UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

GOOGLE LLC,

Petitioner

v.

JAWBONE INNOVATIONS, LLC,

Patent Owner

---

Case IPR2022-01124
U.S. Patent No. 11,122,357

---

**Declaration of Shauna L. Wiest Regarding *McCowan***

I, Shauna L. Wiest, state and declare as follows:

## I. Introduction

1. I have prepared this Declaration in connection with Google LLC's ("Petitioner") Petition for *Inter Partes* Review of U.S. Patent No. 11,122,357, which I understand will be filed concurrently with this Declaration.

2. I am currently a senior research analyst with the Research & Information Services team at Finnegan, Henderson, Farabow, Garrett & Dunner, LLP, located at 901 New York Avenue, NW, Washington, DC 20001-4413.

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

4. I earned a Master of Science in Library Science degree from the University of North Carolina at Chapel Hill in 1999, and a Bachelor of Arts in Political Science degree from the University of California at San Diego in 1989. I have worked as a law librarian for over twenty years. I have been employed in the Research & Information Services Department at Finnegan, Henderson, Farabow, Garrett & Dunner, LLP since 2021. Before that, from 2000-2015, I was employed as a Law Librarian at Stoel Rives LLP, and from 2015-2016, I was employed as a Competitive Intelligence Specialist for Nossaman LLP.

## II. Standard Library Practice for Receiving, Cataloging, and Making Materials, Including Serial Publications, Publicly Available

5. I have knowledge of and experience with standard library practices regarding the receipt, cataloging, shelving, and making materials, including serial publications, available to the public. I am fully familiar with and have knowledge of and experience with the Machine-Readable Cataloging (MARC) system, an industry-wide standard that libraries use to catalog materials.

6. The MARC system was developed during the 1960s to standardize bibliographic records so they could be read by computers and shared among libraries. By the mid-1970s, MARC had become the international standard for the storage of bibliographic data and cataloguing. It is still used today. Many libraries provide public access to their MARC records via the Internet and/or their electronic cataloging systems at the library. In a MARC record, each field provides specific information about the cataloged item, including how materials are held and made available to the public.

## III. Serial Publications

7. A serial publication, often known as a "journal," is a resource that is issued in successive parts and has no predetermined conclusion. These successive parts are commonly referred to as "issues," and 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.

8. There are significant differences between cataloging finite resources (books/monographs) and continuing resources (serials). For serials, the catalog record provides information about the serial as a whole, including the first or earliest available issue. It also provides information as to holdings – the volumes and issues, with dates, received by the library and made available to the public. In serials cataloging, there are identifying characteristics unique to serials that are slightly different from monographs (books). The issue date for a print serial publication, for example, generally appears on the cover (front or back), the masthead page, the title page (if any), the table of contents page(s), or on the pages of the individual articles contained in the issue. More information regarding the unique aspects of cataloguing serials can be found at this link: https://www.loc.gov/aba/pcc/conser/scctppt/Basic-2014/Basic-Trainee-Mannual.pdf

## IV. MARC Records

9. The MARC record system uses a specific three-digit numeric code (from 001-999) to identify each field in a catalog record. For example, field tag 008 provides the six-digit date the item was catalogued (*Date entered on file*). The first six characters of field tag 008 are always in the "**YYMMDD**" format. Field tag 022 provides the International Standard Serial Number (ISSN), a unique identification

number assigned to serial publications (continuing resource). Field tag 245 identifies the full title statement for the work and field tag 260 identifies the place of publication, name of publisher, and copyright date of the publication. Field tag 362 identifies the numbering to be used for chronological cataloguing of individual issues of continuing resources (serials). The designations within field tag 362 determine at what point in time the serial began, which guides how issues are checked in, processed, and added to the library's main collection. Finally, the 9XX field tags denote local holdings information for the resource.

10. Based on standard library practice, when a library receives an item, it stamps (or labels) the item with the library name and often with a date that is within a few days or weeks of receipt. Next, the library will catalog the item within a matter of a few days or weeks of receiving it.

11. Generally, after an item is cataloged, the public may access the item by searching a catalog, browsing the library shelves, and either requesting or electronically accessing the item from the library. Standard library practice is to make the item available to the public within a few days or weeks of cataloging it.

## V. Print Holdings and MARC Records for *McCowan*

12. As detailed below, I have reviewed the print public holdings information and Library of Congress and Iowa State University of Science and

Technology MARC records for Iain A. McCowan, Darren C. Moore, and S. Sridharan, "*Near-field Adaptive Beamformer for Robust Speech Recognition*," Digital Signal Processing, January 2002, Vol. 12, No. 1, pages 87-106 (ISSN 1051-2004) ("*McCowan*"). Exhibit 1007 to the concurrently filed Petition is a true and accurate copy of *McCowan*.

13. **Appendix A** to this declaration is a true and accurate copy of the print journal cover, title pages, table of contents, and library date stamp for the issue of Digital Signal Processing containing *McCowan* held by the Library of Congress. **Appendix A** also includes pages 87 to 106 of this issue, which is the article titled "*Near-field Adaptive Beamformer for Robust Speech Recognition*" (*McCowan*). The date stamp for this issue indicates that *McCowan* was received by the Library of Congress on February 19, 2002. The print journal cover provides directional information so that members of the interested public could access the print issue containing *McCowan* at Call Number TK 5102.5 .D4463 Set 1 with Bar Code 0 020 814 014 0 within a matter of a few days or weeks of February 19, 2002.

14. Based on the information in **Appendix A**, it is clear that the issue of Digital Signal Processing containing *McCowan* was received by the Library of Congress on or before February 19, 2002. Based on standard library practice, *McCowan* would have been processed and catalogued by the Library of Congress within a matter of a few days or weeks of February 19, 2002.

15. Accordingly, *McCowan* would have been made available to the public within a few days or weeks of being checked-in and catalogued. Members of the interested public could have accessed *McCowan* by browsing the Library of Congress shelves or by searching the Library's catalog within a few days or weeks of February 19, 2002.

16. **Appendix B** to this declaration is a true and accurate copy of the Library of Congress MARC record for its holdings of the serial publication Digital Signal Processing containing *McCowan,* which was downloaded from https://lccn.loc.gov/91650983/marcxml on April 5, 2022.

17. **Appendix C** to this declaration is a true and correct copy of the Library of Congress public catalog record for its copy of Digital Signal Processing containing *McCowan*, including holdings and location information, which was downloaded from https://catalog.loc.gov/vwebv/holdingsInfo?searchId=17671&recPointer=0&recCount=25&searchType=2&bibId=11390503 on April 5, 2022.

18. **Appendix D** to this declaration is a true and accurate copy of the print journal cover, title pages, table of contents, and library date stamp for the issue of Digital Signal Processing containing *McCowan* held by the Iowa State University of Science and Technology Library. **Appendix D** also includes pages 87 to 106 of this issue, which is the article titled "*Near-field Adaptive Beamformer for Robust Speech*

*Recognition*" (*McCowan*). The date stamp for this issue indicates that *McCowan* was received by the Iowa State University of Science and Technology Library on February 18, 2002. The print journal cover provides directional information so that members of the interested public could access the print issue containing *McCowan* at Call Number TK 5102.5 .D4463 within a matter of a few days or weeks of February 18, 2002.

19. Based on the information in **Appendix D**, it is clear that the issue of Digital Signal Processing containing *McCowan* was received by the Iowa State University of Science and Technology Library on or before February 18, 2002. Based on standard library practice, *McCowan* would have been processed and catalogued by the Iowa State University of Science and Technology Library within a matter of a few days or weeks of February 18, 2002.

20. Accordingly, *McCowan* would have been made available to the public within a few days or weeks of being checked-in and catalogued. Members of the interested public could have accessed *McCowan* by browsing the Iowa State University of Science and Technology Library shelves or by searching the Library's catalog within a few days or weeks of February 18, 2002.

21. **Appendix E** to this declaration is a true and accurate copy of the Iowa State University of Science and Technology Library MARC record for its holdings of the serial publication Digital Signal Processing containing *McCowan*, which was

downloaded from https://iowa-primo.hosted.exlibrisgroup.com/primo-explore/sourceRecord?vid=01IASU&docId=01IASU_ALMA21233591260002756 on May 10, 2022.

22. **Appendix F** to this declaration is a true and correct copy of the Iowa State University of Science and Technology Library public catalog record for its copy of Digital Signal Processing containing *McCowan*, including holdings and location information, which was downloaded from https://iowa-primo.hosted.exlibrisgroup.com/permalink/f/12tutg/01IASU_ALMA21233591260002756 on April 20, 2022.

23. The Library of Congress MARC record (**Appendix B**) for the serial publication Digital Signal Processing containing *McCowan* confirms the fixed data elements of MARC field tag 008 as 900620c19919999mnubrp0a0engc. As discussed above, the first six characters "900620" are in typical "**YYMMDD**" format and indicate that the serial publication Digital Signal Processing was catalogued by the Library of Congress on June 20, 1990. The publication status code "c" appearing in MARC field tag 008 refers to "continuing resource currently published" indicating that Digital Signal Processing began publication in 1991 and is a continuing published resource.

24. The Library of Congress MARC record field tag 022 denotes the unique International Standard Serial Number (ISSN) for Digital Signal Processing as 1051-

2004. The Library of Congress MARC record field tag 245 denotes the title and statement of responsibility for the work as “Digital Signal Processing.”

25. The Library of Congress MARC record field tag 362 provides a chronological designation of a continuing (serial) resource. In this MARC record for Digital Signal Processing the sequential serial designation begins with Volume 1, no. 1 (Jan. 1991)- with no end date noted.

26. Finally, the Library of Congress MARC record field tags 984 and 991 identify the holdings, location, and call number for its copy of Digital Signal Processing containing *McCowan*. The MARC record lists this information as: General Collection, Call Number TK5102.5 .D4463, Serials Location, with holdings beginning with the issue dated January 9, 2001. This information confirms that the journal issue of Digital Signal Processing containing *McCowan* is publicly available and held in the General Collection at Call Number TK5102.5 .D4463.

27. The Library of Congress’s public catalog record for its copy of Digital Signal Processing (**Appendix C**) sets forth the holdings and onsite location information for members of the public seeking the print issue containing *McCowan*. The public catalog record indicates that the print issue containing *McCowan* should be requested in the Jefferson or Adams Building Reading Rooms and is contained within holdings v.3-v.12:no.1 (1993-2002:Jan.) at Call Number TK5102.5 .D4463.

28. The Iowa State University of Science and Technology Library MARC record (**Appendix E**) for the serial publication Digital Signal Processing containing *McCowan* confirms the fixed data elements of MARC field tag 008 as 900620c19919999mnubr1p0a0engd. As discussed above, the first six characters "900620" are in typical "YYMMDD" format and indicate that the serial publication Digital Signal Processing was catalogued by the Iowa State University of Science and Technology Library on June 20, 1990. The publication status code "c" appearing in MARC field tag 008 refers to "continuing resource currently published" indicating that Digital Signal Processing began publication in 1991 and is a continuing published resource.

29. The Iowa State University of Science and Technology Library MARC record field tag 999 (including the AVA field tag) identifies the local holdings, location, and call number for its copy of *McCowan*. The AVA field tag in this MARC record lists the holdings information as: ##$0990008738310102756$822233 591250002756$a01IASU_INST$bPARKS$cStorage Building$dTK5102.5 D4463 $echeck_holdings$jSGEN$k0$p1$qPARKS$tv. 4 (1994)-v. 15 (2005). This information confirms that the print issue of Digital Signal Processing containing *McCowan*, is available and held in the PARKS Storage Building at Call Number TK5102.5 .D4463, with holdings beginning with Volume 4 (1994) through Volume 15 (2005).

30. The Iowa State University of Science and Technology Library's public catalog record for its copy of Digital Signal Processing (**Appendix F**) sets forth the public holdings and onsite location information for members of the public seeking the print issue containing *McCowan*. The public catalog record indicates that the 2002 Bound Issue (volume 12) of Digital Signal Processing containing *McCowan*, is publicly available and held in the PARKS Storage Building at Bar Code 32792018902886 at Call Number TK5102.5 .D4463.

31. Based on this evidence, it is my opinion that Exhibit 1007 is an authentic document, which would have been made publicly available and publicly accessible within a few days or weeks of February 19, 2002.

## VI. Conclusion

32. In signing this Declaration, I understand it will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I understand I may be subject to cross-examination in this case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

33. I declare that all statements made herein of my knowledge are true, 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 June 14, 2022.

Shauna L. Wiest

# Appendix A

DIGITAL SIGNAL PROCESSING 12 JAN 2002 1

DIGITAL SIGNAL PROCESSING 12 JAN 2002 1

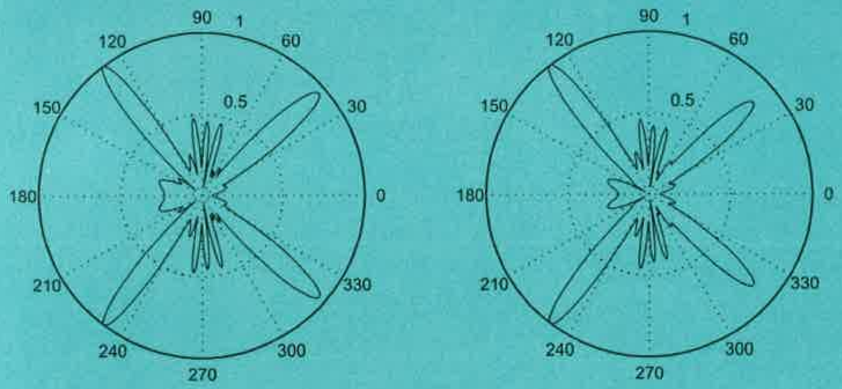
Volume 12, Number 1
January 2002

# Digital Signal Processing

A Review Journal

IDEAL® First
Articles published online first
http://www.idealibrary.com






(a) GSC Lower Path (b) NFAB Lower Path

Editors
**Jim Schroeder**
**Joe Campbell**

ISSN 1051-2004

ACADEMIC PRESS
An Elsevier Science Imprint

# Digital Signal Processing

*A Review Journal*



Cover photo. Lower path directivity pattern at 5000 Hz. See the article by McCowan, Moore, and Sridharan in this issue.

**Digital Signal Processing**

Volume 12, Number 1, January 2002



1051-2004/02 $35.00

MADE IN THE UNITED STATES OF AMERICA

This journal is printed on acid-free paper.

∞

**DIGITAL SIGNAL PROCESSING (ISSN 1051-2004)**

Published quarterly by Elsevier Science.
Editorial and Production Offices: 525 B Street, Suite 1900, San Diego, CA 92101-4495
Accounting and Circulation Offices: 6277 Sea Harbor Drive, Orlando, FL 32887-4900
2002: Volume 12. Price $343.00 U.S.A. and Canada; $374.00 all other countries
All prices include postage and handling

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

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. 

*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

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

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

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

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

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

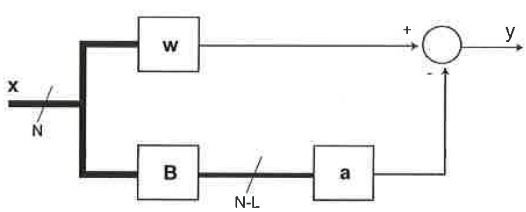



**FIG. 1.** Generalized sidelobe canceler structure.

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

## 2. NEAR-FIELD SOUND PROPAGATION MODEL

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

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

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

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

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

$$d_i = \|\mathbf{p}_s - \mathbf{p}_i\|, \tag{2}$$

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

$$\tau_i = \frac{d_i - d_1}{c}, \tag{3}$$

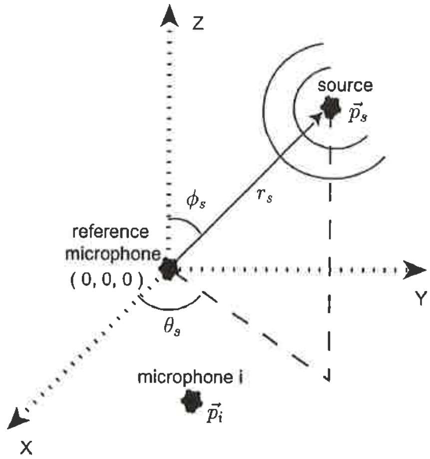



**FIG. 2.** Near-field propagation model.

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 $i$th 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 $(\theta, \phi)$ 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.

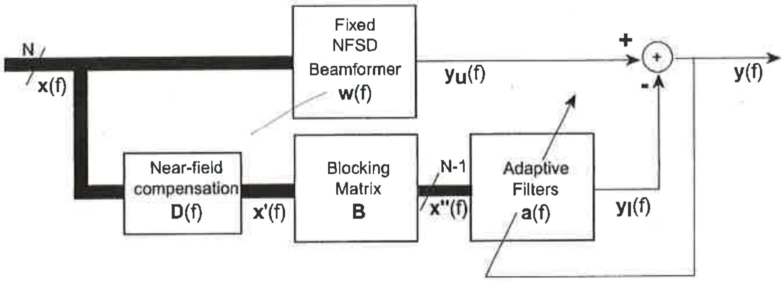



**FIG. 3.** Near-field adaptive beamformer.

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 $(r_s, \theta_s, \phi_s)$, the signal cross-spectral matrix $\mathbf{P}$ simplifies to the propagation vector of the source, and the array gain can be expressed as

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

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 $(\theta, \phi)$.

The superdirectivity problem is thus formulated as:

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

By using a spherical propagation model to formulate the propagation vector, **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, **d**. For a near-field source, the assumption of plane wave (far-field) propagation leads to errors in the array response to the desired signal due to curvature of the direct wavefront. A thorough discussion of the use of a near-field model for superdirective microphone arrays is given by Ryan and Goubran [9].

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

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

as

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

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

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

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

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

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

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

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

### *3.2. Near-field Compensation Unit*

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

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

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

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

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

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

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

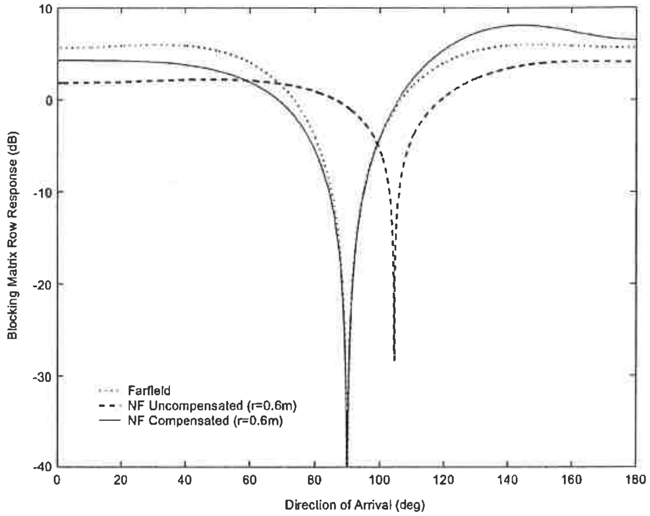



**FIG. 4.** Comparison of blocking matrix row beam-patterns.

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

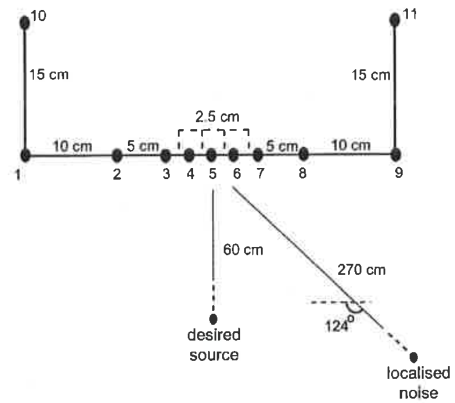



**FIG. 5.** Experimental configuration.

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

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

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

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

### 3.4. Summary of Technique

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

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

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

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

$$y(f) = y_u(f) - y_l(f) \tag{24d}$$

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

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

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

**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 | $\mathbf{w}_o(f) = [\mathrm{diag}(\mathbf{D}(f))]^H - \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$ |
| NFAB | Near-field adaptive beamformer | $\mathbf{w}_o(f) = \mathbf{w}(f) - \mathbf{D}(f)\mathbf{B}\mathbf{a}(f)$ |

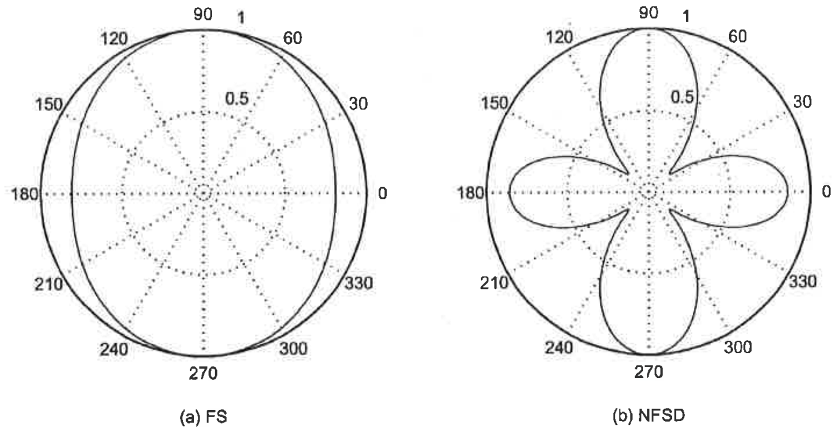
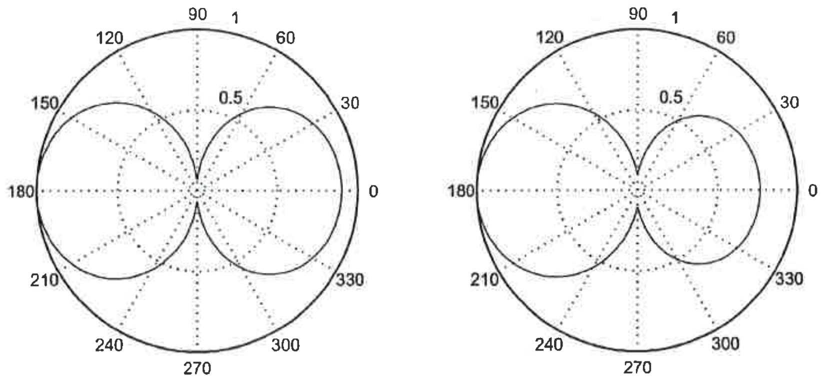



**FIG. 6.** Upper path directivity pattern at 300 Hz.

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

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



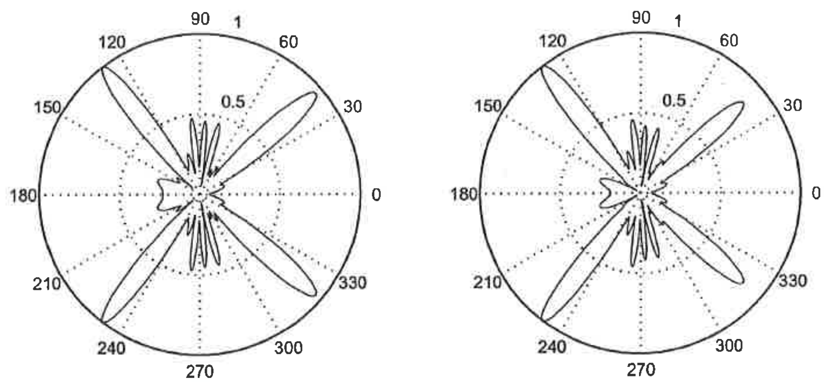

(a) GSC Lower Path (b) NFAB Lower Path

**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




(a) GSC Lower Path (b) NFAB Lower Path

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

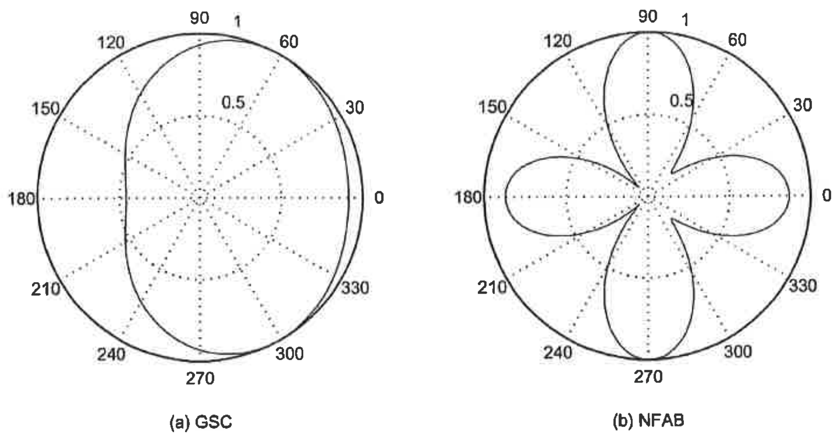



**FIG. 9.** Overall beamformer directivity pattern at 300 Hz.

- 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

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

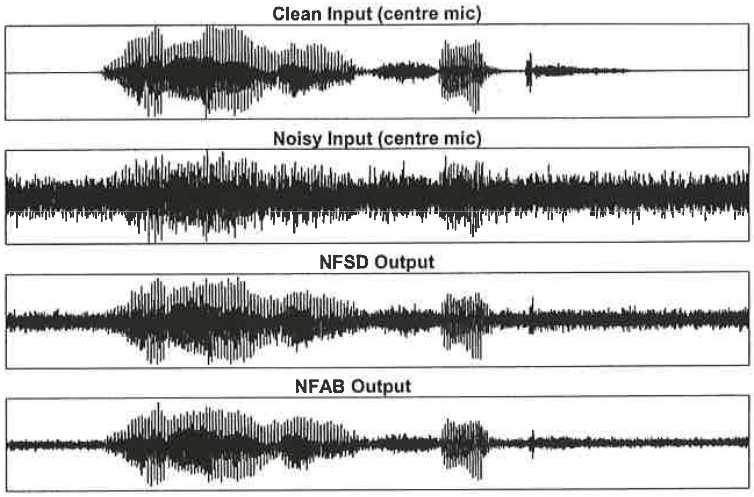



**FIG. 10.** Sample enhanced signal.

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

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

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

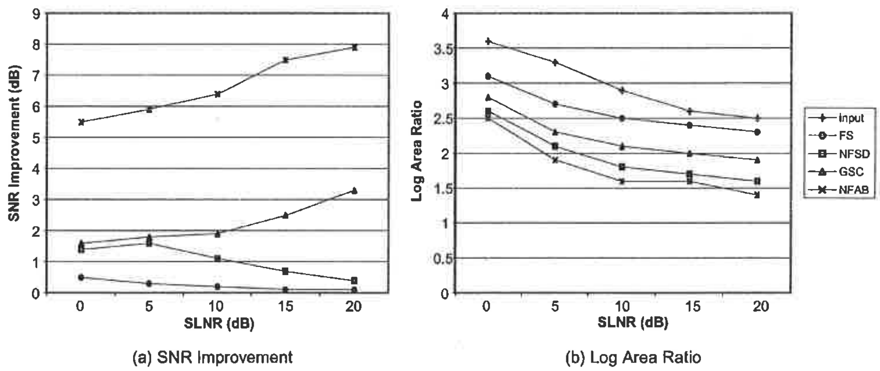
from different speakers in the TIDIGITS database. These results are plotted in Fig. 11.

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

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

## *5.3. Speech Recognition Analysis*

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




(a) SNR Improvement (b) Log Area Ratio

**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

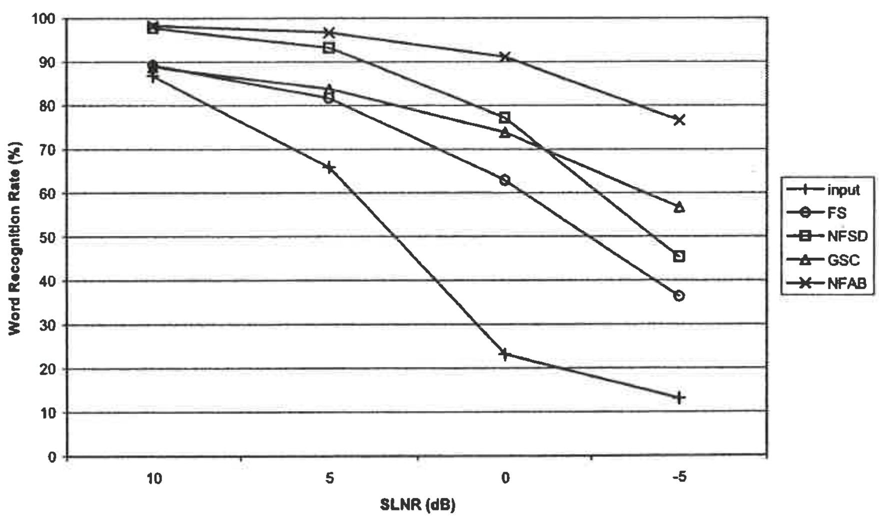



**FIG. 12.** Speech recognition results: Word recognition rates.

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.

# Digital Signal Processing

*A Review Journal*

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

*Contents*



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

1051-2004(200201)12:1;1-E

# Appendix B

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
<record xmlns="http://www.loc.gov/MARC21/slim" xmlns:zs="http://docs.oasis-open.org/ns/search-ws/sruResponse">
  <leader>02696cas a2200541 a 4500</leader>
  <controlfield tag="001">11390503</controlfield>
  <controlfield tag="005">20220221095916.0</controlfield>
  <controlfield tag="008">900620c19919999mnubr p 0 a0eng c</controlfield>
  <datafield ind1=" " ind2=" " tag="010">
    <subfield code="a"> 91650983 </subfield>
    <subfield code="z">sn 90002038 </subfield>
  </datafield>
  <datafield ind1="0" ind2=" " tag="022">
    <subfield code="a">1051-2004</subfield>
    <subfield code="l">1051-2004</subfield>
    <subfield code="2">1</subfield>
  </datafield>
  <datafield ind1=" " ind2=" " tag="030">
    <subfield code="a">DSPREJ</subfield>
  </datafield>
  <datafield ind1=" " ind2=" " tag="035">
    <subfield code="a">(OCoLC)ocm21889588 </subfield>
  </datafield>
  <datafield ind1=" " ind2=" " tag="037">
    <subfield code="b">Academic Press, Inc., 1 E. 1st St., Duluth, MN 55802</subfield>
  </datafield>
  <datafield ind1=" " ind2=" " tag="040">
    <subfield code="a">NSD</subfield>
    <subfield code="b">eng</subfield>
    <subfield code="c">NSD</subfield>
    <subfield code="d">DLC</subfield>
    <subfield code="d">NST</subfield>
    <subfield code="d">CAS</subfield>
    <subfield code="d">NST</subfield>
    <subfield code="d">NSD</subfield>
    <subfield code="d">EYM</subfield>
    <subfield code="d">NSD</subfield>
    <subfield code="d">OCL</subfield>
    <subfield code="d">LYU</subfield>
    <subfield code="d">NSD</subfield>
    <subfield code="d">OCLCQ</subfield>
    <subfield code="d">DLC</subfield>
    <subfield code="d">CIT</subfield>
    <subfield code="d">U9S</subfield>
    <subfield code="d">OCLCQ</subfield>
    <subfield code="d">OCLCF</subfield>
    <subfield code="d">IUL</subfield>
    <subfield code="d">OCLCO</subfield>
  </datafield>
  <datafield ind1=" " ind2=" " tag="042">
    <subfield code="a">nsdp</subfield>
    <subfield code="a">pcc</subfield>
  </datafield>
  <datafield ind1="0" ind2="0" tag="050">
    <subfield code="a">TK5102.5</subfield>
    <subfield code="b">.D4463</subfield>
  </datafield>
  <datafield ind1="0" ind2="0" tag="082">
    <subfield code="a">621.382/2</subfield>
    <subfield code="2">20</subfield>
  </datafield>
  <datafield ind1="0" ind2=" " tag="210">
```

```
    &lt;subfield code="a"&gt;Digit. signal process.&lt;/subfield&gt;
    &lt;subfield code="b"&gt;(Print)&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="0" tag="222"&gt;
    &lt;subfield code="a"&gt;Digital signal processing&lt;/subfield&gt;
    &lt;subfield code="b"&gt;(Print)&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1="0" ind2="0" tag="245"&gt;
    &lt;subfield code="a"&gt;Digital signal processing.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="260"&gt;
    &lt;subfield code="a"&gt;Duluth, MN :&lt;/subfield&gt;
    &lt;subfield code="b"&gt;Academic Press,&lt;/subfield&gt;
    &lt;subfield code="c"&gt;c1991-&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="300"&gt;
    &lt;subfield code="a"&gt;v. :&lt;/subfield&gt;
    &lt;subfield code="b"&gt;ill. ;&lt;/subfield&gt;
    &lt;subfield code="c"&gt;28 cm.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="310"&gt;
    &lt;subfield code="a"&gt;Bimonthly,&lt;/subfield&gt;
    &lt;subfield code="b"&gt;<2005->&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="321"&gt;
    &lt;subfield code="a"&gt;Quarterly,&lt;/subfield&gt;
    &lt;subfield code="b"&gt;<-2003>&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1="0" ind2=" " tag="362"&gt;
    &lt;subfield code="a"&gt;Vol. 1, no. 1 (Jan. 1991)-&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="500"&gt;
    &lt;subfield code="a"&gt;Title from cover.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="588"&gt;
    &lt;subfield code="a"&gt;Latest issue consulted: Vol. 1, no. 2 (Apr. 1991).&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2=" " tag="590"&gt;
    &lt;subfield code="a"&gt;SERBIB/SERLOC merged record&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="0" tag="650"&gt;
    &lt;subfield code="a"&gt;Signal processing&lt;/subfield&gt;
    &lt;subfield code="x"&gt;Digital techniques&lt;/subfield&gt;
    &lt;subfield code="v"&gt;Periodicals.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="0" tag="650"&gt;
    &lt;subfield code="a"&gt;Signal processing&lt;/subfield&gt;
    &lt;subfield code="x"&gt;Digital techniques.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="6" tag="650"&gt;
    &lt;subfield code="a"&gt;Traitement du signal&lt;/subfield&gt;
    &lt;subfield code="x"&gt;Techniques numériques.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="6" tag="650"&gt;
    &lt;subfield code="a"&gt;Traitement du signal&lt;/subfield&gt;
    &lt;subfield code="x"&gt;Techniques numériques&lt;/subfield&gt;
    &lt;subfield code="v"&gt;Périodiques.&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="7" tag="650"&gt;
    &lt;subfield code="a"&gt;Signal processing&lt;/subfield&gt;
    &lt;subfield code="x"&gt;Digital techniques.&lt;/subfield&gt;
    &lt;subfield code="2"&gt;fast&lt;/subfield&gt;
    &lt;subfield code="0"&gt;(OCoLC)fst01118285&lt;/subfield&gt;
  &lt;/datafield&gt;
  &lt;datafield ind1=" " ind2="7" tag="655"&gt;
```

```
      <subfield code="a">Periodicals.</subfield>
      <subfield code="2">fast</subfield>
      <subfield code="0">(OCoLC)fst01411641</subfield>
    </datafield>
    <datafield ind1="0" ind2="8" tag="776">
      <subfield code="t">Digital signal processing (Online)</subfield>
      <subfield code="x">1095-4333</subfield>
      <subfield code="w">(DLC)sn 97006904</subfield>
      <subfield code="w">(OCoLC)36980269</subfield>
    </datafield>
    <datafield ind1=" " ind2=" " tag="850">
      <subfield code="a">CPT</subfield>
      <subfield code="a">CU-I</subfield>
      <subfield code="a">CU-SB</subfield>
      <subfield code="a">DLC</subfield>
      <subfield code="a">InLP</subfield>
      <subfield code="a">MdBJ</subfield>
      <subfield code="a">MiU</subfield>
      <subfield code="a">NN</subfield>
      <subfield code="a">NRU-Mus</subfield>
      <subfield code="a">NcD</subfield>
    </datafield>
    <datafield ind1="7" ind2=" " tag="859">
      <subfield code="u">http://www.idealibrary.com</subfield>
      <subfield code="z">Logon procedure and access to this title is available via the I.D.E.A.L.
      (service provider) homepage</subfield>
      <subfield code="2">http</subfield>
    </datafield>
    <datafield ind1="7" ind2=" " tag="859">
      <subfield code="u">http://firstsearch.oclc.org</subfield>
      <subfield code="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</subfield>
      <subfield code="2">http</subfield>
    </datafield>
    <datafield ind1="7" ind2=" " tag="859">
      <subfield code="u">http://firstsearch.oclc.org/journal=1051-2004;screen=info;ECOIP</subfield>
      <subfield code="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</subfield>
      <subfield code="2">http</subfield>
    </datafield>
    <datafield ind1="2" ind2="0" tag="853">
      <subfield code="8">1</subfield>
      <subfield code="a">v.</subfield>
      <subfield code="b">no.</subfield>
      <subfield code="u">4</subfield>
      <subfield code="v">r</subfield>
      <subfield code="i">(year)</subfield>
      <subfield code="j">(month)</subfield>
      <subfield code="w">q</subfield>
      <subfield code="x">01</subfield>
    </datafield>
    <datafield ind1="4" ind2="1" tag="863">
      <subfield code="8">1.1</subfield>
      <subfield code="a"><11></subfield>
      <subfield code="b"><1></subfield>
      <subfield code="i"><2001></subfield>
      <subfield code="j"><01></subfield>
    </datafield>
    <datafield ind1=" " ind2=" " tag="890">
      <subfield code="a">Digital signal processing.</subfield>
      <subfield code="i">91-650983</subfield>
    </datafield>
    <datafield ind1=" " ind2=" " tag="906">
```

```
      &lt;subfield code="a"&gt;7&lt;/subfield&gt;
      &lt;subfield code="b"&gt;cbc&lt;/subfield&gt;
      &lt;subfield code="c"&gt;serials&lt;/subfield&gt;
      &lt;subfield code="d"&gt;3&lt;/subfield&gt;
      &lt;subfield code="e"&gt;ncip&lt;/subfield&gt;
      &lt;subfield code="f"&gt;19&lt;/subfield&gt;
      &lt;subfield code="g"&gt;n-oclcserc&lt;/subfield&gt;
    &lt;/datafield&gt;
    &lt;datafield ind1=" " ind2=" " tag="920"&gt;
      &lt;subfield code="a"&gt;Keep 1&lt;/subfield&gt;
    &lt;/datafield&gt;
    &lt;datafield ind1=" " ind2=" " tag="984"&gt;
      &lt;subfield code="a"&gt;srvf&lt;/subfield&gt;
      &lt;subfield code="d"&gt;2001-01-09&lt;/subfield&gt;
    &lt;/datafield&gt;
    &lt;datafield ind1=" " ind2=" " tag="985"&gt;
      &lt;subfield code="e"&gt;eserial 200406&lt;/subfield&gt;
    &lt;/datafield&gt;
    &lt;datafield ind1=" " ind2=" " tag="991"&gt;
      &lt;subfield code="b"&gt;c-GenColl&lt;/subfield&gt;
      &lt;subfield code="h"&gt;TK5102.5&lt;/subfield&gt;
      &lt;subfield code="i"&gt;.D4463&lt;/subfield&gt;
      &lt;subfield code="w"&gt;SERIALS&lt;/subfield&gt;
    &lt;/datafield&gt;
    &lt;datafield ind1=" " ind2=" " tag="992"&gt;
      &lt;subfield code="b"&gt;SER&lt;/subfield&gt;
      &lt;subfield code="w"&gt;SERLOC&lt;/subfield&gt;
    &lt;/datafield&gt;
  &lt;/record&gt;
```

# Appendix C

LIBRARY OF CONGRESS CATALOG

## 1. Digital signal processing.

| | |
|---|---|
| **LCCN** | 91650983 |
| **Type of material** | Periodical or Newspaper |
| **Main title** | Digital signal processing. |
| **Published/Created** | Duluth, MN : Academic Press, c1991- |
| **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) |
| **Subjects** | Signal processing--Digital techniques--Periodicals.<br>Signal processing--Digital techniques.<br>Traitement du signal--Techniques numériques.<br>Traitement du signal--Techniques numériques--Périodiques.<br>Signal processing--Digital techniques. |
| **Form/Genre** | Periodicals. |
| **Notes** | Title from cover.<br>Latest issue consulted: Vol. 1, no. 2 (Apr. 1991). |
| **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 |
| **CODEN** | DSPREJ |
| **Dewey class no.** | 621.382/2 |
| **Other system no.** | (OCoLC)ocm21889588 |
| **Invalid LCCN** | sn 90002038 |
| **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<br>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<br>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 |

| | |
|---|---|
| **CALL NUMBER** | TK5102.5 .D4463<br>Set 1 |
| **Request in** | Jefferson or Adams Building Reading Rooms |
| **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.)<br>v.84-v.100 (2019-2020:May) |

| | |
|---|---|
| | 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.)<br>v.84-v.100 (2019-2020:May) |
| **CALL NUMBER** | TK5102.5 .D4463<br>Copy 1<br>Unbound issues |
| **Request in** | Newspaper & Current Periodical Reading Room (Madison LM133) |
| **Latest receipts** | v. 110 (2021 Mar.)<br>v. 109 (2021 Feb.)<br>v. 108 (2021 Jan.)<br>v. 107 (2020 Dec.)<br>v. 105 (2020 Oct.)<br>v. 104 (2020 Sept.)<br>v. 101 (2020 June) |

LC CATALOG
Library of Congress
101 Independence Ave., SE
Washington, DC 20540

Questions? Ask a Librarian:
https://ask.loc.gov/

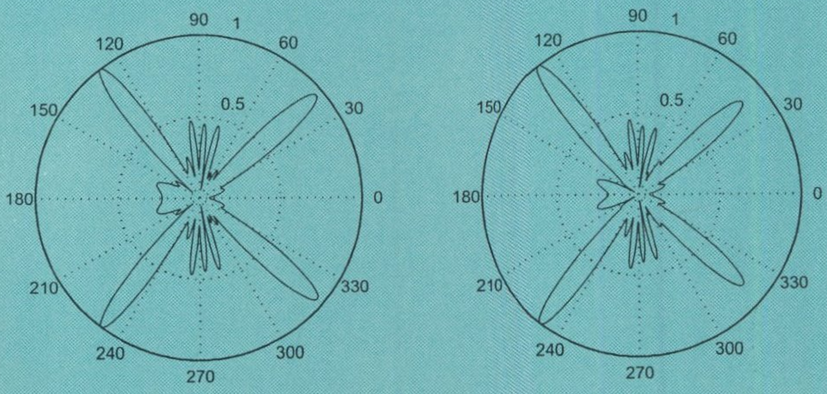
# Appendix D

Volume 12, Number 1
January 2002

# Digital Signal Processing

*A Review Journal*






(a) GSC Lower Path (b) NFAB Lower Path

Editors
**Jim Schroeder**
**Joe Campbell**

ISSN 1051-2004

ACADEMIC PRESS
An Elsevier Science Imprint

# Digital Signal Processing

*A Review Journal*



Cover photo. Lower path directivity pattern at 5000 Hz. See the article by McCowan, Moore, and Sridharan in this issue.

**Digital Signal Processing**

Volume 12, Number 1, January 2002



1051-2004/02 $35.00

MADE IN THE UNITED STATES OF AMERICA

This journal is printed on acid-free paper.

∞

**DIGITAL SIGNAL PROCESSING (ISSN 1051-2004)**

Published quarterly by Elsevier Science.
Editorial and Production Offices: 525 B Street, Suite 1900, San Diego, CA 92101-4495
Accounting and Circulation Offices: 6277 Sea Harbor Drive, Orlando, FL 32887-4900
2002: Volume 12. Price $343.00 U.S.A. and Canada; $374.00 all other countries
All prices include postage and handling

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

# Digital Signal Processing

*A Review Journal*

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

*Contents*



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

1051-2004(200201)12:1;1-E

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. 

*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

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

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

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

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

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

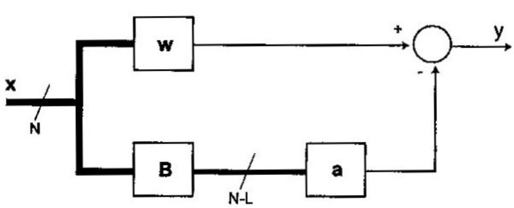



**FIG. 1.** Generalized sidelobe canceler structure.

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

## 2. NEAR-FIELD SOUND PROPAGATION MODEL

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

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

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

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

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

$$\tau_i = \frac{d_i - d_1}{c}, \tag{3}$$

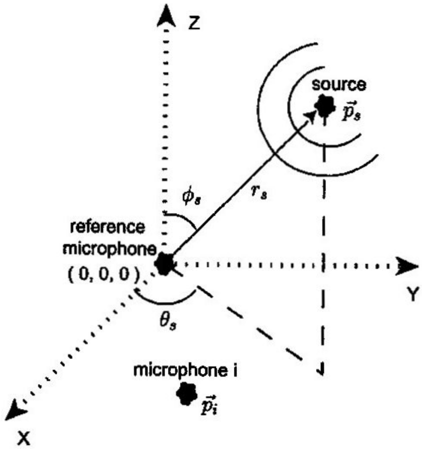



**FIG. 2.** Near-field propagation model.

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

$$\alpha_i = \frac{d_1}{d_i}. \tag{4}$$

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

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

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

$$\mathbf{d}(f, r, \theta, \phi) = \left[\alpha_1 e^{-j2\pi f \tau_1} \ \ldots \ \alpha_i e^{-j2\pi f \tau_n} \ \ldots \ \alpha_N e^{-j2\pi f \tau_N}\right]^T. \tag{6}$$

## 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.

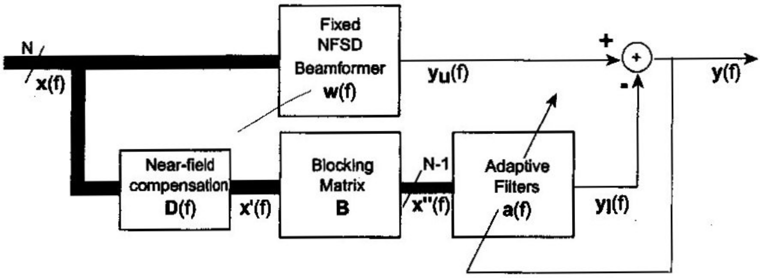



**FIG. 3.** Near-field adaptive beamformer.

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,

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

$(\ )^H$ 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 $(r_s, \theta_s, \phi_s)$, the signal cross-spectral matrix $\mathbf{P}$ simplifies to the propagation vector of the source, and the array gain can be expressed as

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

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

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

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

The superdirectivity problem is thus formulated as:

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

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

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

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

as

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

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

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

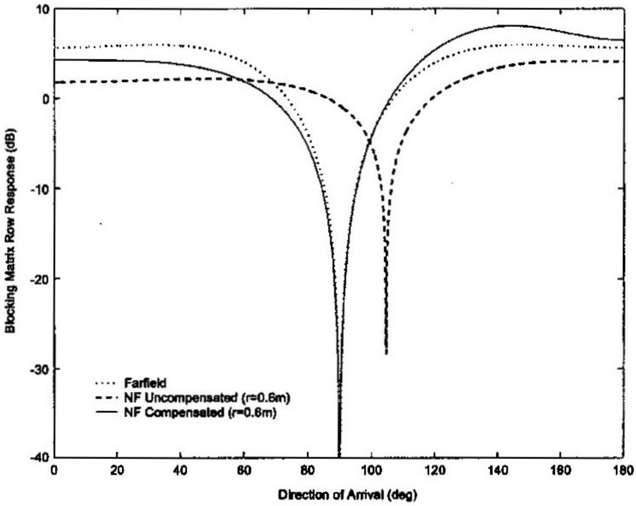



**FIG. 4.** Comparison of blocking matrix row beam-patterns.

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.

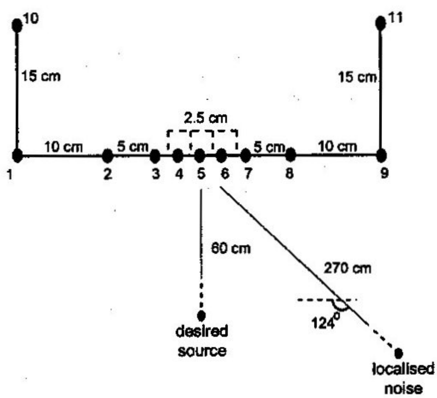



**FIG. 5.** Experimental configuration.

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

### 3.4. *Summary of Technique*

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

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

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

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

$$y(f) = y_u(f) - y_l(f) \tag{24d}$$

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

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

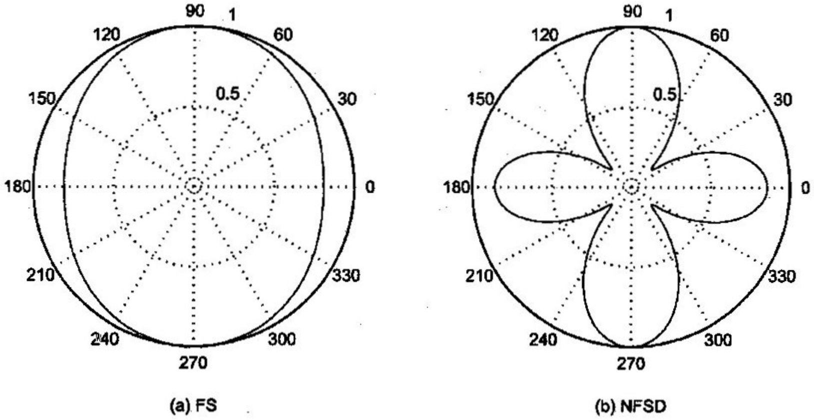
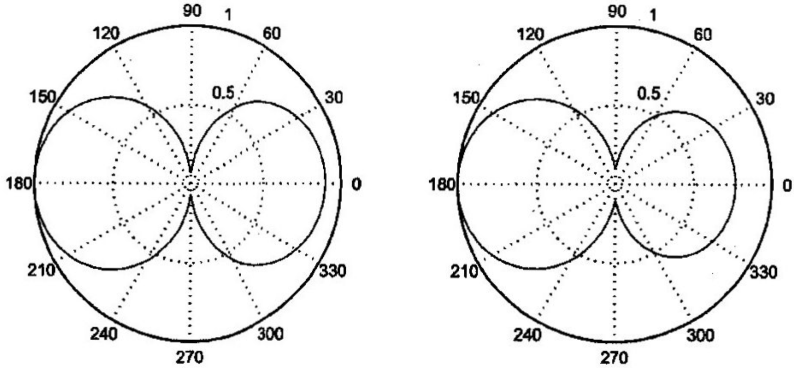



**FIG. 6.** Upper path directivity pattern at 300 Hz.

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

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



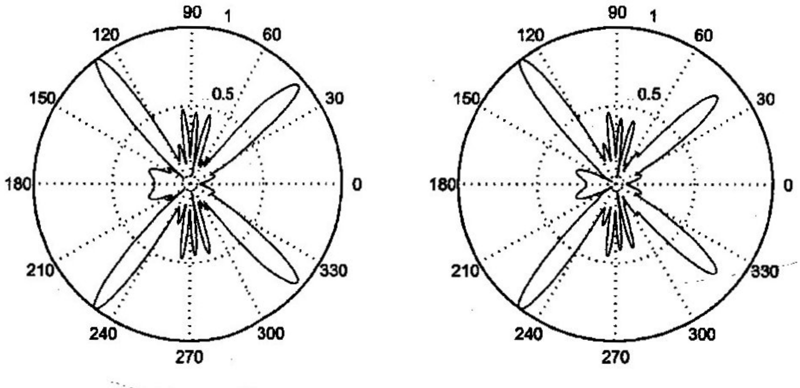

(a) GSC Lower Path (b) NFAB Lower Path

**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




(a) GSC Lower Path (b) NFAB Lower Path

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

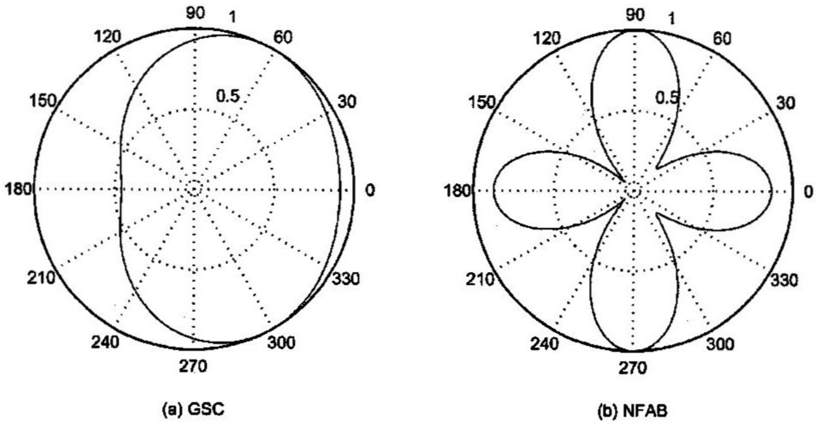



**FIG. 9.** Overall beamformer directivity pattern at 300 Hz.

- 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

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

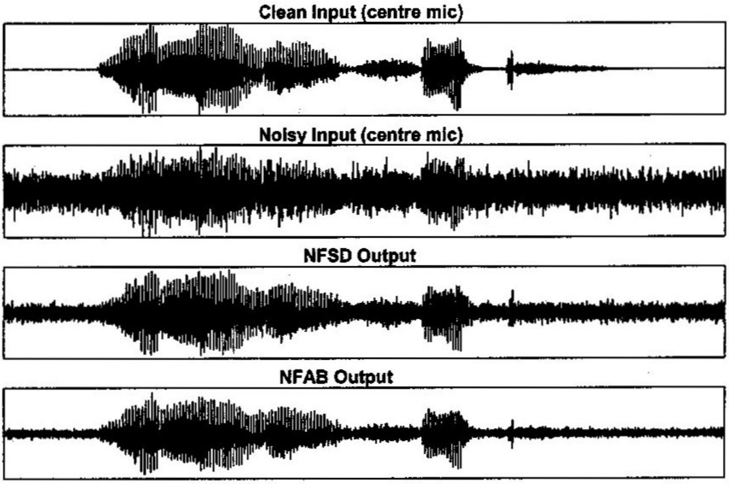



**FIG. 10.** Sample enhanced signal.

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

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

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

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

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

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

### *5.3. Speech Recognition Analysis*

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

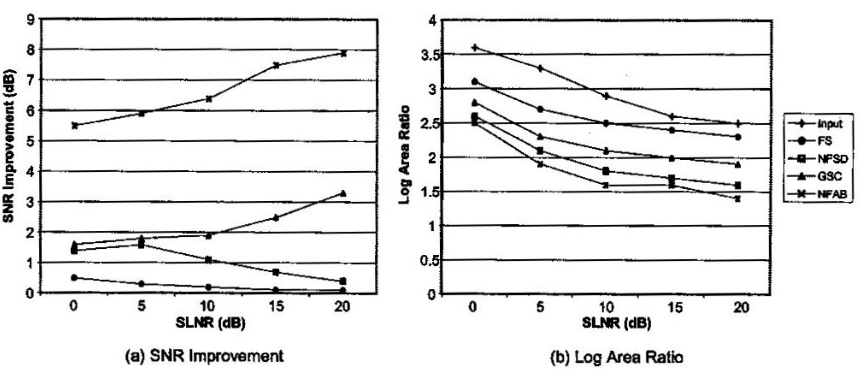



**FIG. 11.** Speech enhancement measures: (a) SNR improvement and (b) LAR.

**TABLE 4**

Speech Recognition Results: Word Recognition Rates

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

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

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

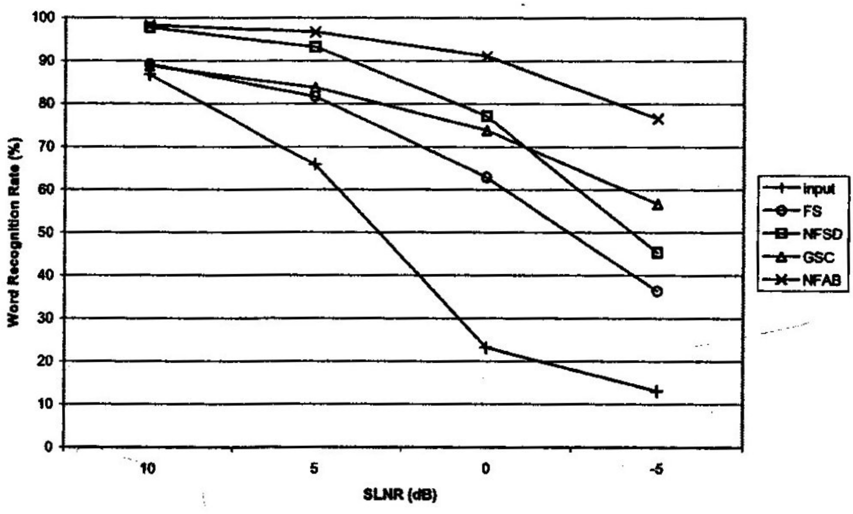



**FIG. 12.** Speech recognition results: Word recognition rates.

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

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

## 6. CONCLUSIONS

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

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

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

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

## ACKNOWLEDGMENTS

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

## REFERENCES

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

---

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

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

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

# Appendix E

```
Leader 01464nas a2200433 a 4500
001    990008738310102756
005    20210426142726.0
008    900620c19919999mnubr1p       0   a0eng d
010    ##$a91650983$zsn 90002038
012    ##$a3$i9111$k1$l1
020    ##$c$110.00
022    0#$a1051-2004
035    ##$a(aleph)000873831ISU01
035    9#$a873831
035    ##$a(OCoLC)21889588
035    ##$a(BIBL)990008738310102756
037    ##$bAcademic Press, Inc., 1 E. 1st St., Duluth, MN 55802
040    ##$aNSD$beng$cNSD$dDLC$dIWA
042    ##$ansdp$alc
050    00$aTK5102.5 D4463
210    0#$aDigit. signal process.
222    #0$aDigital signal processing
245    00$aDigital signal processing.
260    ##$aDuluth, MN :$bAcademic Press,$cc1991-.
300    ##$avolumes :$billustrations ;$c28 cm
310    ##$aBimonthly (varies)
336    ##$atext$btxt$2rdacontent
337    ##$aunmediated$bn$2rdamedia
338    ##$avolume$bnc$2rdacarrier
362    0#$aVol. 1, no. 1 (Jan. 1991)-
500    ##$aTitle from cover.
500    ##$aImprint varies.
530    ##$aAlso issued online.
650    #0$aSignal processing$xDigital techniques$vPeriodicals.
655    #7$aPeriodicals.$2lcgft$0http://id.loc.gov/authorities/genreForms/gf2014026139
776    0#$aDigital signal processing (Online)$c(Online)$d[Amsterdam?] : Elsevier$tDigital signal processing$x1095-4333
890    ##$aDigital signal processing.
911    ##$aMARCIVE 20210811
999    ##$a873831
AVA    ##$0990008738310102756$822233591250002756$a01IASU_INST$bPARKS$cStorage Building$dTK5102.5 D4463$echeck_holdings$jSGEN$k0$p1$qPARKS$tv. 4 (1994)-v. 15 (2005)
CAT    ##$aEngineering_Applied_Sciences$bEngineering$cTelecommunications
INST   ##$a01IASU_INST
INT    ##$aP
PLK    ##$aAdditional Form.$b9920161096502756$mDigital signal processing$g[e-journal].
```

# Appendix F

JOURNAL

## Digital signal processing.

c1991-

Check holdings at Parks Library Storage Building (TK5102.5 D4463 )

View Journal Contents

TOP
TAGS
LOCATIONS / RE...
DETAILS
SEND TO
LINKS
VIRTUAL BROWSE

### Send to

QR | EXPORT TO EXCEL | EXPORT BIBTEX | EXPORT RIS | ENDNOTE | MENDELEY | CITATION | PERMALINK | PRINT | E-MAIL

See something that doesn't look right? Report a Problem

### Locations / request item

Sign in to request, save searches, and more

REQUEST OPTIONS:

Year All | Volume All | Description All

PARKS Storage Building TK5102.5 D4463 LOCATE
Hide Details

v. 4 (1994)-v. 15 (2005)

1 - 10 of 10 Records

| BARCODE | TYPE | POLICY | DESCRIPTION | STATUS | OPTIONS |
|---|---|---|---|---|---|
| 32792021710300 | Bound Issue | Available | v. 15,no.1-2,2005 | On loan until 05/12/2023 17:00:00 CDT (0 requests) | |
| 32792020976910 | Bound Issue | Available | v. 14,2004 | Item in place (0 requests) | |
| 32792019137581 | Bound Issue | Available | v. 13,2003 | Item in place (0 requests) | |
| 32792018902886 | Bound Issue | Available | v. 12,2002 | Item in place (0 requests) | |
| 32792016031555 | Bound Issue | Available | v. 11,2001 | Item in place (0 requests) | |
| 32792015438066 | Bound Issue | Available | v. 10,2000 | Item in place (0 requests) | |
| 32792013061464 | Bound Issue | Available | v. 9,1999 | Item in place (0 requests) | |
| 32792013416494 | Bound Issue | Available | v. 8,1998 | Item in place (0 requests) | |
| 32792013458116 | Bound Issue | Available | v. 6-7,1996-97 | Item in place (0 requests) | |
| 32792012445577 | Bound Issue | Available | v. 4-5,1994-95 | Item in place (0 requests) | |

### Details

| | |
|---|---|
| Title | Digital signal processing. |
| Subjects | Signal processing -- Digital techniques -- Periodicals<br>Periodicals |
| Publisher | Duluth, MN : Academic Press |
| Creation Date | c1991- |
| Format | volumes : illustrations ; 28 cm. |
| Notes | Title from cover.; Imprint varies. |
| Language | English |
| Identifier | ISSN : 1051-2004 |
| Record id | 01IASU_ALMA21233591260002756<br>MMS ID:990008738310102756 |
| Source | 01IASU_ALMA |

### Links

This item in WorldCat®
Staff view

### Virtual Browse

al signal ... | Detection of ... | Digital signal ... | Digital signal ... | Digital signal ... | Digital signal ... | Digital signal ... | Digit...

QR | EXPORT TO EXCEL | EXPORT BIBTEX | EXPORT RIS | ENDNOTE | MENDELEY | CITATION | PERMALINK | PRINT

E-MAIL

TOP
TAGS
LOCATIONS / RE...
DETAILS
SEND TO
LINKS
VIRTUAL BROWSE

See something that doesn't look right? Report a Problem

## Locations / request item

Sign in to request, save searches, and more

REQUEST OPTIONS:

Year All | Volume All | Description All

PARKS Storage Building TK5102.5 D4463
Hide Details

LOCATE

v. 4 (1994)-v. 15 (2005)

1 - 10 of 10 Records

| BARCODE | TYPE | POLICY | DESCRIPTION | STATUS | OPTIONS |
|---|---|---|---|---|---|
| 32792021710300 | Bound Issue | Available | v. 15,no.1-2,2005 | On loan until 05/12/2023 17:00:00 CDT (0 requests) | |
| 32792020976910 | Bound Issue | Available | v. 14,2004 | Item in place (0 requests) | |
| 32792019137581 | Bound Issue | Available | v. 13,2003 | Item in place (0 requests) | |
| 32792018902886 | Bound Issue | Available | v. 12,2002 | Item in place (0 requests) | |
| 32792016031555 | Bound Issue | Available | v. 11,2001 | Item in place (0 requests) | |
| 32792015438066 | Bound Issue | Available | v. 10,2000 | Item in place (0 requests) | |
| 32792013061464 | Bound Issue | Available | v. 9,1999 | Item in place (0 requests) | |
| 32792013416494 | Bound Issue | Available | v. 8,1998 | Item in place (0 requests) | |
| 32792013458116 | Bound Issue | Available | v. 6-7,1996-97 | Item in place (0 requests) | |
| 32792012445577 | Bound Issue | Available | v. 4-5,1994-95 | Item in place (0 requests) | |

## Details

| | |
|---|---|
| **Title** | Digital signal processing. |
| **Subjects** | Signal processing -- Digital techniques -- Periodicals<br>Periodicals |
| **Publisher** | Duluth, MN : Academic Press |
| **Creation Date** | c1991- |
| **Format** | volumes : illustrations ; 28 cm. |
| **Notes** | Title from cover.; Imprint varies. |
| **Language** | English |
| **Identifier** | ISSN : 1051-2004 |
| **Record id** | 01IASU_ALMA21233591260002756<br>MMS ID:990008738310102756 |
| **Source** | 01IASU_ALMA |

## Links

This item in WorldCat®
Staff view

## Virtual Browse

...al signal ...essing : a ...em design ...oach / ...d J. ...

Detection of abrupt changes in signals and dynamical ... 1986?

Digital signal processing / edited by N.B. Jones. ... 1982

Digital signal processing : principles, devices, and applications / ... 1990

Digital signal processing. ... c1991-

Digital signal processing--84 : proceedings of the International ... 1984

Digital signal processing--87 : proceedings of the international ... 1987

Digit... wave... proce... and recog... 1982

## Tags

ADD YOUR OWN TAGS