Speech Enhancement of Single-Source Audio Signal Using Covariance Matrix Adaptation Evolution Strategy

Vincent Fasburg

Department of Computer Science  
Michigan State University  
East Lansing, MI. USA

fasburgv@msu.edu

Joshua Thomas

Department of Computer Science  
Michigan State University  
East Lansing, MI. USA

thom1212@msu.edu

**ABSTRACT**

A Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is a method for efficiently searching through a landscape of multiple real-valued parameters to determine which parameter values give the best solution to a known problem. This paper explores the application of CMA-ES to the problem of speech enhancement, that is, recovering the speech signal from a single microphone signal composed of speech and an unwanted background noise. This research makes use of some of the most popular techniques for speech enhancement, and seeks to find an optimal way to combine these techniques to best recover the original clean speech signal. The results show that the combination of these methods, with parameters tuned using CMA-ES, can outperform the component methods individually, with parameters tuned by experts as would commonly be done.

**CCS Concepts**

• **Information systems➝Database management system engines**   • **Computing methodologies➝Massively parallel and high-performance simulations.**This is just an example, please use the correct category and subject descriptors for your submission*.* The ACM Computing Classification Scheme:

<http://www.acm.org/about/class/class/2012>. Please read the [HOW TO CLASSIFY WORKS USING ACM'S COMPUTING CLASSIFICATION SYSTEM](http://www.acm.org/publications/article-templates/CCS-HOWTO-v6-12Jan2015.docx) for instructions on how to classify your document using the 2012 ACM Computing Classification System and insert the index terms into your Microsoft Word source file.

**Keywords**

Keywords are your own designated keywords separated by semicolons (“;”).

# INTRODUCTION

The field of speech enhancement aims to apply transformations to a digital signal containing both spoken words as well as some type of unwanted noise, and recover the “clean” version of the signal containing speech alone. This is generally done to improve the intelligibility of the speech, either to be listened to by a human, or processed by a voice recognition system.

As mobile and hands-free technology progresses, cell phones are processing voice input not only for phone calls, but as an interface to many other device features. Voice commands are especially likely to be used while the user is driving a car, when it is often unsafe or illegal to use the traditional interface. For this reason a clip of speech corrupted by road noise was used for this research.

In order to produce the cleanest possible speech signal, 3 common methods of speech enhancements are combined using parameters determined using CMA-ES. These methods are Wiener Filtering, Spectral Subtraction, and Noise Gating, which have a combined total of 8 parameters to be tuned by the evolutionary strategy.

# PAGE SIZE

All material on each page should fit within a rectangle of 18 × 23.5 cm (7" × 9.25"), centered on the page, beginning 1.9 cm (0.75") from the top of the page and ending with 2.54 cm (1") from the bottom. The right and left margins should be 1.9 cm (.75"). The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

# Experimental Setup

This section describes the component functions which make up this combined speech enhancement solution, as well as the parameters within these functions that were tuned using the CMA-ES to provide the best enhancement. Finally, the CMA-ES structure that is used for this optimization is described.

## Wiener Filter

SAMPLE: Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Conference’10*, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 …$15.00.

DOI: http://dx.doi.org/10.1145/12345.67890

The Wiener Filter is a classical technique first proposed by Norbert Wiener in 1949[1]. The Wiener Filter is used to estimate a signal of interest through the application of a linear time-invariant filter to a signal consisting of the signal of interest plus additive noise. Although the Wiener filter can be implemented in different ways, with both causal and non-causal forms, the version used in this project is a causal FIR filter, allowing processing to be done in real time, such as during a telephone call. The gains for this filter are determined by taking a noise-only sample of the signal and deconvolving this with noisy signal. Feeding the original noisy signal through this filter results in an estimate of the clean speech signal.

The Wiener Filter has been expanded and improved many times since its original introduction. The specific version that was used in this project includes a two-step noise reduction technique by [3] that uses a decision directed approach to minimize reverberation noise artifacts left over by the original Wiener Filter.

## Spectral Subtraction

The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend phone number (Helvetica 10-point) and e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For more than three authors, you may have to improvise.[[1]](#footnote-1)

## Noise Gate

Please leave 3.81 cm (1.5") of blank text box at the bottom of the left column of the first page for the copyright notice.

## CMA-ES

For pages other than the first page, start at the top of the page, and continue in double-column format. The two columns on the last page should be as close to equal length as possible.

Table 1. Table captions should be placed above the table

|  |  |  |  |
| --- | --- | --- | --- |
| **Graphics** | **Top** | **In-between** | **Bottom** |
| Tables | End | Last | First |
| Figures | Good | Similar | Very well |

## Page Numbering, Headers and Footers

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

# FIGURES/CAPTIONS

Place Tables/Figures/Images in text as close to the reference as possible (see Figure 1). It may extend across both columns to a maximum width of 17.78 cm (7”).

Captions should be Times New Roman 9-point bold. They should be numbered (e.g., “Table 1” or “Figure 2”), please note that the word for Table and Figure are spelled out. Figure’s captions should be centered beneath the image or picture, and Table captions should be centered above the table body.

# SECTIONS

The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub- sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.



Figure 1. Insert caption to place caption below figure.

.

## Subsections

The heading of subsections should be in Times New Roman 12-point bold with only the initial letters capitalized. (Note: For subsections and subsubsections, a word like *the* or *a* is not capitalized unless it is the first word of the header.)

### Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized and 6-points of white space above the subsubsection head.

#### Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized.

#### Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized.

# ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

# REFERENCES

1. Wiener, Norbert. 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York: Wiley. ISBN 0-262-73005-7.
2. Huijun Ding, Ing Yann Soon, Soo Nee Koh, Chai Kiat Yeo, A spectral filtering method based on hybrid wiener filters for speech enhancement. In *Speech Communication,* Volume 51, Issue 3, March 2009, Pages 259-267, ISSN 0167-6393, http://dx.doi.org/10.1016/j.specom.2008.09.003.
3. Plapous, C.; Marro, C.; Scalart, P., "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, Issue 6, pp. 2098 - 2108, Nov. 2006Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.
4. Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.
5. Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.
6. Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology* (Vancouver, Canada, November 02 - 05, 2003). UIST '03. ACM, New York, NY, 1-10. DOI= <http://doi.acm.org/10.1145/964696.964697>.
7. Yu, Y. T. and Lau, M. F. 2006. A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions. *J. Syst. Softw.* 79, 5 (May. 2006), 577-590. DOI= <http://dx.doi.org/10.1016/j.jss.2005.05.030>.
8. Spector, A. Z. 1989. Achieving application requirements. In *Distributed Systems*, S. Mullender, Ed. ACM Press Frontier Series. ACM, New York, NY, 19-33. DOI= <http://doi.acm.org/10.1145/90417.90738>.

Columns on Last Page Should Be Made As Close As Possible to Equal Length

1. If necessary, you may place some address information in a footnote, or in a named section at the end of your paper. [↑](#footnote-ref-1)