each microphone. In practice, the delays relative to a reference microphone are used instead of the absolute delays. This makes all shifting operations causal, which is a requirement of any practical system, and implies that $y(t, q_s)$ will contain an overall delayed version of the desired signal which in practice is not detrimental. The use of a single reference microphone means that the steering delays may be determined directly from the TDOA's (estimated or theoretical) between each microphone and the reference. This implies that knowledge of the TDOA's alone is sufficient for steering the beamformer without an explicit source location.

In the most ideal case with no additive noise and channel effects, the output of the deal-and-sum beamformer represents a scaled and potentially delayed version of the desired signal. For the limited case of additive, uncorrelated, and uniform variance noise and equal source-microphone distances this simple beamformer is optimal. These are certainly very restrictive conditions. In practice, convolutional channel effects are nontrivial and the additive noise is more complicated. The degree to which these noise and reverberation components of the microphone signals are suppressed by the delay-and-sum beamformer is frequently minimal and difficult to analyze. Other methods have been developed to extend the delay-and-sum concept to the more general filter-and-sum approach, which applies adaptive filtering to the microphone

signals before they are time-aligned and summed. Again, these methods tend to not be robust to non-theoretical conditions, particularly with regard to the channel effects.

The output of an N-element, filter-and-sum beamformer can be defined in the frequency domain as

$$Y(\omega, \boldsymbol{q}) = \sum_{n=1}^{N} G_n(\omega) X_n(\omega) e^{j\omega\Delta_n} \tag{8.15}$$

where $X_n(\omega)$ and $G_n(\omega)$ are the Fourier Transforms of the nth microphone signal and its associated filter, respectively. The microphone signals are phase-aligned by the steering delays appropriate for the source location, $\boldsymbol{q}$. This is equivalent to the time-domain beamformer version. The addition of microphone and frequency-dependent filtering allows for some means to compensate for the environmental and channel effects. Choosing the appropriate filters depends on a number of factors, including the nature of the source signal and the type of noise and reverberations present. As will be seen, the strategy used by the PHAT of weighing each frequency component equally will prove advantageous for practical situations where the ideal filters are unobtainable.

The beamformer may be used as a means for source localization by steering the array to specific spatial points of interest in some fashion and evaluating the output signal, typically its power. When the focus corresponds to the location of the sound source, the SRP should reach a global maximum. In practice, peaks are produced at a number of incorrect locations as well. These may be due to strong reflective sources or merely a byproduct of the array geometry and signal conditions. In some cases, these extraneous maxima in the SRP space may obscure the true location and in any case, complicate the search for the global peak. The SRP for a potential source location can be expressed as the output power of a filter-and-sum beamformer by

$$P(\boldsymbol{q}) = \int_{-\infty}^{+\infty} |Y(\omega)|^2 d\omega \tag{8.16}$$

and location estimate is found from

$$\hat{\boldsymbol{q}}_s = \underset{\boldsymbol{q}}{\operatorname{argmax}}\, P(\boldsymbol{q}). \tag{8.17}$$

8.3.5 The SRP-PHAT Algorithm

Given this background, the SRP-PHAT algorithm may now be defined. With respect to GCC-based TDE, the PHAT weighting has been found to provide an enhanced robustness in low to moderate reverberation conditions. While improving the quality of the underlying delay estimates, it is still not sufficient to render TDOA-based localization effective under more adverse conditions.

The delay-and-sum SRP approach requires shorter analysis intervals and exhibits an elevated insensitivity to environmental conditions, though again, not to a degree that allows for their use under excessive multi-path. The filter-and-sum version of the SRP adds flexibility but the design of the filters is typically geared towards optimizing SNR in noise-only conditions and is excessively dependent on knowledge of the signal and channel content. Originally introduced in [5], the goal of the SRP-PHAT algorithm is to combine the advantages of the steered beamformer for source localization with the signal and condition independent robustness offered by the PHAT weighting.

The SRP of the filter-and-sum beamformer can be expressed as

$$P(\boldsymbol{q}) = \sum_{l=1}^{N}\sum_{k=1}^{N}\int_{-\infty}^{\infty} \Psi_{lk}(\omega)X_l(\omega)X_k^*(\omega)e^{j\omega(\Delta_k-\Delta_l)}d\omega \tag{8.18}$$

where $\Psi_{lk}(\omega) = G_l(\omega)G_k^*(\omega)$ is analogous to the two-channel GCC weighting term in (8.7). The corresponding multi-channel version of the PHAT weighting is given by

$$\Psi_{lk}(\omega) = \frac{1}{|X_l(\omega)X_k^*(\omega)|} \tag{8.19}$$

which in the context of the filter-and-sum beamformer (8.15) is equivalent to the use of the individual channel filters

$$G_n(\omega) = \frac{1}{|X_n(\omega)|}. \tag{8.20}$$

These are the desired SRP-PHAT filters. They may be implemented from the frequency-domain expression above. Alternatively, it may be shown that (8.18) is equivalent to the sum of the GCC's of all possible N-choose-2 microphone pairings. This means that the SRP of a 2-element array is equivalent to the GCC of those two microphones. Hence, as the number of microphones is increased, SRP naturally extends the GCC method from a pairwise to a multi-microphone technique. Denoting $R_{lk}(\tau)$ as the PHAT-weighted GCC of the lth and k^{th} microphone signals, a time-domain version of SRP-PHAT functional can now be expressed as

$$P(\boldsymbol{q}) = 2\pi\sum_{l=1}^{N}\sum_{k=1}^{N} R_{lk}(\Delta_k - \Delta_l). \tag{8.21}$$

This is the sum of all possible pairwise GCC permutations which are time-shifted by the differences in the steering delays. Included in this summation is the sum of the N autocorrelations, which is the GCC evaluated at a lag of zero. These terms contribute only a DC offset to the steered response power since they are independent of the steering delays.

Given either method of computation, SRP-PHAT localization is performed in a manner similar to the standard SRP-based approaches. Namely,

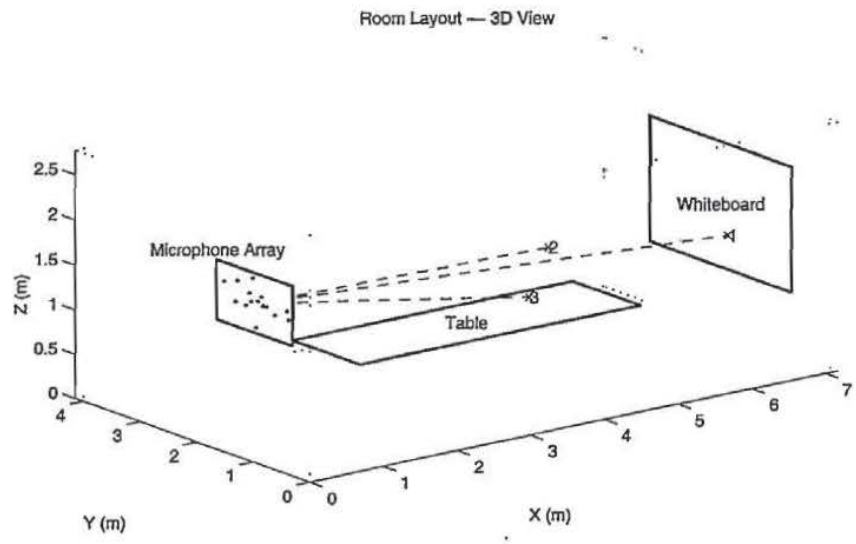


Fig. 8.2. Conference room layout.

$P(q)$ is maximized over a region of potential source locations. As will be shown in the next section, relative to the search space indicative of the standard SRP approach, the SRP-PHAT functional significantly deemphasizes extraneous peaks and dramatically sharpens the resolution of the true peak. These desirable features result in a decreased sensitivity to noise and reverberations and more precise location estimates than the existing localization methods offer. Additionally, this is achieved using a very short analysis interval.

8.4 Experimental Comparison

While more extensive results are available in [5], an experiment is offered here to evaluate and compare the relative characteristics and performance of three different source locators: SRP, SRP-PHAT and ML-TDOA. Five second recordings were made for three source locations in a 7 by 4 by 3 m conference room at Brown University using a 15-element microphone array. Figure 8.2 illustrates the room layout. Pre-recorded speech, which was acquired using a close-talking microphone, was played through a loudspeaker while simultaneously recording the signals from the array. The use of the loudspeaker was preferable to an actual talker since the loudspeaker could be precisely located and would be fixed over the duration of the recordings. The talkers were males uttering a unique string of alpha-digits. Source 1 was most distant from the array and was positioned at standing height in front of a white-board. The other two sources were positioned at a seated level

around a conference table, which was located approximately in the center of the room.

The microphone array was composed of eight omni-directional electret condenser microphones, which were randomly distributed on a plane within a .33 by 0.36 m rectangle. The microphones were attached to a rectangular sheet of acoustic foam, which was supported by an aluminum frame. This frame was mounted on a tripod that was placed parallel to the back wall at a distance of 0.9 m. The acoustic foam damps some of the multi-path reflections from this wall and isolates the microphones from vibrations traveling along the mountings.

The loudspeaker faced the array and the volume level was adjusted at each location to maximize SNR conditions. SNR levels at each microphone averaged about 25 dB for the three source locations. Source 3, with its location the closest to the microphone array, had SNRs as high as 36 dB. With such high SNRs, all microphones signals in the conference room dataset have minimal contributions from the background noise, which was primarily produced by the fans inside the computer equipment.

The measured reverberation time of the room was determined to be 200 ms. This qualifies as a mildly reverberant room. However, the near-end peaks in the impulse responses (as in Figure 8.1) combined with a 200 ms reverberation time do, in fact, have a significant impact on localization. This will be demonstrated by the following performance comparisons.

Given the size of the array aperture relative to the source ranges, all three talkers can be considered to lie in the far field of array. Under such conditions, range estimates are ambiguous, and only the azimuth and elevation angles can be estimated reliably. Accordingly, this experiment will focus on DOA measures as opposed to 3-D Cartesian coordinates. Results obtained with more extensive arrays and near-field sources are available in [5].

The recorded data was segmented into 25 ms frames using a half-overlapping Hanning window. SNR-based speech detection was performed for each frame. All frames where any of the eight microphone channels had SNR within 12dB of the background noise were eliminated. Out of the 399 frames per recording, 313, 340, and 297 were retained for sources 1,2, and 3, respectively. The DOA's of the sources were estimated by minimization of the LS-TDOA error and maximization of SRP and SRP-PHAT evaluated over azimuth and elevation relative to the array's origin. The frequency range used to compute both the steered responses and the GCC's was 300 Hz to 8 kHz. These functions were computed over a range of $-60°$ to $+60°$ for both azimuth and elevation with a $0.1°$ resolution.

By taking all possible combinations, 28 microphone pairs were formed using the 8-element array. Hence, for each data frame, 28 TDOA estimates were made for each of the three speech recordings using GCC-PHAT. Figure 8.3 illustrates the LS-TDOA error as a function of azimuth and elevation for a segment of nine successive frames recorded for source 1. The white point in

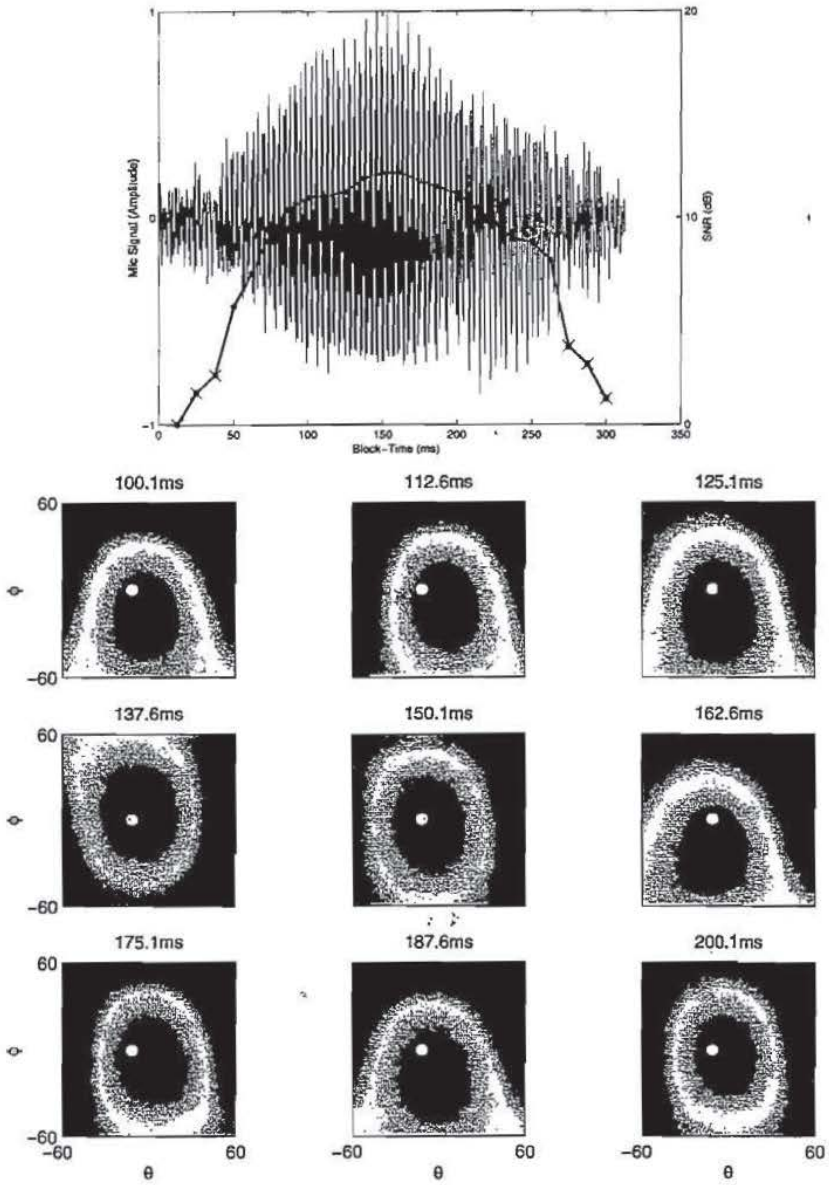


Fig. 8.3. Speech segment (top) with nine frames of the LS-TDOA error surfaces.

each contour plot marks the true DOA. The dark area in the center of the images represents the minima of the LS-TDOA error. At the top of this figure is a plot of the amplitude of the corresponding speech segment, which is the letter "R", spoken as in "Are we there yet?" Superimposed on this speech signal is a curve representing the average power of the signals from the array, with the scale of its vertical axis labeled on the right side of the graph. Each point along this power curve corresponds to the average frame SNR. The three frames at the beginning and end of this speech segment

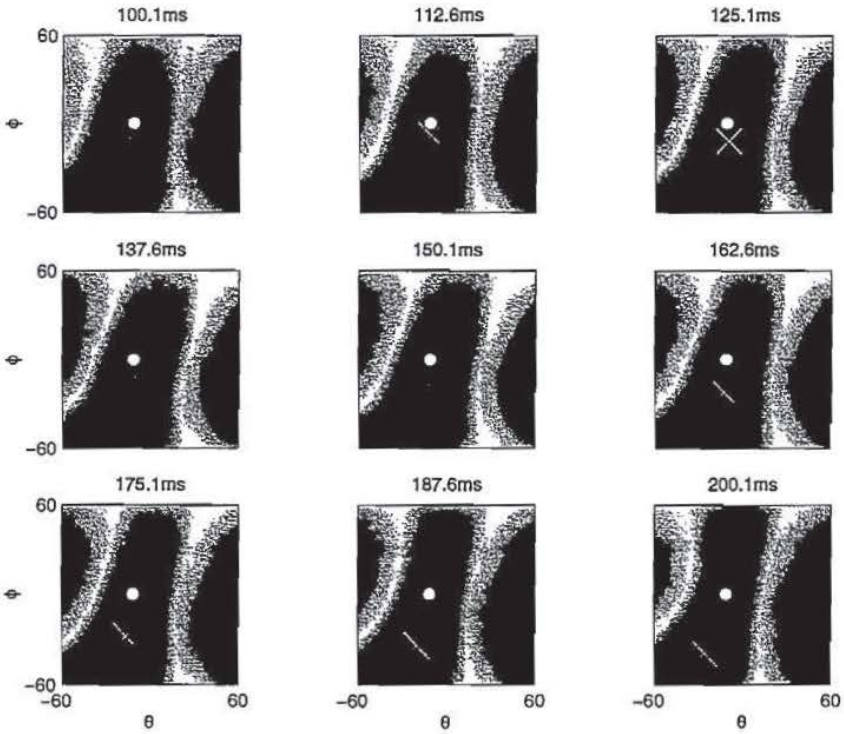


Fig. 8.4. Delay-and-sum beamformer SRP over nine, 25 ms frames.

lacked sufficient SNR to included in the analysis. These plots show that the LS-TDOA error is generally a smooth surface with a global minimum over the angular range of ±60°. However, from frame to frame the minima vary from the true source location. This inaccuracy is caused by erroneous TDOA estimates. Note also that because of the smooth nature of the error space, the resolution of the DOA estimates is considerably limited.

Figures 8.4 and 8.5 illustrate the error spaces of the SRP and SRP-PHAT as evaluated for the same nine 25 ms frames of speech. Relative to the prior figure the contour images are now inverted in darkness to emphasize the maxima. The plots of the delay-and-sum beamformer SRP in Figure 8.4 bear a noticeable similarity in general shape to their LS-TDOA counterparts. The maximum value in each SRP image, marked by an X, occurs at points distant from the actual DOA, indicated by a white dot. The main beam of the delay-and-sum beamformer is broad and fluctuates considerably over the duration of the speech segment. As a result, many inaccurate location estimates are produced by this method. In contrast to the LS-TDOA and SRP cases, the peaks of SRP-PHAT plots in Figure 8.5 match the actual DOA almost exactly. The main beam of the PHAT beamformer is sharp and consistent over each frame. This produces contour images which appear quite different from the LS-TDOA and SRP versions. The PHAT filters, when applied to the filter-and-sum beamformer, yield an error space that is superior to that of

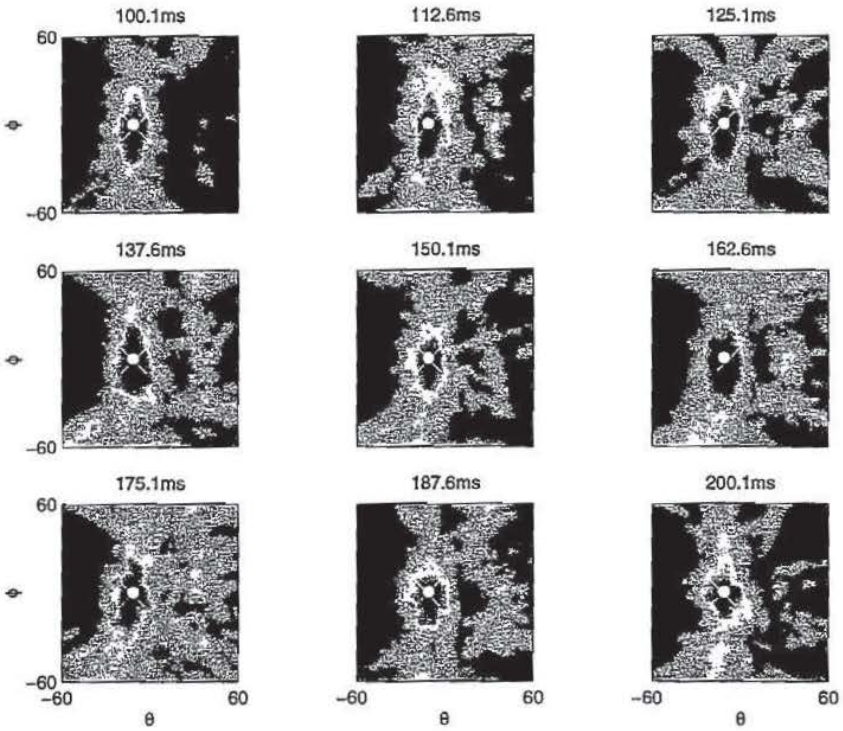


Fig. 8.5. SRP-PHAT response over nine, 25 ms frames.

the delay-and-sum beamformer or the TDOA-based criterion. This qualitative observation will now be corroborated through a numerical performance comparison.

For the DOA estimates produced for each of the three source locations, an RMS DOA error was computed from

$$E_{\mathrm{RMS}}(\hat{\theta}, \hat{\phi}) = \sqrt{(\hat{\theta} - \theta)^2 + (\hat{\phi} - \phi)^2} \tag{8.22}$$

where ϕ and θ are the true azimuth and elevation angles and $\hat{\phi}$ and $\hat{\theta}$ are their estimated counterparts. Figure 8.6 illustrates the results. These plots show the fraction of DOA estimates in each case which exceed a given RMS error threshold. Using this metric, the SRP-PHAT consistently outperforms the other two methods for each of the source locations. The ML-TDOA exhibits definite advantages over the SRP. While the SRP-PHAT's results are nearly identical for all the source locations, including the most distant source 1, the ML-TDOA locator is highly dependent on source location. For example, 60% percent of the estimates from source 1 had error greater than 10° while 50% percent from source 2 and 15% percent from source 3 had error greater 10°. In contrast, nearly all the estimates produced by SRP-PHAT had error less than 10°. About 90% of the estimates from sources 2 and 3, and 80% from source 1 had errors less than 4°.

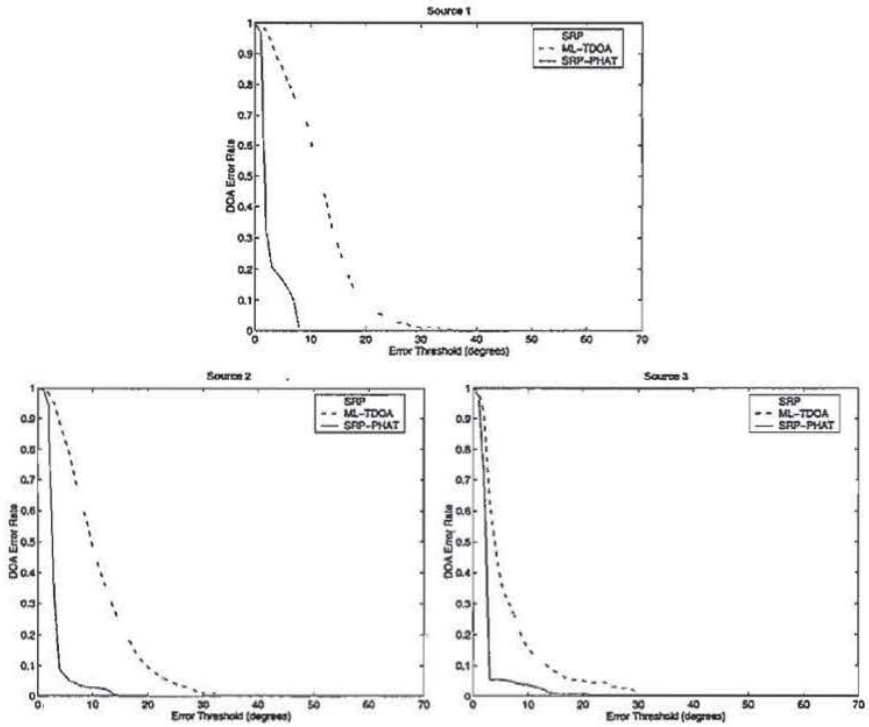


Fig. 8.6. Localizer DOA error rates for three different sources.

The results of this limited experiment illustrate the performance advantages of the SRP-PHAT localizer relative to more traditional approaches for talker localzation with microphone arrays. Other experiments conducted under more general and adverse conditions are consistent with the results here and serve to confirm the utility of combining steered-beamforming and a uniform-magnitude spectral weighting for this purpose.

While the TDOA-based localization method performed satisfactorily for a talker relatively close to the array, it was severely impacted by even the mild reverberation levels encountered when the source was more distant. This result is due to the fact that signal-to-reverberation ratios decrease with increasing source-to-microphone distance. As the reverberation component of the received signal increases relative to the direct path component, the validity of the single-source model inherent in the TDE development is no longer valid. As a result TDOA-based schemes rapidly exhibit poor performance as the talker moves away from the microphones. The SRP-PHAT algorithm is relatively insensitive to this effect. As the results here suggest the proposed algorithm exhibits no marked performance degradation from the near to distant source conditions tested.

The SRP-PHAT algorithm is computationally more demanding than the TDOA-based localization methods. However, its significantly superior performance may easily warrant the additional processing expense. Additionally,

while not discussed here, it is possible to alter the algorithm to dramatically reduce its computational load while maintaining much of its benefit.

References

1. H. Silverman, W. Patterson, J. Flanagan, and D. Rabinkin, "A digital processing system for source location and sound capture by large microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp. 251–254, April 1997.
2. M. Brandstein, J. Adcock, and H. Silverman, "Microphone array localization error estimation with application to sensor placement," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3807–3816, 1996.
3. J. Flanagan and H. Silverman, eds., *International Workshop on Microphone-Array Systems: Theory and Practice*, Brown University, Providence RI, USA, October 1992.
4. B. Radlovic, R. Williamson, and R. Kennedy, "On the poor robustness of sound equalization in reverberant environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-99)*, Phoenix AZ, USA, pp. 881–884, March 1999.
5. J. DiBiase, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments*, PhD thesis, Brown University, Providence RI, USA, May 2000.
6. W. Bangs and P. Schultheis, "Space-time processing for optimal parameter estimation," in *Signal Processing* (J. Griffiths, P. Stocklin, and C. V. Schooneveld, eds.), pp. 577–590, Academic Press, 1973.
7. G. Carter, "Variance bounds for passively locating an acoustic source with a symmetric line array," *J. Acoust. Soc. Am.*, vol. 62, pp. 922–926, October 1977.
8. W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inform Theory*, vol. IT-19, pp. 608–614, September 1973.
9. W. Hahn, "Optimum signal processing for passive sonar range and bearing estimation," *J. Acoust. Soc. Am.*, vol. 58, pp. 201–207, July 1975.
10. M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1210–1217, October 1983.
11. V. M. Alvarado, *Talker Localization and Optimal Placement of Microphones for a Linear MIcrophone Array using Stochastic Region Contraction*. PhD thesis, Brown University, Providence RI, USA, May 1990.
12. H. F. Silverman and S. E. Kirtman, "A two-stage algorithm for determining talker location from linear microphone-array data," *Computer, Speech, and Language*, vol. 6, pp. 129–152, April 1992.
13. D. Johnson and D. Dudgeon, *Array Signal Processing- Concepts and Techniques*, Prentice Hall, 1993.
14. S. Haykin, *Adaptive Filter Theory*, Prentice Hall, second ed., 1991.
15. R. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, PhD thesis, Stanford University, Stanford CA, USA, 1981.
16. J. Krolik, "Focussed wide-band array processing for spatial spectral estimation," in *Advances in Spectrum Analysis and Array Processing* (S. Haykin, ed.), vol. 2, pp. 221–261, Prentice Hall, 1991.

17. H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 823–831, August 1985.
18. K. Buckley and L. Griffiths, "Broad-band signal-subspace spatial-spectrum (BASS-ALE) estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. .36, pp. 953–964, July 1988.
19. A. Vural, "Effects of pertubations on the performance of optimum/adaptive arrays," *IEEE Trans. Aerosp. Electron.*, vol. AES-15, pp. 76–87, January 1979.
20. R. Compton Jr., *Adaptive Antennas*, Prentice Hall, 1988.
21. T. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation in coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806–811, August 1985.
22. M. Brandstein, J. Adcock, and H. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Trans. Speech Audio Proc.*, vol. 5, pp. 45–50, January 1997.
23. P. Svaizer, M. Matassoni, and M. Omologo, "Acoustic source location in a three-dimensional space using crosspower spectrum phase," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp. 231–234, April 1997.
24. Y. Huang, J. Benesty, and G. W. Elko, "Adaptive eigenvalue decomposition algorithm for realtime acoustic source localization system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-99)*, Phoenix AZ, USA, pp. 937–940, March 1999.
25. R. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron.*, vol. AES-8, pp. 821–835, November 1972.
26. J. Smith and J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1661–1669, December 1987.
27. H. Lee, "A novel procedure for accessing the accuracy of hyperbolic multilateration systems," *IEEE Trans. Aerosp. Electron.*, vol. AES-11, pp. 2–15, January 1975.
28. N. Marchand, "Error distributions of best estimate of position from multiple time difference hyperbolic networks," *IEEE Trans. Aerosp. Navigat. Electron.*, vol. .11, pp. 96–100, June 1964.
29. C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-24, pp. 320–327, August 1976.
30. M. Brandstein, J. Adcock, and H. Silverman, "A practical time-delay estimator for localizing speech sources with a microphone array," *Computer, Speech, and Language*, vol. 9, pp. 153–169, April 1995.
31. M. Brandstein and H. Silverman, "A practical methodology for speech source localization with microphone arrays," *Computer, Speech, and Language*, vol. 11, pp. 91–126, April 1997.
32. S. Bédard, B. Champagne, and A. Stéphenne, "Effects of room reverberation on time-delay estimation performance," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-94)*, Adelaide, Australia, pp. II:261–264, April 1994.
33. M. Brandstein and H. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp. 375–378, April 1997.

34. H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp. 187–190, April 1997.
35. M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event localization," *IEEE Trans. Speech Audio Proc.*, vol. 5, pp. 288–292, May 1997.
36. P. Svaizer, M. Matassoni, and M. Omologo, "Acoustic source location in a three-dimensional space using crosspower spectrum phase," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp. 231–234, April 1997.
37. M. Brandstein, "Time-delay estimation of reverberated speech exploiting harmonic structure," *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2914–2919, 1999.
38. A. Stéphenne and B. Champagne, "Cepstral prefiltering for time delay estimation in reverberant environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-95)*, Detroit MI, USA, pp. 3055–3058, May 1995.
39. N. Strobel and R. Rabenstein, "Classification of time delay estimates in reverberant environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-99)*, Phoenix AZ, USA, pp. 3081–3084, March 1999.
40. B. Friedlander, "A passive localization algorithm and its accuracy analysis," *IEEE Jour. Oceanic Engineering*, vol. OE-12, pp. 234–245, January 1987.
41. Y. Chan and K. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Processing*, vol. 42, pp. 1905–1915, August 1994.
42. D. Sturim, M. Brandstein, and H. Silverman, "Tracking multiple talkers using microphone-array measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp. 371–374, April 1997.
43. L. Kinsler, A. Frey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics*, John Wiley & Sons, third ed., 1982.
44. L. Ziomek, *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*, CRC Press, 1995.

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

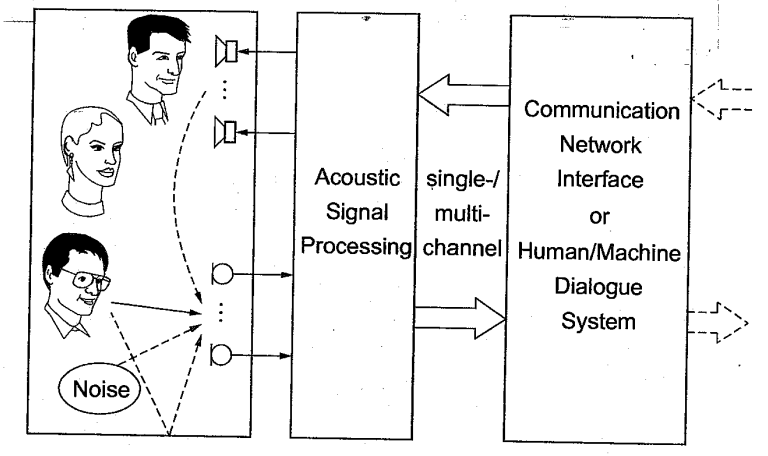


Fig. 13.1. Acoustic interface for natural human–machine communication.

echo cancellation (AEC) algorithms [1–3]. For discussing the combination of AEC with microphone arrays, the concept of AEC is first reviewed in Section 13.2 and beamforming methods are categorized in Section 13.3 with respect to the properties determining the interaction with AEC. Then, generic concepts for the combination of AEC and beamforming are discussed in Section 13.4. Structures for integrating AEC into beamforming are investigated in Section 13.5. Finally, the extension from single-channel reproduction to the case of multiple reproduction channels is outlined.

13.2 Acoustic Echo Cancellation

The concept of AEC is first considered for the case of a single loudspeaker and a single microphone according to Figure 13.2. To remove the echo from the microphone signal $x(n)$ (with n denoting discrete time), AEC aims at generating a replica $\widehat{v}(n)$ for the signal $v(n)$, which is an echoed version of the loudspeaker signal $u(n)$. Aside from the echo $v(n)$, $x(n)$ contains components originating from local desired sources and local interferers, $s(n)$ and $r(n)$, respectively. Introducing the residual echo

$$e(n) = v(n) - \widehat{v}(n), \tag{13.1}$$

the estimate for the desired signal $\widehat{s}(n)$ can be written as:

$$\widehat{s}(n) = x(n) - \widehat{v}(n) = s(n) + e(n) + r(n). \tag{13.2}$$

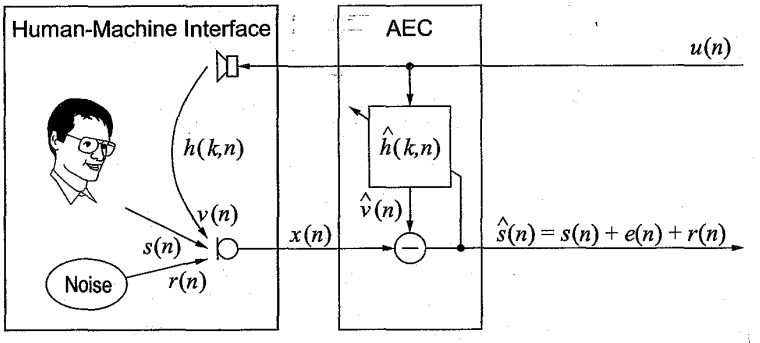


Fig. 13.2. Basic structure for single-channel AEC.

The amount of echo attenuation achieved by AEC is expressed by the *echo return loss enhancement* ($ERLE$)[1]:

$$ERLE_{log}(n) = 10 \cdot \log \frac{\mathcal{E}\left\{v^2(n)\right\}}{\mathcal{E}\left\{e^2(n)\right\}} \quad [\text{dB}], \tag{13.3}$$

with $\mathcal{E}\{\cdot\}$ denoting the expectation operator. As long as potential nonlinearities of the loudspeaker system can be neglected [4], the loudspeaker-enclosure-microphone(LEM) system is completely characterized by its generally time-varying impulse response $h(k,n)$. Indeed, the impulse response may vary drastically and unpredictably over time, as a slight change in position of any object can alter many coefficients significantly [2]. The number of impulse response samples that must be modeled for an $ERLE_{log}$ value of x dB is estimated by [2,5]

$$L_{AEC} \approx \frac{x}{60} \cdot f_s \cdot T_{60}, \tag{13.4}$$

where f_s denotes the sampling frequency, and T_{60} is the reverberation time[2]. Based on this estimate, more than $L_{AEC} = 1000$ impulse response coefficients must be perfectly matched to assure 20 dB of $ERLE_{log}$ for a typical office with $T_{60} = 400$ ms and an echo canceller operating at $f_s = 8$ kHz.

As a model for the LEM system, a digital FIR filter structure with a time-varying impulse response $\hat{h}(k,n)$ of length L_{AEC} is employed, so that the estimated echo $\hat{v}(n)$ is given by

$$\hat{v}(n) = \hat{\mathbf{h}}^T(n) \cdot \mathbf{u}(n) \tag{13.5}$$

[1] As $v(n)$ and $e(n)$ are not accessible in practical situations, $ERLE$ must be estimated from $\hat{s}(n)$ and $x(n)$ [2].

[2] As characteristic parameter of an enclosure, the reverberation time T_{60} is the time until the sound energy decays by 60dB after switching off the source.

where ^{T} denotes transposition and

$$\widehat{\mathbf{h}}(n) = \left[\widehat{h}(0,n), \widehat{h}(1,n), \ldots, \widehat{h}(L_{AEC}-1,n)\right]^T, \tag{13.6}$$

$$\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(n-L_{AEC}+1)]^T. \tag{13.7}$$

The misalignment between the FIR model $\widehat{\mathbf{h}}(n)$ and the LEM system $\mathbf{h}(n)$ is described by the logarithmic system error norm $D_{log}(n)$:

$$D_{log}(n) = 10 \cdot \log \frac{||\mathbf{h}(n) - \widehat{\mathbf{h}}(n)||_2^2}{||\mathbf{h}(n)||_2^2}, \tag{13.8}$$

with $||\cdot||_2$ denoting the l_2 norm[3].

13.2.1 Adaptation algorithms

For identifying the time-varying impulse response $h(k,n)$, adaptive filtering algorithms derive an optimum vector $\widehat{\mathbf{h}}_{opt}(n)$ by minimizing a mean square error criterion based on the input $u(n)$ and the estimation error $e(n)$ (assuming here, for simplicity, $s(n) = r(n) = 0$). Three fundamental algorithms are introduced below for the general case of complex signals (for a comprehensive treatment of adaptive FIR filtering see, e.g., [6,7]). Adaptation control in the context of AEC is addressed and frequency domain implementations are outlined briefly.

Fundamental algorithms. Minimizing the mean squared error $E\left\{|e(n)|^2\right\}$ for (at least) wide-sense stationary signals and a time-invariant echo path $h(k,n) = h(k)$ leads to the Wiener-Hopf equation for the optimum echo canceller $\widehat{\mathbf{h}}_{opt}$ [7]

$$\widehat{\mathbf{h}}_{opt} = \mathbf{R}_{\mathbf{uu}}^{-1} \cdot \mathbf{r}_{\mathbf{uv}} \tag{13.9}$$

with the time-invariant correlation matrix $\mathbf{R}_{\mathbf{uu}}$ and the crosscorrelation vector $\mathbf{r}_{\mathbf{uv}}$ given by

$$\mathbf{R}_{\mathbf{uu}} = E\left\{\mathbf{u}(n)\mathbf{u}^H(n)\right\}, \tag{13.10}$$

$$\mathbf{r}_{\mathbf{u}v} = E\left\{\mathbf{u}(n)v^*(n)\right\}, \tag{13.11}$$

respectively. (* denotes complex conjugation and ^{H} conjugate complex transposition.) For nonstationary environments, iterative or recursive algorithms are required to approach the Wiener solution in (13.9). As the most popular adaptation algorithm, the **NLMS**(*Normalized Least Mean Square*) algorithm [6,7] updates the filter according to

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \alpha \frac{\mathbf{u}(n)}{\mathbf{u}^H(n)\mathbf{u}(n)} e^*(n) \tag{13.12}$$

[3] If the length of $\mathbf{h}(n)$ is greater than L_{AEC}, then $\widehat{\mathbf{h}}(n)$ must be complemented with zeros accordingly.

and may be understood as a stochastic approximation of the steepest descent algorithm, with $\mathbf{u}(n)$ approximating the negative gradient vector, and a step-size parameter α, $0 < \alpha < 2$. Obviously, for correlated signals such as speech, $\mathbf{u}(n)$ will not cover uniformly the L_{AEC}-dimensional vector space, which implies that the convergence to minimum system error $D_{log}(n)$ in (13.8) is slow [7]. The popularity of the NLMS is based on its robust convergence behavior [2] and its low computational complexity (about $2L_{AEC}$ multiplications per sampling interval T (MUL's per T) are needed for implementing (13.1), (13.5), and (13.12)).

To improve the convergence for speech signals, the *Affine Projection Algorithm* (**APA**) uses P previous input vectors

$$\mathbf{U}(n) = [\mathbf{u}(n), \mathbf{u}(n-1), \ldots, \mathbf{u}(n-P+1)] \tag{13.13}$$

to compute an error vector

$$\mathbf{e}(n) = \mathbf{v}(n) - \mathbf{U}^T(n) \cdot \widehat{\mathbf{h}}^*(n), \tag{13.14}$$

where

$$\mathbf{e}(n) = [e(n), e(n-1), \ldots, e(n-P+1)], \tag{13.15}$$

$$\mathbf{v}(n) = [v(n), v(n-1), \ldots, v(n-P+1)]. \tag{13.16}$$

The filter coefficients are then updated according to

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \alpha \mathbf{U}(n) \left[\mathbf{U}^H(n)\mathbf{U}(n) - \delta \mathbf{I}\right]^{-1} \mathbf{e}^*(n), \tag{13.17}$$

with the regularization parameter δ ($\delta \geq 0$) and $\mathbf{I}$ denoting the identity matrix. Thus, the APA can be interpreted as a generalization of the NLMS algorithm, which in turn corresponds to an APA with $P = 1, \delta = 0$. The gradient estimate for the APA is equal to the projection of the system misalignment vector $\mathbf{h}(n) - \widehat{\mathbf{h}}(n)$ onto the P-dimensional subspace spanned by $\mathbf{U}(n)$. Thus, the complementary orthogonal component of the misalignment vector becomes smaller with increasing P. The computational complexity of the APA amounts to approximately $(P+1) \cdot L_{AEC} + O(P^3)$ MUL's per T, where, typically, $P = 2, \ldots, 32$, and L_{AEC} is given by (13.4). Fast versions of the APA reduce the computational load to $2L_{AEC} + 20P$, but require additional measures to assure numerical stability [2,6].

As the most powerful and computationally demanding adaptation method, the **RLS**(*Recursive Least Squares*) algorithm directly minimizes a weighted sum of previous error samples

$$J(\widehat{\mathbf{h}}, n) = \sum_{k=1}^{n} \beta(k) |e(k)|^2, \quad \text{with} \quad 0 < \beta \leq 1. \tag{13.18}$$

The solution has the form of (13.9), however with time-dependent estimates for $\mathbf{R}_{\mathbf{uu}}(n)$, $\mathbf{r}_{\mathbf{u}v}(n)$ given by

$$\widehat{\mathbf{R}}_{\mathbf{uu}}(n) = \sum_{k=1}^{n} \beta(k)\mathbf{u}(k)\mathbf{u}^H(k), \tag{13.19}$$

$$\widehat{\mathbf{r}}_{\mathbf{u}v}(n) = \sum_{k=1}^{n} \beta(k)\mathbf{u}(k)v^*(k). \tag{13.20}$$

The update equation reads here

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \widehat{\mathbf{R}}_{\mathbf{uu}}^{-1}(n)\mathbf{u}(n)e^*(n). \tag{13.21}$$

If an exponential window $\beta(k) = \lambda^{n-k}$ with the forgetting factor $0 < \lambda < 1$ is used, the inversion of $\widehat{\mathbf{R}}_{\mathbf{uu}}(n)$ is avoided by exploiting the matrix inversion lemma that allows recursive update of the inverse [7]. Then, the complexity of the RLS algorithm is on the order of L_{AEC}^2 MUL's per T [6]. Similarly to the APA, fast versions for the RLS algorithm have been proposed which reduce computational complexity to $7L_{AEC}$ MUL's per T. However, the large filter order L_{AEC} and the nonpersistent excitation $u(n)$ require extra efforts to assure stable convergence [6]. A simplified version of fast RLS algorithms is the *Fast Newton Algorithm* [6], which reduces the complexity to $L_{AEC} \cdot P$ MUL's per T, with P being a predictor order that should be matched to the correlation properties of the input $u(n)$. (For speech signals, $P \approx 10$ is a typical value at $f_s = 8$kHz.)

Adaptation control. Adaptation control has to satisfy two contradicting requirements. On one hand, changes in the echo path $h(k,n)$ should be tracked as fast as possible. This requires a large stepsize, α, for the NLMS and APA algorithms in (13.12) and (13.17)), and a rapidly decaying β for the RLS algorithm in (13.21), respectively. On the other hand, the adaptation must be robust to interfering local sources $s(n)$ and noise $r(n)$, which requires a small stepsize, α, and a slowly decaying β, respectively [2,7]. When a local talker is active, adaptation should be stalled immediately to avoid divergence of $\widehat{\mathbf{h}}(n)$. Therefore, a fast and reliable detection of local source activity and estimation of background noise levels is decisive for efficient AEC operation. Correspondingly, a significant amount of computational complexity is invested in monitoring parameters and signals which support adaptation control [2]. With properly tuned adaptation control, acoustic echoes are attenuated by, typically, about 25 dB of $ERLE_{log}$ during steady state using the above adaptation algorithms.

Frequency subband and transform domain structures. To reduce computational load and to speed up convergence of adaptation algorithms

which do not inherently decorrelate $u(n)$ (e.g., the NLMS algorithm), frequency subband and transform domain structures have been developed [1,8]. Subband structures decompose the fullband signals $u(n)$ and $x(n)$ into M subbands which are usually downsampled by $R < M$ [3,9]. The adaptive subband filters operate at a reduced sampling rate and require fewer coefficients which leads to overall computational savings by a factor of close to R^2/M compared to fullband adaptive filtering. After subtraction, the subband signals are synthesized to yield again a fullband signal $\widehat{s}(n)$. While the additional complexity for the analysis/synthesis filter banks is relatively small for large L_{AEC}, the introduced signal delay for $\widehat{s}(n)$ is objectionable in some applications [2,31].

Transform-domain structures draw their computational advantage over direct time-domain implementations from the fast Fourier transform (FFT) and its use for fast convolution [1,6,8]. Block-exact adaptation algorithms, which behave exactly like their time-domain counterparts, have been proposed for all the fundamental algorithms above. For the long impulse responses at issue, the system model $\widehat{h}(k,n)$ is often partitioned into shorter subsystems to reduce the signal delay [2].

13.2.2 AEC for multi-channel sound reproduction

Considering a multi-channel reproduction unit (see Figure 13.1) broadcasting K different sound channels $\mathbf{u}_\kappa(n)$ ($\kappa = 0, \ldots, K-1$) with usually time-varying mutual correlation, any microphone records the sum of K echo signals produced by different echo paths $h_\kappa(k,n)$,

$$v(n) = \sum_{\kappa=0}^{K-1} \mathbf{h}_\kappa(n)^T \cdot \mathbf{u}_\kappa(n), \tag{13.22}$$

with $\mathbf{h}_\kappa(n), \mathbf{u}_\kappa(n)$ being defined according to (13.6) and (13.7). Correspondingly, K echo cancellers, $\widehat{\mathbf{h}}_\kappa(n)$, are needed to model the respective echo paths. As only one error signal, $e(n)$, is available, the K inputs, $u_\kappa(n)$, must be mutually decorrelated without perceptible distortion to allow identification of the individual $\widehat{\mathbf{h}}_\kappa(n)$. This difference to single-channel AEC defines an even more challenging system identification problem, which has been considered only for the stereo case ($K = 2$) so far [1,10–12]. Current adaptation schemes still exhibit slower convergence and multiply computational load by more than K compared to their single-channel AEC counterparts.

13.2.3 AEC for multi-channel acquisition

A straightforward extension of the single-loudspeaker/ single-microphone scenario to an N-microphone acquisition system essentially multiplies the number of adaptive filters by N. The N-channel echo cancellation is captured by

extending the signals in (13.2) to N-dimensional column vectors,

$$\widehat{\mathbf{s}}(n) = \mathbf{x}(n) - \widehat{\mathbf{v}}(n) = \mathbf{s}(n) + \mathbf{r}(n) + \mathbf{e}(n) \tag{13.23}$$

$$= \mathbf{s}(n) + \mathbf{r}(n) + \mathbf{v}(n) - \widehat{\mathbf{H}}^T(n)\mathbf{u}(n) \tag{13.24}$$

with $\mathbf{u}(n)$ according to (13.7), with $\mathbf{e}(n), \mathbf{r}(n), \widehat{\mathbf{s}}(n), \mathbf{s}(n), \mathbf{v}(n), \widehat{\mathbf{v}}(n), \mathbf{x}(n)$ as column vectors of the form

$$\mathbf{x}(n) = [x_0(n), \ldots, x_{N-1}(n)]^T, \tag{13.25}$$

and with $\widehat{\mathbf{H}}(n)$ as a matrix containing the impulse responses $\widehat{\mathbf{h}}_\nu(n)$ as columns according to

$$\widehat{\mathbf{H}}(n) = \left[\widehat{\mathbf{h}}_0(n), \ldots, \widehat{\mathbf{h}}_\nu(n), \ldots, \widehat{\mathbf{h}}_{N-1}(n)\right]. \tag{13.26}$$

While this implies a corresponding multiplication of the computational cost for filtering, the cost for adaptation and its control is not necessarily multiplied by N. All operations depending only on the input data, $u(n)$, have to be carried out only once for all N channels, which would include the matrix inversion in the APA or RLS algorithms, (13.17) and (13.21), respectively. However, some fast versions draw their efficiency from interweaving matrix inversion and update equations [6] and, therefore, do not completely support this separation. Frequency subband and transform domain algorithms [1,6,8,9] support this separation at least by requiring the analysis transform of $u(n)$ only once for all channels.

13.3 Beamforming

This section only aims at categorizing beamforming algorithms with respect to their interaction with AEC. For a comprehensive treatment of fundamental techniques see, e.g., [13,14], while the current state of beamforming technology with microphone arrays is covered in several other chapters of this book.

13.3.1 General structure

Consider a microphone array capturing N real-valued sensor signals, $x_\nu(n)$, which are filtered by linear time-varying systems with impulse responses $g_\nu(k,n)$ and then summed up (Figure 13.3). The resulting beamformer output, $y(n)$, can be written as

$$y(n) = \mathbf{G}^T(n) \cdot \mathbf{X}(n) = \mathbf{G}^T(n) \cdot [\mathbf{S}(n) + \mathbf{R}(n) + \mathbf{V}(n)], \tag{13.27}$$

with the column vector $\mathbf{G}(n)$ representing the concatenated impulse response vectors $\mathbf{g}_\nu(n)$

$$\mathbf{G}(n) = \left[\mathbf{g}_0^T(n), \ldots, \mathbf{g}_{N-1}^T(n)\right]^T, \tag{13.28}$$

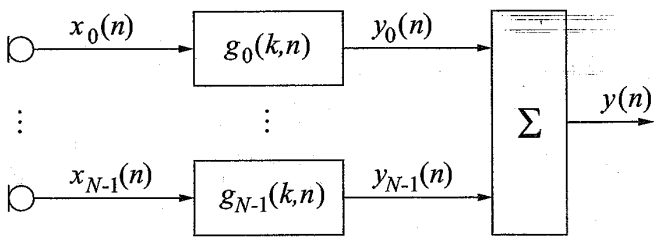


Fig. 13.3. General structure for a beamforming microphone array

where all $\mathbf{g}_\nu(n)$ are of length L_{BF}:

$$\mathbf{g}_\nu(n) = [g_\nu(0,n), \ldots, g_\nu(L_{BF}-1,n)]^T . \tag{13.29}$$

The column vector $\mathbf{X}(n)$ (and, equally, $\mathbf{R}(n), \mathbf{S}(n), \mathbf{V}(n)$) contains the latest L_{BF} signal samples of each microphone signal

$$\mathbf{X}(n) = \left[\mathbf{x}_0^T(n), \ldots, \mathbf{x}_{N-1}^T(n)\right]^T \tag{13.30}$$

with

$$\mathbf{x}_\nu(n) = [x_\nu(n), \ldots, x_\nu(n-L_{BF}+1)]^T . \tag{13.31}$$

In the scenario of Figure 13.1, beamforming aims at spatial filtering to dereverberate the components $\mathbf{s}(n)$ originating from the desired source(s) and to suppress interfering signals $\mathbf{r}(n)$ and echoes $\mathbf{v}(n)$.

For ideal dereverberation of a single source, the desired signal as it is emitted by the source, $s^{(0)}(n)$, should be retrieved except for some delay $n_0 > 0$:

$$\mathbf{G}^T(n) \cdot \mathbf{S}(n) = s^{(0)}(n-n_0). \tag{13.32}$$

Assuming that delayed versions of $s^{(0)}(n)$ are contained in $\mathbf{s}_\nu(n)$ defined by (13.31), the filters $g_\nu(k,n)$ have to equalize the corresponding delays and the sum of the filters has to provide a flat frequency response for all signals arriving from the source direction. Obviously, delay equalization requires knowledge about the location of the desired source(s). For the following, it is assumed that the source location is given by *a priori* knowledge or separately determined by some source localization algorithm (see, e.g., Chapters 8-10). For an anechoic environment and with the desired signal components being delay-equalized by the array geometry, the total impulse response, $g(k,n)$, of the beamformer to the desired source $s^{(0)}(n)$ should ideally fulfill

$$g(k,n) = \sum_{\nu=1}^{N} g_\nu(k,n) \stackrel{!}{=} \delta(k-k_0) \tag{13.33}$$

to assure a constant frequency response with unity gain and constant group delay k_0.

For interference suppression, the beamformer should minimize its response to all undesired signal components, which include here local interferers and loudspeaker echoes. Using, the mean squared error (MSE) as optimization criterion, this reads:

$$\mathcal{E}\left\{\left(\mathbf{G}^T(n)\cdot[\mathbf{R}(n)+\mathbf{V}(n)]\right)^2\right\} \stackrel{!}{=} \min. \tag{13.34}$$

Based on this general concept and with AEC in mind, basic methods for time-invariant or time-varying beamforming are outlined below.

13.3.2 Time-invariant beamforming

Time-invariant beamforming, i.e., $\mathbf{G}(n) = \mathbf{G}$, $\mathbf{g}_\nu(n) = \mathbf{g}_\nu$, is used for applications where the beamformer does not have to change the 'look direction' and where the potential nonstationarity of the involved signals, $\mathbf{s}(n), \mathbf{r}(n), \mathbf{v}(n)$, is not accounted for.

As the most basic beamforming method, the delay-and-sum beamformer (DSB) realizes in its simplest form a tapped delay line with a single non-zero coefficient for each filter $\mathbf{g}_\nu(n)$ [13,14]. If the required delays for the desired 'look direction' do not coincide with integer multiples of the sampling period, interpolation filters are required for realizing fractional delays [15–17]. Accounting for the wideband nature of speech and audio signals, nested arrays are often employed using different sets of sensors for different frequency bands to approximate a constant ratio between aperture width and signal wavelength [17–19]. As a generalization of DSB, filter-and-sum beamforming (FSB) aims for a frequency-independent spatial selectivity within each frequency band as detailed in Chapter 1 and [20]. Both beamforming concepts, DSB and FSB, were first developed on the basis of the far-field assumption [18], but may also be extended to near-field beamforming as described in Chapter 1. Time-invariant DSB and FSB are mostly signal-independent, i.e., no attention is paid to the power spectral densities of the signals $\mathbf{s}(n)$, $\mathbf{r}(n)$, $\mathbf{v}(n)$ and the direction of arrival (DOA) of interferers.

Such 'beamsteering' techniques are obviously appropriate for human–machine interfaces in reverberant environments with a restricted range of movement for a single desired source and where, due to reverberation, unwanted signal components of comparable level must be expected from all directions.

Nevertheless, time-invariant beamforming can incorporate additional spatial information to suppress dominant interferers [21,22]. Moreover, knowledge about long-term statistics of the noise field can be accounted for [23] and may lead to statistically optimum beamformers with superdirective behaviour for low frequencies as described in Chapter 2 and [24].

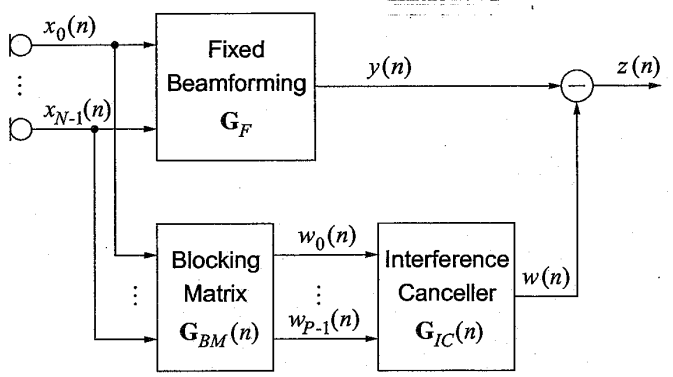


Fig. 13.4. Generalized sidelobe canceller structure for adaptive beamforming.

13.3.3 Time-varying beamforming

For nonstationary environments with both nonstationary signal characteristics and potentially moving sources, the beamformer should be able to track the time-variance of the signal characteristics and the spatial arrangement of the interfering sources. For that purpose, adaptive beamforming methods design filters $g_\nu(k,n)$ which minimize a statistical error criterion based on the array output, $y(n)$, with constraints for the DOA of a desired source (or 'target') such as formulated in (13.33) and (13.34) [13,14,25–27]. See also Chapter 5.

Generalized Sidelobe Canceller (GSC). As an example for an efficient implementation of adaptive beamformers that minimize a mean square error (MSE) criterion subject to a linear constraint, the generalized sidelobe canceller structure [13,25] is considered (Figure 13.4). Here, the adaptive beamforming is separated into two parallel paths: The upper path is a time-invariant, signal-independent beamformer, $\mathbf{G}_F$, steered toward the desired source. In the lower path, the first stage implements a blocking-matrix, $\mathbf{G}_{BM}(n)$, which, ideally, completely suppresses the components of the desired source, $\mathbf{s}(n)$, by a linear combination of the microphone channels [13] or filtering [28]. This topic is also detailed in Chapter 5. The $P \leq N$ outputs, $w_i(n), i = 0, \ldots, P-1$, are then used by the adaptive interference canceller, $\mathbf{G}_{IC}(n)$, to form an estimate for the interference component in $y(n)$. Optimization of $\mathbf{G}_{IC}(n)$ becomes an unconstrained Wiener filtering problem when the MSE criterion of (13.9) is used, and ideally leads to removal of all components in $y(n)$ which are correlated to $w_i(n)$. For identifying the optimum $\mathbf{G}_{IC}(n)$, the same adaptation algorithms as for echo cancellation can be used, i.e., (13.12),(13.17),(13.21), with gradient-type algorithms like the NLMS algorithm being most common.

13.3.4 Computational complexity

For both time-invariant and time-varying beamforming, the computational load is essentially proportional to the number of sensors N. The FIR filter lengths typically do not exceed $L_{BF} = 128$ [17,20,29,30]. With increasing filter length, computational savings are obtained by frequency-domain implementations of the filtering [20,29]. As with AEC, for adaptive beamforming implementations a significant share of computational complexity is dedicated to fast and reliable source activity detection which forms the basis of adaptation control.

13.4 Generic structures for combining AEC with beamforming

First, the combination of AEC with beamforming is motivated by comparing practical requirements with typical performance of AEC and beamforming. Then, the main properties of two generic options for a combination are discussed in some detail.

13.4.1 Motivation

Although AEC and beamforming are two distinct signal processing concepts, their goals meet with regard to acoustic echoes. While AEC subtracts from $\mathbf{x}(n)$ an echo estimate, $\hat{\mathbf{v}}(n)$, based on $u(n)$ as reference information, beamforming suppresses echoes within $\mathbf{x}(n)$ as undesired interference by its spatial filtering capability. With beamforming being undisputed for its effectiveness in suppressing local noise and reverberance of local desired sources, the need for a complementary AEC unit for acoustic echo suppression is discussed in the following.

As a guideline for desired echo suppression for telecommunication, [31] requires $ERLE_{log} \geq 45$ dB during single-talk and at least 30 dB during double-talk, assuming a 'natural' echo attenuation of up to 6 dB between the loudspeaker signal, $u(n)$, and the microphone signal, $x(n)$. Echo suppression methods other than AEC, e.g., noise reduction, loss insertion, or nonlinear devices, impair full-duplex communication and, thus, are only acceptable as supplementary measures [2]. For full-duplex speech dialogue systems employing automatic speech recognition, the echo attenuation requirements are not as well-defined and will depend on the desired recognition rate as well as on the robustness of the speech recognizer with respect to speech-like interference. In view of these requirements, the echo attenuation provided by microphone arrays and the echo path gain for a microphone array are examined below.

Array gain. The echo attenuation provided by a microphone array is usually identified with the array gain for the desired sources relative to echoes as interference. For signal-independent time-invariant beamforming, the directivity index quantifying the array gain of the desired direction over the average of all other directions [26] does typically not exceed 20 dB over a wide frequency range, and is much smaller at low frequencies (< 500 Hz) due to usual geometrical aperture constraints [19,26]. This contrasts with the fact that acoustic echoes usually exhibit their maximum energy at low frequencies [2]. As a remedy, differential beamforming realizes superdirective array gains at low frequencies and allows for a directivity index of up to 12 dB in practical implementations [1,27]. On the other hand, for adaptive beamforming, interference suppression is usually also limited to about 20 dB for reverberant environments if distortion of the desired source signal $s^{(0)}(n)$ should be precluded. See Chapters 2 and 5 as well as [19,32].

Echo path gain. For microphone array applications, the echo path gain between $u(n)$ and the beamformer output, $y(n)$, will often be higher than for single-microphone systems (−6 dB), because the sum of the distances from the loudspeaker to the listener, and from the desired source to the microphone array, will usually be greater (e.g. in teleconferencing). The user will typically increase the gains for the loudspeaker signal and the microphone array correspondingly to compensate for the decay of the sound level ($\approx$ 6 dB per doubling of distance in the far-field). If the microphone array and loudspeaker are relatively close, then the required echo attenuation will be increased accordingly.

13.4.2 Basic options

Restricting the scenario to a single reproduction channel, $u(n)$, and a single acquisition channel, $\widehat{s}(n)$, a combination of AEC and beamforming is obviously conceivable in two fundamentally different ways as shown in Figure 13.5. Here, 'AEC first' realizes one adaptive filter for each microphone in $\widehat{\mathbf{H}}^{(I)}(n)$ of (13.26), whereas 'Beamforming first' uses a single-channel AEC, $\widehat{\mathbf{h}}^{(II)}(n)$, which obviously has to include the beamformer, $\mathbf{G}(n)$, into its echo path model.

13.4.3 'AEC first'

This structure suggests that $\widehat{\mathbf{H}}^{(I)}(n)$ may operate without any repercussions from the beamforming so that the AEC problem corresponds to that described by (13.23). On the other hand, with perfect echo cancellation, the beamforming will be undisturbed by acoustic echoes and will concentrate on suppressing local interferers and reverberation.

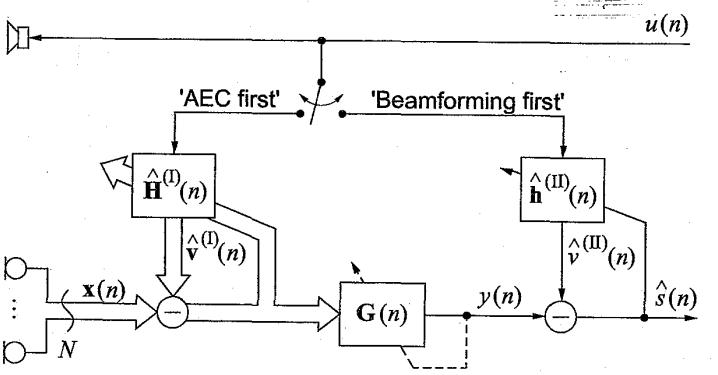


Fig. 13.5. Generic structures for combining AEC with beamforming.

AEC properties. Although AEC could operate independently from the beamforming, synergies with beamforming should be exploited with regard to detection of local source activity and computational complexity.

Local source activity detection. As noted above, the adaptation of $\widehat{\mathbf{H}}^{(I)}(n)$, requires a fast and reliable detection of local source activity to avoid divergence. With single-channel AEC, the detection is based on comparing estimates for

$$Q_\nu(n) = \frac{E\left\{v_\nu^2(n)\right\}}{E\left\{(r_\nu(n) + s_\nu(n))^2\right\}} \tag{13.35}$$

to a given threshold. With subsequent beamforming, this decision can be derived from estimates of

$$Q(n) = \frac{E\left\{\left(\mathbf{G}^T(n)\mathbf{V}(n)\right)^2\right\}}{E\left\{\left(\mathbf{G}^T(n)\left[\mathbf{R}(n) + \mathbf{S}(n)\right]\right)^2\right\}} \tag{13.36}$$

which reflect local source activity much clearer than $Q_\nu(n)$ as $\mathbf{r}_\nu(n), \mathbf{v}_\nu(n)$ are suppressed relative to $\mathbf{s}(n)$ by beamforming. Thus, $Q(n)$ reduces uncertainty in local source activity detection and allows adaptation during time intervals where adaptation might have been stalled if its control was based on $Q_\nu(n)$.

Computational complexity. At least the filtering and the filter coefficient update of the AEC adaptation will require N-fold computational cost compared to a single-channel AEC. Even with continuing growth of the performance-cost ratio of signal processing hardware, this computational load will remain prohibitive in the near future for many cost-sensitive or very large systems employing $N = 5, \ldots, 512$ sensors [17,19,26,30,33,34]. One option to alleviate the computational burden is to reduce the length L_{AEC} in (13.4) of the FIR filter models, $\widehat{\mathbf{h}}_\nu$, and to rely on the beamformer for suppressing the residual

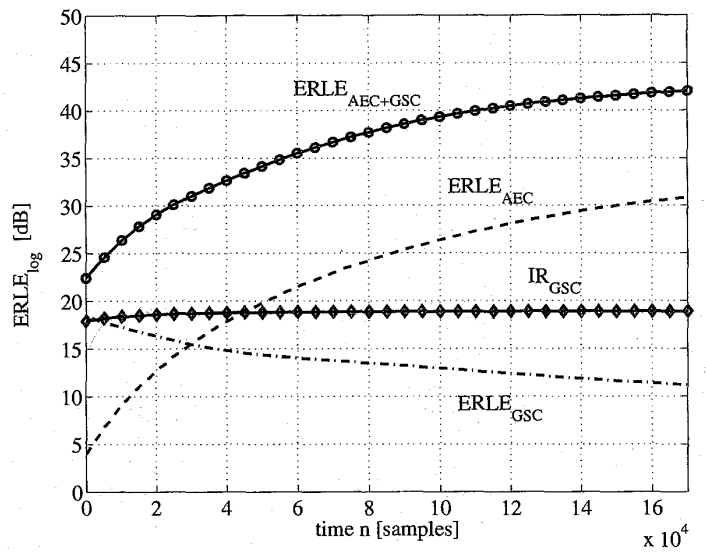


Fig. 13.6. Example for convergence of $ERLE_{log}$ components and local interference suppression(IR) for 'AEC first' structure ($N = 8$, $T_{60} \approx 300$ ms, $f_s = 12$ kHz, $L_{AEC} = 2500$, $L_{BM} = 16$, $L_{IC} = 50$).

echoes, $\mathbf{e}(n)$. Shortening $\widehat{\mathbf{h}}_\nu$ implies however, that the adaptation of the AEC is disturbed by an increased noise component, which is due to the unmodeled tail of the true echo path impulse response, $\mathbf{h}_\nu(n)$ [2].

Beamforming performance. For a signal-independent beamformer, the presence and performance of the AEC has no impact on the beamforming. The signal-independent spatial filtering will increase echo suppression according to its directivity while suppression of local interferers remains unaffected.

Signal-dependent beamformers use $\mathbf{w}(n) = \mathbf{x}(n) - \widehat{\mathbf{v}}^{(I)}(n)$ for optimizing the beamforming filters $\mathbf{G}(n)$. Thereby, at the cost of local interference suppression, the beamformer will concentrate on suppressing echo components, $\mathbf{e}(n)$, if their levels exceed that of local interferers, $\mathbf{r}(n)$, and it will further suppress residual echoes as long as they are not negligibly small compared to the local interferers. For illustration, the typical convergence behaviour for 'AEC first' using a GSC beamformer is shown in Figure 13.6 for $\mathbf{r}(n)$, $\mathbf{s}(n)$, $u(n)$ being white noise signals, and for alternating adaptation of $\mathbf{G}_{IC}(n)$, and $\widehat{\mathbf{H}}^{(I)}(n)$ (see also [32]). Due to its short filters, the beamformer converges almost instantaneously to about $ERLE_{GSC} = 18$ dB, and thereby provides a significant amount of $ERLE_{log}$ long before $\widehat{\mathbf{H}}^{(I)}(n)$ has converged. At the same time, suppression of local interference, IR_{GSC}, remains essentially con-

stant over time, as it converges very rapidly to almost 20 dB and is not allowed to converge much further to preclude distortion of the desired signal.

13.4.4 'Beamforming first'

In this structure, the beamformer is essentially independent from the AEC so that the beamforming performance agrees with Section 13.3 for acoustic echoes being perceived as another source of interference. AEC is realized by a single adaptive filter $\widehat{\mathbf{h}}^{(II)}(n)$ as in Figure 13.5 which is attractive with regard to computational complexity. However, the system identification problem faced by $\widehat{\mathbf{h}}^{(II)}(n)$ requires closer examination.

Echo path for AEC. Incorporating the beamformer, $\mathbf{G}(n)$, into the echo path model means that, ideally, the adaptive filter, $\widehat{\mathbf{h}}^{(II)}(n)$, models the sum of N echo paths from the loudspeaker input, $u(n)$, to the beamformer output, $y(n)$, (see Figure 13.3)

$$\widehat{\mathbf{h}}_{opt}^{(II)}(n) = \mathbf{f}(n) = \sum_{\nu=1}^{N} \mathbf{f}_\nu(n), \tag{13.37}$$

with the impulse responses, $\mathbf{f}(n)$, given by ($\star$ denotes linear convolution):

$$\mathbf{f}_\nu(n) = [f_\nu(0,n), \ldots, f_\nu(L_{AEC+BF}-1,n)]^T, \tag{13.38}$$

$$f_\nu(k,n) = h_\nu(k,n) \star g_\nu(k,n). \tag{13.39}$$

Thus, the impulse response length of $\widehat{\mathbf{h}}^{(II)}(n)$ depends on the beamforming, and, if any $g_\nu(k,n)$ is time-varying, $\widehat{\mathbf{h}}^{(II)}(n)$ has to track this time-variance as well[4]. The required length, L_{AEC+BF}, for $\widehat{\mathbf{h}}^{(II)}(n)$ is essentially the sum of the length L_{BF} and the necessary length for the acoustic path (including loudspeaker and microphone), L_{AEC}:

$$L_{AEC+BF} = L_{AEC} + L_{BF} - 1. \tag{13.40}$$

Note that for a given desired $ERLE_{log}$, L_{AEC} can be chosen smaller than given by (13.4) depending on the expected contribution of beamforming to $ERLE_{log}$ (see also [35]).

Signal-independent, time-invariant beamformers. Due to the time-invariance of $g_\nu(k,n)$, the adaptation of $\widehat{\mathbf{h}}^{(II)}(n)$ only has to track the time-variance of $\mathbf{h}_\nu(n)$ and, thus, the adaptation of $\widehat{\mathbf{h}}^{(II)}(n)$ is identical to the adaptation of one of the N filters $\widehat{\mathbf{h}}_\nu^{(I)}(n)$ in the 'AEC first' structure except for the different filter length L_{AEC+BF}.

[4] Note that the time-varying components $h_\nu(k,n)$ cannot be identified separately, although $g_\nu(k,n)$ is known ('knapsack problem').

Signal-dependent, time-varying beamformers. Here, the main problem is that the adaptation of $\widehat{\mathbf{h}}^{(II)}(n)$ has to track the time-variance of $\mathbf{G}(n)$. As for the adaptation algorithms discussed in Section 13.2.1 an increasing filter order involves a reduced tracking capability [7], the high-order filter, $\widehat{\mathbf{h}}^{(II)}(n)$, cannot follow the time-variance of the low-order filters of $\mathbf{G}(n)$ ($L_{AEC+BF} \gg L_{BF}$). Therefore, $\widehat{\mathbf{h}}^{(II)}(n)$ can find a useful echo path model only when $\mathbf{G}(n)$ remains time-invariant for a sufficiently long time. In Figure 13.7, the adaptation behaviour of the 'beamforming first' structure is analyzed for a speech conversation with a GSC as adaptive beamformer [28,32]. Inspecting the time domain signals $u(n)$ and $s(n)$ in Figures 13.7a and 13.7b shows that a 'double-talk' period occurs for time $n = 3.5 \ldots 4.0 \cdot 10^5$. Figure 13.7c illustrates which component is adapted at a given time. To track slight movements of the desired local source, the blocking matrix, $\mathbf{G}_{BM}(n)$, is adapted if only the local source and noise are present [28,32]. The system error of (13.8) depicted in Figure 13.7d converges monotonically when $\widehat{\mathbf{h}}^{(II)}(n)$ is adapted. When the interference canceller, $\mathbf{G}_{IC}(n)$, or the blocking matrix, $\mathbf{G}_{BM}(n)$, are adapted the system error rises instantaneously ($n = 2 \ldots 3.5 \cdot 10^5$). This is not critical as long as $u(n) = 0$, however, during double-talk ($n = 3.5 \ldots 4.0 \cdot 10^5$), a large residual error, $e(n)$, arises (Figure 13.7e,f) as $\widehat{\mathbf{h}}^{(II)}(n)$ cannot reconverge. Consequently, with the 'beamforming first' structure, the benefits of AEC are missing when they are desired most, i.e., during double-talk and during transitions from far-end activity to local activity and vice-versa (at other times primitive echo suppression methods, such as loss insertion [2], are less objectionable).

13.5 Integration of AEC into time-varying beamforming

As time-varying beamforming, $\mathbf{G}(n)$, cannot be tracked satisfactorily by the adaptation of $\widehat{\mathbf{h}}^{(II)}(n)$, a compromise is desirable for AEC which avoids the computational burden of $\widehat{\mathbf{H}}^{(I)}(n)$ for large N and provides improved echo cancellation compared to $\widehat{\mathbf{h}}^{(II)}(n)$. For this, the beamformer is decomposed into a time-invariant and a time-varying part in the sequel, with AEC acting only on the output of the time-invariant part. Two options for arranging the time-invariant and the time-varying stage are examined: First, a cascade with the time-invariant stage followed by the time-varying stage, and second, a parallel arrangement of the two stages.

13.5.1 Cascading time-invariant and time-varying beamforming

As illustrated in Figure 13.8, instead of a single beamformer output, $y(n)$, (see Figure 13.3), $M < \ldots \ll N$ beamformer output signals $\mathbf{y}(n) =$

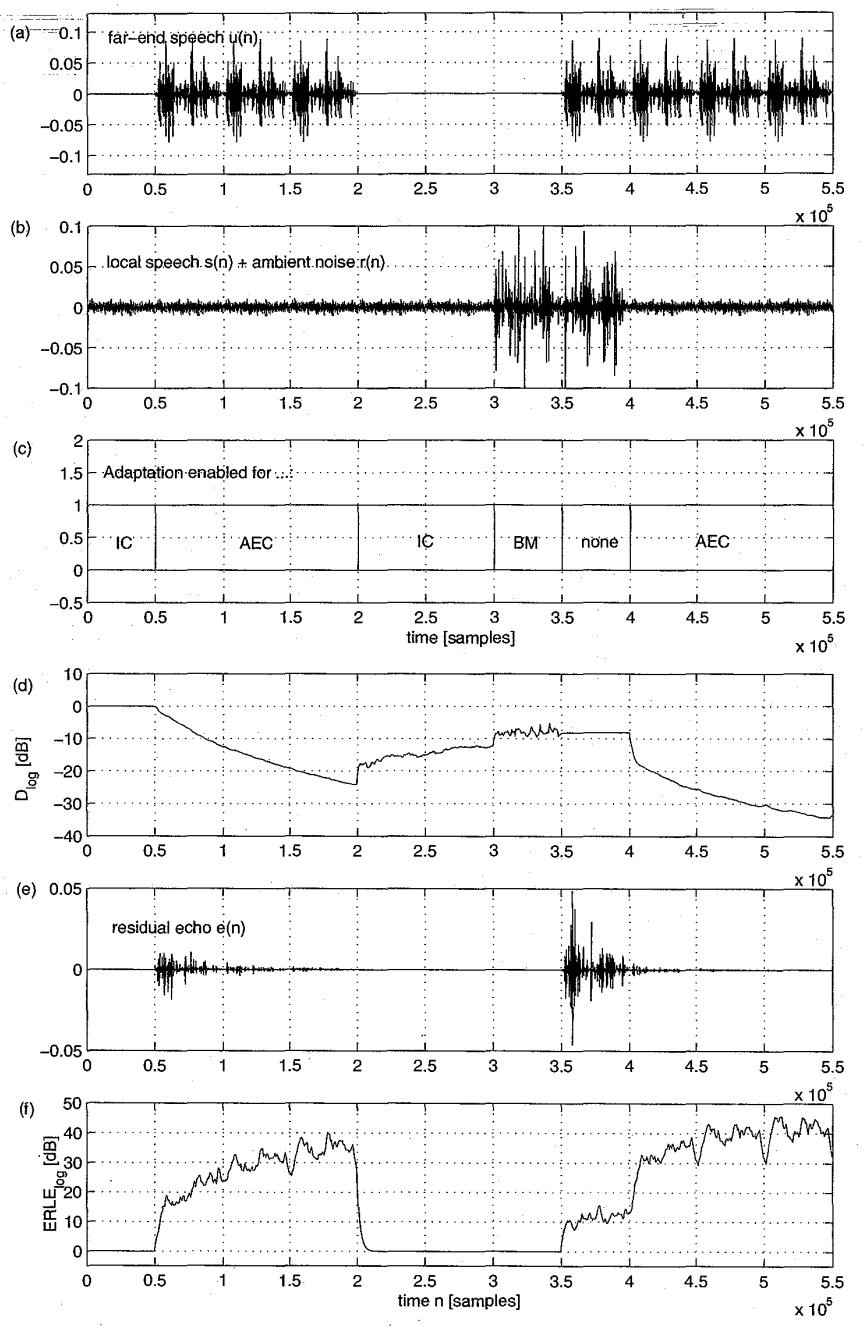


Fig. 13.7. Adaptation of $\widehat{\mathbf{h}}^{(II)}(n)$ in 'beamforming first' structure ($N = 8$, $T_{60} \approx 50$ ms, $f_s = 12$ kHz, $L_{AEC+BF} = 300$, $L_{BM} = 16$, $L_{IC} = 50$, adaptation by NLMS algorithm)

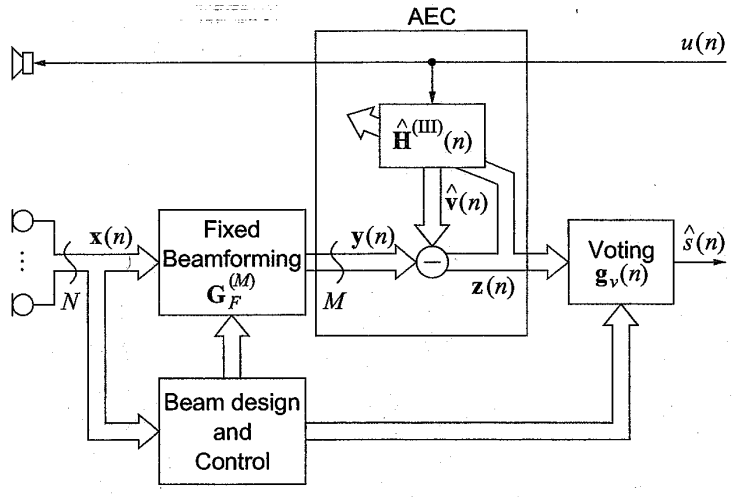


Fig. 13.8. AEC integrated into cascaded beamforming.

$[y_0(n), \ldots y_{M-1}(n)]^T$ are produced by M sets of fixed beamforming filters $\mathbf{G}_F^{(M)}$ according to

$$\mathbf{y}(n) = \mathbf{G}_F^{(M)^T} \cdot \mathbf{X}(n), \tag{13.41}$$

where $\mathbf{X}(n)$ is given by (13.30) and

$$\mathbf{G}_F^{(M)} = \left[\mathbf{G}_{F,0}^T, \ldots, \mathbf{G}_{F,\mu}^T, \ldots, \mathbf{G}_{F,M-1}^T\right] \tag{13.42}$$

with $\mathbf{G}_{F,\mu}$ according to (13.28). For AEC, $\widehat{\mathbf{H}}^{(III)}(n)$ realizes M adaptive echo cancellers $\widehat{\mathbf{h}}_\mu(n), \mu = 0, \ldots, M-1$, which exhibit the same performance as $\widehat{\mathbf{h}}^{(II)}(n)$ with time-invariant $\mathbf{G}(n)$ (see Section 13.4.4). Thus, if $M < N$ and $L_{AEC+BF} \approx L_{AEC}$, AEC operates at a reduced computational cost compared to $\widehat{\mathbf{H}}^{(I)}(n)$ (see Section 13.4.3). The time-varying part of the beamforming implements a weighted sum ('voting') using time-varying weights, $g_{v,\mu}(n)$:

$$\widehat{s}(n) = \mathbf{g}_v^T(n) \cdot \mathbf{z}(n) \tag{13.43}$$

with

$$\mathbf{g}_v(n) = [g_{v,0}(n), \ldots, g_{v,\mu}(n), \ldots, g_{v,M-1}(n)]^T, \tag{13.44}$$

$$\mathbf{z}(n) = [z_0(n), \ldots, z_\mu(n), \ldots, z_{M-1}(n)]^T. \tag{13.45}$$

Fixed beamformer design. The fixed beamformers of $\mathbf{G}_F^{(M)}$ may be designed to account for various situations, for instance, different beamformers could be employed for the presence or absence of echo, $\mathbf{v}(n)$, and of certain

local interferers, $\mathbf{r}(n)$. This concept is easily extended to cover several desired sources or moving desired sources, which is especially attractive for teleconferencing [5,17,18,22,26]. For the actual design of $\mathbf{G}_{F,\mu}$, techniques based on both time-invariant or time-varying beamforming can be applied. Updating may be attractive to allow for long-term flexibility.

$\mathbf{G}_F^{(M)}$ *based on time-invariant beamforming.* As a straightforward approach, $M_0 > M$ signal-independent fixed beams may be formed to cover several possible interference scenarios and/or all possible desired source positions. The output of these M_0 beamformers is monitored and a subset of M beamformers is used for $\mathbf{G}_F^{(M)}(n)$ to produce potentially desired signals $\mathrm{y}(n)$. As an example, in a teleconferencing studio with $M_0 = 7$ seats and three local participants being present, only $M = 3$ beams should produce desired signals (for examples see [17,18,22,26]).

$\mathbf{G}_F^{(M)}$ *based on adaptive beamforming.* Signal-dependent adaptive beamforming can be used to identify fixed beamformers for typical interference scenarios. To this end, an adaptive beamformer operates at a normal adaptation rate with its filter coefficients acting as a training sequence for finding M representative fixed beamformers. *A priori* knowledge of the desired source location(s) for incorporating constraints is necessary as well as initial training [5].

Initializing and updating $\mathbf{G}_F^{(M)}$. The monitoring of M_0 fixed beams, or the learning of optimum beamformers for deciding upon $\mathbf{G}_F^{(M)}$ can be carried out during an initial training phase only, or continuously. Continuous monitoring is recommended when changes are expected that demand the updating of $\mathbf{G}_F^{(M)}$. Monitoring of M_0 beams helps also to establish reliable estimates for background noise levels and supports detection of local talker activity so that convergence speed and robustness of AEC adaptation can be improved. Generally, as long as updating of $\mathbf{G}_F^{(M)}$ occurs less frequently than significant changes in the acoustic path, the model of time-invariant beamforming is justified with respect to AEC behavior. Aiming at minimum computational complexity for AEC, more frequent updates of $\mathbf{G}_F^{(M)}$ may be accepted for reduced M. The update should preferably occur at the beginning of 'far-end speech only' periods, as then, the AEC $\hat{\mathbf{H}}^{(III)}(n)$ can immediately adapt to the new echo path.

Voting. The time-varying weights, $g_{v,\mu}(n)$, in (13.44) must be chosen to allow for a fast reaction ($\leq$ 20 ms) to newly active local sources or changing interference scenarios, while at the same time avoiding the perception of switching, e.g., by applying a sigmoïd-like gain increase over time. For maximum spatial selectivity, only one beam signal should have a nonzero weight,

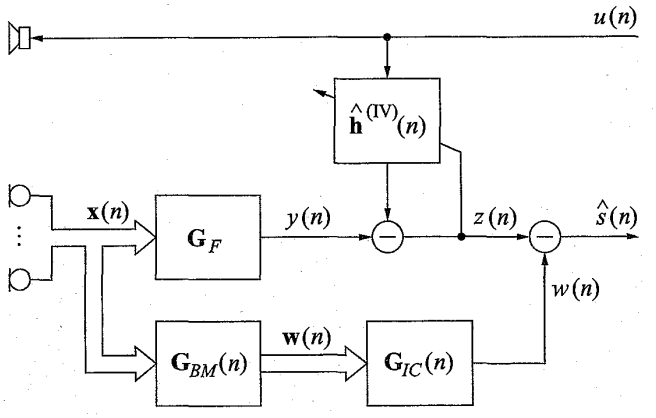


Fig. 13.9. GSC with embedded AEC.

$g_{v,\mu}(n)$, in the stationary case. Frequent toggling between beams is subjectively objectionable and should be prevented by hysteresis mechanisms (see also [17,26]).

13.5.2 AEC with GSC-type beamforming structures

As a popular representative of adaptive beamformers, the GSC (see Section 13.3.3) is also an example for a parallel arrangement of time-varying and time-invariant beamforming. If an integrated AEC should only see time-invariant beamforming in the echo path, it has to act on the output of the fixed beamformer, $y(n)$, as depicted in Figure 13.9 [32]. Obviously, only a single adaptive filter, $\widehat{\mathbf{h}}^{(IV)}(n)$, is necessary which faces the same system identification task as $\widehat{\mathbf{h}}^{(II)}(n)$ for time-invariant beamforming (see Section 13.4.4), which in turn is essentially identical to the plain single-microphone AEC problem. However, residuals of acoustic echoes, $\mathbf{v}(n)$, will also be contained in $w(n)$ unless they are eliminated by $\mathbf{G}_{BM}(n)$ or $\mathbf{G}_{IC}(n)$. Here, leaving echo suppression to the interference canceller, $\mathbf{G}_{IC}(n)$, seems to be the obvious solution. Recall that $\mathbf{G}_{IC}(n)$ minimizes the quadratic norm of $\widehat{s}(n)$ to remove all components from $z(n)$ that are correlated with $\mathbf{w}(n)$. If $\widehat{\mathbf{h}}^{(IV)}(n)$ is perfectly adjusted, no echo components remain in $z(n)$ and the echo estimate within $w(n)$ should be zero. On the other hand, local interference components in $\mathbf{w}(n)$ should be linearly combined using nonzero filter coefficients, so that $w(n)$ can remove interference residuals from $z(n)$. Clearly, a conflict in the design of $\mathbf{G}_{IC}(n)$ arises [32].

For illustration, consider a stationary situation for a given frequency, ω_0, in a 2-D plane containing a linear beamforming array with time-invariant $\mathbf{G}_F$, $\mathbf{G}_{BM}$, and $\mathbf{G}_{IC}$. A local interferer, $\mathbf{r}(n)$, arrives as a planar wave from θ_0 and passes the blocking matrix which is transparent for $\mathbf{r}(n)$ ($\mathbf{G}_{BM}^T \cdot \mathbf{r}(n) = \mathbf{r}(n)$).

Then, for perfect interference cancellation, $\mathbf{G}_{IC}(n)$ has to model the response of the fixed beamformer, $\mathcal{F}\{\mathbf{G}_{IC}\}(\theta_0,\omega_0) = \mathcal{F}\{\mathbf{G}_F\}(\theta_0,\omega_0)$, with $\mathcal{F}\{\mathbf{G}_{(\cdot)}\}(\theta,\omega)$ denoting the frequency response for a plane wave of frequency ω arriving from θ. If, on the other hand, an acoustic echo arrives from the same direction, θ_0, with nonzero spectral support at ω_0, this should be perfectly suppressed if $z(n)$ contains no echo, which means $\mathcal{F}\{\mathbf{G}_{IC}\}(\theta_0,\omega_0) = 0$. Obviously, this conflict requires a compromise at the cost of either local interference suppression or echo attenuation. Here, adaptation algorithms will automatically favor the dominant signal component in $\mathbf{w}(n)$. Even if echo and local interference do not arrive from the same direction, the finite number of degrees of freedom limits the ability of $\mathbf{G}_{IC}$ to suppress echo and local interference simultaneously. This is especially true for reverberant environments which possess a very large (if not infinite) number of DOAs for both echoes and local interference.

To avoid the conflict of interests within $\mathbf{G}_{IC}$, a suppression of the acoustic echoes, $\mathbf{v}(n)$, using $\mathbf{G}_{BM}(n)$ seems attractive. Considering the options, it is obvious that the output, $\mathbf{w}(n)$, should be freed from $\mathbf{v}(n)$ without suppressing $\mathbf{r}(n)$ or impairing the suppression of $\mathbf{s}(n)$. This means that no additional filtering on $\mathbf{x}(n)$ is allowed. As an alternative, estimates for the echoes, $\mathbf{v}(n)$, could be subtracted from $\mathbf{w}(n)$, which requires one adaptive filter for each of the $P \leq N$ channels and is similar to the generic concept of Section 13.4.3.

13.6 Combined AEC and beamforming for multi-channel recording and multi-channel reproduction

Multi-channel recording means that the output of the acquisition part of the acoustic interface in Figure 13.1 consists of several ($L > 1$) channels which, e.g., are necessary to convey spatial information for remote multi-channel sound reproduction, but may also support local signal processing. In Figures 13.5, 13.6, and 13.7 this translates to an L-dimensional output vector $\widehat{\mathbf{s}}(n)$. With respect to the beamforming, it means a duplication of the filtering and adaptation for each channel using the techniques outlined in Section 13.3. Both, time-invariant and adaptive beamforming will typically use L different 'look directions.' Regarding the generic methods to combine AEC with beamforming (Section 13.4), this means that for the 'AEC first' structure, the AEC part, $\widehat{\mathbf{H}}^{(I)}(n)$, remains unchanged while only the beamforming has to be duplicated. For the 'beamforming first' structure, the AEC realized by $\widehat{\mathbf{h}}^{(II)}(n)$ has to be duplicated as well.

When AEC is integrated into cascaded beamforming (see Section 13.5.1) the extension to the multi-channel recording case is included in the concept. The number of parallel fixed beams simply must equate or exceed the number of recorded channels, $M \geq L$, and the voting must be chosen accordingly. The AEC structure, $\widehat{\mathbf{H}}^{(III)}(n)$, remains unchanged. If the AEC is embedded into

a GSC-like structure, both the beamforming, $\mathbf{G}(n)$, and the AEC structure, $\widehat{\mathbf{h}}^{(IV)}(n)$, have to be implemented L times. However, removal of the acoustic echoes in the blocking matrix is necessary only once if performed directly on the microphone signals, $\mathbf{x}(n)$.

Multi-channel reproduction introduces a K-channel AEC problem (as described in Section 13.2.2), wherever a single echo cancellation filter is employed for single-channel reproduction, regardless of whether echo is to be removed from a microphone output or from a beamformer output. Essentially, this deteriorates convergence behavior and increases computational complexity for all structures discussed in Sections 13.4 and 13.5, accordingly.

Finally, for a system with both multi-channel reproduction and multi-channel recording as suggested in Figure 13.1, the complexity for combined AEC and beamforming obeys the superposition principle with respect to filtering and filter adaptation. Synergies are obtained by the common use of control information for several channels. The nature of the problems, however, does not change compared to the basic scenarios studied in Sections 13.2.2, 13.4, and 13.5 so that the corresponding results remain meaningful.

13.7 Conclusions

Beamforming and acoustic echo cancellation have been shown to jointly contribute to the suppression of acoustic feedback occurring in hands-free acoustic man-machine interfaces. While for time-invariant beamforming a single adaptive AEC filter suffices in the case of single-channel reproduction and single-channel recording, time-varying beamformers demand multiple adaptive filters if echo cancellation performance is not to degrade severely. However, realizing a time-varying beamformer as a cascade of time-invariant beamforming and time-varying voting requires only a few adaptive echo cancellers even for microphone arrays with many sensors. Implementing a combination of AEC and beamforming for a multi-channel reproduction and multi-channel recording system involves a corresponding increase in computational complexity. Signal processing performance, however, is still determined by the solutions for the elementary problems.

Acknowledgement

The author wishes to thank Wolfgang Herbordt for providing the simulation results and Susanne Koschny for preparing the illustrations.

References

1. S.L. Gay and J. Benesty, eds., *Acoustic Signal Processing for Telecommunication*, Kluwer, 2000.

2. C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, July 1999.
3. M.M. Sondhi and W. Kellermann, "Echo cancellation for speech signals," in *Advances in Speech Signal Processing*, (S. Furui and M.M. Sondhi, eds.), Marcel Dekker, 1991.
4. A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, pp. 1747–1760, 2000.
5. W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp.219–222, Apr. 1997.
6. G.-O. Glentis, K. Berberidis, and S. Theodoridis. "Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 13–41, July 1999.
7. S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3rd edition, 1996.
8. J.J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
9. W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-88)*, New York NY, USA, pp.2570–2573, Apr. 1988.
10. M.M. Sondhi, D.R. Morgan, and J.L. Hall, "Stereophonic echo cancellation: An overview of the fundamental problem," *IEEE Signal Processing Letters*,vol. 2, no. 8, pp. 148–151, Aug. 1995.
11. S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-95)*, Detroit MI, USA, pp.3059–3062, May 1995.
12. J.Benesty, D.R. Morgan, and M.M. Sondhi, "A hybrid mono/stereo acoustic echo canceler," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 5, pp. 468–475, Sept. 1998.
13. B.D. Van Veen and K.M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
14. D.H. Johnson and D.E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, 1993.
15. R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice Hall, 1983.
16. R.G. Pridham and R.A. Mucci, "Digital interpolation beamforming for low-pass and bandpass signals," *Proceedings of the IEEE*, vol. 67, no. 6, pp. 904–919, June 1979.
17. W. Kellermann, "A self-steering digital microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-91)*, Toronto, Canada, pp.3581–3584, May 1991.
18. J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Am.*, vol. 78, no. 5, pp. 1508–1518, Nov. 1985.
19. C. Marro, Y. Mahieux, and K.U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, May 1998.

20. T. Chou, "Frequency-independent beamformer with low response error," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-95)*, Detroit MI, USA, pp.2995–2998, May 1995.
21. Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 34, no. 6, pp. 1391–1400, Dec. 1986.
22. P. Chu, "Desktop mic array for teleconferencing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-95)*, Detroit MI, USA, pp.2999–3002, May 1995.
23. M. Dahl, I. Claesson, and S. Nordebo, "Simultaneous echo cancellation and car noise suppression employing a microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp.239–242, Apr. 1997.
24. P. Chu, "Superdirective microphone array for a set-top videoconferencing system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp.235–238, Apr. 1997.
25. L.J. Griffiths and C.W. Jim, "An alternative approach to linear constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.
26. J.L. Flanagan, D.A. Berkley, G.W. Elko, J.E. West, and M.M. Sondhi, "Autodirective microphone systems," *Acustica*, vol. 73, pp. 58–71, 1991.
27. G. Elko, "Microphone array systems for hands-free telecommunication," *Speech Communication*, vol. 20, pp. 229–240, 1996.
28. O. Hoshuyama and A. Sugiyama, "A robust adaptive beamformer for microphone arrays with a blocking matrix unsing constrained adaptive filters, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-96)*, Atlanta GA, USA, pp.925–928, May 1996.
29. I. Claesson, S.E. Nordholm, B.A. Bengtsson, and P. Eriksson, "A multi-DSP implementation of a broad-band adptive beamformer for use in a hands-free mobile radio telephone," *IEEE Trans. on Vehicular Technology*, vol. 40, no. 1, pp. 194–202, Feb. 1991.
30. S. Oh, V. Viswanathan, and P. Papamichalis, "Hands-free voice communication in an automobile with a microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-92)*, San Francisco CA, USA, pp. I:281–I:284, Mar. 1992.
31. ITU-T recommendation G.167, *Acoustic echo controllers*, Mar. 1993.
32. W. Herbordt and W. Kellermann, "GSAEC - Acoustic echo cancellation embedded into the generalized sidelobe canceller," in *Signal Processing X: Theories and Applications (Proceedings of EUSIPCO-2000)*, Tampere, Finland, vol.3, pp.1843–1846, Tampere, Finland, Sept. 2000.
33. S. Nordebo, S. Nordholm, B. Bengtsson, and I. Claesson, "Noise reduction using an adaptive microphone array in a car – a speech recognition evaluation," in *Conference Recordings of the ASSP Workshop on Application of Digital Signal Processing to Audio and Acoustics*, New Paltz NY, USA, Oct. 1993.
34. H. Silverman, W. R. Patterson, J.L. Flanagan, and D. Rabinkin, "A digital processing system for source location and sound capture by large microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-97)*, Munich, Germany, pp.251–254, Apr. 1997.

35. W. Kellermann, "Some properties of echo path impulse responses of microphone arrays and consequences for acoustic echo cancellation," in *Conf. Rec. of the Fourth International Workshop on Acoustic Echo Control*, Røros, Norway, June 1995.

EXHIBIT B

BOOK

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

Brandstein, Michael, 1967-; Ward, Darren, 1970-

©2001

Request through Interlibrary Loan (UC Affiliates)

TOP
SEND TO
GET IT
DETAILS
VIRTUAL BROWSE
LINKS

Send to

EXPORT TO EXCEL | EMAIL RECORD | PERMALINK | CITATION | PRINT

EASYBIB | ENDNOTE BASIC | EXPORT RIS

Get It

Please sign in to check if there are any request options. Sign in

BACK TO LOCATIONS

LOCATION ITEMS

Out of library **Science Library ; TK5986 .M53 2001**
(1 copy, 0 available, 0 requests)

On loan until 12/12/2022 5:00:00 PM PST
Sign in for borrowing details
Location: Science Library TK5986 .M53 2001
Barcode: 31970021907909

Call Number Locator
Report a Problem

OTHER UC LIBRARIES

University of California Los Angeles
Available in institution

University of California, Santa Barbara
Available in institution

University of California Berkeley
Available in institution

Details

Title	Microphone arrays : signal processing techniques and applications / Michael Brandstein, Darren Ward (eds.).
Format	xviii, 398 pages : illustrations ; 25 cm.
Publication Date	©2001
Publisher	New York : Springer
Series	Digital signal processing Digital signal processing (Springer-Verlag)
Contents	pt. I. Speech Enhancement. 1. Constant Directivity Beamforming / Darren B. Ward, Rodney A. Kennedy and Robert C. Williamson. 2. Superdirective Microphone Arrays / Joerg Bitzer and K. Uwe Simmer. 3. Post-Filtering Techniques / K. Uwe Simmer, Joerg Bitzer and Claude Marro. 4. Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields / Gary W. Elko. 5. Robust Adaptive Beamforming / Osamu Hoshuyama and Akihiko Sugiyama. 6. GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement / Simon Doclo and Marc Moonen. 7. Explicit Speech Modeling for Microphone Array Speech Acquisition / Michael Brandstein and Scott Griebel -- pt. II. Source Localization.
Subject	Microphones microphones Microphone Microfones Processamento de sinais
Contributor	Brandstein, Michael, 1967- Ward, Darren, 1970-
Identifier	ISBN : 3540419535 ISBN : 9783540419532 ISBN : 3642075479 ISBN : 9783642075476
Language	English
Source	Library Catalog

Virtual Browse

neto-
tive and
valve
ls :
amentals ...
02

Magnetic tape storage and handling : a guide for libraries and ...
1995

Design and analysis of magnetoresistive recording heads ...
©2001

Microphone arrays : signal processing techniques and ...
©2001

NIJ standa
for emerge
vehicle sire
a voluntar
national
1981

Links

Display Source Record

EXHIBIT C

```
leader  03185cam a22006614a 4500
001     9912716521506531
005     20220627213022.0
008     010417s2001    nyua     b    001 0 eng
010     ##$a  2001031428
015     ##$aGBA146095 $2bnb
016     7#$a961270640 $2DE-101
020     ##$a3540419535 $q(alk. paper)
020     ##$a9783540419532 $q(alk. paper)
020     ##$a3642075479
020     ##$a9783642075476
035     ##$a(CLU)4401276-ucladb-Voyager
035     ##$a(OCoLC)46791090 $z(OCoLC)47777122 $z(OCoLC)710021781
035     ##$a(OCoLC)ocm46791090
040     ##$aDLC $beng $cDLC $dC#P $dOHX $dUKM $dPUL $dBAKER $dBTCTA $dYDXCP $dOCLCG $dUBA $dIG# $dILU $dBDX $dOCLCF $dOCLCO $dOCLCQ $dDHA $dOCLCQ $dZ
042     ##$apcc
049     ##$aMAIN
050     00$aTK5986 $b.M53 2001
072     #7$aTK $2lcco
080     ##$a688.22
082     00$a621.382/84 $221
245     00$aMicrophone arrays : $bsignal processing techniques and applications / $cMichael Brandstein, Darren Ward (eds.).
260     ##$aNew York : $bSpringer, $c©2001.
300     ##$axviii, 398 pages : $billustrations ; $c25 cm.
336     ##$atext $btxt $2rdacontent
337     ##$aunmediated $bn $2rdamedia
338     ##$avolume $bnc $2rdacarrier
490     1#$aDigital signal processing
504     ##$aIncludes bibliographical references and index.
505     00$gpt. I. $tSpeech Enhancement. $g1. $tConstant Directivity Beamforming / $rDarren B. Ward, Rodney A. Kennedy and Robert C. Williamson. $g2.
650     #0$aMicrophone.
650     #6$aMicrophones.
650     #7$amicrophones. $2aat
650     #7$aMicrophone. $2fast $0(OCoLC)fst01019979
650     #7$aMicrofones. $2larpcal
650     #7$aProcessamento de sinais. $2larpcal
700     1#$aBrandstein, Michael, $d1967-
700     1#$aWard, Darren, $d1970-
830     #0$aDigital signal processing (Springer-Verlag)
908     ##$aWorldCat Daily Updates 2022-06-27 $bWorldCat record variable field(s) change: 650
962     ##$aGKM $9LOCAL
996     ##$a991027958079704701 $9LOCAL
996     ##$a.b28511517 $9LOCAL
```

EXHIBIT D

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	Brandstein, Michael, 1967- Ward, Darren, 1970-
LC Subjects	Microphone.
Browse by shelf order	TK5986
Notes	Includes bibliographical references and index.
Series	Digital signal processing Digital signal processing (Springer-Verlag)
LCCN	2001031428
Dewey class no.	621.382/84
Type of material	Book

Item Availability

CALL NUMBER	TK5986 .M53 2001 Copy 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

EXHIBIT E

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 Copy 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

EXHIBIT F

JOURNAL

IEEE transactions on signal processing : a publication of the IEEE Signal Processing Society.

c1991-

Check holdings **SEL/8270 Boelter Hall Stacks; Current issues in Current Jnls (TK5981 .I59t)**

Available Online

View Journal Contents

TOP
SEND TO
SEARCH INSIDE
ONLINE ACCESS
HOW TO GET IT
LINKS
DETAILS
RELATED RESOURCES

Send to

QR | PERMALINK | EMAIL | CITATION | EXPORT BIBTEX | ENDNOTE BASIC | EXPORT TO EXCEL | MENDELEY | REFWORKS | EXPORT RIS | PRINT

Search inside

Search for articles within this journal

Article title or keyword

Online access

Full availability (Access issues? Try using the VPN)

IEEE Electronic Library (IEL) SHOW LICENSE
Available from 1991 volume: 39 issue: 1.

IEEE Electronic Library (IEL) Journals
Available from 1991 volume: 39 issue: 1.

IEEE Xplore All Journals
Available from 1991 volume: 39 issue: 1.

How to get it

Please sign in to check if there are any request options. MY LIBRARY ACCOUNT

BACK TO LOCATIONS

LOCATION ITEMS

May be available. Contact the Library for access. **SEL/8270 Boelter Hall Stacks; Current issues in Current Jnls ; TK5981 .I59t**

Holdings: v.39-48(1991-2000)v.49:no.1-5,7-12(2001),v.50-55(2002-2007)v.56:no.1-4,6-10,12(2008),v.57:no.4-9,11-12(2009)

Note: v.57:no.7-9(2009) sent to bindery 4/13/10. Library does not have current subscription

Item in place (0 requests) v.57 no.11-12 yr.2009
Non-circulating

Item in place (0 requests) v.57 no.7-9 yr.2009
Non-circulating

Item in place (0 requests) v.57 no.4-6 yr.2009
Non-circulating

Item in place (0 requests) v.56 no.12 yr.2008
Non-circulating

Item in place (0 requests) v.56 no.10 yr.2008
Non-circulating

SHOW MORE ITEMS

OTHER UC LIBRARIES

Links

MARC Record

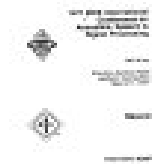
Details

Author / Contributor	IEEE Signal Processing Society.
Title	IEEE transactions on signal processing : a publication of the IEEE Signal Processing Society.
Other Title	Institute of Electrical and Electronics Engineers transactions on signal processing Transactions on signal processing Signal processing
Publication Information	New York, NY : Institute of Electrical and Electronics Engineers Vol. 39, no. 1 (Jan. 1991)-
Creation Date	c1991-
Related titles	Earlier title : IEEE transactions on acoustics, speech, and signal processing Available in other form: Online version: IEEE transactions on signal processing Available in other form: Online version: IEEE transactions on signal processing (Online)
Type	Journal
Physical Description	v. : ill. ; 28 cm.
Notes	Chemical abstracts 0009-2258 Jan. 1991- Title from cover.
Language	English
Subject	Signal processing -- Periodicals Speech processing systems -- Periodicals

Electro-acoustics -- Periodicals
Traitement du signal
Électroacoustique -- Périodiques
Traitement du signal -- Périodiques
Traitement automatique de la parole -- Périodiques
Electro-acoustics
Signal processing
Speech processing systems
Signaalverwerking

Genre	Periodicals.
Frequency	Monthly
Identifier	LC : 91642053 ISSN : 1053-587X OCLC : (OCoLC)22582582 OCLC : (OCoLC)ocm22582582 ISSN : 1941-0476
MMS ID	99246023606533
Source	Library Catalog
Former System Number	24602-ucladb

Related Resources



signal
essing
azine. ...
91-

IEEE ASSP magazine : a publication of the IEEE Acoustics, ...
©1984-1990

1977 IEEE International Conference on Acoustics, Speech, & ...
©1977

IEEE transactions on signal processing : a publication of ...
c1991-

IEEE transactio on audio a electroaco cs.
©1966-©1

EXHIBIT G

```
leader  04110cas a2200865 a 4500
001     9913117329106531
005     20220810213025.0
008     901025c19919999nyumr p       0   a0eng c
010     ##$a   91642053  $zsn 90004861
012     ##$i9104 $k1 $l1 $m1
016     7#$a9885223 $2DNLM
016     7#$aSR0070324 $2DNLM
022     0#$a1053-587X $l1053-587X $21
030     ##$aITPRED
032     ##$a256620 $bUSPS
035     ##$a(CLU)24602-ucladb-Voyager
035     ##$a(OCoLC)22582582 $z(OCoLC)173448432 $z(OCoLC)1012002384 $z(OCoLC)1013229445 $z(OCoLC)1029222376
035     ##$a(OCoLC)ocm22582582
037     ##$bIEEE Service Center, P.O. Box 1331, 445 Hoes Lane, Piscataway, NJ 08855-1331
040     ##$aNSD $beng $cNSD $dWAU $dNSD $dNYG $dDLC $dCOO $dDLC $dNST $dNSD $dELM $dHUL $dNST $dDLC $dCAS $dDLC $dMYG $dIUL $dNST $dIUL $dNST $dDLC $
042     ##$ansdp $apcc
049     ##$aMAIN
050     00$aTK5981 $b.I2
082     00$a621.382/2/05 $220
210     0#$aIEEE trans. signal process.
222     #0$aIEEE transactions on signal processing
245     00$aIEEE transactions on signal processing : $ba publication of the IEEE Signal Processing Society.
246     30$aInstitute of Electrical and Electronics Engineers transactions on signal processing
246     30$aTransactions on signal processing
246     30$aSignal processing
260     ##$aNew York, NY : $bInstitute of Electrical and Electronics Engineers, $cc1991-
300     ##$av. : $bill. ; $c28 cm.
310     ##$aMonthly
362     0#$aVol. 39, no. 1 (Jan. 1991)-
510     2#$aChemical abstracts $x0009-2258 $bJan. 1991-
588     ##$aLatest issue consulted: Vol. 39, no. 6 (June 1991.
500     ##$aTitle from cover.
599     ##$aUPDATED $bUclaCollMgr $bSerials1 $bMaster record variable field(s) change: 530, 588, 776 $9LOCAL
650     #0$aElectro-acoustics $vPeriodicals.
650     #0$aSignal processing $vPeriodicals.
650     #0$aSpeech processing systems $vPeriodicals.
650     #0$aSpeech processing systems $xPeriodicals.
650     #0$aSignal processing.
650     #0$aElectro-acoustics $xPeriodicals.
650     #6$aTraitement automatique de la parole $xPériodiques.
650     #6$aTraitement du signal.
650     #6$aÉlectroacoustique $xPériodiques.
650     #6$aÉlectroacoustique $vPériodiques.
650     #6$aTraitement du signal $vPériodiques.
650     #6$aTraitement automatique de la parole $vPériodiques.
650     #7$aElectro-acoustics. $2fast $0(OCoLC)fst00906339
650     #7$aSignal processing. $2fast $0(OCoLC)fst01118281
650     #7$aSpeech processing systems. $2fast $0(OCoLC)fst01129243
650     17$aSignaalverwerking. $2gtt
655     #7$aPeriodicals. $2fast $0(OCoLC)fst01411641
655     #7$aPeriodicals. $2lcgft $0http://id.loc.gov/authorities/genreForms/gf2014026139 $0(uri) http://id.loc.gov/authorities/genreForms/gf20140261
710     2#$aIEEE Signal Processing Society.
776     08$iOnline version: $tIEEE transactions on signal processing (Online) $x1941-0476 $w(DLC)  2008212918 $w(OCoLC)44608166
776     08$iOnline version: $tIEEE transactions on signal processing $w(OCoLC)623586529
780     00$tIEEE transactions on acoustics, speech, and signal processing $x0096-3518 $w(DLC)   74646881 $w(OCoLC)1156938
891     00$9853 $81 $av. $bno. $u12 $vr $i(year) $j(month) $wm $x01
891     40$9863 $81 $a<39>- $i<1991>- $xprovisional
850     ##$aAAP $aArU $aAzTeS $aCSt $aCU-A $aCU-I $aCU-RivP $aCU-SB $aCoFS CoU $aDGW $aDLC $aDeU $aICarbS $aIaU $aInLP $aInU $aMBU $aMCM $aMH-GM $aMc
908     ##$aWorldCat Daily Updates 2022-08-10 $bWorldCat record variable field(s) change: 040, 655
951     ##$aUclaCollMgr Serials1 20151030 $acdl 090811 $acdl 060324 $acdl 041212 $a20 Mar 91 reb $aMARS $9LOCAL
992     ##$aTC0208302 $cACB-3362 $9LOCAL
992     ##$c07-AOU-4792 $9LOCAL
996     ##$a24602-ucladb $9LOCAL
```

EXHIBIT H

PERIODICAL OR NEWSPAPER

IEEE transactions on signal processing : a publication of the IEEE ...

Full Record MARC Tags

Main title

IEEE transactions on signal processing : a publication of the IEEE Signal Processing Society.

Published/Created

New York, NY : Institute of Electrical and Electronics Engineers, c1991-

Request this Item LC Find It

More Information

LCCN Permalink	https://lccn.loc.gov/91642053
Publication history	Vol. 39, no. 1 (Jan. 1991)-
Description	v. : ill. ; 28 cm.
Current frequency	Monthly
ISSN	1053-587X
Linking ISSN	1053-587X
LC classification	TK5981 .I2

Continues	IEEE transactions on acoustics, speech, and signal processing 0096-3518 (DLC) 74646881 (OCoLC)1156938
Portion of title	Institute of Electrical and Electronics Engineers transactions on signal processing Transactions on signal processing Signal processing
Serial key title	IEEE transactions on signal processing
Abbreviated title	IEEE trans. signal process.
Related names	IEEE Signal Processing Society.
LC Subjects	Electro-acoustics--Periodicals. Signal processing--Periodicals. Speech processing systems--Periodicals. Speech processing systems--Periodicals. Signal processing. Electro-acoustics--Periodicals.
Other Subjects	Electro-acoustics. Signal processing. Speech processing systems. Signaalverwerking.
Form/Genre	Periodicals. Periodicals.
Browse by shelf order	TK5981
Notes	Title from cover.
Indexed selectively by	Chemical abstracts 0009-2258 Jan. 1991-
Reproduction no./Source	IEEE Service Center, P.O. Box 1331, 445 Hoes Lane, Piscataway, NJ 08855-1331
Additional formats	Online version: IEEE transactions on signal processing (Online) 1941-0476 (DLC) 2008212918 (OCoLC)44608166 Online version: IEEE transactions on signal processing (OCoLC)623586529
LCCN	91642053

Invalid LCCN	sn 90004861
CODEN	ITPRED
Dewey class no.	621.382/2/05
Postal reg no.	256620 USPS
National bib agency no.	9885223 SR0070324
Other system no.	(OCoLC)ocm22582582
Additional Links	http://ieeexplore.ieee.org/servlet/opac?punumber=78
Type of material	Periodical or Newspaper
Source of description	Latest issue consulted: Vol. 39, no. 6 (June 1991.

Item Availability

CALL NUMBER	TK5981 .I2 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	c.1 v. 58/, no. 3-4 2010 Mar-Apr In Process 06-11-2012
Older receipts	v.39-v.44 (1991-1996:Apr.), v.44:no.4-v.54:no.1 (1996:June-2006:Jan.) v.54:no.3-v.56:no.3 (2006:Mar.-2008:Mar.), v.56:no.7-v.62:no.1/4 (2008:July-2014:Jan./Feb.), v.62:no.9/12-v.65:no.9/12 (2014:May/June-2017:May/June), v.65:no.17/20 (2017:Sept./Oct.), v.66:no.1/4-v.67:no.4 (2018:Jan./Feb.-2019:Feb.)

CALL NUMBER	TK5981 .I2 Copy 1 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)
Latest receipts	v. 56, no. 6 (2008 June)

v. 56, no. 4 (2008 Apr.)

Supplements

"Content Gazette" (2010:June)

EXHIBIT I

PERIODICAL OR NEWSPAPER

IEEE transactions on signal processing : a publication of the IEEE ...

Full Record MARC Tags

000 04049cas a2200829 a 4500

001 11385724

005 20220812105052.0

008 901025c19919999nyumr p 0 a0eng c

010 __ |**a** 91642053 |**z** sn 90004861

016 7_ |**a** 9885223 |**2** DNLM

016 7_ |**a** SR0070324 |**2** DNLM

022 0_ |**a** 1053-587X |**l** 1053-587X |**2** 1

030 __ |**a** ITPRED

032 __ |**a** 256620 |**b** USPS

035 __ |**a** (OCoLC)ocm22582582

037 __ |**b** IEEE Service Center, P.O. Box 1331, 445 Hoes Lane, Piscataway, NJ 08855-1331

040 __ |**a** NSD |**b** eng |**c** NSD |**d** WAU |**d** NSD |**d** NYG |**d** DLC |**d** COO |**d** DLC |**d** NST |**d** NSD |**d** ELM |**d** HUL |**d** NST |**d** DLC |**d** CAS |**d** DLC |**d** MYG |**d** IUL |**d** NST |**d** IUL |**d** NST |**d** DLC |**d** MYG |**d** CUS |**d** MYG |**d** SYS |**d** IEEEX |**d** NLM |**d** OCLCQ |**d** NLGGC |**d** NSD |**d** OCLCQ |**d** OCLCF |**d** OCLCO |**d** OCLCA |**d** IUL |**d** OCLCO |**d** OCLCA |**d** CNMON |**d** OCLCO |**d** PUL

042 __ |**a** nsdp |**a** pcc

050 00 |**a** TK5981 |**b** .I2

082 00 |**a** 621.382/2/05 |**2** 20

210 0_ |**a** IEEE trans. signal process.

222 _0 |**a** IEEE transactions on signal processing

245 00 |**a** IEEE transactions on signal processing : |**b** a publication of the IEEE Signal Processing Society.

246 30 |**a** Institute of Electrical and Electronics Engineers transactions on signal processing

246 30 |**a** Transactions on signal processing

246 30 |**a** Signal processing

249 0_ |**i** ACQUIRE Title: |**a** IEEE transactions on signal processing : a publication of the IEEE Signal Processing Society

260 __ |**a** New York, NY : |**b** Institute of Electrical and Electronics Engineers, |**c** c1991-

300 __ |**a** v. : |**b** ill. ; |**c** 28 cm.

310 __ |**a** Monthly

362 0_ |**a** Vol. 39, no. 1 (Jan. 1991)-

500 __ |**a** Title from cover.

510 2_ |**a** Chemical abstracts |**x** 0009-2258 |**b** Jan. 1991-

588 __ |**a** Latest issue consulted: Vol. 39, no. 6 (June 1991.

590 __ |**a** SERBIB/SERLOC merged record

592 __ |**a** ACQN: aq 94020572

650 _0 |**a** Electro-acoustics |**v** Periodicals.

650 _0 |**a** Signal processing |**v** Periodicals.

650 _0 |**a** Speech processing systems |**v** Periodicals.

650 _0 |**a** Speech processing systems |**x** Periodicals.

650 _0 |**a** Signal processing.

650 _0 |**a** Electro-acoustics |**x** Periodicals.

650 _6 |**a** Traitement automatique de la parole |**x** Périodiques.

650 _6 |**a** Traitement du signal.

650 _6 |**a** Électroacoustique |**x** Périodiques.

650 _6 |**a** Électroacoustique |**v** Périodiques.

650 _6 |**a** Traitement du signal |**v** Périodiques.

650 _6 |**a** Traitement automatique de la parole |**v** Périodiques.

650 _7 |**a** Electro-acoustics. |**2** fast |**0** (OCoLC)fst00906339

650 _7 |**a** Signal processing. |**2** fast |**0** (OCoLC)fst01118281

650 _7 |**a** Speech processing systems. |**2** fast |**0** (OCoLC)fst01129243

650 17 |**a** Signaalverwerking. |**2** gtt

655 _7 |**a** Periodicals. |**2** fast |**0** (OCoLC)fst01411641

655 _7 |**a** Periodicals. |**2** lcgft |**0** http://id.loc.gov/authorities/genreForms/gf2014026139 |**0** (uri) http://id.loc.gov/authorities/genreForms/gf2014026139

710 2_ |**a** IEEE Signal Processing Society.

776 08 |**i** Online version: |**t** IEEE transactions on signal processing (Online) |**x** 1941-0476 |**w** (DLC) 2008212918 |**w** (OCoLC)44608166

776 08 |**i** Online version: |**t** IEEE transactions on signal processing |**w** (OCoLC)623586529

780 00 |**t** IEEE transactions on acoustics, speech, and signal processing |**x** 0096-3518 |**w** (DLC) 74646881 |**w** (OCoLC)1156938

850	__	|**a** AAP |**a** ArU |**a** AzTeS |**a** CSt |**a** CU-A |**a** CU-I |**a** CU-RivP |**a** CU-SB |**a** CoFS CoU |**a** DGW |**a** DLC |**a** DeU |**a** ICarbS |**a** IaU |**a** InLP |**a** InU |**a** MBU |**a** MCM |**a** MH-GM |**a** MdBJ |**a** MdU |**a** MiEM |**a** MiU |**a** MnU |**a** MoKL |**a** MoU |**a** MsSM |**a** NBuU |**a** NN |**a** NNC |**a** NNE |**a** NRU |**a** NSyU |**a** NcD |**a** NcRS |**a** NcU |**a** NjP |**a** NjR |**a** NmU |**a** OCP |**a** OCU |**a** OU |**a** OrCS |**a** PPD |**a** PPiM |**a** PU |**a** RPB-S |**a** TU |**a** TxHR |**a** ViBlbV
859	41	|**u** http://ieeexplore.ieee.org/servlet/opac?punumber=78
853	00	|**8** 1 |**a** v. |**b** no. |**u** 12 |**v** r |**i** (year) |**j** (month) |**w** m |**x** 01
863	40	|**8** 1 |**a** <39>- |**i** <1991>- |**x** provisional
890	__	|**a** IEEE transactions on signal processing. |**i** 91-642053
906	__	|**a** 7 |**b** cbc |**c** serials |**d** u |**e** ncip |**f** 19 |**g** n-oclcserc
920	__	|**a** Keep 1
923	__	|**d** 20020726 |**s** Group Registration
984	__	|**a** srvf
985	__	|**e** eserial 200406
991	__	|**b** c-GenColl |**h** TK5981 |**i** .I2 |**w** SERIALS
991	__	|**b** o-NLS/REF |**w** ACQUIRE
992	__	|**b** 93000437 |**w** ACQUIRE
992	__	|**b** SER |**w** SERLOC

Request this Item

LC Find It

Item Availability

CALL NUMBER	TK5981 .I2 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	c.1 v. 58/, no. 3-4 2010 Mar-Apr In Process 06-11-2012
Older receipts	v.39-v.44 (1991-1996:Apr.), v.44:no.4-v.54:no.1 (1996:June-2006:Jan.) v.54:no.3-v.56:no.3 (2006:Mar.-2008:Mar.), v.56:no.7-v.62:no.1/4 (2008:July-2014:Jan./Feb.), v.62:no.9/12-v.65:no.9/12 (2014:May/June-2017:May/June), v.65:no.17/20 (2017:Sept./Oct.), v.66:no.1/4-v.67:no.4 (2018:Jan./Feb.-2019:Feb.)

CALL NUMBER	TK5981 .I2 Copy 1 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)
Latest receipts	v. 56, no. 6 (2008 June) v. 56, no. 4 (2008 Apr.)
Supplements	"Content Gazette" (2010:June)

EXHIBIT J

JOURNAL

IEEE transactions on antennas and propagation.

1963

Available at **SEL/8270 Boelter Hall Stacks; Current issues in Current Jnls (TK7872.A6 I59ta)**

Available Online

View Journal Contents

TOP
SEND TO
SEARCH INSIDE
ONLINE ACCESS
HOW TO GET IT
LINKS
DETAILS
RELATED RESOURCES

Send to

QR | PERMALINK | EMAIL | CITATION | EXPORT BIBTEX | ENDNOTE BASIC | EXPORT TO EXCEL | MENDELEY | REFWORKS | EXPORT RIS | PRINT

Search inside

Search for articles within this journal

Article title or keyword

Online access

Full availability (Access issues? Try using the VPN)

IEEE Electronic Library (IEL) SHOW LICENSE
Available from 1963 volume: 11 issue: 1.

IEEE Electronic Library (IEL) Journals
Available from 1963 volume: 11 issue: 1.

IEEE Xplore All Journals
Available from 1963 volume: 11 issue: 1.

How to get it

Please sign in to check if there are any request options. MY LIBRARY ACCOUNT

BACK TO LOCATIONS

LOCATION ITEMS

Available **SEL/8270 Boelter Hall Stacks; Current issues in Current Jnls ; TK7872.A6 I59ta**
Holdings: v.11-50(1963-2002)v.51:no.1-8,10-

12(2003),v.52(2004)v.53:no.1-7,9-12(2005)v.54-55(2006-2007)v.56:no.1-3,5-6,9-12(2008),v.57:no.2,4-12(2009)

Note: Library does not have current subscription. v.33:no.4(1985) includes the Symposia Cumulative Index 1963-1984.

Indexes: Cumulative index: 1952-1968;1969-1973;1974-1979;1963-1984;1952-84;1985-1990

Item in place (0 requests) v.57 no.10-12 yr.2009
Non-circulating

Item in place (0 requests) v.57 no.7-9 yr.2009
Non-circulating

Item in place (0 requests) v.57 no.4-6 yr.2009
Non-circulating

Item in place (0 requests) v.57 no.2 yr.2009
Non-circulating

Item in place (0 requests) v.56 no.9-12 yr.2008
Non-circulating

SHOW MORE ITEMS

OTHER UC LIBRARIES

Links

MARC Record

Details

Author / Contributor IEEE Antennas and Propagation Society, issuing body.
Institute of Electrical and Electronics Engineers. Antennas and Propagation Group, issuing body.

Title IEEE transactions on antennas and propagation.

Other Title Transactions on antennas and propagation
Antennas and propagation

Publication Information New York, Institute of Electrical and Electronics Engineers
v. AP-11- Jan. 1963-

Creation Date 1963

Related titles Earlier title : Institute of Radio Engineers. Professional Group on Antennas and Propagation. IRE transactions on antennas and propagation
Available in other form: Online version: IEEE transactions on antennas and propagation
Available in other form: Online version: IEEE transactions on antennas and propagation (Online)

Type Journal

Physical Description volumes illustrations 28 cm

Notes "A publication of the IEEE Antennas and Propagation Society" (called 1963-May 1964, Institute of Electrical and Electronics Engineers, Professional

Technical Group on Antennas and Propagation; 1964- Institute of Electrical

and Electronics Engineers, Antennas and Propagation Group).
Chemical abstracts 0009-2258

Additional forms Also issued in microform.

Language English

Subject
Radio -- Periodicals
Radio -- Antennas -- Periodicals
Antennas (Electronics) -- Periodicals
Radio waves -- Periodicals
Radio -- Antennes -- Périodiques
Ondes radioélectriques -- Périodiques
Radio -- Périodiques
Antennes (Électronique) -- Périodiques
53.74 antennas
Antennas (Electronics)
Radio
Radio -- Antennas
Radio waves
Electricity And Electrical Engineering

Genre Periodicals.

Indexes/Finding Aid Vols. 1-16, 1953-1968, 1 v.; v.17-21, 1969-1973, 1 v.

Frequency Monthly, <Jan. 1984->

Identifier
LC : 57039363
ISSN : 0018-926X
OCLC : (OCoLC)01752540
OCLC : (OCoLC)ocm01752540
ISSN : 1558-2221

MMS ID 99245553606533

Source Library Catalog

Former System Number 24555-ucladb

Related Resources

sactions.
- 1954

IEEE transactions on reliability ...
©1963-

Transactions of the I.R.E. Professional Group on Antennas and ...
©1952-©1955

IEEE transactions on antennas and propagation. ...
1963

IRE transactio on antenn and propagatic
©1955-©1

EXHIBIT K

```
leader  04449cas a2200913   4500
001     9913117297506531
005     20220810213025.0
008     751101c19639999nyumr p       0   a0eng d
010     ##$a   57039363
016     7#$aI03585000 $2DNLM
016     7#$a009844966 $2Uk
022     0#$a0018-926X $l0018-926X $21
030     ##$aIETPAK
032     ##$a256600 $bUSPS
035     ##$a(CLU)24555-ucladb-Voyager
035     ##$a(OCoLC)01752540 $z(OCoLC)255560521 $z(OCoLC)605926506 $z(OCoLC)977676120 $z(OCoLC)1011796876 $z(OCoLC)1013325270 $z(OCoLC)1063260080 $z(C
035     ##$a(OCoLC)ocm01752540
037     ##$bIEEE Service Center, 445 Hoes La., Piscataway, NJ 08854
040     ##$aMUL $beng $cMUL $dNSD $dYUS $dHUL $dOCL $dRCS $dNSD $dAIP $dDLC $dNST $dOCL $dNSD $dNST $dMYG $dCUS $dMYG $dDLC $dLYU $dIEEEX $dOCLCQ $dC
042     ##$ansdp
049     ##$aMAIN
050     00$aTK7800 $b.I2
210     0#$aIEEE trans. antennas propag. $b(Print)
222     #0$aIEEE transactions on antennas and propagation $b(Print)
245     00$aIEEE transactions on antennas and propagation.
246     30$aTransactions on antennas and propagation
246     30$aAntennas and propagation
260     ##$a[New York], $b[Institute of Electrical and Electronics Engineers]
300     ##$avolumes $billustrations $c28 cm
310     ##$aMonthly, $b<Jan. 1984->
321     ##$aBimonthly
336     ##$atext $btxt $2rdacontent
337     ##$aunmediated $bn $2rdamedia
338     ##$avolume $bnc $2rdacarrier
362     0#$av. AP-11- Jan. 1963-
550     ##$a"A publication of the IEEE Antennas and Propagation Society" (called 1963-May 1964, Institute of Electrical and Electronics Engineers, Pr
530     ##$aAlso issued in microform.
510     0#$aChemical abstracts $x0009-2258
555     ##$aVols. 1-16, 1953-1968, 1 v.; v.17-21, 1969-1973, 1 v.
599     ##$aUPDATED $bUclaCollMgr $bSerials1 $bMaster record variable field(s) change: 710 $9LOCAL
650     #0$aRadio $vPeriodicals.
650     #0$aRadio $xAntennas $vPeriodicals.
650     #0$aAntennas (Electronics) $vPeriodicals.
650     #0$aRadio waves $vPeriodicals.
650     #6$aRadio $xAntennes $vPériodiques.
650     #6$aOndes radioélectriques $vPériodiques.
650     #6$aRadio $vPériodiques.
650     #6$aAntennes (Électronique) $vPériodiques.
650     #7$a53.74 antennas. $0(NL-LeOCL)077605020 $2bcl
650     #7$aAntennas (Electronics) $2fast $0(OCoLC)fst00810043
650     #7$aRadio. $2fast $0(OCoLC)fst01087053
650     #7$aRadio $xAntennas. $2fast $0(OCoLC)fst01087055
650     #7$aRadio waves. $2fast $0(OCoLC)fst01087617
653     ##$aElectricity And Electrical Engineering
655     #7$aPeriodicals. $2fast $0(OCoLC)fst01411641
655     #7$aPeriodicals. $2lcgft $0http://id.loc.gov/authorities/genreForms/gf2014026139 $0(uri) http://id.loc.gov/authorities/genreForms/gf20140261
710     2#$aIEEE Antennas and Propagation Society, $eissuing body.
710     2#$aInstitute of Electrical and Electronics Engineers. $bAntennas and Propagation Group, $eissuing body.
776     08$iOnline version: $tIEEE transactions on antennas and propagation (Online) $x1558-2221 $w(DLC)  2005215237 $w(OCoLC)44459909
776     08$iOnline version: $tIEEE transactions on antennas and propagation $w(OCoLC)565405967
780     00$aInstitute of Radio Engineers. Professional Group on Antennas and Propagation. $tIRE transactions on antennas and propagation $x0096-1973
891     20$9853 $81 $av. $bno. $u12 $vr $i(year) $j(month) $wm
891     41$9863 $81.1 $a<49> $b<1> $i<2001> $j<01>
908     ##$aWorldCat Daily Updates 2022-08-10 $bWorldCat record variable field(s) change: 040, 655
936     ##$aUnknown $aJan. 1984
951     ##$aUclaCollMgr Serials1 20200306 $acdl 190814 $aUclaCollMgr Serials1 20181123 $acdl 160406 $aUclaCollMgr Serials1 20160401 $acdl 090811 $acc
986     ##$aUnknown $aJan. 1984 $9LOCAL
992     ##$aTC0034007 $cACB-3288 $9LOCAL
992     ##$c07-AOU-4718 $9LOCAL
996     ##$a24555-ucladb $9LOCAL
```

EXHIBIT L

PERIODICAL OR NEWSPAPER

IEEE transactions on antennas and propagation

Full Record MARC Tags

Main title

IEEE transactions on antennas and propagation.

Published/Created

[New York], [Institute of Electrical and Electronics Engineers]

Request this Item LC Find It

More Information

LCCN Permalink	https://lccn.loc.gov/57039363
Publication history	v. AP-11- Jan. 1963-
Description	volumes illustrations 28 cm
Current frequency	Monthly, <Jan. 1984->
Former frequency	Bimonthly
ISSN	0018-926X
Linking ISSN	0018-926X
LC classification	TK7800 .I2

Continues	Institute of Radio Engineers. Professional Group on Antennas and Propagation. IRE transactions on antennas and propagation 0096-1973 (DLC)sn 80000246 (OCoLC)2007254
Portion of title	Transactions on antennas and propagation Antennas and propagation
Serial key title	IEEE transactions on antennas and propagation (Print)
Abbreviated title	IEEE trans. antennas propag. (Print)
Related names	IEEE Antennas and Propagation Society, issuing body. Institute of Electrical and Electronics Engineers. Antennas and Propagation Group, issuing body.
LC Subjects	Radio--Periodicals. Radio--Antennas--Periodicals. Antennas (Electronics)--Periodicals. Radio waves--Periodicals.
Other Subjects	53.74 antennas. Antennas (Electronics) Radio. Radio--Antennas. Radio waves.
Subject keywords	Electricity And Electrical Engineering;
Form/Genre	Periodicals. Periodicals.
Browse by shelf order	TK7800
Notes	"A publication of the IEEE Antennas and Propagation Society" (called 1963-May 1964, Institute of Electrical and Electronics Engineers, Professional Technical Group on Antennas and Propagation; 1964- Institute of Electrical and Electronics Engineers, Antennas and Propagation Group).
Indexed by	Chemical abstracts 0009-2258
Indexes	Vols. 1-16, 1953-1968, 1 v.; v.17-21, 1969-1973, 1 v.
Reproduction no./Source	IEEE Service Center, 445 Hoes La., Piscataway, NJ 08854

Additional formats	Also issued in microform. Online version: IEEE transactions on antennas and propagation (Online) 1558-2221 (DLC) 2005215237 (OCoLC)44459909 Online version: IEEE transactions on antennas and propagation (OCoLC)565405967
LCCN	57039363
CODEN	IETPAK
Postal reg no.	256600 USPS
National bib agency no.	I03585000 009844966
Other system no.	(OCoLC)ocm01752540
Additional Links	v.36(1988)- http://ieeexplore.ieee.org/servlet/opac?punumber=8 v.3(1955)-v.35(1987). http://ieeexplore.ieee.org/servlet/opac?punumber=8234
Type of material	Periodical or Newspaper
Content type	text
Media type	unmediated
Carrier type	volume

Item Availability

CALL NUMBER	TK7800 .I2 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	c.1 v. 67, no. 1 2019 Jan In Process 03-03-2021 c.1 v. 69, no. 1 2021 Jan At Bindery - 09-13-2022 c.1 v. 69, no. 2 2021 Feb At Bindery - 09-13-2022 c.1 v. 69, no. 3 2021 Mar At Bindery - 09-13-2022 c.1 v. 69, no. 5 2021 May At Bindery - 09-13-2022 c.1 v. 69, no. 6 2021 June At Bindery - 09-13-2022
Older receipts	

v.AP1 (1952), v.AP2-v.AP50:no.3 (1953:July-2002:Mar.), v.50:no.8-v.52 (2002:Aug.-2004), v.53:no.2-v.53:no.3 (2005:Feb.-2005:Mar.),v.53:no.5-v.54:no.1 (2005:May-2006:Jan.)
v.54:no.3-v.56:no.3 (2006:Mar-2008:Mar.), v.56:no.6-v.66:no.10 (2008:June-2018:Oct.), v.66:no.12-v.69:no.3 (2018:Dec.-2021:Mar.), v.69:no.5-v.69:no.6 (2021:May-2021:June)

Supplements

"Spec. suppl." 1959:Dec.

Indexes

"Cumulative Index" 1952/1984
v.AP39 (1991)

CALL NUMBER

TK7800 .I2
Copy 1
Unbound issues

Request in

Newspaper & Current Periodical Reading Room (Madison LM133)

Latest receipts

v. 70, no. 9 (2022 Sept.)
v. 70, no. 8 (2022 Aug.)
v. 70, no. 7 (2022 July)
v. 70, no. 6 (2022 June)
v. 70, no. 5 (2022 May)
v. 70, no. 4 (2022 Apr.)
v. 70, no. 3 (2022 Mar.)
v. 70, no. 1 (2022 Jan.)
v. 69, no. 12 (2021 Dec.)
v. 69, no. 11 (2021 Nov.)
v. 69, no. 10 (2021 Oct.)
v. 69, no. 9 (2021 Sept.)
v. 69, no. 8 (2021 Aug.)
v. 69, no. 7 (2021 July)
v. 69, no. 4 (2021 Apr.)
v. 57, no. 4, pt. 2 (2009 Apr.)

CALL NUMBER

TK7800 .I2

Request in

Newspaper & Current Periodical Reading Room (Madison LM133)

EXHIBIT M

PERIODICAL OR NEWSPAPER

IEEE transactions on antennas and propagation

Full Record | **MARC Tags**

000 03990cas a2200805 4500
001 11150579
005 20220812095006.0
008 751101c19639999nyumr p 0 a0eng d
010 __ |**a** 57039363
016 7_ |**a** I03585000 |**2** DNLM
016 7_ |**a** 009844966 |**2** Uk
022 0_ |**a** 0018-926X |**l** 0018-926X |**2** 1
030 __ |**a** IETPAK
032 __ |**a** 256600 |**b** USPS
035 __ |**a** (OCoLC)ocm01752540
037 __ |**b** IEEE Service Center, 445 Hoes La., Piscataway, NJ 08854
040 __ |**a** MUL |**b** eng |**c** MUL |**d** NSD |**d** YUS |**d** HUL |**d** OCL |**d** RCS |**d** NSD |**d** AIP |**d** DLC |**d** NST |**d** OCL |**d** NSD |**d** NST |**d** MYG |**d** CUS |**d** MYG |**d** DLC |**d** LYU |**d** IEEEX |**d** OCLCQ |**d** CUS |**d** OCLCQ |**d** NSD |**d** LVB |**d** OCLCQ |**d** OCLCF |**d** OCLCO |**d** IUL |**d** OCLCO |**d** QAQUL |**d** EUX |**d** RCE |**d** OCLCQ |**d** UKMGB |**d** L2U |**d** IUL |**d** OCLCO |**d** PUL
042 __ |**a** nsdp
050 00 |**a** TK7800 |**b** .I2
210 0_ |**a** IEEE trans. antennas propag. |**b** (Print)
222 _0 |**a** IEEE transactions on antennas and propagation |**b** (Print)
245 00 |**a** IEEE transactions on antennas and propagation.
246 30 |**a** Transactions on antennas and propagation
246 30 |**a** Antennas and propagation
260 __ |**a** [New York], |**b** [Institute of Electrical and Electronics Engineers]

300 __ |**a** volumes |**b** illustrations |**c** 28 cm

310 __ |**a** Monthly, |**b** <Jan. 1984->

321 __ |**a** Bimonthly

336 __ |**a** text |**b** txt |**2** rdacontent

337 __ |**a** unmediated |**b** n |**2** rdamedia

338 __ |**a** volume |**b** nc |**2** rdacarrier

362 0_ |**a** v. AP-11- Jan. 1963-

510 0_ |**a** Chemical abstracts |**x** 0009-2258

530 __ |**a** Also issued in microform.

550 __ |**a** "A publication of the IEEE Antennas and Propagation Society" (called 1963-May 1964, Institute of Electrical and Electronics Engineers, Professional Technical Group on Antennas and Propagation; 1964- Institute of Electrical and Electronics Engineers, Antennas and Propagation Group).

555 __ |**a** Vols. 1-16, 1953-1968, 1 v.; v.17-21, 1969-1973, 1 v.

590 __ |**a** SERBIB/SERLOC merged record

650 _0 |**a** Radio |**v** Periodicals.

650 _0 |**a** Radio |**x** Antennas |**v** Periodicals.

650 _0 |**a** Antennas (Electronics) |**v** Periodicals.

650 _0 |**a** Radio waves |**v** Periodicals.

650 _6 |**a** Radio |**x** Antennes |**v** Périodiques.

650 _6 |**a** Ondes radioélectriques |**v** Périodiques.

650 _6 |**a** Radio |**v** Périodiques.

650 _6 |**a** Antennes (Électronique) |**v** Périodiques.

650 _7 |**a** 53.74 antennas. |**0** (NL-LeOCL)077605020 |**2** bcl

650 _7 |**a** Antennas (Electronics) |**2** fast |**0** (OCoLC)fst00810043

650 _7 |**a** Radio. |**2** fast |**0** (OCoLC)fst01087053

650 _7 |**a** Radio |**x** Antennas. |**2** fast |**0** (OCoLC)fst01087055

650 _7 |**a** Radio waves. |**2** fast |**0** (OCoLC)fst01087617

653 __ |**a** Electricity And Electrical Engineering

655 _7 |**a** Periodicals. |**2** fast |**0** (OCoLC)fst01411641

655 _7 |**a** Periodicals. |**2** lcgft |**0** http://id.loc.gov/authorities/genreForms/gf2014026139 |**0** (uri) http://id.loc.gov/authorities/genreForms/gf2014026139

710 2_ |**a** IEEE Antennas and Propagation Society, |**e** issuing body.

710 2_ |**a** Institute of Electrical and Electronics Engineers. |**b** Antennas and Propagation Group, |**e** issuing body.

776 08 |**i** Online version: |**t** IEEE transactions on antennas and propagation (Online) |**x** 1558-2221 |**w** (DLC) 2005215237 |**w** (OCoLC)44459909

776 08 |**i** Online version: |**t** IEEE transactions on antennas and propagation |**w** (OCoLC)565405967

780 00 |**a** Institute of Radio Engineers. Professional Group on Antennas and Propagation. |**t** IRE

transactions on antennas and propagation |**x** 0096-1973 |**w** (DLC)sn 80000246 |**w** (OCoLC)2007254

859 41 |**3** v.36(1988)- |**u** http://ieeexplore.ieee.org/servlet/opac?punumber=8

859 41 |**3** v.3(1955)-v.35(1987). |**u** http://ieeexplore.ieee.org/servlet/opac?punumber=8234

853 20 |**8** 1 |**a** v. |**b** no. |**u** 12 |**v** r |**i** (year) |**j** (month) |**w** m

863 41 |**8** 1.1 |**a** <49> |**b** <1> |**i** <2001> |**j** <01>

890 __ |**a** IEEE transactions on antennas and propagation. New York. |**i** sv86-621

906 __ |**a** 7 |**b** cbc |**c** serials |**d** u |**e** ncip |**f** 19 |**g** n-oclcserc

920 __ |**a** Keep 1

923 __ |**d** 2003-04-18 |**s** Group Registration

984 __ |**a** srvf |**d** 2001-04-11

985 __ |**e** eserial 200406

991 __ |**b** c-GenColl |**h** TK7800 |**i** .I2 |**w** SERIALS

992 __ |**b** SER |**w** SERLOC

Request this Item | LC Find It

Item Availability

CALL NUMBER	TK7800 .I2 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	c.1 v. 67, no. 1 2019 Jan In Process 03-03-2021 c.1 v. 69, no. 1 2021 Jan At Bindery - 09-13-2022 c.1 v. 69, no. 2 2021 Feb At Bindery - 09-13-2022 c.1 v. 69, no. 3 2021 Mar At Bindery - 09-13-2022 c.1 v. 69, no. 5 2021 May At Bindery - 09-13-2022 c.1 v. 69, no. 6 2021 June At Bindery - 09-13-2022
Older receipts	v.AP1 (1952), v.AP2-v.AP50:no.3 (1953:July-2002:Mar.), v.50:no.8-v.52 (2002:Aug.-2004), v.53:no.2-v.53:no.3 (2005:Feb.-2005:Mar.),v.53:no.5-v.54:no.1 (2005:May-2006:Jan.) v.54:no.3-v.56:no.3 (2006:Mar-2008:Mar.), v.56:no.6-v.66:no.10 (2008:June-2018:Oct.), v.66:no.12-v.69:no.3 (2018:Dec.-2021:Mar.), v.69:no.5-v.69:no.6 (2021:May-2021:June)
Supplements	

"Spec. suppl." 1959:Dec.

Indexes

"Cumulative Index" 1952/1984
v.AP39 (1991)

CALL NUMBER

TK7800 .I2
Copy 1
Unbound issues

Request in

Newspaper & Current Periodical Reading Room (Madison LM133)

Latest receipts

v. 70, no. 9 (2022 Sept.)
v. 70, no. 8 (2022 Aug.)
v. 70, no. 7 (2022 July)
v. 70, no. 6 (2022 June)
v. 70, no. 5 (2022 May)
v. 70, no. 4 (2022 Apr.)
v. 70, no. 3 (2022 Mar.)
v. 70, no. 1 (2022 Jan.)
v. 69, no. 12 (2021 Dec.)
v. 69, no. 11 (2021 Nov.)
v. 69, no. 10 (2021 Oct.)
v. 69, no. 9 (2021 Sept.)
v. 69, no. 8 (2021 Aug.)
v. 69, no. 7 (2021 July)
v. 69, no. 4 (2021 Apr.)
v. 57, no. 4, pt. 2 (2009 Apr.)

CALL NUMBER

TK7800 .I2

Request in

Newspaper & Current Periodical Reading Room (Madison LM133)

EXHIBIT N

Menu

Digital signal processing

Imprint
Duluth, MN : Academic Press, c1991-

Physical description
v. : ill. ; 28 cm.

Online

Available online

Elsevier ScienceDirect Journals
- Available from 1991/01/01 volume: 1 issue: 1

Elsevier SD Backfile Computer Science
- Available from 1991/01/01 volume: 1 issue: 1 until 1994/10/31 volume: 4 issue: 4

Elsevier SD Backfile Engineering and Technology
- Available from 1991/01/01 volume: 1 issue: 1 until 1994/10/31 volume: 4 issue: 4

Elsevier SD Freedom Collection Journals [SCFCJ]

See the full **find it @ Stanford** menu

Report a connection problem

At the library

SAL3 (off-campus storage)

No public access

Stacks
Library has: v.7(1997)-v.19(2009)

TK5102.5 .D4463 V.19 Available
2009

TK5102.5 .D4463 V.18 Available
2008

TK5102.5 .D4463 V.17 Available
2007

TK5102.5 .D4463 V.16 Available
2006

Description

Contents/Summary

Supplemental links
library.stanford.edu

Subjects

Subjects
Signal processing > Digital techniques > Periodicals.

Bibliographic information

Beginning date
1991

Frequency
Quarterly

Vol/date range
Vol. 1, no. 1 (Jan. 1991)-

Note
Title from cover.

Referenced in
Computer & control abstracts 0036-8113 Jan. 1991-

Electrical & electronics abstracts 0036-8105 Jan. 1991-

Physics abstracts 0036-8091 Jan. 1991-

Note
Also available on the World Wide Web.

ISSN
1051-2004

Key title
Digital signal processing

TK5102.5 .D4463 V.15 Available
2005

TK5102.5 .D4463 V.14 Available
2004

TK5102.5 .D4463 V.13 Available
2003

TK5102.5 .D4463 V.12 In transit
2002

TK5102.5 .D4463 V.11 Available
2001

TK5102.5 .D4463 V.10 Available
2000

TK5102.5 .D4463 V.9 Available
1999

TK5102.5 .D4463 V.8 Available
1998

TK5102.5 .D4463 V.7 Available
1997

show less

More options

Find it at other libraries via WorldCat

| Catkey: 3471

EXHIBIT O

```
LEADER 03577cas a2200589 a 4500001    a3471549
003    SIRSI
005    20221116003001.0
007    cr||||
008    900620c19919999mnuqr1p       0   a0eng d
010       91650983  |zsn 90002038
022 0  1051-2004
030    DSPREJ
035    (OCoLC-M)21889588
035    (OCoLC-I)275468790
037     |c$110.00
037     |bAcademic Press, Inc., 1 E. 1st St., Duluth, MN 55802
040    NSDP |cNSDP |dDLC |dNST |dOUCA |dNST |dNSDP |dMiU |dCSt
042    nsdplc
049    STF
050 00 TK5102.5 |b.D4463
082 00 621.382/2 |220
210 0  Digit. signal process.
222  0 Digital signal processing
245 00 Digital signal processing.
260    Duluth, MN : |bAcademic Press, |cc1991-
300    v. : |bill. ; |c28 cm.
310    Quarterly
362 0  Vol. 1, no. 1 (Jan. 1991)-
500    Title from cover.
510 2  Computer & control abstracts |x0036-8113 |bJan. 1991-
510 2  Electrical & electronics abstracts |x0036-8105 |bJan. 1991-
510 2  Physics abstracts |x0036-8091 |bJan. 1991-
530    Also available on the World Wide Web.
596    31
650  0 Signal processing |xDigital techniques |vPeriodicals. |=^A1061119
856 41  |zAvailable to Stanford-affiliated users at: |uhttps://stanford.idm.oclc.org/login?
url=https://www.sciencedirect.com/science/journal/10512004
916    DATE CATALOGED |b19971216
956 41  |uhttp://library.stanford.edu/sfx?url%5Fver=Z39.88-2004&ctx%5Fver=Z39.88-
2004&ctx%5Fenc=info:ofi/enc:UTF-
8&rfr%5Fid=info:sid/sfxit.com:opac%5F856&url%5Fctx%5Ffmt=info:ofi/fmt:kev:mtx:ctx&sfx.ignore%5Fdate%5
Fthreshold=1&rft.object%5Fid=954922650164&svc%5Fval%5Ffmt=info:ofi/fmt:kev:mtx:sch%5Fsvc&
|xeLoaderURL |xmr4 |xmr954922650164
852    CSt |bSAL3 |cSTACKS
866 31  |80v.7(1997)-v.19(2009)
852    CSt |bMUSIC |cR-STACKS |zAll holdings transferred to SAL3. |=7053
853 2   |82v. |bno. |u4 |vr |i(year) |j(month)
853 2   |83v. |bno. |u6 |vr |i(year) |j(month)
866 31  |80
863  1  |82.19 |b1 |i1999 |j01
863  1  |82.29 |b2 |i1999 |j04
863  1  |82.39 |b3 |i1999 |j07
863  1  |82.49 |b4 |i1999 |j10
863  1  |82.510 |b1/3 |i2000 |j01/07
863  1  |82.610 |b4 |i2000 |j10
863  1  |82.711 |b1 |i2001 |j01
863  1  |82.811 |b2 |i2001 |j04
863  1  |82.911 |b3 |i2001 |j07
863  1  |82.1011 |b4 |i2001 |j10
863  1  |82.1112 |b1 |i2002 |j01
863  1  |82.1212 |b2/3 |i2002 |j04/07
863  1  |82.1312 |b4 |i2002 |j10
863  1  |82.1413 |b1 |i2003 |j01
863  1  |82.1513 |b2 |i2003 |j04
863  1  |82.1613 |b3 |i2003 |j07
863  1  |82.1713 |b4 |i2003 |j10
863  1  |83.114 |b1 |i2004 |j01
863  1  |83.214 |b2 |i2004 |j03
```

```
863  1  |83.314 |b3 |i2004 |j05
863  1  |83.414 |b4 |i2004 |j07
863  1  |83.514 |b5 |i2004 |j09
863  1  |83.614 |b6 |i2004 |j11
863  1  |83.715 |b1 |i2005 |j01
863  1  |83.815 |b2 |i2005 |j03
863  1  |83.915 |b3 |i2005 |j05
863  1  |83.1015 |b4 |i2005 |j07
863  1  |83.1115 |b5 |i2005 |j09
863  1  |83.1215 |b6 |i2005 |j11
863  1  |83.1316 |b1 |i2006 |j01
863  1  |83.1416 |b2 |i2006 |j03
863  1  |83.1516 |b3 |i2006 |j05
863  1  |83.1616 |b4 |i2006 |j07
863  1  |83.1716 |b5 |i2006 |j09
863  1  |83.1816 |b6 |i2006 |j11
863  1  |83.1917 |b1 |i2007 |j01
863  1  |83.2017 |b2 |i2007 |j03
863  1  |83.2117 |b3 |i2007 |j05
863  1  |83.2217 |b4 |i2007 |j07
863  1  |83.2317 |b5 |i2007 |j09
863  1  |83.2417 |b6 |i2007 |j11
863  1  |83.2518 |b1 |i2008 |j01
863  1  |83.2618 |b2 |i2008 |j03
863  1  |83.2718 |b3 |i2008 |j05
863  1  |83.2818 |b4 |i2008 |j07
863  1  |83.2918 |b5 |i2008 |j09
863  1  |83.3018 |b6 |i2008 |j11
863  1  |83.3119 |b1 |i2009 |j01
863  1  |83.3219 |b2 |i2009 |j03
863  1  |83.3319 |b3 |i2009 |j05
863  1  |83.3419 |b4 |i2009 |j07
863  1  |83.3519 |b5 |i2009 |j09
863  1  |83.3619 |b6 |i2009 |j12
```

EXHIBIT P

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://lccn.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	Signal processing--Digital techniques--Periodicals. Signal processing--Digital techniques.
Other Subjects	Signal processing--Digital techniques.
Form/Genre	Periodicals.
Browse by shelf order	TK5102.5
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 (OCoLC)36980269
LCCN	91650983
Invalid LCCN	sn 90002038
CODEN	DSPREJ
Dewey class no.	621.382/2
Other system no.	(OCoLC)ocm21889588
Additional Links	http://www.idealibrary.com Logon procedure and access to this title is available via the I.D.E.A.L. (service provider) homepage http http://firstsearch.oclc.org 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 http http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP 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 http
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. 101 2020 June In Process 10-24-2022 c.1 v. 104-105 2020 Sept-Oct In Process 10-24-2022 c.1 v. 106 2020 Nov In Process 10-24-2022 c.1 v. 107 2020 Dec In Process 10-24-2022
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.107 (2020:Sept.-2020:Dec.)

CALL NUMBER	TK5102.5 .D4463 Copy 1 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)
Latest receipts	v. 114 (2021 July) v. 113 (2021 June) v. 112 (2021 May) v. 110 (2021 Mar.) v. 109 (2021 Feb.) v. 108 (2021 Jan.)

EXHIBIT Q

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** LYU |**d** N SD |**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 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** http://www.idealibrary.com |**z** Logon procedure and access to this title is available via the I.D.E.A.L. (service provider) homepage |**2** http

859 7_ |**u** http://firstsearch.oclc.org |**z** 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 |**2** http

859 7_ |**u** http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP |**z** 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 |**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

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. 101 2020 June In Process 10-24-2022
c.1 v. 104-105 2020 Sept-Oct In Process 10-24-2022
c.1 v. 106 2020 Nov In Process 10-24-2022
c.1 v. 107 2020 Dec In Process 10-24-2022

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.107 (2020:Sept.-2020:Dec.)

CALL NUMBER

TK5102.5 .D4463
Copy 1
Unbound issues

Request in

Newspaper & Current Periodical Reading Room (Madison LM133)

Latest receipts

v. 114 (2021 July)
v. 113 (2021 June)
v. 112 (2021 May)
v. 110 (2021 Mar.)
v. 109 (2021 Feb.)
v. 108 (2021 Jan.)

EXHIBIT R

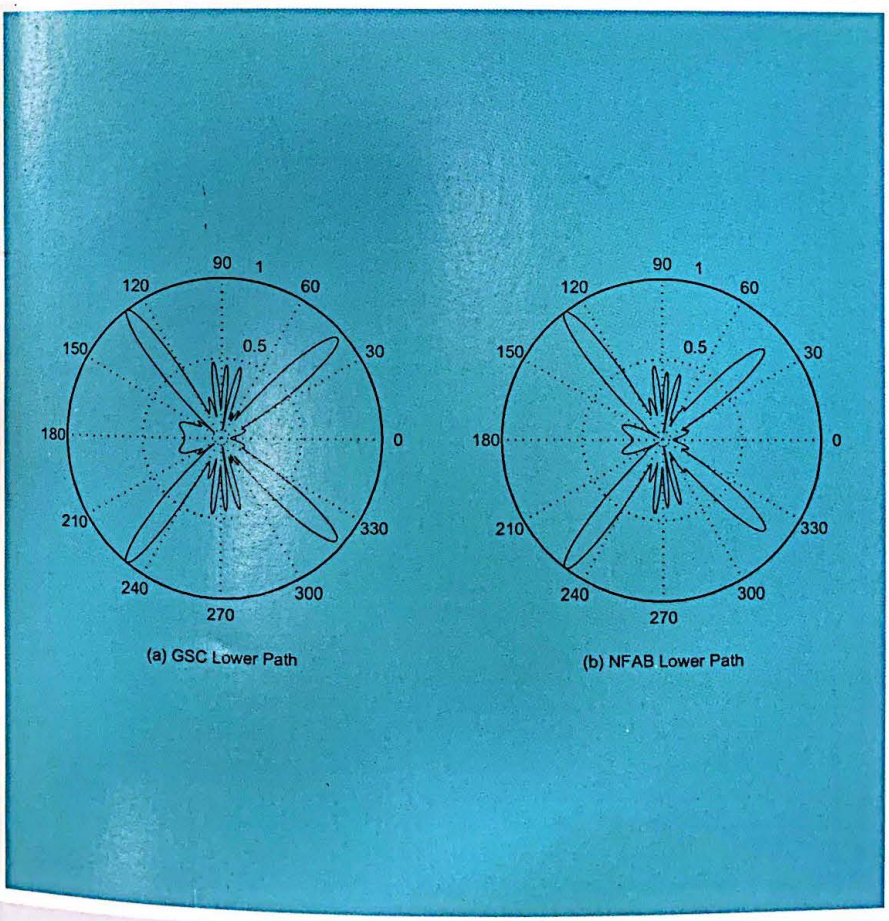
CCRMR

Digital Signal Processing

Volume 12,
January 2002

IDEAL® First
Articles published online first
http://www.idealibrary.com

A Review Journal



Editors
Jim Schroeder
Joe Campbell

ISSN 1051-2004

ACADEMIC PRESS
An Elsevier Science Imprint

Digital Signal Processing

A Review Journal

Editors

Jim Schroeder
SPRI/CSSIP
Adelaide, SA, Australia
E-mail: schroeder@cssip.edu.au

Joe Campbell
M.I.T. Lincoln Laboratory
Lexington, Massachusetts
E-mail: 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.
Yoqneam, Isreal

Rao Yarlagadda
Oklahoma State University
Stillwater, Oklahoma

Cover photo. Lower path directivity pattern at 5000 Hz. See the article by McCowan, Moore, and Sridharan in this issue.

Digital Signal Processing **12**, 87–106 (2002)
doi:10.1006/dspr.2001.0414, available online at http://www.idealibrary.com on IDEAL®

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; moore@idiap.ch; 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

1051-2004/02 $35.00
© 2002 Elsevier Science (USA)
All rights reserved.

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, **w**, with L constraints on the desired signal response, and an adaptive path, consisting of a blocking matrix, **B**, and a set of adaptive filters, **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

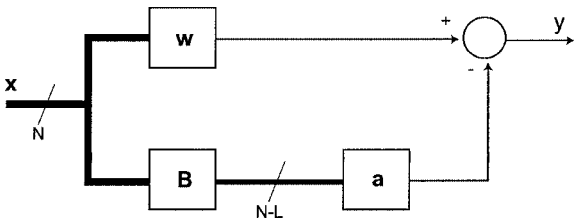


FIG. 1. Generalized sidelobe canceler structure.

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 λ 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 (θ_s, ϕ_s), at distance r_s from the reference microphone, is denoted $\mathbf{p}_s$ and is given by:

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

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

$$d_i = \|\mathbf{p}_s - \mathbf{p}_i\|, \tag{2}$$

where $\|\ \|$ 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

$$\tau_i = \frac{d_i - d_1}{c}, \tag{3}$$

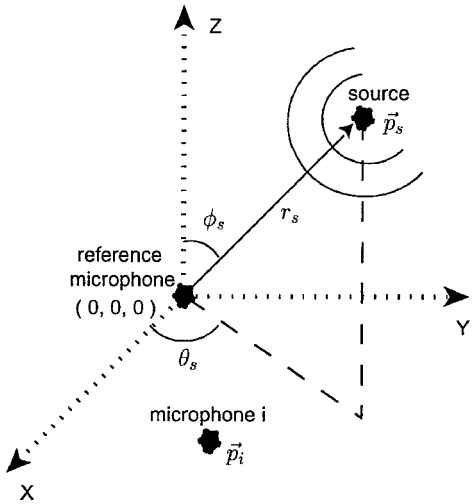


FIG. 2. Near-field propagation model.

where $c = 340\ \mathrm{ms}^{-1}$ 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

$$\alpha_i = \frac{d_1}{d_i}. \tag{4}$$

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

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

Consequently, we define the near-field propagation vector for a source at distance r and direction (θ, ϕ) as

$$\mathbf{d}(f, r, \theta, \phi) = \left[\alpha_1 e^{-j2\pi f \tau_1} \ \ldots \ \alpha_i e^{-j2\pi f \tau_n} \ \ldots \ \alpha_N e^{-j2\pi f \tau_N}\right]^T. \tag{6}$$

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.

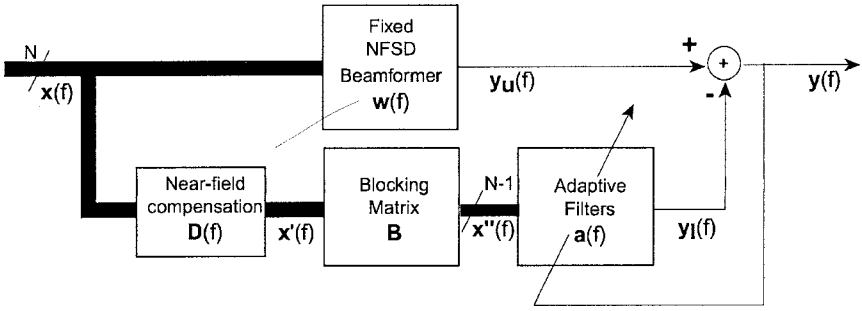


FIG. 3. Near-field adaptive beamformer.

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]

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

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

$$\mathbf{w}(f) = \left[w_1(f) \ \ldots \ w_i(f) \ \ldots \ w_N(f) \right]^T, \tag{8}$$

$(\ )^H$ 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, θ_s, ϕ_s), the signal cross-spectral matrix $\mathbf{P}$ simplifies to the propagation vector of the source, and the array gain can be expressed as

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

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

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

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

The superdirectivity problem is thus formulated as:

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

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 δ^2 is the desired white noise gain.

as

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

where ϵ 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 ϵ and the desired white noise gain δ^2, shows that the white noise gain increases monotonically with increasing ϵ. One possible means of obtaining the desired value of ϵ is thus an iterative technique employing a binary search algorithm between a specified minimum and maximum value for ϵ. 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 by simplified by substituting $\mathbf{C}(f) = \mathbf{d}(f)$ and $\mathbf{g}(f) = 1$, giving

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

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

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

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

$$\mathbf{x}(f) = \left[x_1(f) \ \ldots \ x_i(f) \ \ldots \ x_N(f)\right]^T. \tag{15}$$

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

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

where $\mathbf{d}(f)$ is the near-field propagation vector from Eq. (6). In this paper we define the diagonal operator, 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

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

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

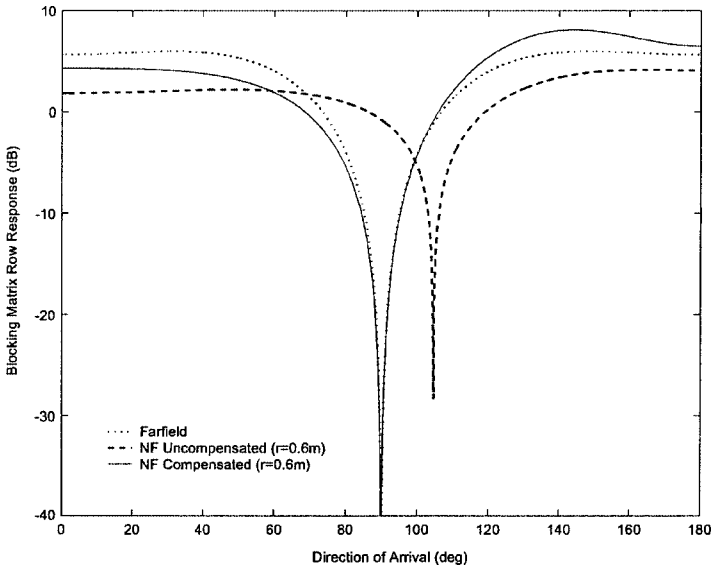


FIG. 4. Comparison of blocking matrix row beam-patterns.

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.

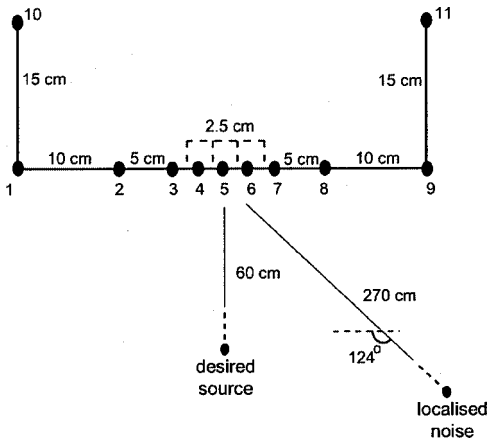


FIG. 5. Experimental configuration.

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:

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

The output of the blocking matrix is calculated as

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

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

$$\mathbf{a}(f) = \left[a_1(f) \ \ldots \ a_i(f) \ \ldots \ a_{N-1}(f)\right]^T, \tag{20}$$

the output of the lower path is given as

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

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

$$y(f) = y_u(f) - y_l(f) \tag{22}$$

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

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

where μ 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

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

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

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

$$y(f) = y_u(f) - y_l(f) \tag{24d}$$

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

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 kHz $< f <$ 2 kHz): microphones 1, 2, 5, 8, and 9;
- (2 kHz $< f <$ 4 kHz): microphones 2, 3, 5, 7, and 8; and
- (4 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° 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

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

TABLE 1

Beamforming Techniques in Evaluation

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

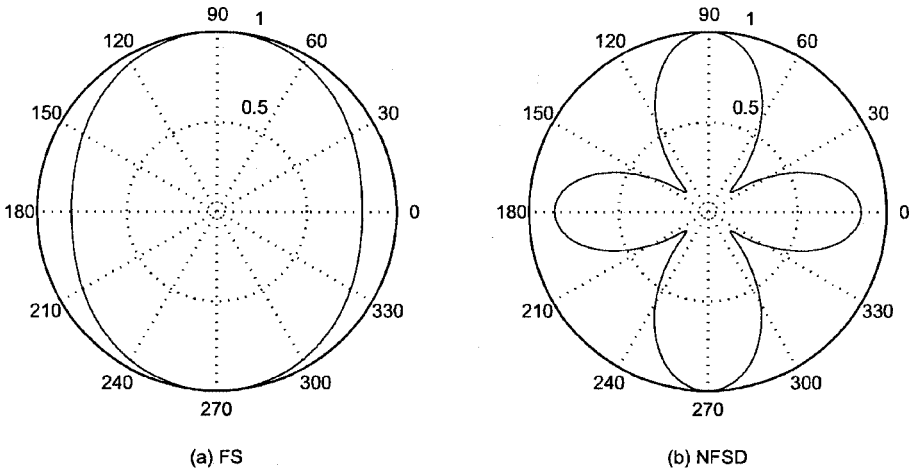


FIG. 6. Upper path directivity pattern at 300 Hz.

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 $\mathbf{w}_o(f) = [\mathrm{diag}(\mathbf{D}(\mathbf{f}))]^H$. 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 ($\approx 124°$ 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

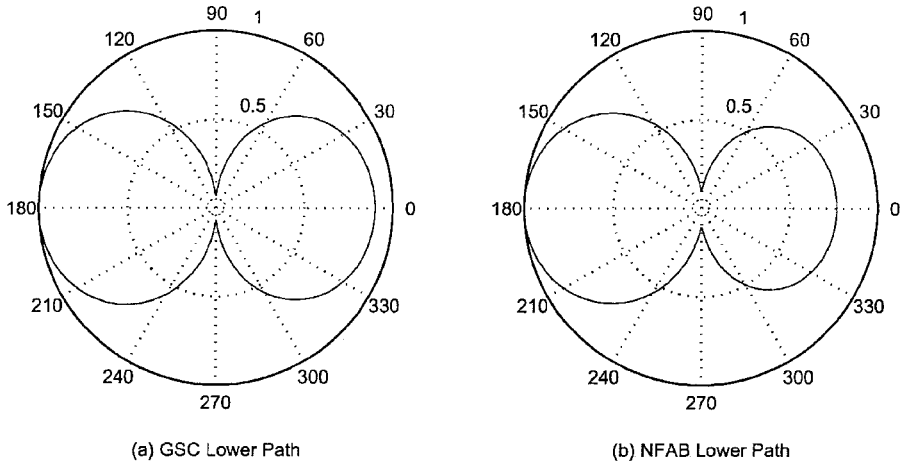


FIG. 7. Lower path directivity pattern at 300 Hz.

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

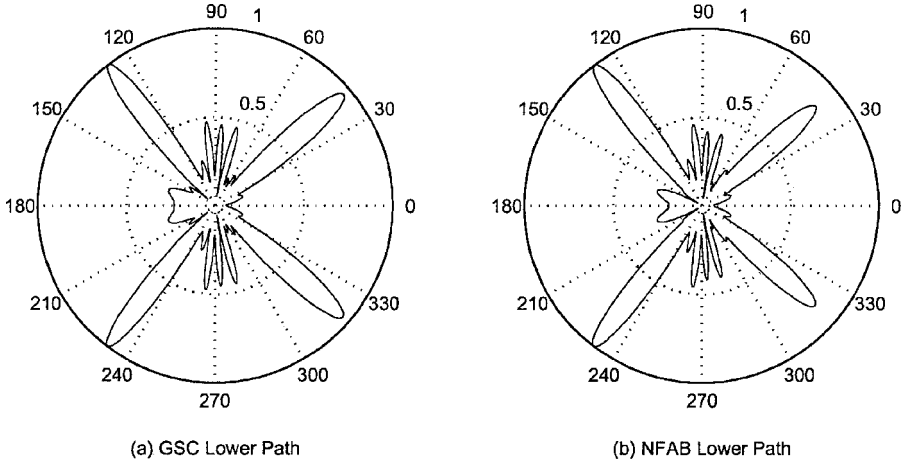


FIG. 8. Lower path directivity pattern at 5000 Hz.

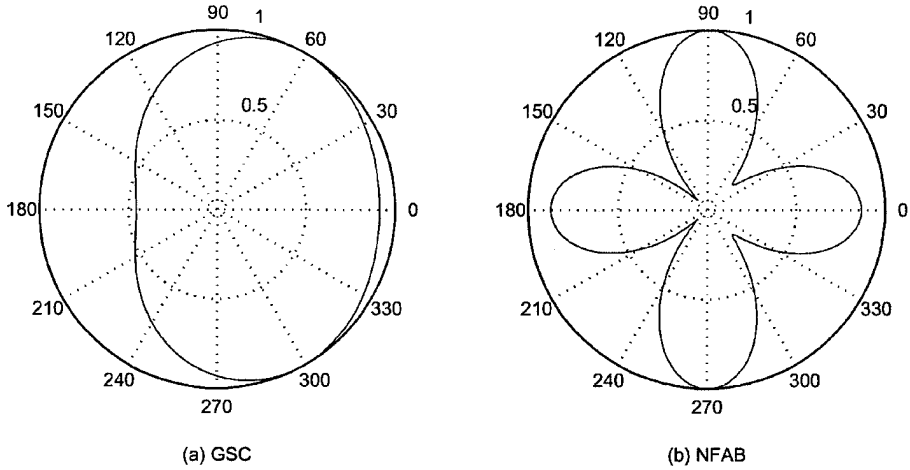


FIG. 9. Overall beamformer directivity pattern at 300 Hz.

• 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

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

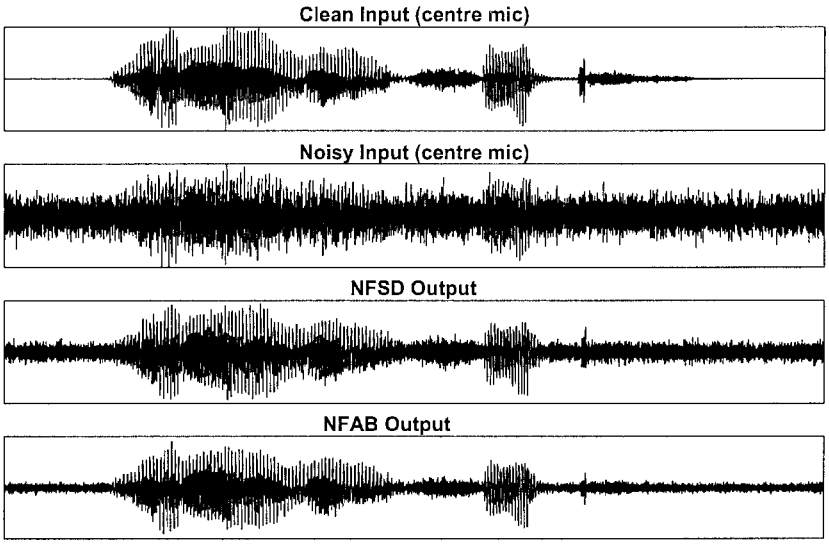


FIG. 10. Sample enhanced signal.

where n is the frame number, and r_o and r_p are the original and processed Pth-order linear predictive coefficients of the nth 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.[1] The measures have been averaged over 10 randomly chosen speech segments taken

TABLE 2

Signal to Noise Ratio Improvement (SANR = 10 dB)

	SLNR (dB)				
Technique	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

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

TABLE 3

Log Area Ratio: (SANR = 10 dB)

	SLNR (dB)				
Technique	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-

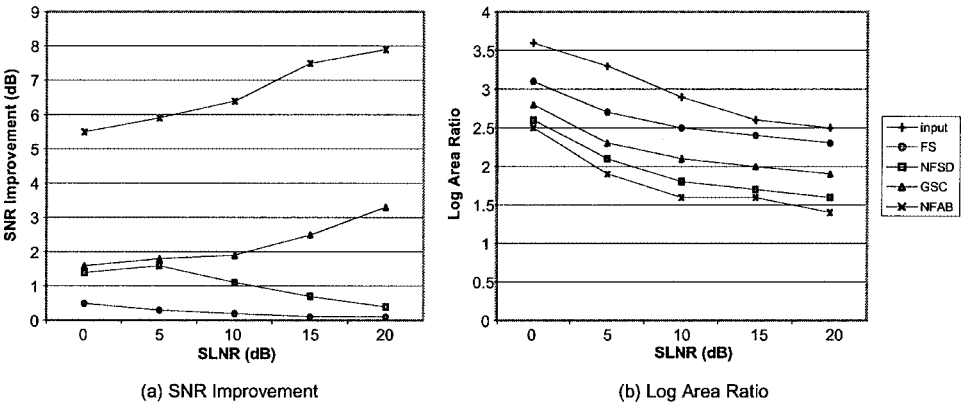


FIG. 11. Speech enhancement measures: (a) SNR improvement and (b) LAR.

TABLE 4

Speech Recognition Results: Word Recognition Rates

	SLNR (dB)			
Technique	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

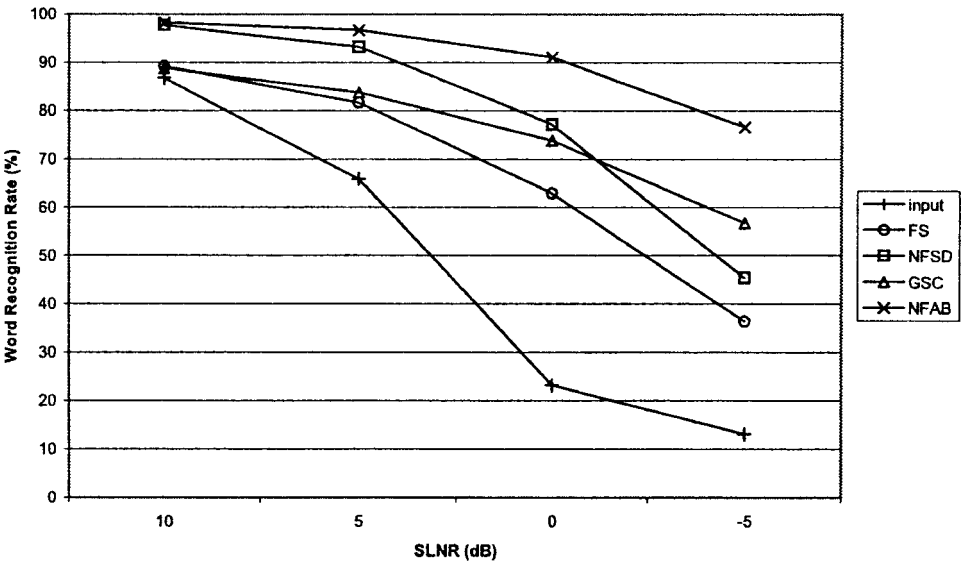


FIG. 12. Speech recognition results: Word recognition rates.

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 enhancment 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.