

Hearing Aid Fitting with Genetic Algorithms

by

Eric Alan Durant

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2002

Doctoral Committee:

Associate Professor Gregory H. Wakefield, Chair
Professor David J. Anderson
Associate Professor Yili Liu
Dr. Dianne J. VanTasell

ABSTRACT

Hearing Aid Fitting with Genetic Algorithms

by

Eric Alan Durant

Chair: Gregory H. Wakefield

Hearing aids are controlled by numerous parameters, presenting the audiologist with a difficult fitting task. While many of these parameters can be set near their optimal values using prescriptive formulas based on simple measurements of the patient's hearing loss, others are a matter of comfort and other qualities that the patient must evaluate subjectively. Thus, there is much interest in including the patient's preferences as a key element of the fitting process.

To accomplish this, various researchers have looked to mathematical optimization theory. As the number and complexity of hearing aid parameters increase, the shortcomings of these methods become burdensome. To overcome these shortcomings, we investigate the genetic algorithm (GA). The GA is a search procedure that borrows many concepts from biology, including natural selection and genetic crossover and mutation. The GA maintains a population of solutions (hearing aid parameter sets) and repeatedly replaces the least fit solutions with the offspring of better performing solutions. In this thesis, methods are developed for applying the GA to hearing aid fitting. These include an efficient procedure to determine the relative quality of solutions in the population by repeatedly asking the patient to select the better of two alternatives.

To evaluate our approach, we conducted two experiments with eight normal hearing and eight hearing-impaired subjects. In the first, three parameters were varied to control the cancellation of feedback, a common problem in which objects near the aid

and certain motions cause it to squeal. In the second, six parameters were varied to fit three-band dynamic range expansion, which controls tradeoffs between suppressing unwanted background sounds and sufficiently amplifying desired sounds such as speech.

We found that the GA worked very well for fitting the feedback cancellation system, using both objective and subjective measures. In addition, we learned that patients have greatly differing preferences for feedback cancellation parameters and that these preferences do not change much when subjects are retested. The results for the expansion system were also positive, but highlighted some problems and suggested changes that might be made to improve performance when fitting more complicated parameter sets.

© Eric Alan Durant 2002
All Rights Reserved

To my parents, for their complete confidence in me

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Gregory Wakefield. He regularly challenged me to think in new ways and to critically evaluate my work. I will forever appreciate the ardent effort he invested in my education and research.

I would also like to express my gratitude to the Advanced Research group at Starkey Laboratories, Inc. in Eden Prairie, MN for their outstanding collaboration and support. Ms. Melanie Gregan was essential to the success of the work; among other things, she ran the experiments for all of our hearing-impaired subjects. Mr. Stephen MacLennan and Dr. Jon Kindred have my thanks for their work in many areas, especially in integrating the GA and audio processing systems. My many discussions with Dr. Martin Rickert on the design of the genetic operators and statistical methods for evaluating the experimental results were invaluable. I cannot sufficiently express my appreciation for the many ways in which Dr. Dianne VanTasell supported my work — she was enthusiastic about the project from its earliest stages and continues to be an outstanding mentor. Other members of the Advanced Research group at Starkey made contributions to this work that are too numerous to mention here.

I am also grateful to Dr. David Anderson and Dr. Yili Liu, the other members of my committee, for their thorough review and suggestions for improving this thesis.

The first stages of my research were funded by a grant from the Office of Naval Research and by the Department of Otolaryngology at the University of Michigan. I would like to thank Dr. Paul Kileny, Director of Audiology and Electrophysiology and Professor with the Department of Otolaryngology of the University of Michigan Medical Center, who provided insight on the challenges of the hearing aid fitting problem and current clinical practice. I would also like to thank Mr. Thomas O'Connor, a clinical audiologist with the University of Michigan Health System's Otorhinolaryngology Clinic, who graciously allowed me to observe several patients having their

hearing aids fitted.

Also, my appreciation goes to Dr. Corey Cheng and Dr. Maureen Mellody, former members of Professor Wakefield's lab, for their frequent advice on my work and the voice of experience that they provided over the last few years.

I am grateful to the Motorola Foundation, which provided financial support for the middle stages of my research.

My sincerest thanks also go to Dr. Steven Reyer, Professor with the Department of Electrical Engineering and Computer Science at the Milwaukee School of Engineering, my *alma mater*. While I was an undergraduate, he taught digital signal processing with such skill and enthusiasm that learning more about it became one of my major motivations for attending graduate school. When it came time to consider and apply to graduate school, Dr. Reyer provided much valuable counsel. Over the years, he has continued to be a good friend and valued mentor.

Finally, my appreciation goes to Dr. Steven Bomba, retired Vice President of Technology at Johnson Controls, Inc. in Glendale, WI. Working for him while I was an undergraduate, first in the Business Systems Research Group and later on the Phoenix Project, were extremely rewarding experiences. Dr. Bomba was a strong supporter of my plan to attend graduate school and has provided much invaluable advice and insight over the nearly seven years during which I have known him.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTERS	
1 Introduction	1
1.1 The Problem: Growing Complexity	2
1.2 Dissertation Overview	3
2 Background	4
2.1 Hearing Aid Capabilities	4
2.1.1 Current Capabilities	4
2.1.2 Future Capabilities	7
2.2 Hearing Aid Fitting	8
2.2.1 Prescriptive Rationales	9
2.2.2 Adaptive Rationales	11
2.3 Genetic Algorithms	12
2.3.1 Definitions	13
2.3.2 Graph vs. Parameter Space Applications	15
2.3.3 Encoding	15
2.3.4 Selecting Genes for Reproduction	16
2.3.5 Population Update: Steady-state vs. Generational . .	18
2.3.6 Reproduction Operators: Crossover and Mutation . .	18
2.3.7 Previous Perceptual Applications	20
3 Methods	22
3.1 Parametric Model	22
3.2 Inferencing via Paired Comparisons	23
3.3 Initial Populations	24
3.4 Data Collected	25

3.5	Stopping Criteria	25
3.6	Portable Prototyping Device	26
3.7	User Interface	26
3.8	Measures of Success	27
4	Experiment 1: Acoustic Feedback Cancellor Fitting	31
4.1	Methods	32
4.1.1	Parameters	32
4.1.2	Subject Instructions	33
4.1.3	Initial Populations	34
4.2	Results	35
4.2.1	Simulated and Experimental Results vs. Round	35
4.2.2	Gain Margin Measurements	40
4.2.3	Distance from Favorite to Progenitor	40
4.2.4	Decision Bias	42
4.3	Discussion	43
5	Experiment 1 Follow-up: Mutation Strength	46
5.1	Methods	46
5.2	Results	47
5.3	Discussion	49
6	Experiment 1 Follow-up: Cross-subject Preferences	53
6.1	Methods	53
6.1.1	Subject Instructions	55
6.1.2	Presentation Order	55
6.2	Results	57
6.3	Discussion	57
7	Experiment 2: Multiband Expansion Fitting	60
7.1	Background	61
7.2	Methods	62
7.2.1	Parameters	62
7.2.2	Subject Instructions	64
7.2.3	Follow-up Survey	65
7.2.4	Initial Populations	66
7.3	Results	66
7.3.1	Simulated and Experimental Results vs. Round	67
7.3.2	Follow-up Survey	72
7.3.3	Repeatability	72
7.3.4	Decision Bias	73
7.4	Discussion	73
8	Population Variation Metrics and Their Bounds	77
8.1	Notation	78

8.2	Binary Population Variation Metric	79
8.2.1	Exact Lower Bound	80
8.2.2	Approximate Lower Bound	82
8.2.3	Discussion	82
8.3	Parametric Population Variation Metric	82
8.3.1	Definitions	84
8.3.2	Hypotheses	85
8.3.3	Examples	88
9	Paired Comparison Scheduling and Rank Inferencing Algorithm . . .	95
9.1	Background	96
9.2	Notation	97
9.3	Algorithm Components	97
9.3.1	Choosing the Next Pair	98
9.3.2	Inferences in the Dominance Matrix	99
9.3.3	Determining Completion	99
9.3.4	Ranks from the Partial Dominance Matrix	100
9.4	Comparison with Efficient Sorting Algorithms	101
10	Non-binary Operators	103
10.1	First Order Analysis of the Binary Mutation Operator	104
10.2	New Operators	106
10.3	Summary	109
11	Summary and Conclusions	110
	BIBLIOGRAPHY	112

LIST OF TABLES

Table

3.1	User interface.	27
4.1	Experiment 1 pilot study parameter values.	32
4.2	Experiment 1 parameter values.	33
4.3	Experiment 1 initial population indexes.	34
4.4	Experiment 1 favorite solutions for 8 hearing-impaired subjects. . . .	41
4.5	Experiment 1 gain margins for 8 hearing-impaired subjects.	41
4.6	Analysis of decision bias in experiment 1.	43
5.1	Asymptotic diversity for mutation follow-up.	47
6.1	Representative solutions for cross-subject preference follow-up.	54
6.2	Presentation order for cross-subject preference follow-up.	57
6.3	Most likely rankings from 20 paired comparisons.	58
7.1	Experiment 2 parameter values.	63
7.2	Experiment 2 initial population indexes.	67
7.3	Experiment 2 survey results.	72
7.4	Repeatability for four subjects in experiment 2.	73
7.5	Analysis of repeatability for four subjects in experiment 2.	74
7.6	Analysis of decision bias in experiment 2.	75
8.1	Number of connected forms in 3-space.	87
8.2	Minimum norm values for the 2-dimension, 5-item problem.	89
8.3	Forms giving minimum norms as a function of norm order for the 3-D, 5 item case.	90
8.4	Forms giving minimum norms as a function of norm order for the 3-D, 5 item case with an optimal central shift.	91
8.5	Minimum radii for all 5-node 3-D forms.	92
8.6	Minimum variations and radii for two forms containing eight 6-D mem- bers (experiment 2 configuration).	93
10.1	Probability that a binary mutation is of order 0 or 1 for various gene lengths and mutation probabilities.	105
10.2	Probability that any given mutation is of order 1 for various gene lengths and mutation probabilities.	106

LIST OF FIGURES

Figure

2.1	Dynamic range compression.	6
2.2	Hypergeometric distributions for a population of size 8.	17
2.3	Examples of binary crossover and mutation operators.	19
2.4	Conceptual view of mutation and crossover.	20
4.1	Experiment 1: diversity vs. round.	36
4.2	Experiment 1: minimum distance to favorite vs. round.	38
4.3	Experiment 1: probability of improvement being indicated vs. round.	39
4.4	Experiment 1 results: city-block distance from favorite to progenitor.	42
5.1	Experiment 1 mutation strength follow-up: diversity vs. round.	48
5.2	Experiment 1 mutation strength follow-up: minimum distance to favorite vs. round.	50
5.3	Experiment 1 mutation strength follow-up: probability of improvement being indicated vs. round.	51
6.1	Experiment 1 cross-subject preference follow-up: Representative solutions.	54
6.2	All possible orderings of $\binom{5}{2} = 10$ pairs that satisfy the non-adjacency condition.	56
7.1	Experiment 2: diversity vs. round.	68
7.2	Experiment 2: minimum distance to favorite vs. round.	70
7.3	Experiment 2: probability of improvement being indicated vs. round.	71
8.1	Divisions for computing the lower bound on the bitstring variation metric.	81
8.2	Minimum variation of genes of length 10 vs. population size.	83
8.3	p -balls of dimension 2.	85
8.4	Unique 3-D forms with 5 or fewer nodes.	89
9.1	Partial dominance matrix and corresponding graph.	98
9.2	Partial dominance submatrix and corresponding subgraph for inferencing.	99
9.3	Relative efficiency of sorting algorithm.	102
10.1	Zeroth and first order binary mutation distance probabilities for a gene length of 4 and a bit mutation probability of 0.1.	105

10.2	New crossover operator.	108
10.3	New mutation operator.	108

CHAPTER 1

Introduction

Fitting modern DSP-based hearing aids, which have numerous parameters, is a challenge for even the experienced audiologist. Simple but proven fitting rationales are used, limiting the prescribed setting to a subset of the hearing aid’s capabilities [7]. Given the success of the genetic algorithm (GA) in tuning simpler systems with subjective feedback [55, 58, 66], we propose it as an efficient and robust method for fitting hearing aids. GAs borrow biological concepts such as natural selection and mutation and apply them to non-biological search and optimization problems. Many previous studies of genetic algorithms for perceptual tuning [54, 59] and other adaptive hearing aid fitting methods¹ [46] were conducted in tightly controlled laboratory environments, limiting their applicability to real-world use. In contrast, our approach uses a simple paired comparison input method, suitable for both clinical and unsupervised field use. This input method provides an easy way to integrate patient input into the fitting procedure, which is recognized as a key to successful fitting by the hearing aid research community [10].

We first investigate whether a GA can guide a hearing-impaired subject towards perceptually good feedback canceller parameters in everyday conditions. Feedback occurs when hearing aid gain is increased so far that the portion of the signal at the microphone due to the hearing aid output is greater than the portion from the environment. Since the output is fed back and amplified further, the hearing aid quickly reaches its maximum output level. This results in a loud squealing sound.

¹An adaptive method refines the user task depending upon user responses. A non-adaptive method presents a fixed series of tasks (ranking, paired comparisons, etc.) regardless of user input.

An adaptive digital filter can be used to model the feedback path and partially cancel its effects, allowing more gain to be applied successfully. This filter can also produce unwanted artifacts, such as mechanical sounding output or unwanted attenuation of a pure tone, such as a telephone dial tone.² The adaptation of the filter is controlled by a number of parameters — 3 in our implementation. Changing these parameters produces varying degrees of feedback cancellation and artifacts. While there is an objective measure of feedback cancellation, gain margin, only the patient can determine when the cost of the artifacts outweighs the benefit of the feedback cancellation.

Next, we investigate the performance of the GA with a larger and more complicated set of hearing aid parameters. Specifically, we configure the GA to vary six parameters that control a frequency-selective dynamic range expansion system. The purpose of this system is to provide less gain to unwanted background sounds, such as motor hum and breathing, while amplifying speech and other desired sounds sufficiently.

1.1 The Problem: Growing Complexity

Increasing processing power in DSP-based hearing aids has enabled increasingly sophisticated hearing impairment compensation algorithms. This has led to a corresponding increase in the number of system parameters and confusion over how to best set them. For example, after multiband dynamic range compression became feasible, several conflicting rationales for using it were proposed [13, 14]. With more sophisticated processing capabilities such as noise cancellation and frequency transposition [62, 65] on the horizon, the complexity of fitting a hearing aid will increase, while available clinical resources are not likely to increase correspondingly.

Some researchers have already made the leap from prescriptive rationales with *ad hoc* fine-tuning to more efficient adaptive fitting methods, such as the modified simplex method [33]. However, we argue that a more general and powerful search method is required for modern hearing aid fitting [§2.2.2]. Given its success in other perceptual domains, and the limitations of other adaptive hearing aid fitting methods,

²In some countries, including the United States, a dial tone is actually the sum of two pure tones at standard frequencies.

we believe the GA is such a method.

1.2 Dissertation Overview

There are 10 chapters in the remainder of this dissertation. Chapter 2 presents background information on hearing aids and genetic algorithms. Chapter 3 gives an overview of the experimental methods used. Chapters 4 through 7 present the experiments, follow-up experiments, and results. Chapters 8 through 10 discuss various aspects of our system of using a GA for perceptual tuning: measuring convergence, using paired comparisons as input to the GA, and genetic operators for efficient search of the parameter space. Finally, Chapter 11 summarizes the dissertation and presents some conclusions.

CHAPTER 2

Background

The dissertation draws from disciplines that have been mostly disjoint until now. The first of these disciplines covers hearing aids and fitting techniques while the second covers genetic algorithms. The present chapter introduces several aspects of hearing disorders and the hearing aid features that address them. These features are controlled by parameters that must be set for individual hearing aid users. We summarize some methods, or *rationales*, used to choose these parameters. One class of rationales addresses *prescriptive* parameters, which can be somewhat accurately determined from direct measurements of an individual's hearing loss. Another class of rationales, *adaptive methods*, seeks both to fine-tune prescriptive parameters and to set other parameters that are more a matter of individual preference than of objective optimization. The final section of this chapter introduces a genetic algorithm (GA) as a potentially efficient, robust means for adaptive hearing aid fitting.

2.1 Hearing Aid Capabilities

2.1.1 Current Capabilities

Both analog and DSP-based hearing aids are in current use. We focus on DSP-based hearing aids, which have more capabilities and parameters to fit. The most basic DSP-based hearing aid provides fixed gain in one or more frequency bands, with two and three configurable bands being common.

The sound intensity scale commonly used in the hearing aid literature is the sound

pressure level (SPL) scale. On this scale, 0 dB SPL corresponds to $10^{-12} \text{ W/m}^2 = 20 \text{ } \mu\text{Pa}$. A gunshot at close range is approximately 140 dB SPL, a loud rock concert 120 dB SPL, and normal conversation 70 dB SPL [42, p. 11]. Hearing aids typically provide maximum output power in the range of 120 to 140 dB SPL.

When a subject suffers from a hearing loss, loudness is not merely reduced by some factor. Instead, the hearing thresholds as a function of frequency are elevated, but loud sounds, and uncomfortably loud sounds, still sound loud and uncomfortably loud, respectively. For example, while a normal hearing subject can detect a 1 kHz tone at 0 dB SPL, a hearing-impaired subject might require 30 dB SPL. However, an 80 dB SPL tone may seem just as loud to both listeners. Between the threshold and the 80 dB SPL tone, both subjects perceive the same increase in loudness, even though the normal subject perceives it over a range of 80 dB while the hearing-impaired subject perceives it over a range of 50 dB. This effect is called *loudness recruitment*.

To compensate for loudness recruitment, many hearing aids go beyond fixed gain by providing dynamic range compression in each frequency band. Operation is typically linear up to some level, after which gain is reduced with increasing input level to compress the output range [Figure 2.1]. Such a system has several parameters, including compression ratio, threshold, and attack and release time constants. More complex I/O gain curves also exist, including various piecewise linear and first derivative continuous functions. There are several rationales for applying dynamic range compression, some of which we overview in §2.2.1.

A common problem in hearing aids is feedback, which is a whistling or howling sound that can occur when the closed loop gain of the hearing aid exceeds unity for some frequency. The closed loop gain is the gain from the microphone, through the hearing aid processing circuitry, to the receiver¹, and back to the microphone. The path from the receiver back to the microphone, the external feedback path, is primarily through bone and soft tissue surrounding the hearing aid. Hearing aids typically provide gains in excess of 20 dB, so feedback may occur despite much attenuation in the external feedback path. The external feedback path is extremely sensitive to chewing, yawning, and bringing objects such as a telephone near the ear. These

¹The receiver converts electrical signals to sound waves.

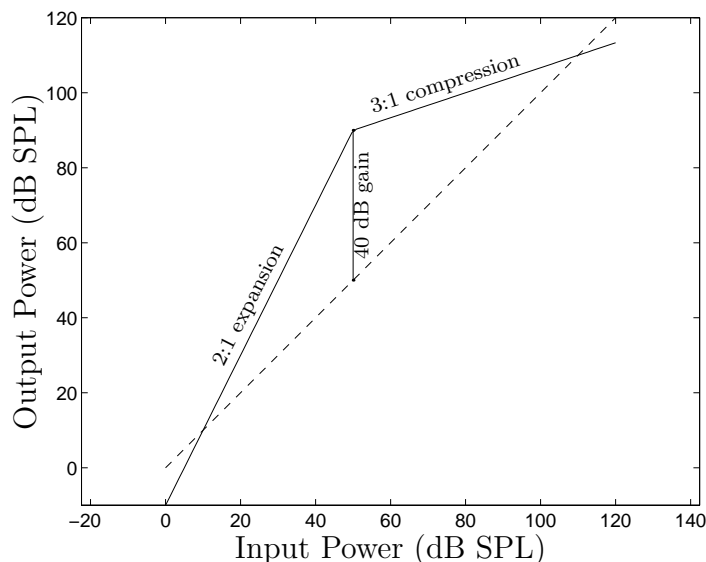


Figure 2.1: Dynamic range compression. A gain of 40 dB at an input kneepoint of 50 dB SPL with 3:1 compression above the kneepoint and 2:1 expansion below the kneepoint is illustrated. The soft to loud speech range of 50 to 80 dB SPL is mapped onto the compressed but louder output range of 90 to 100 dB SPL.

stimuli often initiate feedback in an otherwise stable system. Once initiated, drastic means, such as manually lowering the hearing aid volume, are sometimes required to eliminate feedback. Because feedback is such a common problem, many hearing aids include feedback management that reduces frequency-specific gain when the output level exceeds a frequency-specific threshold. Another approach, and that used in our first experiment, is to use LMS filtering² to cancel signal components with specific correlation properties.

Some hearing aids also manage noise, both external (*i.e.*, breathing noise, wind noise, and any unwanted signal present in the environment) and internal (*i.e.*, hearing aid hardware and software). Most noise management approaches, including the one used in our second experiment, are based on spectral content and involve frequency specific gain and compression adjustment, sometimes including dynamic range expansion on the low end to make breathing and other noises inaudible.

All the aforementioned hearing aid capabilities increase the number of interacting

²Kates [30] discusses various methods for feedback cancellation in hearing aids, including an adaptive least-mean square (LMS) filter. Joson *et al.* [29], Maxwell and Zurek [37], Rafaely and Roccasalva-Firenze [50], Siqueira *et al.* [56], and Wang and Harjani [68] investigate other adaptive feedback cancellation methods for hearing aids.

parameters that must be set for each patient. The fitting rationales discussed in §2.2.1 address some of these, but do not cover the gamut of hearing aid capabilities. Thus, we look to adaptive methods [§2.2.2] to address larger parameter sets.

Patients also have varied reactions to processing artifacts caused by such hearing aid features as feedback cancellation and noise management systems. These reactions are based on both personal preference and expectations from past experience. Since such reactions vary greatly between patients and are expensive to quantify, they cannot be incorporated into prescriptive fitting methods and are instead in the domain of adaptive methods [§2.2.2], which consider preferences of individual users.

Further, prescriptive rationales do not address the need to customize fitting for differing environments such as offices, restaurants, and theaters. These environments have unique ambient noise characteristics (*e.g.*, spectral content and variability). They also have unique reverberation characteristics that affect the coloration of all sounds, including desired signals such as speech. A hearing aid that adjusts its gain and noise management features in response to the environment can provide the user with a better listening experience.

2.1.2 Future Capabilities

Loss of spectral resolution, distorted perception, and other effects have a variety of causes, including conductive losses and sensorineural damage. This means the change in perceptions caused by hearing loss cannot be fully explained by loudness recruitment. So, simple amplification and dynamic range compression are insufficient. Although current hearing aids do not address these deficiencies, future aids will attempt to remedy them through frequency transposition [62, 65] and other techniques.

Like simplex³, but without its limitations, our method is quite general in that it is readily adaptable to such future capabilities. This is because our approach treats the hearing aid as a “black box” with several controlling parameters. It can discover

³For a discussion of this local optimization method originally published in 1965, see, for example, [34, p. 112], which explains, “The Nelder–Mead [simplex] method attempts to minimize a scalar-valued nonlinear function of n real variables using only function values, without any derivative information (explicit or implicit). [...] [The simplex method] maintain[s] at each step a non-degenerate simplex, a geometric figure in n dimensions of nonzero volume that is the convex hull of $n + 1$ vertices.”

the ranges of these parameters as part of the search process, but knowledge about the problem domain can be used to formulate a mapping of the parameters to improve performance, as will be done for our second experiment [§7.2.1].

2.2 Hearing Aid Fitting

Fitting is the process of setting a hearing aid's parameters to give the maximum benefit to the patient. The fitting process typically begins with the audiologist measuring the patient's pure tone hearing thresholds at several standard frequencies in each ear. One common set of frequencies begins at 250 Hz and proceeds to 8 kHz in octave steps. An *audiogram* is then generated by plotting these thresholds on an inverted decibel SPL scale with a logarithmic frequency axis. Since the scale is inverted, normal hearing (thresholds between about 0 and 20 dB SPL, depending upon age and frequency) is represented by points near the top of the audiogram, while severe losses (thresholds greater than about 70 dB SPL) are represented by points near the bottom of the audiogram. After taking the audiogram, the audiologist may take additional measurements, such as MCLs (most comfortable levels) or UCLs (uncomfortable levels, the sound pressure level at which discomfort first occurs as a function of frequency). Other tests may also be performed, such as word recognition tests at various presentation levels. The audiologist then takes an impression of the concha and the outer portion of the canal of each ear being fitted. Typically, the audiologist uses a computer program that implements a prescriptive rationale to derive hearing aid parameters. A *prescriptive rationale* is a method of determining hearing aid parameters for a group of devices such as linear hearing aids based on audiometric data. The parameters determine which of the available aids are suitable for the patient. The audiologist then selects and orders the aid(s) from a manufacturer, specifying several parameters and providing the ear mold impression(s).

The patient has another appointment with the audiologist after the aid(s) arrive from the manufacturer. During this appointment, the audiologist checks that the aids are comfortable for the patient, both physically and in the sound they produce, and that excessive feedback does not occur. The audiologist may then evaluate the patient's performance, either informally or by using standard intelligibility tests. The

audiologist then fine-tunes the aid parameters and schedules a follow-up appointment with the patient during which additional fine-tuning may be performed. As the patient becomes acclimatized to the hearing aid, additional follow-up appointments and fine-tunings may be needed. *Adaptive rationales*, which are procedures for refining hearing aid parameters in response to user inputs, seek to model, in part, this fine-tuning process performed by audiologists.

2.2.1 Prescriptive Rationales

Linear Rationales

Byrne and Dillon [8] present a linear rationale (prescription of relative gain as a function of frequency), the revised NAL (Australian National Acoustic Laboratories) procedure, based only on the audiogram. Their work is based on clinical research and analysis of earlier linear fitting methods. It prescribes gain as a function of frequency having a slope of 0.31 times the audiogram slope, with frequency-specific offsets designed to maximize speech intelligibility. The slope factor was determined from analysis of experimental data and partially compensates for broadening of the auditory filters⁴ and loudness recruitment⁵. Their research showed that using only the audiogram was at least as effective as using MCLs and more effective than using UCLs.

Compressive Rationales

Byrne [6] summarizes four ends to which dynamic range compression is applied:

- Noise reduction — Low frequency components dominate much noise. Simple noise reduction systems take advantage of this by cutting low frequency gain as overall signal level increases.
- Improving audibility — A quiet but desired sound, such as a soft voice, is often below hearing-impaired thresholds. Dynamic range compression can amplify

⁴Both sensorineural hearing loss and increase in sound level broaden the auditory filters, which results in a loss of spectral resolution. Thus, amplifying in direct proportion to the hearing loss would cause greater broadening in regions of greater hearing loss and reduce intelligibility.

⁵The compensation can be only partial since loudness recruitment is a non-linear effect.

such a sound, making it audible to the hearing-impaired listener, while not making normal to loud sounds uncomfortably loud.

- Loudness normalization (recruitment compensation) — Sensorineural hearing impairments lead to elevations in threshold, but normally have little effect on the sound level required for discomfort. Thus, the perceived loudness for a hearing-impaired subject increases more rapidly with sound pressure level than for a subject with normal hearing. This is called loudness recruitment. Compression may compensate for loudness recruitment by providing the hearing-impaired subject with the same loudness perceived by an unimpaired subject [14].
- Automatic volume control — When compressive time constants are much longer than syllabic durations, system behavior is nearly linear during syllables and words. In this case, dynamic range compression effectively varies overall volume and may adjust the relative amplification of frequency bands. In some hearing aids, such a system is used to eliminate the need for manual volume control.

Byrne comments on the conflicting premises and goals of various compressive rationales. He notes that many focus on loudness normalization, which has not been shown to be directly beneficial to the patient. Additionally, fast-acting syllabic compression systems reduce consonant-vowel loudness contrasts. These contrasts are very important for understanding speech, so reducing them often reduces speech intelligibility.

NAL-NL1 [14] is a recently developed non-linear rationale. It seeks to maximize the speech intelligibility index and is based on optimal linear prescriptions across a wide range of audiograms and speech levels. It takes only the audiogram and number of compression channels as input and prescribes crossover frequencies, compression ratios, and compression thresholds. NAL-NL1 does not prescribe compression time constants. However, given that NAL-NL1 is derived by considering speech at various fixed levels, the prescribed settings seem most appropriate for a long time constant (automatic volume control) system and may not apply to a syllabic compression system. The NAL is currently investigating the effect of time constants with this rationale.

2.2.2 Adaptive Rationales

Adaptive rationales are procedures for refining system parameters in response to user inputs. Some rationales make many assumptions about the “response surface”, while others make none. The *response surface* is a multidimensional objective function. Its inputs are the system parameters and its output quantifies the quality of the system as perceived by the user. Estimating a perceptual response surface is a complicated task requiring extensive user input [3, 61]. It requires a disproportionately large effort relative to the benefits to be had in the perceptual tuning problem. Tournament-like strategies, such as double elimination and iterative round robin, make very few assumptions about the response surface, but require many comparisons. The simplex method assumes that the response surface is unimodal and requires fewer comparisons.

Kuk and Pape [33] used a modified simplex method under various speech stimulus conditions to guide patients to good linear gain settings from a grid of four low-frequency gains by four high-frequency gains. This method was found to be reliable for this experiment and Neuman *et al.* [46] found it to be efficient when compared to two other adaptive hearing aid fitting methods, iterative round robin and double elimination. The original simplex method of Box and Wilson [4] was modified for a fixed rectangular lattice and fixed step size of one lattice increment. In addition, the modified method tested all active pairs around the pivot solution, even if some of the pairs had been tested in previous iterations. This is less efficient than the original simplex method, but gives the modified method some robustness to human judgment error and is also the basis of its stopping criterion. The modified method stops after three reversals in each dimension. Thus, it will stop at a peak in the response surface, and may also stop on a plateau. This method does not work well when there are multiple peaks in the response surface⁶, and we shall see that it lacks the ability of the GA to combine information from multiple solutions.

⁶Indeed, Neuman *et al.* [46] foresaw this problem, but noted that it did not seem to exist for then-current hearing aid parameter spaces.

2.3 Genetic Algorithms

GAs are a class of algorithms for function maximization. They do not rely on the existence of a continuous parameter space or differentiability of the error function, but they do exploit features of the error function. Assumptions regarding continuity and critical points⁷ are required for efficient use of many optimization strategies, such as gradient descent, response surface modeling, and simplex methods [34]. However, many practical design problems involve nonlinear systems and multimodal or discontinuous error surfaces. GAs are attractive for addressing such problems.

A genetic algorithm performs a parallelized stochastic search. It is parallelized in that many solutions (“genes”) are considered simultaneously (constituting a “population”). It is stochastic in that (i) solutions are chosen for refinement randomly, with the probability of selection enhanced by the solution quality (“fitness”) and (ii) the search direction about a solution (or between solutions) is chosen randomly.

The search problem may be posed in either continuous or discrete space. In the present work, we pose it as a discrete-space⁸ optimization problem. This follows from Holland’s original work ([27] provides a concise summary), which used bitstrings exclusively to represent solutions. Discrete-space does not mean that we are confined to bitstrings, however. For now, many of our examples focus on bitstrings for simplicity, but our experiments use a multilevel parameter representation as discussed in §3.1.

Regardless of how the problem is posed, many search algorithms besides the GA may be applied. Most of these suffer from at least one of the following drawbacks that GAs avoid:

- They use a fixed, objective error criterion.
- They perform only local optimization.
- They cannot combine good features of multiple solutions.

The GA overcomes the first drawback by incorporating the error criterion into its fitness function. The GA’s parallelism, that is, its ability to work with many solutions

⁷At a critical point, the first derivative is either zero or undefined.

⁸Kawamata *et al.* [31] used a discrete-space approach to the 2-D filter design problem with a GA. Many researchers, including Michalewicz [38], have since used continuous-space parameter representations of other design spaces.

at once, allows it to overcome the other two drawbacks.

For an example of an algorithm that clearly suffers from the first drawback, consider the 1965 work of Steiglitz and McBride [57]. The authors present a method to solve the system identification problem using the linear regression equations that result from minimizing squared error. Their method performs well over a wide range of test cases, but it is not clear how well it avoids local optima that may be globally suboptimal. Also, since their method depends on properties of squared error and the resulting regression equations, it is not suitable for minimizing other error functions.

Simulated annealing (see, for example, [49]) is a stochastic search method that begins by using large random perturbations to search for improvements to an initial solution. Analogous to the lowering of temperature during the annealing of metal or glass, simulated annealing gradually lowers the temperature of the search. This temperature controls the size of the random perturbations. Since simulated annealing only considers the maximum objective value seen, it can be used with arbitrary error criteria. Also, when the initial perturbations are sufficiently large and the cooling schedule is carefully chosen, simulated annealing is effective at not becoming stranded in local optima. However, since it only considers perturbations about a single point, simulated annealing cannot combine features from multiple solutions.

After defining several terms, the remainder of this section explains aspects of GAs relevant to our work.

2.3.1 Definitions

- child – a gene that is output by a reproduction operator
- competition – the relative advantage in selection probability given to genes with high fitness over genes with low fitness
- crossover – a type of reproduction operator that exchanges information between 2 genes to produce 2 new genes
- decode – convert a genotype to the corresponding phenotype
- direct encoding – encoding in which there is a linear relationship between parameters and their representation in the genotype; when used of binary genotypes,

this also means that the bits are interpreted as a binary number (as opposed to using a Gray code, for example)

- encode – convert a phenotype to the corresponding genotype
- fitness – a number representing the quality of a particular gene; a larger value represents a better gene
- fitness function – a method of determining fitness given a gene
- gene – the fundamental structure manipulated by the GA; represents a particular solution to the problem being solved
- generational update – replacing the entire population before recalculating fitness and allowing new genes to reproduce, in contrast to steady-state update
- genotype – the representation of a gene, such as a bitstring or list of values
- mutation – a type of reproduction operator that modifies a single gene slightly to produce a new gene
- parent – a gene that is input to a reproduction operator
- phenotype – solution parameters corresponding to a particular gene
- population – a collection of a fixed number of genes under consideration by the GA
- reproduction – generating new genes from old genes
- reproduction operator – a method of reproduction, such as mutation or crossover
- selection probability – the probability that a gene is selected for reproduction, based either on the gene's fitness or the rank of its fitness in the population
- steady-state update – replacing only a small number of genes before recalculating fitness and allowing new genes to reproduce, in contrast to generational update

2.3.2 Graph vs. Parameter Space Applications

Classic GA techniques use bitstrings and operators that work at the level of individual bits [18]. This makes them extremely flexible and only imposes the most general of geometries on the search space. For many problem classes, with appropriate encoding (*e.g.*, of trees or other graph theoretic constructs), the classic techniques perform quite well. This is in contrast to problems in which the fundamental element is the parameter, which may take on a range of continuous or discrete values. These problem spaces have an inherent geometry much less general than that imposed by the bitstrings of the classic GA. This is not to say that the spaces are continuous, simply described, or have a simple mapping to the corresponding perceptual space. However, in such spaces, there typically is a greater importance on the ordering of parameter values. Also, multilevel concepts (such as Minkowski norms), not binary concepts (such as Hamming distances), are of primary relevance. Thus, we investigate diversity metrics (Chapter 8) and operators appropriate for parameter space applications (Chapter 10), such as those arising in hearing aid fitting problems.

2.3.3 Encoding

GAs, which were originally developed by Holland [26], are described in several textbooks [12, 20, 21, 24, 40, 41, 70] and articles [18, 19, 25]. Briefly, a GA operates on genes. In many GA applications, genes are fixed-length bitstrings that encode solution parameters. Direct encoding is seldom optimal, but it gives good performance for many problems.

We can enhance the GA's performance with domain knowledge, such as smoothness and modality of the error surface. For example, when using a mutation operator [§2.3.6] with a smooth error surface, we desire that a single bit change in the genotype lead to a small change in the phenotype. We can increase the probability of the change in phenotype being small by using Gray code order⁹. Alternatively, we might

⁹In a Gray code [22], codewords are assigned such that only one bit differs between any two adjacent codewords. Such an assignment requires no more bits than direct encoding, which requires $n = \lceil \log_2 N \rceil$ bits. Neglecting the values at each end of the range and assuming that N is a power of 2, a Gray code assignment results in a probability of $2/n$ that a 1-bit change is a 1-step change, while direct encoding results in a probability of $1/n$ that a 1-bit change is a 1-step change. Hollstien [28] investigated using Gray codes for parameter encoding with GAs.

operate directly on parameters instead of their bit-by-bit representations [12] or, as we do in Section 10.2, design an operator that works at a higher level of abstraction than that of individual bits.

2.3.4 Selecting Genes for Reproduction

A selection operator chooses pairs of genes to reproduce based on fitness. It transforms a population's set of fitness values, $\{f_n\}$, to a set of selection probabilities, $\{p_n\}$, and uses these probabilities to select two genes. The simplest selection operator uses probabilities directly proportional to fitness:

$$\begin{aligned}\gamma &= \sum_{k=1}^N f_k \\ p_n &= \frac{f_n}{\gamma}\end{aligned}\tag{2.1}$$

There are many approaches to enhancing selection operators based on fitness value transformations. We may add a bias to the fitness or perform another non-linear operation (such as squaring, taking a logarithm, or range limiting; see [12]) prior to evaluation to emphasize either the best solutions (to speed convergence) or the less fit solutions (to explore more of the solution space).

Like Runkle [54], we use a hypergeometric selection operator [Equation 2.2]. This operator works with the ranks of the genes instead of working directly with their fitness values. Since the dynamic range of fitness values tends to decrease as the GA progresses, the fixed dynamic range of the hypergeometric operator tends to decrease competition early in the search and increase it late in the run. This solves the problem discussed by [20], in which too much competition early in the search can allow a superior solution to quickly dominate the population, causing premature convergence.

$$\gamma = \sum_{k=1}^N \frac{1}{k^q}$$

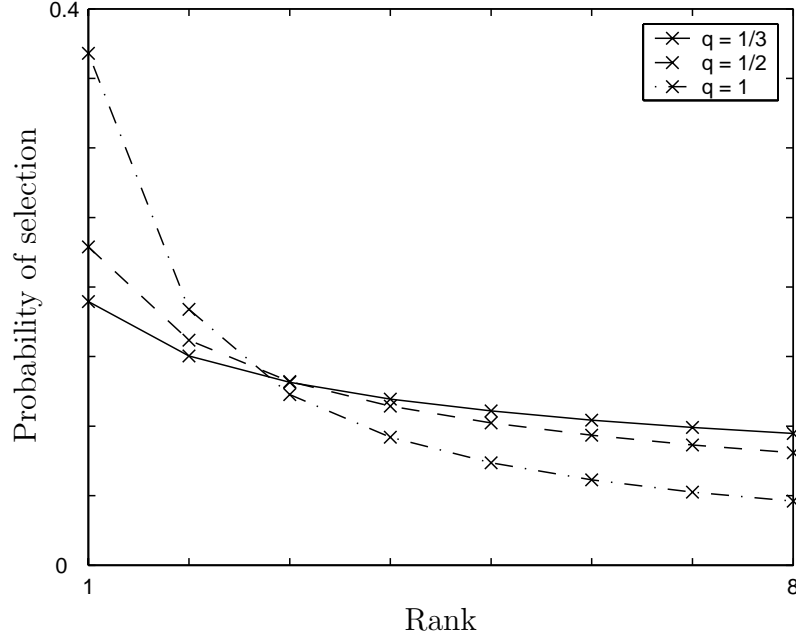


Figure 2.2: Hypergeometric distributions for a population of size 8. Using rank-based selection probabilities such as these results in a fixed dynamic range. This tends to decrease competition early in the run and increase it late in the run relative to using fitness directly for selection. Choosing $0 \leq q < 1$ decreases the dynamic range, reducing competition and biasing the GA towards a global search.

$$p_n = \frac{1}{\gamma n^q} \quad (2.2)$$

Specifically, the hypergeometric operator sorts the fitness values in descending order ($f_1 \geq \dots \geq f_N$) and assigns a selection probability inversely proportional to the rank. The parameter γ is chosen such that the probabilities sum to one. The parameter q is set to one for basic hypergeometric selection. Choosing q on $[0, 1)$ flattens the distribution. This selection operator mitigates the “super chromosome” problem [20] in which a superior solution quickly dominates the population, causing premature convergence.

The resulting selection probabilities are illustrated for a population of size 8 in Figure 2.2. We use $q = 1/2$, which results in slower convergence but a more thorough search (*i.e.*, a global search bias) than the basic hypergeometric operator.

2.3.5 Population Update: Steady-state vs. Generational

The population of solutions is updated in the GA's reproduction stage. In a generational GA, the entire population is replaced with a new population. That is, if there are N genes, $N/2$ mating pairs are chosen for the next generation. To guarantee that the maximum fitness value never decreases between updates, an elitist policy is usually applied. When using an elitist policy, the M most fit genes always proceed unaltered (only $N - M$ new genes enter the population during each generation). In contrast to a generational GA, a steady-state GA, which we use, selects a single mating pair and uses its progeny to replace the two least fit genes. The steady-state GA is inherently elitist (the $N - 2$ most fit genes remain in the population after each update) and requires fewer fitness evaluations before new genes have an opportunity to reproduce. Another advantage of the steady-state GA, investigated in simulations by [64], is that it is more robust to environmental changes than the generational GA.

2.3.6 Reproduction Operators: Crossover and Mutation

A genetic operator takes a small set of genes (usually 1 or 2) and produces an updated set of genes. There are two classes of genetic operators — a crossover operator interchanges bits between two genes and a mutation operator toggles bits of a single gene with a specified probability. Various crossover operators and a mutation operator are demonstrated in Figure 2.3.

Fixing the number of crossover points makes performance sensitive to bit encoding order. That is, nearby bits are more likely to be kept together than far apart bits during crossover. So, operators with a fixed number of crossover points tend to maintain intact certain regions of the gene that may not be relevant for any particular problem. On the other hand, allowing crossover at each point with probability $1/2$ (uniform crossover) eliminates the effects of bit encoding order, but is likely to break apart regions of the gene that are best kept together for a particular problem. In Chapter 10, we develop operators better suited to the perceptual tuning problem in consideration of these issues.

Roughly speaking, regardless of any particular genotype geometry, the purpose of crossover is to search globally (between solutions), while the purpose of mutation is to

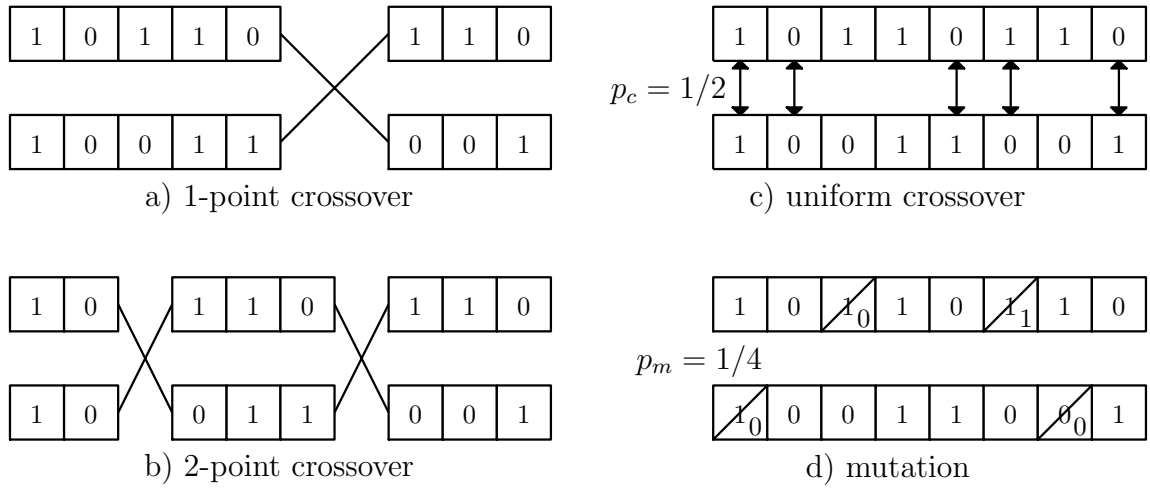


Figure 2.3: Examples of binary crossover and mutation operators applied to the parent genes 10110110 and 10011001. In a), the 1-point crossover operator chooses a random crossover point and swaps the bits after that point. The crossover point is between the 5th and 6th bits in this example. A 2-point crossover operator is illustrated in b). When using a uniform crossover operator as in c), each bit position has a p_c probability of being swapped. A mutation operator is illustrated in d). Each bit has a p_m chance of being mutated. The mutation value is chosen randomly and is not necessarily different from the original value.

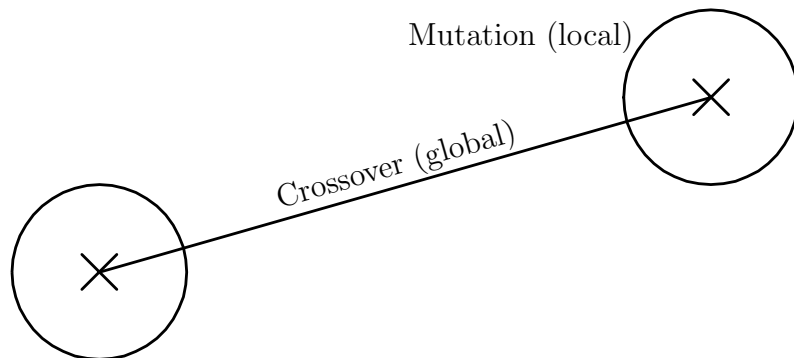


Figure 2.4: Conceptual view of mutation and crossover. The ‘x’s represent parent genes, while the line represents potential children resulting from crossover and the circles represent potential children resulting from mutation.

search locally (around a solution). Figure 2.4 illustrates this. Specifically, crossover is only truly global for a random population, but we use the term “global” more generally to mean the space about a set of solutions (for a bitstring encoding, this is a binary N -space), as opposed to “local”, which we use to mean the space about a particular point. Stated another way, the purpose of the crossover operator is to combine features of different genes. For this reason we do not allow the same gene to be chosen for both members of the pair.

2.3.7 Previous Perceptual Applications

One of the earliest systematic applications of a GA in a perceptual tuning problem is by Sterian, Runkle and Wakefield [58]. They applied a GA with subjective feedback to tune a 6-band equalizer used to generate simulated automobile windnoise to subjects’ preferences. The authors used a population of size 12 containing length 24 bitstrings with adjacent groups of 4 bits used to directly encode gains. Termination was controlled by the system using a population homogeneity measure. Prior research indicated that the optimal equalizer gains would monotonically decrease with increasing frequency, so the authors sorted the gains encoded by the genes before applying them to the equalizer. Therefore, any particular equalizer could have up to $6! = 720$ bitstring representations. Despite the inefficiency of this encoding, the system’s results were repeatable within limits in agreement with prior work on wind-

noise preferences. Like us, the authors recognized that paired comparisons are an efficient and relatively easy method of incorporating user preferences into a GA, but the authors did not discuss the details of their inferencing method (ours is explained in Chapter 9). The authors used a generational update, which was used by most other researchers at the time and was the original method used by Holland.

Runkle, Blommer and Wakefield [55] investigate several perceptual tuning methods, but focus on a genetic algorithm, for approximating a directional transfer function (DTF, the directional, space-varying component of a head-related transfer function) using all zero and pole-zero models. The authors use paired comparisons to incorporate user input and an insertion sort for inferencing.¹⁰ A generational update was used.

Takagi and Ohsaki [59] applied a GA to hearing aid fitting. They took a more traditional approach to getting fitness data than we did. Subjects interacted with a computer program to rank 20 population members using a 5-level absolute scale. We believe their input method is too complicated for field use. Also, besides complexity, it requires that subjects maintain a stable mental image of the scale so that, for example, an item that is clearly judged a “2” early in the run remains a “2” throughout the run. Although Takagi and Ohsaki saw the potential of the GA as a tool for both clinical and in-field fitting, their method, with its complicated user task, does not demonstrate the latter.

Our work builds on these methods, focusing on designing a system suitable for use by inexperienced users in perceptual tuning problems such as hearing aid fitting. Key improvements include gathering fitness data with a simplified two-interval forced choice response paradigm (Chapter 9), non-binary operators (Chapter 10), and using a steady-state population update method to reduce the number of user inputs required between successive searches.

¹⁰We present our inferencing method in Chapter 9. In Figure 9.3, our method is compared with various sort algorithms. Quicksort uses insertion sort on its subpopulations, so for populations smaller than the subpopulation size, the efficiency of quicksort is equivalent to that of insertion sort. We use the threshold $M = 7$, which is the maximum subpopulation size, as suggested by an authoritative reference on sorting algorithms [48].

CHAPTER 3

Methods

In the previous chapter, we looked at the problem of hearing aid fitting and suggested genetic algorithms as a solution. Now we will develop a method of applying GAs to the hearing aid fitting problem. This chapter lays the foundation for experiments in the coming chapters, but defers finer details to the latter chapters of the dissertation. We discuss the parametric model used to link the GA to fitting a hearing aid. We then introduce our method for inferring relations between the population members using paired comparisons. Next, we discuss how we choose initial populations, what data we collect, and stopping criteria. We then overview the portable prototyping device used to conduct the experiments and present the user interface. Finally, we discuss how we will measure success.

3.1 Parametric Model

As discussed in §2.3.2, the classic approach of using genes and operators based on bitstrings has the benefit of being extremely flexible and general, but is inefficient in domains more naturally described by several parameters that each take on a range of continuous or discrete values.¹ The response surface is often well behaved (nearly continuous and not excessively oscillatory) throughout most of these parametric do-

¹Both Davis [12] and Nagao [43] discuss the importance of designing appropriate encodings and operators for each particular problem domain. Wolpert and Macready [69] formalize the tradeoffs between generality and specialization for a particular problem class. The goal of our parametric model and the related GA apparatus is to specialize and improve performance on hearing aid fitting problems. According to the theorems of Wolpert and Macready, if we do this correctly, the tradeoff is merely a loss in performance on problems *outside* the domain for which we are optimizing.

mains. Thus, if a parameter can take one of 16 ordered values, the intervals between the values are more directly comparable than the significance between any of the 4 bits that could be used to encode the parameter value.² That is, the parameter represents a single dimension over which performance varies, not 4 separate dimensions. Hearing aid settings are primarily parametric, so we use a parametric representation instead of a bitstring representation. Further, although parameters usually have non-linear interactions that warp the mapping between the parametric and perceptual space, something is known about the general form of the perceptual scale of many classes of parameters. To take advantage of this knowledge of scaling, our representation consists of indexes that are mapped to actual parametric values. Each index step corresponds to roughly the same amount of perceptual change in the parameter.³ For example, time constant parameter values may be spaced exponentially (with equal ratios), while gain parameter values specified in decibels may be spaced linearly. In Chapter 10, we propose genetic operators designed to efficiently manipulate such lists of indexes.

3.2 Inferencing via Paired Comparisons

A major goal was to make the system easy for subjects to interact with, so we needed a simple input method. The paired comparison preference judgment (presenting the subject with two alternatives and asking which they prefer) is widely recognized as an efficient and reliable method of gathering user input (see, *e.g.*, [46]). A common alternative is to ask the user to make a numerical judgment of quality. A drawback with the latter approach is that it requires the user to maintain a consistent mental image of the scale. In contrast, assuming the user’s objective does not change, preference judgements are repeatable to within his ability to discriminate between the

²Nagao [43] proposes a “homogeneous encoding” method for mapping discrete parametric values to bitstrings that gives all bits nearly equal significance. However, his method does not address the low dimensionality/high cardinality property of parametric values.

³Ideally, each index step corresponds to a just noticeable difference (JND) so that alternatives are discriminable and useful information can be learned from comparing any two alternatives. However, the various parameters are not necessarily orthogonal or even components of a linear vector space, so approximations must be made. Kuk [10] gives reasons for erring on the side of a tighter grid when the precise JND is unknown. We follow this advice in our work when choosing the number of levels for each parameter.

alternatives.⁴

Given that the input method is paired comparisons and the goal is to estimate the rank order of the genes in the population so that hypergeometric selection may be used, an approach resembling a sort algorithm would be appropriate. The main factor used to evaluate sort algorithms is the distribution of the number of comparisons required to sort a list of a given size. Similarly, we seek to estimate the rank order with as few comparisons as possible for a given accuracy. In contrast to the objective sorting problem, human subjects are prone to error. Also, when using hypergeometric selection, distinctions among the more fit genes are more significant than distinctions among the less fit genes. In Chapter 9, we propose a method of choosing the paired comparisons and deciding when enough information has been gathered to enter the population update phase of the GA.

From the user’s point of view, a simple input method results. Users need not be concerned with the occurrence of population updates. The inferencing algorithm decides that and continually presents appropriate pairs.

3.3 Initial Populations

So that results are not dependent upon a particular initial population, four initial populations were used for each experiment. Using only four allows sufficient diversity, but still lets multiple subjects be assigned to each. This allows more direct measures of inter-subject behavior to be made than if a unique population had been used for each subject. The starting populations are designed to cover a large portion of the parameter space to increase the probability that the crossover operator will be able to make useful improvements. The other extreme would be to seed the initial population with very similar genes, but this would require much mutation before searching other areas of the parameter space, increasing the number of comparisons required to find a solution not near any of the initial genes.

⁴Other elements of the system can address changes in the user’s objective. For example, increasing the mutation strength increases the diversity of solutions in the population, making it more likely that an improvement under the new objective will be found. In the current work, we do not investigate methods for determining when the mutation strength should be increased.

3.4 Data Collected

Ideally, the following information would be automatically gathered during each experiment. The motivations for gathering each piece of information are listed below it:

- Duration the subject listened to each pair of alternatives
 - May be correlated with setting quality, or difficulty of discriminating settings
- Decisions made by the subject (including when each decision was made)
 - Allows some estimation of consistency of decisions
 - Allows correlation between the time to make a decision and the similarity of alternatives
- Internal state of the GA (which settings are being tested, etc.)
 - Allows *post hoc* determination of why a particular setting was presented
- Environment (loud/quiet, noisy, telephone, music/speech, etc.)
 - Allows analysis of the relationship between the environment and parameter settings

The data collected are more limited than this, mainly because the portable prototyping device does not have a day/date clock. We record the time required to make each decision instead of the absolute time of each event. Also, we do not attempt to automatically characterize the environment. The data stored in the internal log provide the remainder of the information listed above.

3.5 Stopping Criteria

In the experiments, we instructed subjects to listen for at least some minimum number of comparisons. Based on preliminary simulations, that minimum was chosen to be large enough to observe convergence behavior of the GA for various observer

models. Subjects were instructed to listen longer if they could readily differentiate between many of the pairs, which would indicate that the GA had not converged. In a practical application, this might suffice – subjects would continue making judgments until they found a setting they wanted to keep. Alternatively, more rigorous stopping criteria might be used. In GA applications, convergence of the population is often used as a stopping criterion. *Convergence*, roughly speaking, measures how similar the members of the population are to each other. We propose metrics and derive lower bounds for diversity, the inverse of convergence, for genes made of both bitstrings and multilevel indexes in Chapter 8. Asking subjects to run the experiment to convergence, which, we shall see, they did in many cases, provides performance data that may be used to develop better stopping criteria for future applications. For example, if experiments show that diversity asymptotes at 2 but that improvements are rarely made once diversity decreases below 3, the GA could stop after the diversity decreases below 3 and “lock in” the subject’s favorite setting at that point.

3.6 Portable Prototyping Device

Starkey has developed a portable prototyping device for testing new hearing aid algorithms in real-world environments. It is based on the Motorola 56303 DSP, which runs at up to 66 MHz and has $8\text{ k} \times 24\text{ b}$ of on-chip RAM. An external non-volatile memory of 32 kB is used for code, parameter, and log storage. The device casing measures $5.5 \times 9.2 \times 1.4\text{ (cm}^3\text{)}$ and has a toggle switch and three buttons for user input. A two-color LED provides visual feedback of user actions and system state. Wires run to the subject’s custom ear modules. The device is powered by three AAA batteries that last between two and several hours depending upon the power requirements of the loaded software. This device is used for both in-house and field trials.

3.7 User Interface

A total of five buttons allows subjects to control the overall volume of the hearing aid, swap the two alternatives under consideration, indicate a preference for the active

Table 3.1: User interface.

Button	Purpose	LED indication
+	Volume up	1 green blink
−	Volume down	1 red blink
Toggle up	Current solution preferred	2 green blinks
Toggle down	Solutions of equal quality †	2 red blinks
Bar	Switch between 2 solutions	1 or 2 green blinks ‡

†: The equality decision is not allowed in all experiments.
‡: When switching to the first alternative, there is 1 green blink. For the second alternative, there are 2 green blinks.

solution, and indicate that the solutions are of equal quality. The functions of the buttons and the visual feedback given for each button press are shown in Table 3.1.

When the device is power-cycled, the last population of solutions is recovered from the log. Results of previous comparisons are not recovered for inferencing purposes, but they are available for analysis. In the first experiment, users were allowed to decide that two solutions presented were of equal quality. This was done to lessen the task load, but proved cumbersome during analysis, so this option was removed from all subsequent experiments.

3.8 Measures of Success

In all of our experiments we investigate parameters with largely subjective effects, so we require multiple classes of data to come to informed conclusions about the success of the GA for perceptual tuning. The main classes are:

1. Objective measures

- Examples: AI/SII⁵, speech intelligibility⁶, gain margin⁷

⁵An articulation index (AI) quantifies the amount of speech information transmitted in a signal. The speech intelligibility index (SII) is a particular AI that considers spectral characteristics of speech and noise signals, a subject’s hearing thresholds, and relative importance of speech information at various frequencies. Eisenberg *et al.* [16] and Fabry and VanTasell [17] discuss these measures in more depth.

⁶Speech intelligibility is the percent of speech stimuli that a subject correctly understands. The stimuli, speech level, noise level, and method of determining subject understanding are experiment-specific.

⁷Gain margin is defined in §4.2.2.

- Pro: direct and mathematically rigorous measures of success
- Con: Do not apply to, or do not tell enough of the story for parameters primarily affecting subjective quantities (such as comfort and naturalness), such as those investigated in all experiments in the current work

2. Indirect measures – ensemble statistics

- Examples: population variation, distance to favorite, probability of improvement, etc., all as functions of round or decisions made
- Pro: can be derived from objective experimental data; can be simulated for various observer models; concisely summarize behavior across several subjects
- Con: abstract and indirect; some of these measures are especially noisy (high variance); not sufficient for inferring successful use of the GA

3. Anecdotal measures

- Examples: subject survey, subject debriefing, follow-up preference judgments
- Pro: direct measure of subject satisfaction
- Con: less rigorous, open to bias

The experiment-specific measures, such as gain margin for the feedback canceller experiment and a survey following the multiband expansion experiment, are discussed in the chapters covering those experiments. The indirect measures we consider, however, are properties of the genetic algorithm, not of the specific hearing aid parameters being optimized. So, we provide some details about these indirect measures here. In all our experiments in which the GA was run,⁸ we also ran simulations of the experiment with three simple observer models. Observer 0 (named “Random”) makes random decisions and is meant to approximate worst case performance. Observers 1 and 2 seek a specific point in the index space. They never make a mistake, always preferring the point that is closer, in the Euclidean sense, to the objective. When points

⁸This includes *all* of our experiments except a follow-up to the feedback cancellation experiment in which subjects compared 5 fixed alternative settings to investigate repeatability and preferences across subjects.

are equidistant from the objective, these observers randomly choose one. Observer 1 (named “Central”) prefers each index to be at its middle value, rounded down. Observer 2 (named “Corner”) prefers each index to be at its maximum value. Observers 1 and 2, collectively called the rational observers, are meant to approximate best case performance. We expect real users to perform somewhere between the extremes of the random observer and the rational observers.

The 4 GA performance quantities for which we compare simulated and experimental distributions vs. round for each experiment are:

1. Diversity
2. Minimum distance to favorite
3. Probability of a child improving upon the previous favorite
4. Decisions

Diversity measures how different genes in the population are from one another. A high diversity indicates that crossover is likely to yield genes significantly different from (and hopefully superior to) genes already in the population. A low diversity indicates that crossover will be able to make few significant improvements. Also, low diversity can be used as a criterion for terminating the run. Diversity is based on the index values and assumes appropriate choices of parameter scales. Also, our diversity measure treats each parameter as equally important. By using a Euclidean metric, the diversity measure is constrained to a specific method of estimating the importance of interactions of multiple parameters and neglects any warping between the parameter and perceptual spaces.

The minimum distance to favorite is the distance from the subject’s favorite solution to the nearest gene in the population. The subject’s favorite is known for the rational observers and is taken to be the top ranked solution at the end of the run for experimental subjects. We do not plot this value for the random observer since it has little meaning. This “distance” makes the same simplifying assumptions about the parameter space as the diversity measure.

The probability of a child improving upon the previous favorite is a measure of the value of running subsequent rounds. This probability is easily calculated for the

rational observers since their cost functions and objectives are fixed. For experimental subjects, it is taken as the probability that a child is chosen over the previous favorite when making paired comparisons.

The number of decisions is a partial indication of how much effort the subject must exert to complete each round. In the initial round, the ranks must be estimated with no *a priori* knowledge, so more decisions are required than in subsequent rounds, in which only 1 or 2 genes are new and the relations between the old genes are already known. In the experiments, if a subject power cycles the device, he continues from the population and round where he left off, but the relationships between the genes are built up from scratch.

We shall see that these quantities, especially the probability of a child improving upon the previous favorite, are noisier for human subjects than for simulated subjects. Nonetheless, informative trends can be found in both the simulated and experimental results.

CHAPTER 4

Experiment 1: Acoustic Feedback Canceller Fitting

For our first experiment, we investigated whether a GA can guide a hearing-impaired subject towards perceptually good feedback canceller parameters in everyday conditions. A feedback canceller system was chosen because of our desire to investigate an area with little prior research and because of the system's strong bearing on subjective measures of patient satisfaction. Indeed, the goal of using the GA is to leverage the subject's desire to give input to provide greater satisfaction with the fitting, especially with parameters that cannot be readily prescribed using objective rationales.

A drawback of designing the experiments this way is that objective measures of success, which are the most compelling, are either not applicable, or are critically limited in scope with respect to the optimized parameters. In addition to presenting several indirect measures of success, we consider one direct and objective but limited measure of success, gain margin [§4.2.2]. A follow-up experiment in Chapter 6 investigates repeatability, another objective measure of success.

Table 4.1: Experiment 1 pilot study parameter values.

L [samples]	α [s ⁻¹]
16	$18 \cdot 10^{-6}$
32	$32 \cdot 10^{-6}$
64	$56 \cdot 10^{-6}$

4.1 Methods

4.1.1 Parameters

Each subject’s three-band linear gain prescription was programmed into the portable prototyping device and was not modified during the experiment. A normalized least mean squares filter (NLMS) similar to the one discussed by Kates [30] was used to cancel acoustic feedback. The GA modified the following three variables that control the NLMS system:

- L – filter length; a longer filter length allows a longer feedback delay to be modeled at the cost of additional processing power
- α – time constant; controls how rapidly the filter adapts
- ρ – update rate decimation factor; increasing the decimation factor decreases the frequency of filter coefficient updates, reducing the processing power required; if the decimation rate is increased too far, filter quality suffers and stability of the filter coefficients may be compromised

Pilot Study

Members of the Starkey Advanced Research group evaluated a range of α and L values in situations especially susceptible to feedback in order to identify appropriate parameter ranges for the experiment. Table 4.1 shows the values tested. Group members had direct control over each parameter; no GA or other fitting algorithm was involved. Also, update decimation was not activated, so $\rho = 1$.

$(L, \alpha) = (32 \text{ samples}, 56 \cdot 10^{-6} \text{ s}^{-1})$ was the most-preferred setting. Further analysis of the data and available clock cycles showed that the filter length should not become

Table 4.2: Experiment 1 parameter values.

L [samples]	8	10	13	16	20	25	32	40
α [s ⁻¹] ($\cdot 10^6$)	10	14	19	27	37	52	72	100
ρ [samples/sample]	10	20	40	80				

much longer than this and that a slower adaptation rate may give fewer artifacts. The sample rate is 19.97 kHz. Based on this analysis, the parameter tiling given in Table 4.2 was chosen.

4.1.2 Subject Instructions

Subjects received an instruction sheet to read before beginning the experiment. This sheet contained a diagram of the hearing aid prototyping device with the function of each button listed. The text described how feedback can be induced and suggested criteria for evaluating feedback cancellation. . .

Listening Environments: Your portable processor has been programmed to control the temporary feedback that may occur when something is placed close to the microphone. Examples of such situations include:

- Cupping your hand over your ear
- Holding the telephone receiver up to your ear
- Leaning your head against a wall (ear toward the wall)
- Jaw movements caused by yawning, chewing, etc.

How to listen: The processor is programmed with two different feedback canceller programs to evaluate. Create a situation that results in feedback (see above) as you switch back and forth between the two programs. You may listen to each program as many times as is necessary to determine which program you prefer.

Please select your favorite feedback canceller program using the following criteria:

- Speed with which feedback is removed
- Completeness of feedback removal

Table 4.3: Experiment 1 initial population indexes.

Population Solution	1					2				
	0	1	2	3	4	0	1	2	3	4
L	1	0	3	7	6	4	6	1	6	1
α	1	7	4	0	6	3	1	6	6	1
ρ	1	3	2	1	0	2	0	1	2	3
Population Solution	3					4				
	0	1	2	3	4	0	1	2	3	4
L	0	7	1	6	3	7	6	0	4	1
α	0	1	5	7	2	0	7	2	3	6
ρ	1	1	0	2	3	2	1	0	0	3

- Sound quality of your own voice while feedback is being canceled (*i.e.*, while speaking on telephone)
- Sound quality of other voices/sounds while canceller is active

Please note: Although the two programs will always be different from each other, the difference may not be detectable to you as a listener. This is expected and perfectly acceptable. See the following section for more information.

Making your preference ratings: Once you have determined which of the two programs you prefer, please indicate your choice by pressing the appropriate button on the processor. If you have been unable to determine a difference between the two programs after intensive listening, you may indicate this by pressing the appropriate button on the processor. Either action will result in a new set of feedback canceller programs being loaded into the processor for your continued evaluation.

4.1.3 Initial Populations

As discussed in §3.3, four initial populations were used so that results were not dependent upon particular starting conditions. Each subject was assigned to one of the four cohorts at the beginning of the experiment. The populations are shown in Table 4.3.

4.2 Results

In this section, we first compare simulated and experimental results to arrive at several indirect ensemble measures of success [§4.2.1]. We then present the solution chosen by each subject, along with the measured gain margin, which is an objective and partial measure of success [§4.2.2]. Next, we calculate the distance from each subject’s chosen solution to the nearest progenitor (gene from the initial population) and compare the distribution to a theoretical distribution [§4.2.3]. Finally, we look at first/second alternative decision bias [§4.2.4].

4.2.1 Simulated and Experimental Results vs. Round

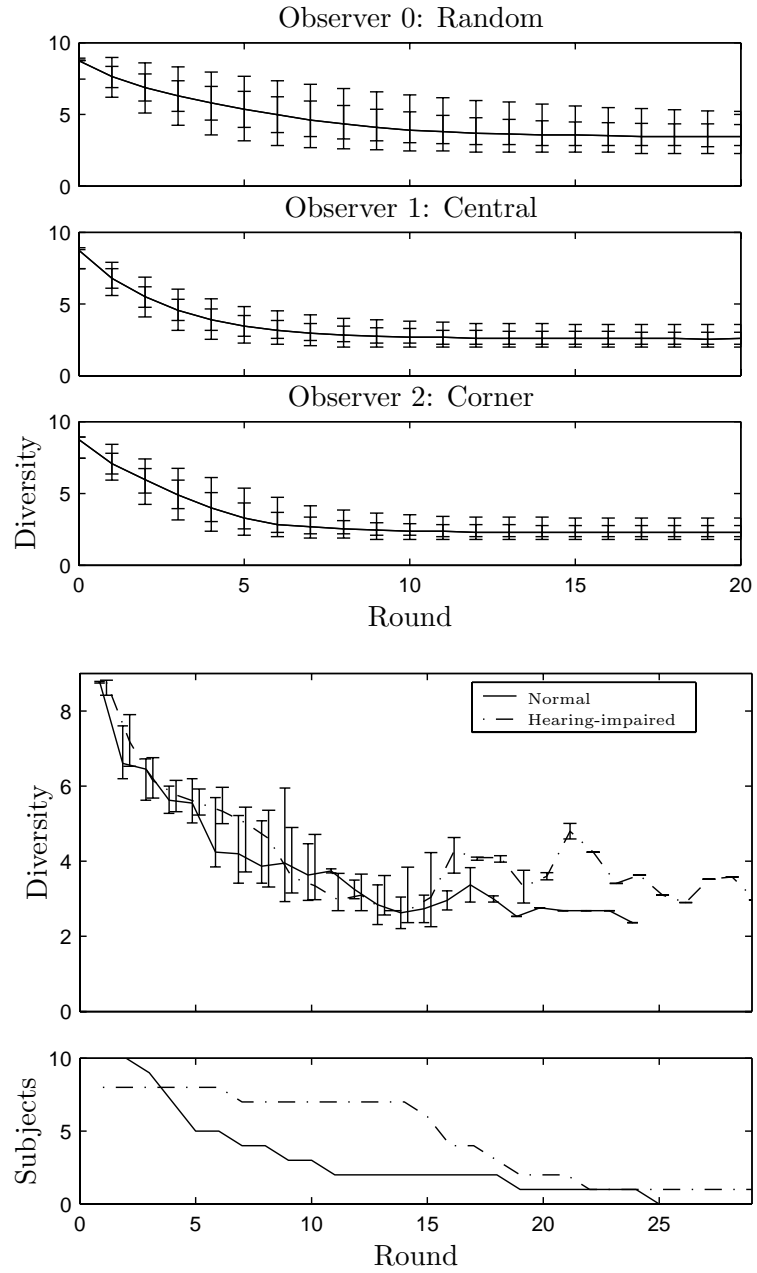
Figure 4.1 shows the diversity¹ vs. round, both simulated and experimental. It has a downward trend. From the experimental results, we see that when subjects drop out of the run the distribution often moves towards greater diversity. Thus, subjects having low diversity (closer to convergence) are more likely to end the experiment than subjects having high diversity. Indeed, subjects were instructed to stop the experiment if they could not hear a difference for many consecutive comparisons.

Also in Figure 4.1, we see that the asymptotic experimental diversity is about 2.5. For a simulated random observer, the diversity asymptotes at about 3.3, while for a perfect central observer, it asymptotes at about 2.4, and, for a perfect corner observer, it asymptotes at about 2.2. Thus, the decisions of the experimental subjects resulted in diversity trends much closer to rational simulated observers than random simulated observers. That experimental subjects were allowed to make equality decisions confounds things somewhat. The primary effect we would expect to see from equality decisions is an *increase* in diversity due to a flattening of the distribution of selection probabilities. Nonetheless, the experimental asymptotic diversity was relatively low, further supporting comparisons to the rational observer models.

Figure 4.2 shows the minimum distance to the favorite solution as a function of round, both simulated and experimental. For the rational simulated observers, the favorite solution is known. For the experimental subjects, the favorite solution is approximated by the top-ranked solution at the end of the run. Also, each subject

¹The population diversity measure is discussed in §8.3.

Figure 4.1: Experiment 1 simulation (top) and experimental results (bottom): diversity vs. round. Median with interquartile (for all) and interdecile (simulation only) ranges indicated by bars.



decided how long to run the experiment, while simulated observers, regardless of when and whether they found their objective, always completed exactly 20 rounds. The median number of rounds for the normal hearing subjects was 5, while the median number of rounds for the hearing-impaired subjects was 16. The number of rounds completed is one of the very few significant differences between the normal and hearing-impaired subjects. The median number of rounds required by the rational observers to reach each of their goals was 14. For the experimental subjects, the median distance began at 2.1 and decreased to 0.5 by round 14. Because of the favorite solution approximation and variable number of rounds for the experimental subjects, the distance to favorite data are not directly comparable. Nonetheless, the similar downward trend in distance to favorite vs. round between the simulated and actual subjects supports the hypothesis that the benefits of the system to the experimental subjects are similar to the benefits to the ideal observers. That is, the occasional indiscriminable differences and judgment errors are not preventing the experimental subjects from experiencing most of the fitting system's benefits.

Figure 4.3 shows the probability that a new solution was preferred over the previous favorite as a function of round for both simulated observers and experimental subjects. The random observer has a 50% probability of preferring a new child over the previous favorite. The rational observers have about a 20% chance of preferring a new child for the first 6 rounds. After the 6th round, the probability steadily decreases to about 7% when the final decisions are made before the 20th round. The experimental data are much noisier due to the small number of samples. Also, since subjects occasionally make errors, we expect their measured probability of improvement to be slightly higher. And it is — the probability of a subject indicating that a child is better than the previous favorite is about 40% through round 6 and about 20% thereafter. The random observer can be considered to model a maximally error-prone subject, that is, one that has a 50% chance of making an incorrect decision. The figure shows that the experimental subjects performed roughly between the rational and random simulated observers.

Simulation showed that the rational simulated observers required between 5 and 8 decisions in the first round, with 6 and 7 decisions required for the vast majority of runs. Fewer decisions were required in subsequent runs due to information car-

Figure 4.2: Experiment 1 simulation (top) and experimental results (bottom): minimum distance to favorite vs. round. Median with interquartile (for all) and interdecile ranges (simulation only) indicated by bars. For the experimental results, the favorite setting is taken as the top-ranked value at the end of the run. Since the subjects were allowed to indicate that they could not discriminate between two alternatives, there may be multiple favorites. When this occurs, the minimum distance to each is averaged.

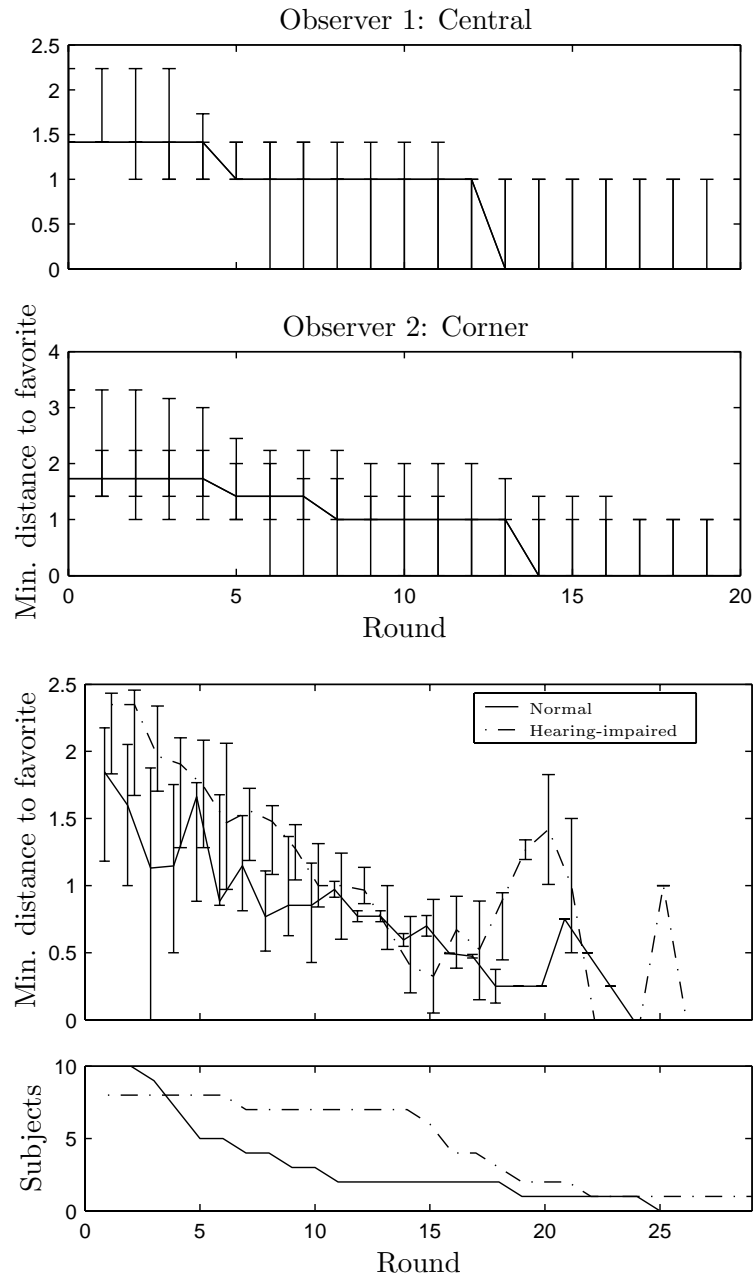
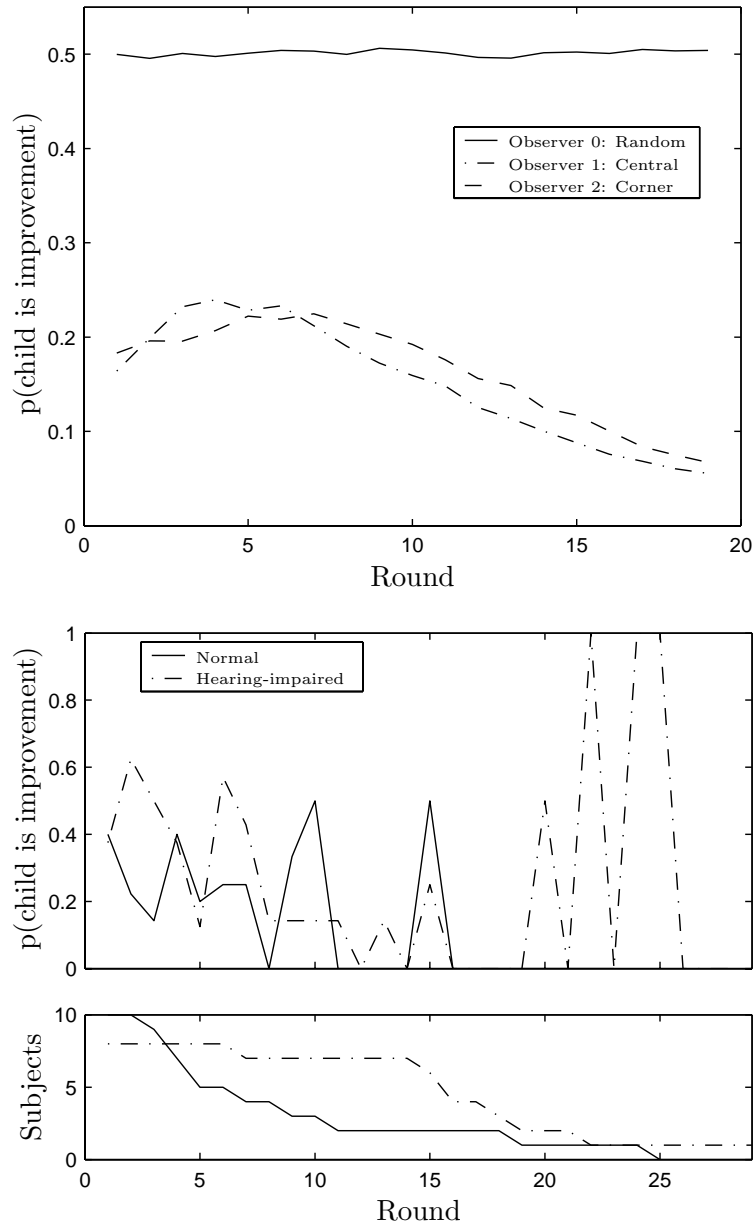


Figure 4.3: Experiment 1 simulation (top) and experimental results (bottom): probability of improvement being indicated vs. round. For the objective simulated observers the probability of indicating improvement is the probability of actual improvement since these observers never make an error. For the random observer, the indicated probability is very near 50% since our inferencing algorithm compares the better of the new solutions each round with the previous best solution. The random response to this single comparison determines whether an improvement is counted for the random observer. For the experimental subjects, the probability of the user preferring a new solution over the previous favorite solution is shown. Since there is some uncertainty in the subject responses, we expect the curves shown to be slightly higher than the actual probability of improvement.



ried forward by the inferencing algorithm. The median and 90th percentile number of decisions required was 4, while the 10th percentile value was 3. The experimental subjects required a similar number of decisions in the first round, and an average of about 1 1/2 fewer decisions than the simulated rational subjects in the subsequent rounds. The difference between the simulated and experimental results can be explained by the lack of an equality option for the simulated observers.² From Table 4.6, we see that subjects chose the equality option $211/(211 + 211 + 162) \approx 36\%$ of the time. There was no significant difference between normal and hearing-impaired subjects.

4.2.2 Gain Margin Measurements

Gain margin is an objective measure of feedback management system performance. It is the amount by which hearing aid gain may be increased before feedback occurs after a feedback management system is activated. Gain margin may be taken at a specific frequency, averaged across frequencies of interest, or taken as the minimum margin in a frequency range. For example, if feedback occurs at gains of 40 dB and 30 dB at frequencies of 1 kHz and 2 kHz with feedback cancellation disabled, but gains can be increased to 53 dB and 39 dB with feedback cancellation active, the gain margin may be stated as a) 13 dB at 1 kHz and 9 dB at 2 kHz, b) 11 dB averaged over 1 kHz and 2 kHz, or c) at least 9 dB between 1 kHz and 2 kHz³. The favorite solutions at the end of the experiment for each hearing-impaired subject are shown in Table 4.4. The gain margins achieved by one of these solutions for each subject are shown in Table 4.5.

4.2.3 Distance from Favorite to Progenitor

Another measure of performance is how far the final solution found by the GA falls from the nearest value in the initial population, the nearest progenitor. If this

²As discussed in greater detail in Chapter 9, our inferencing method makes some simplifying assumptions about the perceptual space, including that equality is transitive. Since equality is reflexive, but dominance is not, an equality decision may allow more inferences to be made. For example, if we know that A and B are equivalent and wish to determine the relative ranking of a new alternative, C, only one comparison is required. On the other hand, if A is superior to B, either one or two comparisons will be required to insert C into the ranking.

³When gain margin is characterized as in c), a much denser sampling across frequency is normally used.

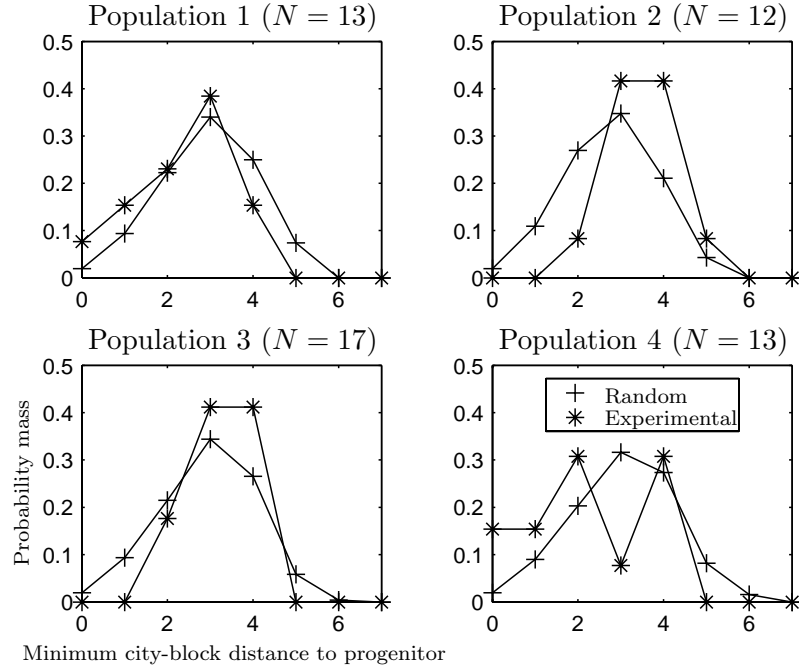
Table 4.4: Experiment 1 favorite solutions for 8 hearing-impaired subjects. The three parameters were discussed in §4.1.1. When multiple solutions were tied for the favorite, a single favorite solution was arbitrarily chosen for gain margin measurement. The chosen solutions are first in the lists below.

Subject	Cohort	Favorite solution(s)				
CB	1	[32 72 20]				
CC	1	[16 27 20]	[20 19 20]	[16 14 20]	[16 19 20]	[13 72 20]
DA	4	[20 37 10]				
GD	2	[25 27 10]	[16 100 40]	[16 19 10]	[16 27 20]	[16 14 10]
GG	2	[25 10 40]	[25 37 40]	[20 10 80]	[25 10 80]	[25 19 80]
GJ	4	[13 72 10]	[20 100 20]	[32 27 10]		
KP	3	[10 27 80]	[13 37 80]	[10 37 80]	[13 52 80]	[16 37 80]
LW	3	[25 37 40]	[20 52 80]	[25 27 40]	[25 27 80]	[20 37 80]

Table 4.5: Experiment 1 gain margins for 8 hearing-impaired subjects. A Frye probe microphone system was used to determine the gain margin as a function of frequency. Then, per the standard definition of gain margin in use at Starkey Laboratories, the gain margins at 1.0 kHz, 1.6 kHz, and 2.5 kHz were averaged to arrive at the overall gain margin value. All values are in relative dB SPL.

Subject	Gain Margin			
	1.0 kHz	1.6 kHz	2.5 kHz	Mean
CB	6.1	13.3	10.6	10.0
CC	10.0	10.1	13.3	11.1
DA	16.8	11.2	7.1	11.7
GD	11.6	14.4	10.9	12.3
GG	15.0	19.4	18.0	17.5
GJ	12.4	12.9	13.9	13.1
KP	12.0	14.0	11.0	12.3
LW	15.3	14.3	12.5	14.0
Mean	12.4	13.7	12.2	12.8

Figure 4.4: City-block distance from favorite to progenitor.



distance is small, the GA is doing little better than choosing the best from the set of initial solutions. If we assume that favorite solutions are distributed independently of the initial populations, we would expect the actual distance distribution to be the same as the distribution for a set of randomly chosen values. Figure 4.4 shows both the experimental distribution and the distribution for randomly chosen favorites for each of the four populations. The distributions are quite similar considering the small experimental sample size.

4.2.4 Decision Bias

Table 4.6 shows that there was a small bias for the first solution presented. We eliminated one source of bias by randomizing presentation order. A is the number of times a subject preferred the first solution presented, N is the number of times he registered no preference, and B is the number of times he preferred the second solution presented. For simplicity we neglect the “no preference” decisions.

When the observations are summed across all runs of the experiment, the lower end of the 95% confidence interval is at $p = 0.514 > 0.5$. Since we are potentially

Table 4.6: Analysis of decision bias in experiment 1.

A	N	B	95% CI>	$E(p) = A/(A + B)$	95% CI<
6	5	4	0.262	0.600	0.878
7	11	8	0.213	0.467	0.734
8	25	6	0.289	0.571	0.823
24	3	11	0.507*	0.686	0.831
6	2	13	0.126	0.316	0.566
9	8	9	0.260	0.500	0.740
14	17	12	0.334	0.538	0.734
16	17	8	0.447	0.667	0.844
9	6	6	0.323	0.600	0.837
13	23	23	0.208	0.361	0.538
12	17	4	0.476	0.750	0.927
13	13	15	0.275	0.464	0.661
12	18	8	0.361	0.600	0.809
8	0	6	0.289	0.571	0.823
18	6	6	0.533*	0.750	0.902
11	0	4	0.449	0.733	0.922
15	36	15	0.313	0.500	0.687
10	4	4	0.419	0.714	0.916
211	211	162	0.514*	0.566	0.617

combining data from subjects with opposing biases, we can say with at least 95% certainty that there is a decision bias. Although the analysis was run at 95% confidence, it holds up at even 98% confidence.

4.3 Discussion

As mentioned at the beginning of the chapter, we optimize for perceptual quality of feedback cancellation, an area with little prior research and one that has a strong bearing on subjective measures of patient satisfaction. Indeed, the goal of using the GA is to leverage a subject's desire to give input to provide greater satisfaction with the fitting, especially with parameters that cannot be readily prescribed using objective rationales. The drawback of designing the experiments this way is that objective measures of success, which are the most compelling, are either not applicable, or are critically limited in their scope with respect to the parameters for which we optimize.

So, we use both these objective measures of success, such as gain margin, and indirect measures of success based on simulated performance.

Comparing overall subject and simulated results using several measures showed similar performance [§4.2.1]. This is promising since it shows that the potentially more complicated objectives and occasional decision errors by the human subjects did not adversely affect their ability to use the system. Still, it does not speak directly to the quality of the solution. For an objective measure of quality, we measured gain margin. For a subjective measure of success, subjects were asked about the comfort and quality of the solution and generally responded positively. For a more rigorous measure of subjective quality, we conducted a follow-up experiment in which each subject compared his optimal solution with solutions representative of the favorites across all subjects [Chapter 6].

Maxwell and Zurek [37] concisely summarize the limitations of using gain margin to characterize acoustic feedback cancellation. Their work is representative of much of the research on notch filters, LMS filtering, and other feedback cancellation techniques, such as those discussed by Siqueira *et al.* [56]. Despite the rigorous mathematical approach to the feedback cancellation system common in the literature, characterization normally includes gain margin (objective component) and, only sometimes, attempts to characterize effects on speech intelligibility (and even that is done subjectively in most works, including those cited here). The authors typically average ratings of subjective quality across subjects. This neglects the hypothesis, explicit in our work, that different subjects may have vastly different preferences for acceptability of particular processing artifacts. Although gain margin does not provide a comprehensive quantization of performance, it is a well understood measure of the objective aspects of feedback cancellation performance, so we have included it for our hearing-impaired subjects in Table 4.5. The average gain margin of 12.8 dB is consistent with the findings of Maxwell and Zurek that various adaptive filtering techniques yielded gain margins of 10 to 15 dB.

The experimental distribution of distance from the favorite solution to the nearest progenitor was shown along with the expected distribution for a uniform distribution of favorite solutions in Figure 4.4. Although the similarity of the distributions is not sufficient to show that the GA is finding an optimal solution, it is necessary if

we assume that subject favorite solutions are distributed independently of the initial populations. Good matches between the distributions could also occur if the GA were performing a random walk, but our other data, such as the population diversity measure [§4.2.1], do not support this.

The data presented in Table 4.6 showed that subjects had a slight bias for the first presented alternative. If a subject were merely trying to complete the experiment quickly, we would expect a bias towards the *second* solution presented since the software forces the subject to listen to both alternatives before making a decision. While listening to the first alternative for the first time during a comparison, the subject is not allowed to make a decision. The subject is only allowed to make a decision after pressing the toggle button at least once. While significant, the most likely value for the bias (6.6%) is not large.

CHAPTER 5

Experiment 1 Follow-up: Mutation Strength

Several parameters control the behavior of the GA. One of these is mutation strength, which determines how much, on average, parameters in new genes are perturbed after the crossover operator is applied but before the genes enter the population and are presented to the subject. In general, the optimum mutation strength cannot be determined analytically and must be found experimentally. If the strength is too low, the population converges before the GA sufficiently searches promising areas of the parameter space. If the strength is too high, the algorithm performance approaches that of a random search. The present experiment compares performance under two mutation strengths.

5.1 Methods

In the original experiment [Chapter 4], crossover was performed in some updates, and mutation in all the others. This cautious approach was taken to allow better analysis of algorithm performance, especially in the event of poor results. Since the algorithm performed well, the standard approach of applying both operators for all updates was used in the present and all subsequent experiments. In addition, while subjects were allowed to indicate that they could not discriminate between two settings in the first experiment, this option was removed and 2-interval forced choice was used for this and all subsequent experiments.

Four normal hearing subjects repeated the first experiment with the modifications discussed above, providing a performance baseline. As in the first experiment, the

Table 5.1: Asymptotic diversity for mutation follow-up.

	Weak mutation	Strong mutation
Observer 0: Random	3.9	5.7
Observer 1: Central	2.8	4.1
Observer 2: Corner	2.4	3.4
Experimental data	2.6	< 3.4

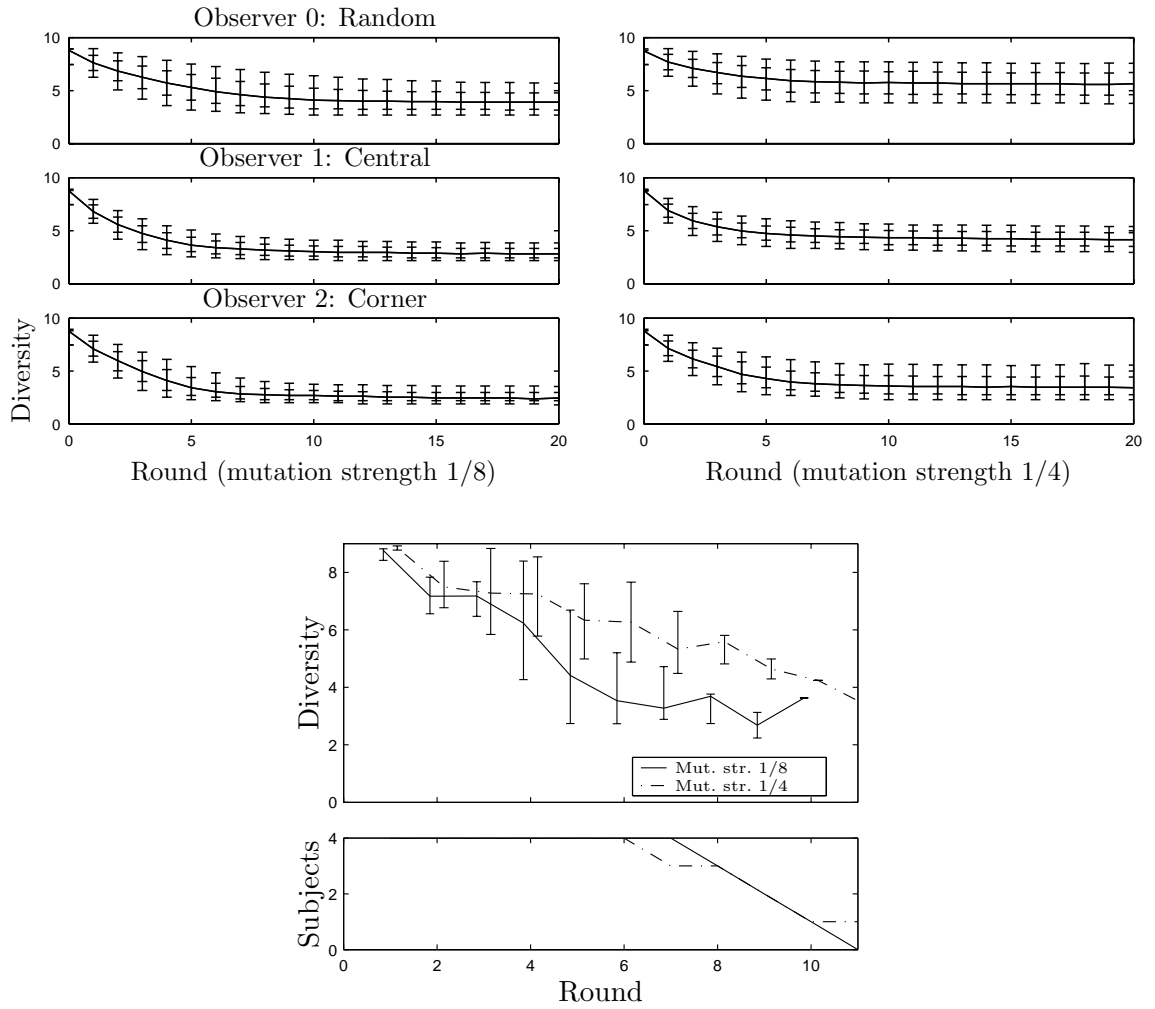
mutation strength used corresponded to perturbing each eight-level parameter by a Gaussian random variable with a standard deviation, σ , of 1 and the four-level parameter by a Gaussian random variable with $\sigma = 1/2$. The baseline results suggested that larger σ 's would be more interesting than smaller ones, so the σ 's were doubled and the experiment was repeated with the same subjects.

5.2 Results

Figure 5.1 shows the diversity vs. round, both simulated and experimental. In all cases, diversity began at about 9.0 and steadily decreased. Also, for both the simulations and experimental data, stronger mutation lead to a greater asymptotic value. The asymptotic diversity values are summarized in Table 5.1.

Figure 5.2 shows the minimum distance to the favorite solution vs. round, both simulated and experimental. The favorite solutions for the simulated observers are known, while the favorite solutions for the experimental subjects are taken to be the solutions at the end of each subject's run. For the central observer with weak mutation, 50% of the runs reached the objective by the 18th round, whereas it took more than 20 rounds when the mutation strength was increased. For the corner observer with weak mutation, 50% of the runs reached the objective by the 14th round, while it took only 10 rounds when the mutation strength was increased. Since the mutation operator truncates at the edge of the parameter space [§10.2], increasing mutation strength improves performance for the corner observer since there is no possibility of overshoot. Increasing mutation strength to a very high value would result in nearly all genes being at a random corner of the 3-D parameter space, giving nearly a $(1/2)^3 = 1/8$ chance that any mutation would yield the optimum value. The

Figure 5.1: Experiment 1 mutation strength follow-up simulation (top) and experimental results (bottom): diversity vs. round. Median with interquartile (for all) and interdecile (simulation only) ranges indicated by bars.



experimental subjects traveled similar distances, with most subjects ending their run after the 8th or 9th round.

Figure 5.3 shows the probability that a subject preferred a new solution above the previous favorite in a given round. As a result of the small subject pool, the probability of improvement, which averages about 50% regardless of mutation strength, is quite noisy for the experimental subjects. On the one hand, for the central observer, increasing the mutation strength decreased the probability of improvement for all rounds. Taken together with the slower decrease in minimum distance to the favorite solution, this indicates that the increased mutation decreased performance for the central observer. On the other hand, the increased mutation increased the probability of improvement for the corner observer for the same reason that it also improved the minimum distance to favorite performance. Thus, the corner observer clearly benefited from the increased mutation strength while the central observer clearly suffered.

5.3 Discussion

Although the population variation data for the experimental subjects show convergence behavior similar to the simulated results, the size of the subject pool is too small to come to strong conclusions based on the minimum distance to favorite and probability of improvement plots. That the increase in mutation yielded an increase in performance for the corner observer but a decrease in performance for the central observer is interesting. This highlights the interdependence of parameter space representation, user objective, and GA control parameters. In this case, the increase in mutation benefited a specialized observer at the cost of decreased performance in the general case, an objective not at the edge of the parameter space. Since our approach is to design problem space constraints, such as interdependence between hearing aid parameters, into the parameter encoding, we do not desire that the GA be tuned to favor objectives in particular regions of the encoded parameter space at the cost of generality. So, in this case, the weaker mutation strength is preferred. In some cases, however, the least reliable subjects may benefit from greater mutation strength, which would enable quicker recovery from errors.

Figure 5.2: Experiment 1 mutation strength follow-up simulation (top) and experimental results (bottom): minimum distance to favorite vs. round. Median with interquartile (for all) and interdecile (simulation only) ranges indicated by bars. For the experimental subjects, the favorite setting is taken as the top-ranked value at the end of the run.

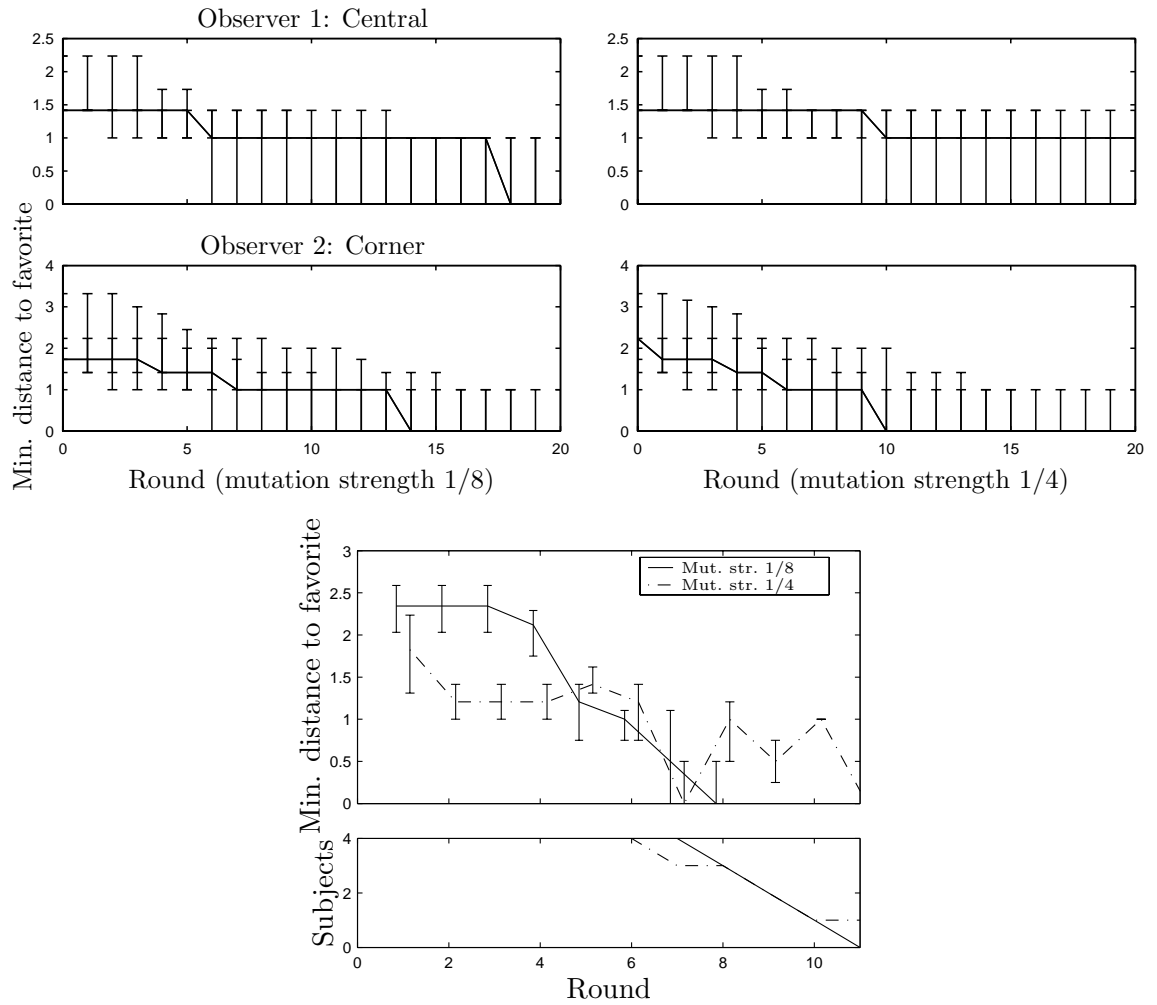
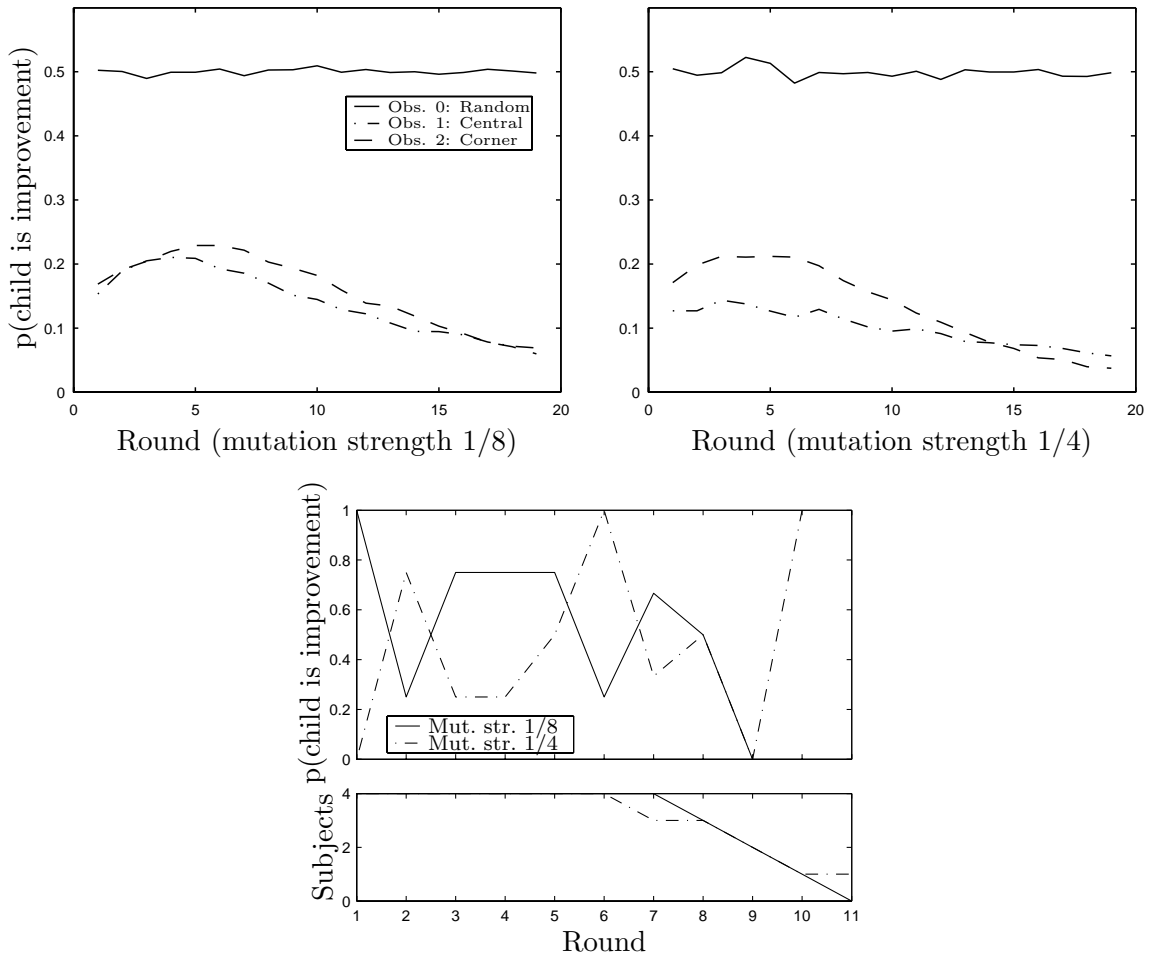


Figure 5.3: Experiment 1 mutation strength follow-up simulation (top) and experimental results (bottom): probability of improvement being indicated vs. round. For the objective simulated observers the probability of indicating improvement is the probability of actual improvement since these observers never make an error. For the random observer, the probability of indicated improvement is very near 50% since our inferencing algorithm compares the better of the 1 or 2 new solutions each round with the previous best solution. The random response to this single comparison determines whether an improvement is counted for the random observer. For the experimental subjects, the probability of the user preferring a new solution over the previous favorite solution is shown. Since there is some uncertainty in the subject responses, we expect the curves shown to be slightly higher than the actual probability of improvement.



This uncertainty about mutation strength points to more fundamental issues, however. First, the sampling density of the parameter space may need to be adjusted on a per-subject basis if JNDs vary greatly between subjects. Second, to be robust to subjects who give hasty and error-prone responses even when alternatives are separated on the order of a JND, the inferencing method discussed in Chapter 9 could be extended to incorporate redundant information by occasionally verifying that inferred relationships are correct.

CHAPTER 6

Experiment 1 Follow-up: Cross-subject Preferences

The results of the first experiment indicated widely varying preferences between subjects. Subject debriefings and a few re-runs of the experiment suggested that these preferences were perceived as having very good quality and were repeatable, implying that the GA had successfully found subject-specific preferences. We desired more solid data to substantiate that conclusion and therefore conducted the present follow-up experiment.

6.1 Methods

The results of the first experiment were analyzed to find the four settings most representative of the most preferred setting across all subjects [Figure 6.1 and Table 6.1]. For each subject, his or her own most preferred setting was added to this set. The subjects then compared all settings against all other settings, using a balanced presentation order [§6.1.2]. In addition to indicating preference, subjects were asked to record the degree of similarity among the members of each pair by placing an ‘X’ on a line between 1 (indicating “clearly different”) and 5 (indicating “identical”). This required $2\binom{5}{2} = 20$ comparisons.

Figure 6.1: Representative solutions. Any solution that was a favorite of any subject is indicated with an ‘x’. No indication is made here of how many other solutions were tied with the favorite, or whether any other subjects arrived at the same favorite. A cluster analysis of subject favorites from experiment 1, using a metric that accounted for the number of ties for the favorite solution, suggested 4 solutions representative of the set of favorite solutions across all subjects. We then searched for the 12-D (3 parameters \cdot 4 solutions) local minimizer of 1-norm error (of the subject’s favorite to the nearest item in the representative population), again weighted to compensate for some subjects having more ties than others. This was first done with a quick optimizer (which considered the $2 \cdot 12 = 24$ sets within one parameter increment of the pivot set), and, after the quick optimizer could do no more, a comprehensive local optimizer that considered all $3^{12} = 531441$ sets in the local neighborhood of the pivot set. The optimal set is represented by stars and its parameter values are listed in Table 6.1. The local minimum found by this method had an error nearly 26% smaller than the set first suggested by the cluster analysis.

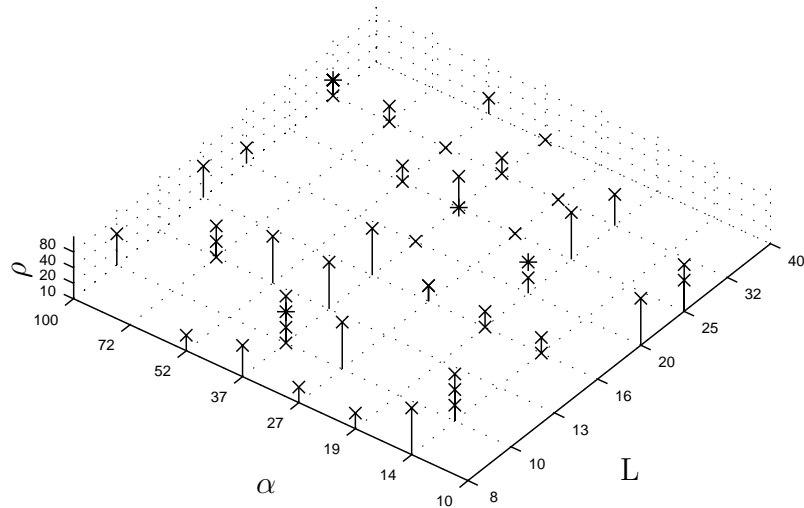


Table 6.1: Representative solutions. Each subject’s individual favorite was added to this set (gene E) and then a complete set of comparisons among the solutions was performed.

Parameter	Gene				
	A	B	C	D	E
L	10	20	25	32	?
α ($\cdot 10^6$)	37	19	37	100	?
ρ	40	40	10	20	?

6.1.1 Subject Instructions

The subject instructions were similar to those used in the main experiment. Subjects were also given a data sheet showing a linear scale from 1 (clearly different) to 5 (identical) for each comparison. The sheet contained the following instructions...

Please place an X on the line to indicate how similar the feedback canceller settings sound in each pair. A “1” indicates that canceller settings sound clearly different. A “5” indicates that settings sound identical.

After you have made 20 decisions, the green light will blink continually indicating that you have completed the evaluation.

6.1.2 Presentation Order

To control for changes in subject preferences during the run, a presentation order was designed such that if a solution appears in pair N , it may not appear in pair $N + 1$. We call this the “non-adjacency condition”. For 5 alternatives, as we have in the current experiment, there are several orders that satisfy this condition.

In addition, since subjects in the main experiment showed a bias for the first solution presented [§4.2.4], it was decided to perfectly balance the presentation and present each pair in both possible orders. This was accomplished by selecting for the first $\binom{5}{2} = 10$ pairs a non-adjacent ordering and randomizing the presentation order of each pair. Then, for the final 10 pairs, the 5 alternatives were randomly permuted, with a constraint, and the pairs were reversed as necessary to be perfectly balanced against the first 10 pairs. The random permutation was constrained so that the non-adjacency condition was not violated between pairs 10 and 11. The presentation order used is shown in Table 6.2.

Figure 6.2 enumerates the possible orderings of the first 10 pairs. It is interesting that none of the presentation orders satisfy the non-adjacency condition after any cyclic shift. That is, the tenth and final pair always has one element in common with the first pair.

Figure 6.2: All possible orderings of $\binom{5}{2} = 10$ pairs without respect to presentation order within the pair that satisfy the non-adjacency condition. For brevity, the labels “A” through “E” are omitted from the diagonals. Obvious permutations are omitted. For example, the positions of “1” and “2” in the head node are immaterial as long as they do not contain any of the same elements. Also, “3” was placed in “AE”, but it could also have been placed in “BE”, which would merely have been a permutation. Further, “4” was placed in “BC”, but swapping labels “C” and “D” would have put it in “BD”, the other legal location. Arrows point from parent nodes to child nodes. When there is only one possible subsequent pair, multiple nodes are collapsed. “X” indicates that a branch has ended without producing any valid presentation orders.

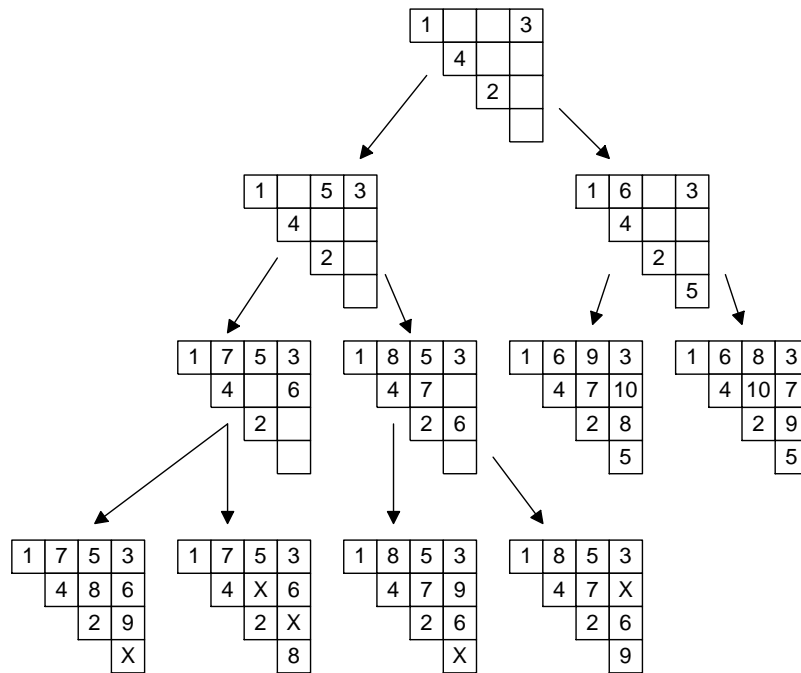


Table 6.2: Presentation order for cross-subject preference follow-up.

First Half	Second Half
AB	DA
DC	BE
EA	CD
BC	AE
ED	CB
CA	DE
BD	BA
EC	CE
AD	DB
EB	AC

6.2 Results

Table 6.3 summarizes the results of the experiment. Potential rankings based on subject preference and similarity judgments are evaluated with an objective function. The objective function is the sum over all comparisons of the dissimilarity (6 minus the similarity) multiplied by +1 if the comparison agrees with the ranking and -1 if the comparison disagrees with the ranking.

6.3 Discussion

Of 16 subjects, 5 gave responses resulting in an order with their favorite from the first run being at the top. Another 5 gave responses that were in moderate to strong agreement with the favorite not changing. Given that an average of 2 months passed between the experiments (during which time subject criteria may have changed as subjects gained familiarity with the system and its artifacts), that the subjects often felt uncertain (low values in the “Objective Max.” column in the table), and that they gave somewhat inconsistent responses (indicated by a large difference between “Objective Max.” and “Objective Max. Real.” in Table 6.3), the results indicate a solid amount of repeatability and avoidance of local maxima. If more subjects had arrived at merely local maxima in the initial experiment, we would have expected fewer to have chosen their own favorite a second time. For subjects with a high

Table 6.3: Most likely rankings from 20 paired comparisons with similarity judgments. The most likely ranking maximizes the objective function defined in the text. Subjects are ordered according to the maximum realized value of the objective function for their decisions, which is an indicator of subject-specific discrimination ability and consistency. The maximum possible objective value is 100 (20 comparisons \cdot 5 (maximum dissimilarity)), but this is limited by the similarity judgments made to the value shown in the “Max.” column. The maximum objective value for an order with item 4 (the subject-specific favorite from the main experiment) in the lead, which resulted from moving 4 to the lead of the listed order, is also shown. Strong agreement between objective values indicates a high likelihood that the favorite from the main experiment was closer to the subject’s actual favorite than any of the representative solutions. An asterisk (*) is used to highlight each subject whose most likely ranking put his or her putative favorite above all the representative solutions (100% agreement). For 7 subjects there were between 1 and 3 additional rankings that also maximized the objective function.

Subject	Rank						Objective			
	1	2	3	4	5	Max.	Max. Real.	4 Leading	Agreement	
MJG	2	0	4	1	3	67.85	65.75	58.55	89%	
EAD	4	3	2	1	0	62	60	*	*	
SLL	2	1	0	4	3	63.1	50.2	39.1	78%	
CC	2	3	4	1	0	56.9	44.9	16.7	37%	
WSW	4	1	0	2	3	54	40	*	*	
SAM	1	0	4	3	2	47	39	11	28%	
GD	0	2	4	1	3	45	37	33	89%	
GJ	1	0	2	4	3	60	36	18	50%	
MER	1	2	3	4	0	62.25	35.45	2.25	6%	
JSK	2	0	3	1	4	55.3	31.8	9.25	29%	
KP	4	2	1	0	3	40.7	30.5	*	*	
DVT	4	3	1	2	0	58.7	26.8	*	*	
GG	1	0	3	4	2	42	26	14	54%	
LW	1	2	3	0	4	52	26	2	8%	
CB	0	1	2	4	3	40.25	25.75	11.45	44%	
DA	4	1	3	2	0	28	9	*	*	

degree of consistency in the follow-up, but with a low objective value for preferring their own initially indicated solution, such as CC and SAM, either their criterion changed, or they were indeed caught in a local optimum in the initial run. Subject SAM noted that he became more familiar with the system after running the initial experiment and was aware of additional artifacts during the follow-up experiment, so it is likely that his criterion function changed.

CHAPTER 7

Experiment 2: Multiband Expansion Fitting

For our second main experiment, we used the GA to investigate subject preferences for three-band dynamic range expansion of low-level sounds. With 6 parameters, this system is more complex than the 3-parameter feedback cancellation system used in Chapters 4 through 6. The goal of this multiband expansion system is to put unwanted quiet environmental background sounds (such as breathing, fan noise, etc.) below the threshold of hearing while providing sufficient amplification for desired sounds. Part of the motivation for investigating expansion preferences is that little research has been published in this area. A summary of previous work is included in the Background below.

In experiment 1, we had a partially satisfactory objective measure of success, gain margin, along with a number of indirect measures. The cross-subject preferences follow-up experiment gathered additional direct measures by investigating repeatability and, through similarity judgments, strength of preference. In the current experiment, we again have various indirect ensemble measures consistent with good simulated performance. In addition, we had four subjects repeat the experiment from different starting conditions to investigate repeatability and sensitivity to initial conditions. Also, these four subjects completed a survey that investigated their perceptions of solution quality and the fitting system.

7.1 Background

Most of the published work on expansion concerns *fast acting* expansion ($\tau \ll 1$ s), which is put forth as a method to increase the modulation depth and, hopefully therefore, the intelligibility of speech. In contrast, our work is on *slow acting* expansion (τ between about 0.5 and 10 s), which seeks to suppress unwanted low level sounds (breathing, motor hum, etc.) to provide more comfortable listening without sacrificing intelligibility.

Byrne and Dillon [5] use paired comparisons for sound quality (“pleasantness and naturalness”) judgments of compression/expansion systems, citing strong evidence in earlier works for the potential efficiency of methods based on paired comparisons. The authors consider fast acting expansion, and so their work is not directly applicable to ours. Also, the authors review literature that discusses cases in which increased subjective quality could be gained at the expense of intelligibility, supporting the hypothesis that objective measures alone do not provide an adequate evaluation of system performance.

Walker *et al.* [67] address fast acting expansion for simultaneous presentation of speech and noise, which is much different from what we address in our experiment. The authors use a nonsense syllable test as their objective measure, but make no use of subjective preference measures.

Dillon [13] provides an excellent introduction to the research on compression. As in most of the compression literature, the author focuses on speech intelligibility. Unfortunately, speech intelligibility tells such a limited part of the story for our experiment; good intelligibility could be had with much discomfort (*e.g.*, background and breathing noises amplified to the point of unacceptably annoying the subject).

van Buuren *et al.* [63] use time constants on the scale of syllabic durations. In contrast to our motivation for using expansion, the authors were using it in an attempt to increase intelligibility by increasing modulation depth. They stated, “As opposed to analysis of speech intelligibility, sound quality has not been evaluated very often for amplitude compression and expansion; we will mention three examples known to us.” The examples are [5, 44, 45]. Byrne and Walker [5] was discussed above. Both papers by Neuman and others [44, 45], however, considered compression alone. Most

other experiments of subjective quality use 5- or 10-level rating scales.

7.2 Methods

7.2.1 Parameters

In this experiment, the GA system varied six parameters. Three of the parameters were expansion ratios in the three bands. The bands, gains at threshold, and compression ratios above threshold came from each subject's existing prescription. The non-compressive prescription used for normal hearing subjects was the same as in prior experiments. The fourth parameter was the time constant used for estimating signal power in each band.

The two remaining parameters controlled the expansion/compression thresholds in the three bands. Given that the main signal of interest to hearing aid users is speech, the LTASS (long-term average speech spectrum) data of Byrne *et al.* [9] were used along with the band locations to calculate nominal thresholds in each of the three bands. One of the parameters was used as an offset to the overall level of all thresholds, while the other served to enhance or diminish the slope of the thresholds across the bands. Thus, the LTASS was used as a prototype, adjusted for overall speech level and high or low frequency emphasis as demanded by the subject's listening environment.

Specifically, given the two crossover frequencies, integration of the LTASS yields nominal values for all three kneepoints between the expansion and compression regions, T_{Ni} . We express these as deviations from the first band nominal kneepoint

$$\begin{aligned}\Delta T_{N1} &= T_{N1} - T_{N1} = 0 \\ \Delta T_{N2} &= T_{N2} - T_{N1} \\ \Delta T_{N3} &= T_{N3} - T_{N1}\end{aligned}\tag{7.1}$$

Then, our parameterization consists of a slope, s , and a level, l . These two parameters are varied by the GA and are used to calculate the actual kneepoints

Table 7.1: Experiment 2 parameter values. Expansion ratios, x_i , $i \in \{1, 2, 3\}$, are given as the change in input decibels required to obtain a 1 dB change in output level. A ratio of 0.07 gives extreme expansion while a ratio of 1 gives linear operation. The time constant, τ , is given in seconds. The inter-kneepoint slope, s , is unitless and the offset from the nominal kneepoint levels, l , is in decibels.

x_i	0.07	0.08	0.095	0.12	0.15	0.21	0.35	1
τ	0.5	0.767	1.18	1.81	2.77	4.25	6.52	10
s	0.6	0.8	1.0	1.2	1.4			
l	0	2.5	5	7.5	10	12.5	15	17.5

$$T_i = T_{N1} + \Delta T_{Ni} \cdot s + l \quad (7.2)$$

The parameter values used in the experiment are shown in Table 7.1. The x_i were chosen to cover the range from no expansion (a change of 1 dB output for every 1 dB change in input) to extreme expansion (a change of $0.07^{-1} \approx 14$ dB output for every 1 dB change in input). The time constants, τ , were chosen to cover the range over which subjects can discriminate, as suggested by previous experiments done at Starkey. The slopes of the kneepoints vs. frequency, s , were chosen to cover a wide range from shallow (0.6) through nominal (1.0) up to steep (1.4). Through a pilot experiment, it was discovered that if the offset from the nominal kneepoint levels, l , was negative, then expansion would seldom become active even in quiet situations. Thus, a minimum offset of 0 dB was used. In order to provide a wide performance range over which to optimize, this offset had a maximum value of 17.5 dB.

To allow for the additional complexity of the parameter space in this experiment, the population size was increased from five to eight. When choosing a population size, a balance must be struck. If the population size is too small, the GA will not have sufficient memory to perform an efficient search. If the population is too large, so many decisions will be required that the user would become fatigued and perhaps stop using the system before enough information was gathered to begin searching.

7.2.2 Subject Instructions

Subjects received an instruction sheet to read before beginning the experiment that contained a diagram of the hearing aid prototyping device and a description of the function of each button. The sheet also contained the following text that describes the purpose of the expansion system and suggests environments and qualities for which to listen in order to critically evaluate the system's performance...

Listening Environments: Your portable processor has been programmed to reduce background noise in situations with very little external sound while still allowing you to comfortably hear soft sounds that may occur. Therefore, you will need to conduct your listening evaluation in an environment that you can alternately make very quiet and introduce soft sounds into, as needed. Here are some suggestions:

Begin listening in a very quiet environment with as close to no external sound as possible (*e.g.*, sitting alone in your living room, *without* the TV or radio). Then...

Introduce some soft sounds into the environment (*e.g.*, have a quiet conversation with one other person, turn the TV or radio on at a low volume, etc.).

Which program do you like best overall?

To determine which program you like best ask yourself the following:

Very Quiet Situation: How much does the background noise fade away during the first 30 to 90 seconds you are in the quiet situation? Too much? Too little? Just enough? Not at all?

"Soft Sound" Situation: Is the loudness of soft sounds too loud? Too soft? Just right?

You may switch back and forth between programs and listening environments as many times as necessary to determine which of the two programs you prefer.

Indicating your preferences: Once you have determined which of the two programs you prefer, please indicate your choice by pressing the appropriate button on the processor. If you have been unable to determine a difference between the two programs after intensive listening (this is okay and may happen many times), select either one. After you choose one of the programs, two *new* programs will be loaded into the processor. Repeat the above procedure to determine which of these two new programs you prefer in the two listening environments.

Continue this process until you have listened to approximately 50 to 75 pairs of

programs (about 1 hour to 1.5 hours of listening). If you cannot hear any difference for about 20 consecutive choices, you may stop earlier. Please continue longer if you *can* hear the difference between most of the sets of two programs — even if you have already met or exceeded the above time span.

If possible, please do not turn the processor off during your listening session. If you need to stop listening for more than about 20 minutes, you may turn the processor off, but please try to listen to at least about 30 pairs of programs each time the processor is turned on.

7.2.3 Follow-up Survey

Four subjects who participated in two runs of the experiment also completed the following survey after their second run. This survey was designed to gather information about each subject's perception of the GA fitting system and of the solutions presented. . .

1. Please select the term for each statement below that best characterizes your experience with the portable processor during your most recent listening evaluation. You may write in your own term if none of those provided accurately describes your experience.

(a) I found that, for most of the pairs, determining a difference between them was:

- Extremely easy
- Fairly easy
- Slightly difficult, but possible
- Difficult
- Very difficult
- Other (please describe)

(b) As I continued listening, the task of determining the difference between the settings generally:

- Got easier

- Stayed about the same
 - Became more difficult
 - Other (please describe)
2. Which of the following statements best characterizes your experience with the portable processor during your most recent listening evaluation?
- When I started listening, things sounded bad. As I continued listening, things sounded better and better.
 - When I started listening, I liked the way things sounded. As I continued listening, things sounded worse.
 - Things sounded good when I started listening, and did not sound any better or worse as I continued listening.
 - Things sounded bad when I started listening, and did not get any better or worse as I continued listening.
 - Other (please describe)

7.2.4 Initial Populations

As in the first experiment, four initial populations were used so that results were not dependent upon a particular starting configuration. Each subject was assigned to one of the four cohorts at the beginning of the experiment. The populations are shown in Table 7.2.

7.3 Results

In this section, we first compare simulated and experimental results to arrive at several indirect ensemble measures of success [§7.3.1]. We then present the survey results for four subjects [§7.3.2]. These results provide a subjective measure of success. Next, we analyze the repeatability of results for the same four subjects [§7.3.3]. Finally, we look at first/second alternative decision bias [§7.3.4].

Table 7.2: Experiment 2 initial population indexes.

Population Solution	1								2							
	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
x_1	0	4	3	7	7	3	1	6	2	2	2	4	6	2	5	5
x_2	3	4	6	7	6	1	3	5	5	4	2	3	6	4	4	5
x_3	0	3	7	0	7	6	7	2	7	1	4	7	6	2	7	2
τ	5	1	4	2	1	0	1	1	0	7	1	1	7	5	5	2
s	4	2	0	2	3	4	3	1	4	3	2	0	3	4	3	2
l	3	5	3	6	6	5	0	4	6	4	4	5	3	3	1	1

Population Solution	3								4							
	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
x_1	5	3	7	2	1	3	3	5	1	1	1	0	0	5	6	7
x_2	7	5	1	5	3	5	2	6	2	0	4	2	3	2	3	0
x_3	5	0	4	2	4	1	1	3	2	1	6	6	3	7	1	5
τ	3	4	0	5	0	0	2	0	0	3	6	5	7	6	1	7
s	0	4	0	4	2	4	1	4	1	3	3	0	0	0	0	1
l	3	5	2	3	0	5	5	1	3	7	3	6	3	3	3	0

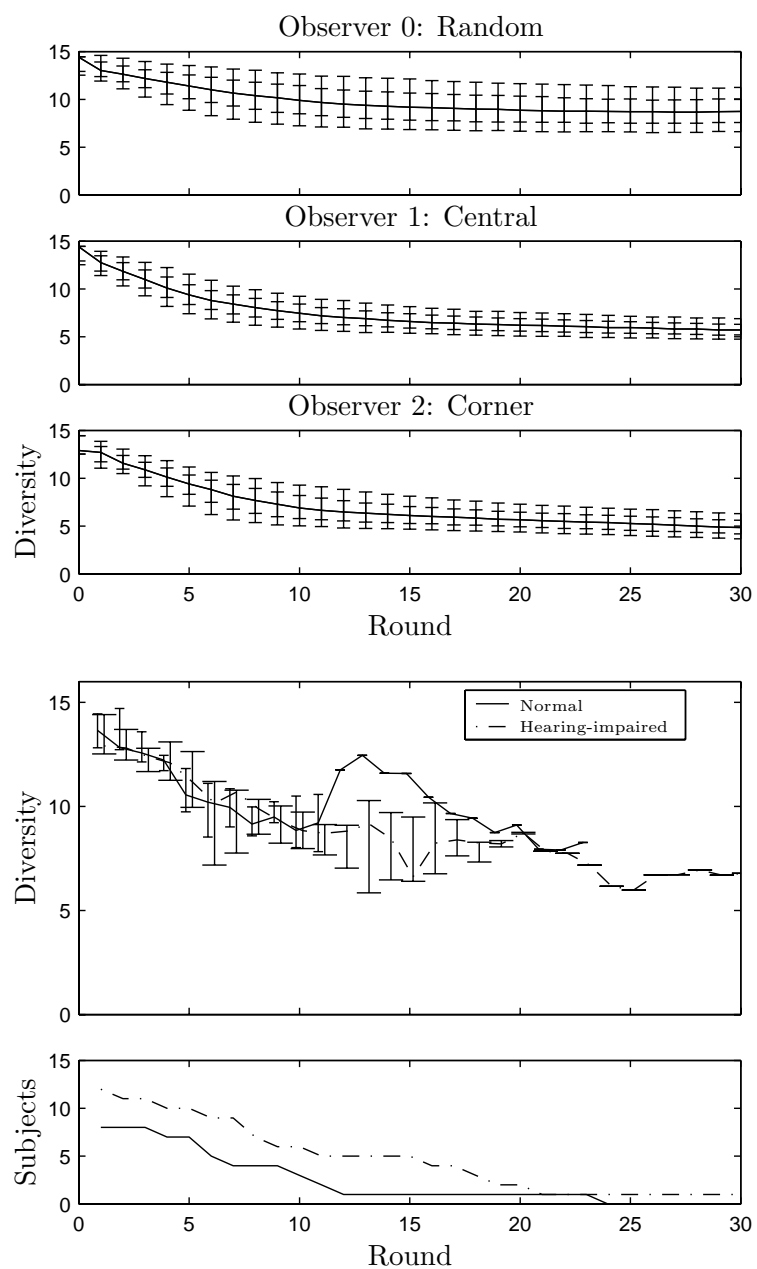
7.3.1 Simulated and Experimental Results vs. Round

Figure 7.1 shows the diversity¹ as a function of round, for both the simulation and the experiment. Diversity is initially between 14 and 15 for all experimental and simulated subjects and has a general downward trend. The large increase in diversity for normal hearing subjects after round 10 is due to subjects who had converged and ended their runs. Those subjects who continued had greater diversity, which increased the median value. The asymptotic experimental diversity is about 7, while the random, central, and corner observers had asymptotic diversities of 9, 5, and 5, respectively.

Figure 7.2 shows the minimum distance to the favorite solution vs. round for both simulated and experimental subjects. The favorite solutions for the simulated observers are known, while the favorite solution for each experimental subject is approximated by the top-ranked solution from the final round of the run. Simulated subjects completed exactly 30 rounds, while the median number of rounds completed was 15 for hearing-impaired subjects and 9 for normal hearing subjects. Both simulated observers show a steady decrease in the distance as a function of round. The

¹The population diversity measure is discussed in §8.3.

Figure 7.1: Experiment 2 simulation (top) and experimental results (bottom): diversity vs. round. Median with interquartile (all) and interdecile (simulation only) ranges indicated by bars.



experimental data show a similar trend for the first 5 rounds, but vary much in latter rounds. This was likely caused by a combination of some subjects ending the experiment around the 5th round and some parameters being neglected by the subjects (see Table 7.5 and the associated discussion). This is in contrast to the simulated subjects, which were affected by all the parameters equally and consistently.

Figure 7.3 shows the probability that a new child solution was preferred over the previous favorite solution vs. round for both simulated observers and experimental subjects. As in the first experiment, the probability of the random observer preferring a new solution over the previous favorite is 50%. For the central observer, the chance of preferring a child decreases from about 20% in the early rounds to about 10% in the 30th and final round. For the corner observer, the probability quickly increases from about 20% to 30% and then steadily decreases to about 20% by the end of the run. Both normal hearing and hearing-impaired subjects had nearly a 50% chance of preferring a child solution over the previous favorite. This is quite close to the performance of the random observer. However, most subjects indicated that they had difficulty discriminating between the alternatives in a moderate number of pairs, so, when considered with the diversity and forthcoming repeatability data, it seems likely that a large majority of decisions indicating an improvement were either for actual improvements or among nearly indiscriminable alternatives. In the latter case, the subject's decision has little bearing on the outcome. That is, when solutions are indiscriminable or nearly indiscriminable, their relative rankings are unimportant.

Simulation showed that the rational simulated observers required between 14 and 21 decisions in the first round, with between 16 and 18 decisions required for the vast majority of runs. Fewer decisions were required in subsequent runs due to information carried forward by the inferencing algorithm. The median and 90th percentile number of decisions required was 7, while the 10th percentile value was 5. The real subjects required a similar number of decisions as the simulated rational subjects in the first round and in the subsequent rounds. There was no significant difference between normal and hearing-impaired subjects.

Figure 7.2: Experiment 2 simulation (top) and experimental results (bottom): minimum distance to favorite vs. round. Median with interquartile (all) and interdecile (simulation only) ranges indicated by bars. For the experimental subjects, the favorite setting is taken as the top-ranked value at the end of the run.

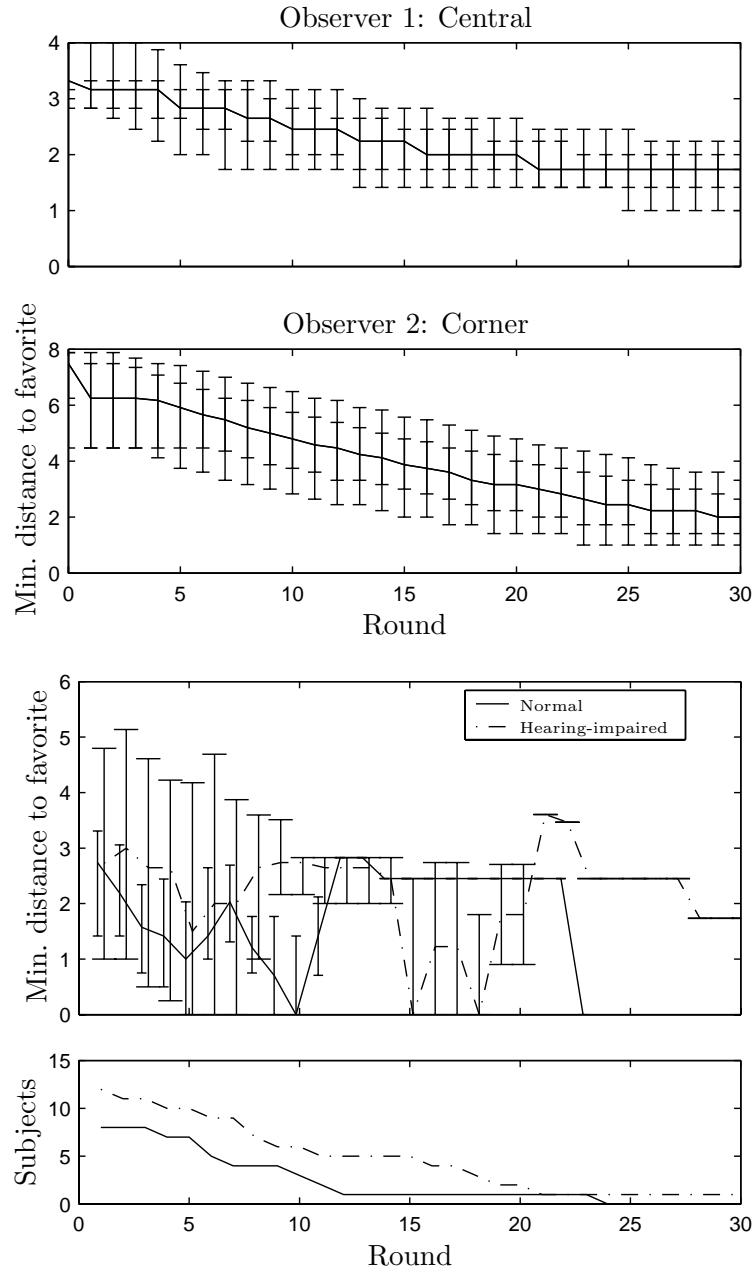


Figure 7.3: Experiment 2 simulation (top) and experimental results (bottom): probability of improvement being indicated vs. round. For the objective simulated observers the probability of indicating improvement is the probability of actual improvement since these observers never make an error. For the random observer, the probability of indicated improvement is very near 50% since our inferencing algorithm compares the better of the 1 or 2 new solutions each round to the previous best solution. The random response to this single comparison determines whether an improvement is counted for the random observer. For the experimental subjects, the probability of the user preferring a new solution over the previous favorite solution is shown. Since there is some uncertainty in the subject responses, we expect the curves shown to be slightly higher than the actual probability of improvement.

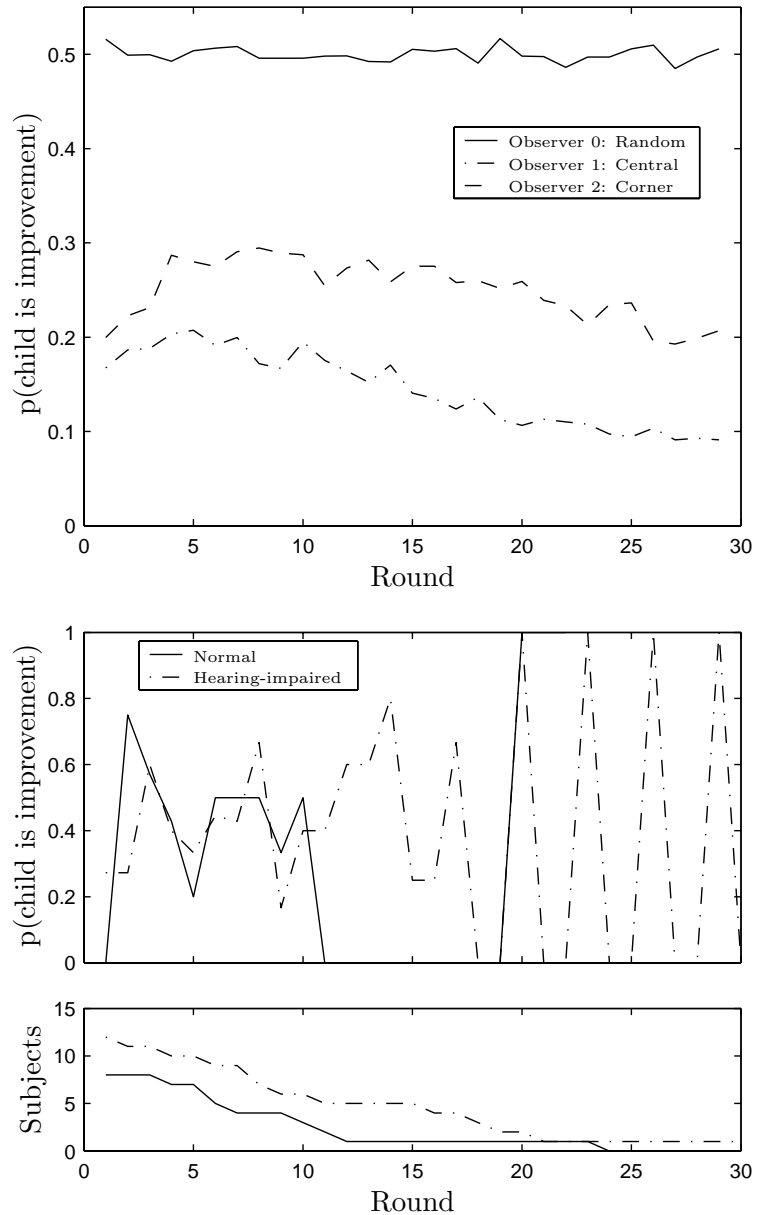


Table 7.3: Experiment 2 survey results.

	1a	1b	2
CC	3: slightly difficult but possible	3: became more difficult	O: *
GD	2: fairly easy	3: became more difficult	1: bad to good
GJ	O: some obvious, some very difficult	3: became more difficult	O: *
KP	O: one extreme or the other	O: *	O: indecisive – faster response at end
*: Subject provided comments, but did not answer the question asked.			

7.3.2 Follow-up Survey

The survey results are presented in Table 7.3. In response to the first question, two subjects noted that discriminating between alternatives was at times very easy and at other times very difficult. Another subject indicated that discriminating was slightly difficult, while the fourth subject indicated that the task was fairly easy. Responding to the second question, three subjects noted that discriminating between alternatives became more difficult as the run progressed. The remaining subject provided comments in response to the question, but did not give a clear answer. The final question asked subjects about the change in solution quality. Two subjects provided detailed responses but did not answer the question. Another subject was uncertain if the solutions became better, but noted that the response was faster at the end — this was stated as an observation, not as a complaint, so we consider it a slight indication that solution quality improved. The remaining subject indicated that solution quality seemed to increase throughout the run.

7.3.3 Repeatability

Table 7.4 shows the best solution found by the GA and the number of rounds completed for four hearing-impaired subjects. Each of these subjects completed two runs with different initial populations. These data are analyzed in Table 7.5, which contains the weighted absolute parameter index deviations between the initial and follow-up favorite solutions. Results from subjects completing more rounds in both

Table 7.4: Repeatability for four subjects. The parameter indexes correspond to the values shown in Table 7.1. In addition to each subject’s favorite at the end of the first and second run, the number of rounds completed in each run is shown.

	Main							Rerun						
	x_1	x_2	x_3	τ	s	l	Rounds	x_1	x_2	x_3	τ	s	l	Rounds
CC	7	4	7	4	2	3	5	3	4	0	5	2	5	3
GD	7	3	6	6	4	2	19	6	6	0	2	2	6	7
GJ	1	1	1	3	3	7	10	3	5	0	4	4	6	7
KP	3	4	1	5	1	6	20	6	4	7	5	3	7	15

experiments are considered more significant since there was more time to converge. So, subject deviations are weighted in proportion to the geometric mean of rounds completed. Given the short length of many of the runs and that six parameters were being simultaneously optimized, these values indicate that subject responses were repeatable. In addition to some short run lengths, the power of the statistical test is limited by having only four subjects. Nonetheless, two parameters were controlled at the 80% confidence level. The weighted absolute deviations also give an indication of the relative importance of each parameter; τ , the signal power estimation time constant, was the most quickly and consistently controlled, which is consistent with the observation many subjects made regarding the importance of the adaptation speed.

7.3.4 Decision Bias

As can be seen from Table 7.6, there is a very slight bias for the second solution presented. We eliminated one source of bias by randomizing presentation order. It is interesting that in the first experiment there was a slight bias for the *first* solution presented. The data were analyzed in the same manner as in experiment 1 [§4.2.4].

7.4 Discussion

While the ensemble performance measures and statistical baselines (variation, distance to favorite, and probability of improvement vs. round) suggest that subjects

Table 7.5: Analysis of repeatability for four subjects. The second column shows values of weighted absolute deviation across subjects. Subjects are given greater weight for completing more rounds in both the main and follow-up experiment. The table also includes lower bounds for a single-ended hypothesis test at various confidence levels. The hypotheses are $H_0 : x \geq r$ and $H_1 : x < r$, where x is the absolute deviation weighted across four subjects and r is the weighted absolute deviation for parameter sets selected using a uniform distribution.

Parameter	Deviation	p (for CIs)
8-level parameters		
	1.30	0.9
τ	1.42	
x_2	1.66	
	1.69	0.8
l	1.94	
	2.00	0.7
	2.28	0.6
x_1	2.33	
	2.56	0.5
x_3	5.08	
5-level parameter		
	1.60	0.5
s	1.61	

Table 7.6: Analysis of decision bias in experiment 2.

A	B	95% CI>	$E(p) = A/(A + B)$	95% CI<
176	164	0.463	0.518	0.572
24	26	0.337	0.480	0.626
28	23	0.403	0.549	0.689
55	66	0.364	0.455	0.548
12	10	0.322	0.545	0.756
36	45	0.334	0.444	0.559
29	30	0.359	0.492	0.625
50	75	0.313	0.400	0.491*
43	49	0.363	0.467	0.574
82	75	0.441	0.522	0.603
27	26	0.368	0.509	0.649
7	54	0.047	0.115	0.222*
11	11	0.282	0.500	0.718
24	17	0.421	0.585	0.737
51	47	0.417	0.520	0.622
18	33	0.224	0.353	0.499*
24	26	0.337	0.480	0.626
35	36	0.372	0.493	0.614
57	66	0.373	0.463	0.556
35	46	0.322	0.432	0.547
824	925	0.448	0.471	0.495*

had only moderate success with the system, the survey results and, even more, the repeatability results show that subject responses were consistent enough to guide the GA to similar solutions in multiple runs. Arriving at these similar solutions when using different starting conditions supports the hypotheses that the GA was not trapped by local maxima and that settings for each subject were not merely the result of a random walk (which is a legitimate concern when considering the probability of improvement data in isolation).

The difficulty that many subjects had discriminating between alternatives together with the large deviations between reruns for some parameters suggests that the JNDs may be much larger for those parameters than the spacings in our samplings. That is, the overly fine samplings increase the likelihood that the GA will generate indistinguishable settings that degrade the performance of the fitting procedure. The results suggest that this oversampling had little effect in the first 5 rounds as many improvements were made. However, the slowed performance in the latter rounds is likely due to oversampling of the perceptual space. While this results in slower objective convergence after the initial rounds, we must be careful not to overstate the detrimental effects of perceptual oversampling — that subjects reported much more difficulty discriminating between solutions late in the run suggests that the remaining variation in the parameter space was simply not significant. So, the search had converged perceptually even though the objective diversity was somewhat higher (7 instead of 5) than that predicted by the simulations, which used infinitely discerning observer models.

CHAPTER 8

Population Variation Metrics and Their Bounds

In this chapter, we develop variation metrics for populations of binary and multi-level values and investigate their lower bounds given that all members of the population are unique.¹ These metrics are useful in detecting convergence and determining when the GA should be terminated, that is, when the effort required to continue the GA is not worth the probable improvements. The terms “convergence”, “diversity”, and “variation” have many meanings in the literature, so it will be helpful to define them here.

- convergence – a state of the GA, characterized by a low diversity, in which most genes in the population are very near a single solution
- diversity – a property of a population that indicates how varied its genes are; diversity is equal to 0 if and only if all genes in the population are identical, otherwise diversity is positive
- variation – a synonym of diversity

Some researchers, such as Leung *et al.* [36] and Rudolph [52], define convergence differently than we do. To them, convergence indicates whether the global optimum is reached, and how quickly it is approached. Rudolph used a Markov chain analysis to show that with probability one, without an elitist operator a GA never converges

¹Without the uniqueness constraint, the lower bound is 0, which is achieved for a population containing the same gene for all its members. We impose a uniqueness constraint to maintain some minimum level of variation and to ensure that the population slots are used efficiently.

to the global optimum. Since our goal is not to seek the global optimum at all costs, but to efficiently make improvements to parameter values, the above definition of convergence, which is similar to Fogel’s [18], serves us better.

Nagao [43] is one of the few GA researchers to use the term “diversity”. He uses it much as we do and proposes a bitstring encoding method, different from both direct and Gray codes, which maintains genotypic diversity even as the phenotypes converge. Although we do not consider this dichotomy between diversities to be useful in our work, we do agree that it is important to monitor diversity for two reasons. First, rapidly decreasing diversity indicates premature convergence. That is, it indicates that an insufficient amount of effort is being applied to global search. Second, regardless of how quickly it decreases, when diversity becomes too low, the crossover operator has little effect, so the primary method of search becomes mutation. After this occurs, the GA tends to be very inefficient, so the GA should either be terminated or a local search algorithm should be applied [18].

8.1 Notation

The following symbols are used throughout this chapter.

- C – ball center location
- i – gene number within the population
- k – number of bits in current sub-block
- L – length of bitstring or gene
- M – dimensionality
- N – population size or number of Bernoulli trials
- p – norm number or probability of success in Bernoulli trials
- P – population
- P_i – the i th gene of population P

- Q – gene shift vector
- r – p -norm radius
- v – variation
- v_m – minimum variation

8.2 Binary Population Variation Metric

We now propose a variation metric for a population of bitstrings. Our metric is the expected² proportion of bits that differ between distinct members of the population. Let P be a population with N members P_i of length L . Then, the variation of the population is³

$$v(P) = \frac{\mathbb{E}_{i \neq j} \|P_i \oplus P_j\|_1}{L} \quad (8.1)$$

This metric is invariant to bit-order permutations and inversions of any set of bits and is bounded on $[0, 1]$. Often, we put a uniqueness constraint on the population.⁴ With a uniqueness constraint, the population cannot converge to a single solution. Instead, it may converge to a set of closely related solutions, differing in a small number of bits. Thus, we desire a lower bound on the variation metric so that we can measure the convergence of a uniqueness-constrained population. We first derive the tight lower bound as a function of gene length L and population size N . We then give a simple intuitive approximation to this bound that turns out to lie strictly below the actual bound (except for the trivial case of a population of size 1).

²We use expectation since we are interested in the statistical properties of the variation as a function over the set of possible populations. However, for a specific population, variation is a deterministic quality, so we replace the expectation operator with a summation over all pairs of distinct population members.

³ $\|\cdot\|_1$ indicates the 1-norm, or simple summation, of the elements of the vector argument.

⁴Without a uniqueness constraint, the algorithm could crossover identical genes, yielding no improvement. This would be a minor problem for a generational GA, but could be catastrophic in the steady-state GA since the parent genes remain in the population. Duplicate genes of moderately good fitness would quickly cause extinction of inferior genes. Thus the algorithm would tend to settle on a solution near one of its starting points, neglecting to search much of the space spanned by its initial population.

8.2.1 Exact Lower Bound

Since the measure is invariant to inversions of any set of bits, we may consider the all-zeros gene to be a representative member without loss of generality. We then add $N - 1$ genes to the population so as to minimize variation. Assigning these genes in lexical order⁵ starting with the all-zeros gene results in the minimum variation⁶.

We first consider the case in which the number of genes in the population is a power of 2: $N = 2^k$, $1 \leq k \leq L$. Then, for any given gene in the population, 2^{k-1} of all other the genes in the population ($2^k - 1$) differ in any given bit position. k positions out of a total of L may differ, so we have

$$v_m(L, 2^k) = \frac{2^{k-1}}{2^k - 1} \cdot \frac{k}{L} \quad (8.2)$$

The above function is undefined for a population of size one ($k = 0$). We define the variation of any population of size one to be 0.⁷

We now consider populations of sizes other than powers of two using a divide-and-conquer approach. Given a population of size N , choose k such that $2^k < N < 2^{k+1}$. Then the population consists of two chunks having sizes 2^k and $N - 2^k$. We assign the members of the population (again chosen in lexical order) that have zeros in the first $L - k$ bits to the first group and the remaining members, which have zeros in the first $L - (k + 1)$ bits and a one in the following bit, to the second group. As illustrated in Figure 8.1, we consider the cross product without replacement of these groups in three cases. After determining the variation in each cross product, we take

⁵Lexical order means that each subsequent gene is determined by adding 1 to the least significant bit of the previous gene and carrying as necessary.

⁶That this assignment results in the minimum possible variation is a hypothesis. It has been shown to be true by an exhaustive search for small N . Also, if we neglect that we are considering 2 values *without* replacement, this can be modeled as a Bernoulli process. In any case, the effect of replacement is negligible for large N . The variance of a Bernoulli process is well known: $\sigma^2 = Np(1-p)$, where p is the probability of success of a trial and is equal to the proportion of population members that have a value of 1 in any particular bit position. $\frac{d\sigma^2}{dp} = N(1-2p)$ passes through 0 at $p = 1/2$, so the variance is least sensitive and maximum at $p = 1/2$ and grows increasingly sensitive and smaller as p approaches 0 or 1. This suggests that, to minimize variation, it is more important to keep the maximum number of bit positions homogeneous (which assignment in lexical order accomplishes) since σ^2 is most sensitive there than to spread the changes across the maximum number of bits.

⁷This simplifies the derivation for non-power-of-two cases in which we consider subpopulations of size one as having variation 0 since the element in isolation cannot contribute to the variation of the larger population.

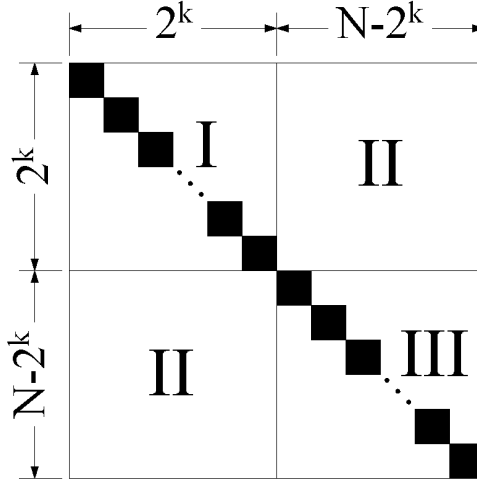


Figure 8.1: Divisions for computing the lower bound on the bitstring variation metric given a uniqueness constraint on the population. The solid squares indicate that only pairs of *unique* genes are chosen. Case I, representing a power-of-two population size, is solved directly. Case II is straightforward as there is no overlap between choices for each member of the pair. Case III is solved by recursion.

a weighted average to arrive at the tight lower bound of population variation with a uniqueness constraint.

Case I has a population size that is a power of two. We have already found the lower bound for this case [(8.2)].

Case II genes always differ in the $k+1$ bit and differ in each of the k least significant bits with probability $\frac{1}{2}$. Thus, these genes contribute $\frac{1+k/2}{L}$ to the minimum variation.

Case III is simply a smaller version of the original problem, so we solve it by recursion. This causes the original N to be broken up into a sum of powers of two, possibly ending with $2^0 = 1$, which we have defined as a special case of (8.2).

We weight each of these three cases by the number of pairings from Figure 8.1

$$\begin{array}{ll}
 \text{weight :} & \text{variation :} \\
 v_m(L, N) = \left[\begin{array}{ll} 2^k(2^k - 1) & v_m(L, 2^k) \quad \text{(I)} \\ + 2 \cdot 2^k(N - 2^k) & \frac{1+k/2}{L} \quad \text{(II)} \\ + (N - 2^k)(N - 2^k - 1) v_m(L, N - 2^k) & \text{(III)} \end{array} \right] / (N(N - 1))
 \end{array} \quad (8.3)$$

Equation 8.3 agrees with the lower bound found by brute force calculation. It is

possible to express the bound in a non-recursive form, but that neither simplifies the formula nor lessens the computational cost.

8.2.2 Approximate Lower Bound

Given our lexical order assignment rule, we can state a simple approximation formula for the lower bound on variation that serves as a check on the derivation of the exact bound. Neglecting that genes are distinct in each pair and interpolating between powers of two, we see that the approximate proportion of bits that are used is $(\log_2 N)/L$. If exactly half of the bits in each used position are ones, as is the case when the population size is a power of 2, we can multiply this proportion by $1/2$ since exactly half of the pairs of bits in each position contain different values

$$\hat{v}_m(L, N) = \frac{\log_2 N}{2L} \quad (8.4)$$

8.2.3 Discussion

Figure 8.2 shows the exact and approximate lower bounds for genes of length 10. Except for a population of size one, the exact bound is strictly above the approximate lower bound. As expected, the exact bound is close to the approximate bound, consistent with the assumption that the exact bound was derived correctly.

8.3 Parametric Population Variation Metric

In this section, we propose a variation metric for populations containing genes with multilevel parameters. Our motivation for calculating variation based on the multilevel parameter representation instead of on a bitstring representation is the same as for using the parametric representation instead of a bitstring representation [§3.1] — the parametric model is more appropriate to the problem domain being considered.

Recall the definition of the p -norm (or Minkowski norm) of a vector x of length M

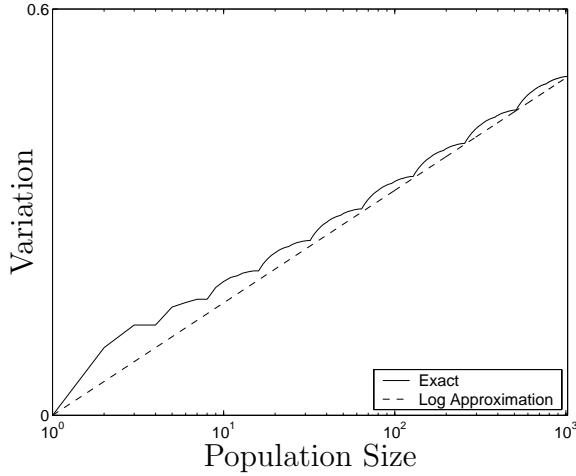


Figure 8.2: Minimum variation of genes of length 10 vs. population size given a uniqueness constraint. The scalloping is caused by tighter packing when the population size is near a power of 2. When the population contains the maximum number of unique genes, the approximate bound is $1/2$, since we expect half of the bits to differ between any two randomly chosen bitstrings. The actual variation (not merely a bound when the population is of maximum size) is slightly larger than this since we select a random pair *without* replacement.

$$\|x\|_p = \left(\sum_{i=1}^M |x_i|^p \right)^{1/p} \quad (8.5)$$

We define our multilevel parameter population variation metric to be the minimum value that can be obtained by shifting the population by some central value, which need not consist of integer coordinates, and then taking the p -norm of all coordinates of all population members

$$v(P) = \min_Q \left\| \left(\|P_i - Q\|_p \right) \right\|_p \quad (8.6)$$

$$= \min_Q \left(\sum_i \left(\|P_i - Q\|_p \right)^p \right)^{1/p} \quad (8.7)$$

In (8.6), the inner norm is over the index values in each gene while the outer norm is over all genes. The index i selects individual genes from the population. This is stated another way in (8.7), in which the outer norm has been expanded. In both

equations, P is the population and Q is the vector by which the population is shifted before calculating the norm.

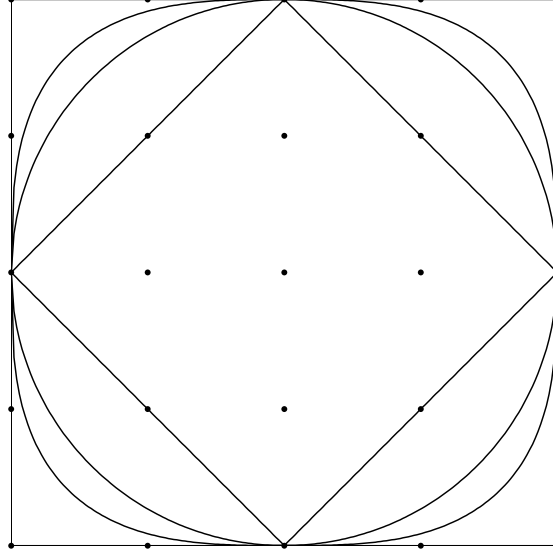
We consider the lower bound on the multilevel population variation metric as a function of dimension and population size. We generalize beyond the 2-norm. After defining several terms and proposing four hypotheses regarding the minimum variation, we find the lower bounds for the configuration used in both hearing aid experiments.

8.3.1 Definitions

Before we investigate the lower bound on the variation metric, we need to define several terms.

- ball – see p -ball
- compact – describes a set containing all the points within the convex hull of the points belonging to the set; all compact sets are also connected
- connected – describes a set in which any point can be reached from any other point by repeatedly moving one lattice step to another point in the set; describes points that belong to a connected set
- density – a rough characterization of how close a set is to being compact
- disjoint – describes a set or points that are not connected
- generalized Tetris problem – the extension of the Tetris problem in which s and q take on arbitrary integer values and in which *all* isomorphic sets are considered equivalent
- n -disjoint – describes a set that contains n groups of points such that each point belongs to exactly one group, each group is a connected set, and each pair of points drawn from separate groups is not connected
- p -ball – a surface in M -dimensional space that is the locus of all points having a p -norm distance of r (the radius) from a point C (the center of the ball); for example, for $(p, M) = (2, 2)$, the p -ball is a circle, for $(p, M) = (2, 3)$, the p -ball

Figure 8.3: p -balls of dimension $M = 2$ and radius $r = 2$ for $p \in \{1, 2, 3, \infty\}$. As p increases, the area of the ball increases. For $p \in \{1, \infty\}$, the ball is a square and for $p = 2$, the ball is a circle.



is a sphere, and for $(p, M) = (\infty, 4)$, the p -ball is a tesseract; some 2-D p -balls are illustrated in Figure 8.3

- set – a group of points belonging to an M -dimensional lattice
- Tetris problem – enumerating the connected sets in $s = 2$ -space consisting of $q = 4$ points; isomorphic sets, with the exception of reflections, are considered equivalent

8.3.2 Hypotheses

The following hypotheses are helpful in determining the lower bound on our variation metric as function of dimension, population size, and norm order for populations that satisfy the uniqueness constraint.

Hypothesis 1 *The set having minimum diversity is compact.*

We do not attempt to prove this here. A weaker version of this hypothesis, with a proof outline, is presented below.

Hypothesis 2 *The set having minimum diversity is connected.*

Proof outline: Consider a 2-disjoint set. Since our diversity metric is a Minkowski norm, it is minimized by subtracting some proper choice of central point from all solutions.⁸ For example, the second moment (in any number of dimensions) is minimized by subtracting the mean. For a 2-disjoint set having connected subsets with means M_1 and M_2 , the mean is a weighted average of M_1 and M_2 . By moving the two regions closer together so that they are no longer disjoint, there is less of a compromise on the shared mean, hence the diversity decreases. This argument extends to n -disjoint sets.

Recall that shifting a sequence by its negative mean minimizes the 2-norm (variance), but not, in general, the 3-norm, 4-norm, etc. The values that minimize those norms are solutions of quadratic, cubic, etc., equations. Antsaklis [1, p. 42] gives the following theorems for Minkowski norms

$$\|x\|_\infty \leq \|x\|_1 \leq l\|x\|_\infty \quad (8.8)$$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{l}\|x\|_\infty \quad (8.9)$$

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{l}\|x\|_2 \quad (8.10)$$

In order to apply the norm theorems to our variation metric, the shift by the central point must be constant across all orders (*e.g.*, 2-norm, 3-norm, ∞ -norm). However, the shift that minimizes the metric varies with p , the order of the norm. Thus, since we define our metric as the shift-minimized p -norm of the population, we cannot apply the norm theorems to compare variations as a function of p . We investigate using a constant shift to satisfy the theorems, which allows more powerful statements about bounds on the norm, but ultimately move to a shift optimized for each p .

Hypothesis 3 does not hold for the constant shift metric. It appears to hold for the metric that uses an optimal shift for each p . Indeed, as shown by exhaustive

⁸To find the minimizing offset, we consider (8.5). We replace x_i with $x_i - c_i$, where c is the offset. The $1/p$ exponent can be dropped without changing the location of the minimizing value. We then take the partial derivative with respect to each c_i . This results in a polynomial equation for each dimension. One of the roots of each polynomial is the location of the minimizing offset in that particular dimension.

Table 8.1: Number of connected forms in 3-space.

q	forms
1	1
2	1
3	2
4	7
5	23

enumeration of connected forms and calculation of several norms (with a heuristic-seeded gradient search for the norm-minimizing shift, which may be a multimodal function), Hypothesis 3 *does* hold for the 3-5 (3 dimensional, size 5) problem.

Hypothesis 3 *For 2 p -balls, B_1 and B_2 , if $p > 1$, $|B_1|$ (the number of items in B_1) $\leq q$ (the number of genes in the population), and $|B_2| \geq q$, the diversity-minimizing population, calculated using the p -norm, consists of all the elements in B_1 plus $q - |B_1|$ of the elements from B_2 . It is not required that $M_1 = M_2$, but the points in B_1 must be a subset of those in B_2 .*

Discussion: Given a putative central point, M_i , the points that minimize the diversity are those that are closest to M_i in the sense that those points minimize the p -norm. This follows from the extension of a Minkowski norm to multiple dimensions. By considering ball-shaped boundaries, we avoid the need to consider the fine geometric structure of the optimal populations. This structure is a tedious extension of the Tetris problem and is enumerated in Figure 8.4 for the 3-5 problem. That the M_i may be anywhere (they need not be a lattice points, halfway between lattice points, etc.) complicates matters somewhat (*e.g.*, the possible values for $|B_1|$ depend on the choice of M_1). The restriction $p > 1$ is required since the 1-norm behaves much differently than all the higher order norms⁹. Nonetheless, this is a powerful hypothesis, especially for larger sets, since the number of connected forms grows rapidly with set size (Table 8.1). An example with the 3-5 problem below demonstrates a case in which the M_i differ.

⁹Specifically, the 1-norm is not affected by tradeoffs among coordinates. For example, using the 1-norm, (1, 1) and (2, 0) are equidistant from the origin, but (1, 1) is closer for all higher order norms.

Hypothesis 4 *For the $j - k$ (j dimension, k point) Tetris problem and for $p > 1$, if there is a p -ball containing exactly k points, the points it contains constitute a set that minimizes the deviation calculated using the p -norm. There may be other sets that give the same norm for a particular choice of p , but no set gives a smaller norm.*

Proof: Follows from Hypothesis 3.

Often, especially for small population sizes, the same set of points gives the minimum variation for several values of p . This is because there is little difference in the shape of the ball (and often no difference in covered lattice points) as p increases.

8.3.3 Examples

We begin with two simple examples to demonstrate the use of the hypotheses. We then apply the hypotheses to finding lower bounds on diversity for the configurations used in the experiments.

Consider the standard Tetris problem and let $p = 2$. If we position the circle with radius $1/\sqrt{2}$ with its center at $(1/2, 1/2)$, it covers exactly 4 points. By Hypothesis 4, these 4 points, which are configured in a 2-by-2 square, constitute the population of size 4 that minimizes the 2-norm variation in 2 dimensions.

Now, let us apply Hypothesis 4 to the 2-5 Tetris problem. A circle with radius 1 and centered on a lattice point will cover 5 points. By Hypothesis 4, the covered points, which form a 3-by-3 cross (S_1), minimize the 2-norm variation in 2 dimensions. Attempting to find a counterexample to Hypothesis 4, we consider a promising potential minimizer, the 2-by-2 square plus a single point (S_2). Indeed, by some measures (of a different class than the Minkowski norms we are considering here), S_2 is denser than S_1 . For example, consider the area of the convex hull. For S_1 , it is 2, while for S_2 , it is only 1.5. However, the minimum radius for S_2 , $\sqrt{5}/2$, is greater than that of S_1 . From the diversity measures in Table 8.2, we see that $\|S_1\|_p \leq \|S_2\|_p$, which is consistent with Hypothesis 3.

For the 3-5 problem, which is the case in Experiment 1, we have the connected sets shown in Figure 8.4.

Considering a simplified version of the general case of our metric, we shift each population form by the mean of the population and compute several norms. Shifting

Table 8.2: Minimum norm values for the 2-dimension, 5-item problem.

p	$\ S_1\ _p$	$\ S_2\ _p$
1	4	5
2	2	$\sqrt{14/5}$
∞	1	1

Figure 8.4: Unique 3-D forms with 5 or fewer nodes. Isomorphisms (reflections and rotations) are considered equivalent and are presented as single nodes. The first three levels correspond to forms with 2, 3, and 4 nodes, respectively. The remaining levels contain all the 5-node forms. Arrows indicate which forms of $N + 1$ nodes can be derived from a form of N nodes by connecting a single additional node. All such possible connections are made. If the graph is consistent, then the set of parents of each form of $N + 1$ nodes contains all the forms of N nodes that result from pruning a single node such that the form remains connected.

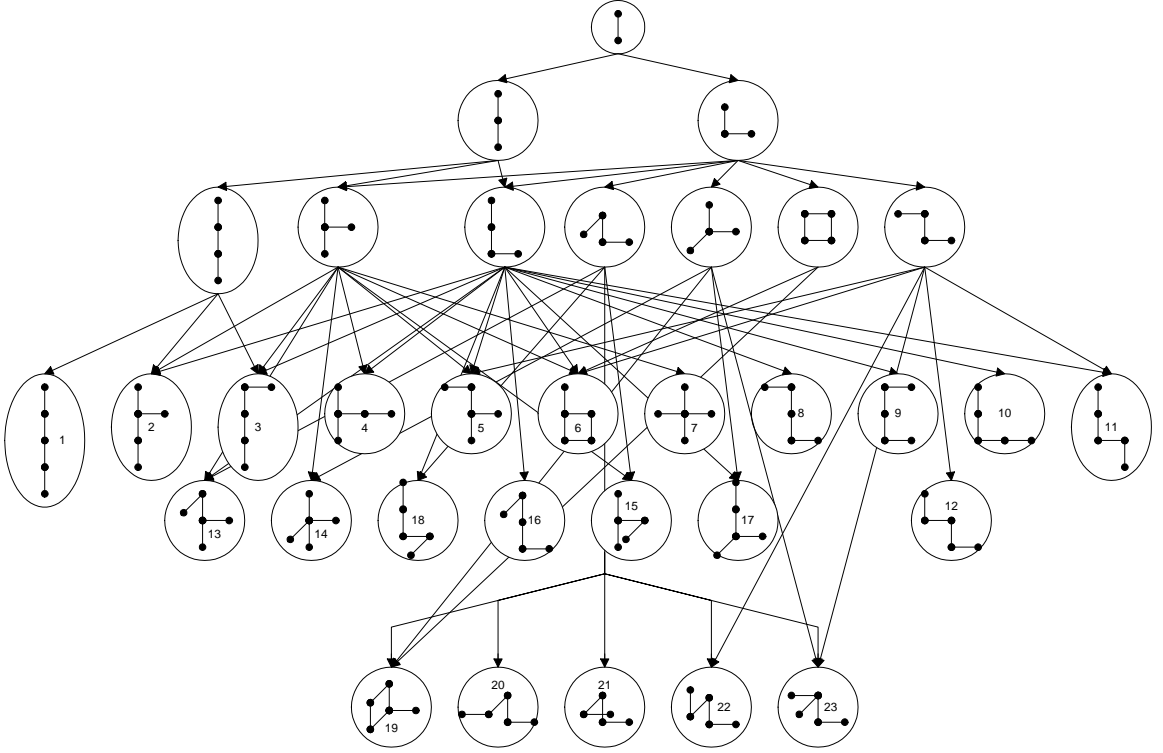


Table 8.3: Forms giving minimum norms as a function of norm order for the 3-D, 5 item case. These are the norms resulting from shifting by the mean, which, except for the 2-norm, is generally not the minimizing shift. There were no ties between forms for minimizing the norms computed thusly.

p	form	norm
1	7	4
2	19	1.7889
3	19	1.2146
4	21	1.0020
5	21	0.8901
6	21	0.8256
\vdots		
∞	21	0.6

by the mean minimizes the 2-norm, but does not necessarily minimize any of the other norms. For brevity, we only show the norms for the minimizing forms (Table 8.3), but it was verified that all the norms satisfy Antsaklis' theorems.

Now, consider the metrics minimized by the optimal choice of central offset. Proving that the error function is unimodal does not seem trivial. For order p , the potential minimizing values for each dimension are roots of a polynomial of degree $p - 1$. Trying to show uniqueness by contradiction (that is, by showing that distinct real roots would result in a contradiction) results in a Vandermonde polynomial matrix problem of order $p - 1$ (*e.g.*, for $p = 4$, we must find solutions to $\Sigma x_i = a_1$, $\Sigma x_i^2 = a_2$ and $\Sigma x_i^3 = a_3$). Nonetheless, we assume unimodality and use Nelder and Mead's simplex method to find the optimal central point. Even without unimodality, a reasonable choice of starting position (we use the population mean) should have a fair chance of being in the neighborhood of the global minimum. The results are presented in Table 8.4.

In order to perform an empirical verification of Hypothesis 4, we find the minimum radii of all forms. There is no obvious direct way to calculate this.¹⁰ Delaunay's triangularization method solves a related problem, finding the maximal triangles with circumcircles containing no points in a given set, but is not readily adaptable to this

¹⁰Halving the maximum interpoint distance works for several of the forms, but is not a general solution; it merely gives a lower bound on the radius. In form 4, for example, the minimum radius is $5/4$, while half of the minimum interpoint distance is $\sqrt{5}/2$.

Table 8.4: Forms giving minimum norms as a function of norm order for the 3-D, 5 item case with an optimal central shift. In some cases there were ties, but form 19 minimized norms of all orders except 1. P_i represents the form with index i .

p	i	$\ P_i\ _p$
1	7 14	4
2	19	$1.7889 = 4/\sqrt{5}$
3	19	$1.1833 = \sqrt{7/5}$
4	19	0.9571
5	19	0.8417
6	19	0.7722
\vdots		
∞	19 21	0.5

problem.¹¹ So, we define a function of a point that gives the minimum radius required for an encompassing 2-ball. We then minimize this function using Nelder and Mead’s method. The minimum radii are presented in Table 8.5.

We see that forms 19 and 21 have the same radius ($r = \sqrt{3}/2$) from the center point $C = (1/2, 1/2, 1/2)$, but form 19 (the 2-by-2 square plus one point in the third dimension) is present in all the minimized norms except the 1-norm. The set of points on the sphere of radius r centered at C contains the 8 points having each ordinate equal to either 0 or 1. Neglecting isomorphisms, there are two 5-point connected subsets; they are forms 19 and 21. Applying Hypothesis 3, we take B_1 to be the sphere of radius $\sqrt{2}/2$ centered at $(1/2, 1/2, 0)$ and B_2 to be the sphere of radius $\sqrt{3}/2$ centered at $(1/2, 1/2, 1/2)$. Then $|B_1| \leq p \leq |B_2|$ ($4 \leq 5 \leq 8$). Also, the

¹¹A related line of reasoning suggests an algorithm that may solve this problem, but, due to its complexity, it is not pursued further herein. . . Three points determine a circle. So, in 2 dimensions, there is a subset of 3 or fewer points that determine a circle that circumscribes the entire set. The “or fewer” is important. If we have two points, for example, we take them as a diameter. Also, “or fewer” addresses the problem of colinearity (consider form 1) and the more general “flattening” problem (consider form 2). In more than 2 dimensions, the generalization is more complex. Recall that in 3-space, 4 points determine a sphere and in M -space, $M + 1$ points determine a ball. Consider 3-space with 4 or more points. Three of the points (assume all 3 are needed to determine the circumcircle for those 3 points – as in form 4) determine a circle whose diameter can be taken as the diameter of the sphere (*i.e.*, the 3 points lie on a great circle). If all the other points are strictly inside this sphere, we have found the circumsphere. This suggests an algorithm: for dimension M , let $i = 2$. Consider all subsets of i points. If one of the subsets defines a circumball, we are done. If not, increment i and try again. A circumball will be found for $i \leq M + 1$. By starting with smaller i , the smallest balls are searched first. The algorithm can be made more efficient by eliminating points from consideration that are not on the convex hull of the overall set.

Table 8.5: Minimum radii for all 5-node 3-D forms.

Form	Minimum radius
1	$2.0000 = 2$
2	$1.5000 = 3/2$
3	$1.5811 = \sqrt{5/2}$
4	$1.2500 = 5/4$
5	$1.1785 = 5/(3\sqrt{2})$
6	$1.1180 = \sqrt{5}/2$
7	$1.0000 = 1$
8	$1.4142 = \sqrt{2}$
9	$1.1180 = \sqrt{5}/2$
10	$1.4142 = \sqrt{2}$
11	$1.5811 = \sqrt{5/2}$
12	$1.4142 = \sqrt{2}$
13	$1.1180 = \sqrt{5}/2$
14	$1.0000 = 1$
15	$1.0607 = 3/(2\sqrt{2})$
16	$1.2247 = \sqrt{3/2}$
17	$1.1832 = \sqrt{7/5}$
18	$1.2247 = \sqrt{3/2}$
19	$0.8660 = \sqrt{3}/2$
20	$1.2247 = \sqrt{3/2}$
21	$0.8660 = \sqrt{3}/2$
22	$1.2247 = \sqrt{3/2}$
23	$1.1280 = ?$

Table 8.6: Minimum variations and radii for two forms containing eight 6-D members (experiment 2 configuration). Note that $p > 1 \rightarrow \|T_1\|_p \leq \|T_2\|_p$, which is likely to be true for a small set.

p	$\ T_1\ _p$	$\ T_2\ _p$
1	12	8
2	2.4494	2.5495
3	1.4422	1.5486
4	1.1066	1.1827
5	0.9440	1.0010
6	0.8491	0.8941
\vdots		
∞	0.5	0.5
r	$\sqrt{3}/2$	1.0076

points in B_1 are subset of those in B_2 , as required. Form 21 does not contain all the points in B_1 , but form 19 does. Thus, Hypothesis 3 correctly predicts that form 19 is the minimizer of the 2-norm variation metric. Form 19 also happens to minimize all variation metrics based on higher order norms; this was likely to happen due to the population size being small and similarity in shapes of the p -balls for the higher order norms. In Table 8.4, form 19 is in all sets (except that of the 1-norm), but for some of the norms, there are additional forms that result in a tie for minimum variation. So, for experiment 1, the minimum 2-norm variation is $4/\sqrt{5} = 1.7889$.

For experiment 2, we have the 6-8 problem. Consider the 2-ball centered at $(1/2, 1/2, 1/2, 0, 0, 0)$ with a radius of $\sqrt{3}/2$. There are only 8 points within this ball. They are on the cube with edge length 1 in the first 3 dimensions that is degenerate in the other 3 dimensions (T_1). Hypothesis 4 tells us that T_1 must be the population that minimizes the variation metric for the 2-norm. For comparison, let us consider another configuration that would likely be put forth as a potential variation minimizer. This configuration (T_2) contains the zero point, the 6 units, and, to complete the set of 8 members, any one sum of two of the units. The minimum variations for several norms and radii for these configurations are shown in Table 8.6. It turns out that T_1 minimizes all norms of order 2 and greater, but T_2 has a lower variation than T_1 for the 1-norm. From the table, the minimum variation for experiment 2 is $2.4494 \dots = \sqrt{6}$.

Together with the simulated and experimental results, these minimum bounds

on variation, $4/\sqrt{5} = 1.7889$ for experiment 1 and $\sqrt{6} = 2.4494$ for experiment 2, indicate how close the GA came to convergence. Knowing the bounds along with the experimental results, we may be able to generalize and develop methods for preventing premature convergence. For example, we might gradually increase the mutation strength when the variation drops to within 3 times the lower bound and gradually decrease it when the variation rises above 6 times the lower bound. That is, we have a rational basis for changing the global/local search bias during a run.

CHAPTER 9

Paired Comparison Scheduling and Rank Inferencing Algorithm

In this chapter, we present an efficient algorithm for estimating the rank order of solutions from a population. An estimate of the rank order is needed so that selection probabilities may be determined for the GA. Our algorithm adaptively determines a series of paired comparisons to present to the user. Research has shown [46] that paired comparisons are reliable and correlate well with aided intelligibility scores. Also, paired comparisons have the added benefit of resulting in a simple user interface. Once the algorithm determines ranks, they are mapped to fitness values using a hypergeometric distribution [§2.3.4]. Given *a priori* partial ranking information, our algorithm requires fewer comparisons to estimate ranks. This is especially useful in our steady-state GA, because, with only one or two new genes each round, much partial ranking data is carried forward.

The algorithm does not handle intransitivities in the perceptual space [11]. Also, equivalence is transitive. Inconsistencies are avoided by not making comparisons when the dominance-equivalence relationship can be inferred from previous responses. Properties of the steady-state GA mitigate the effects of these simplifications. Specifically, only one or two new solutions enter the population and are ranked in each round. Even when the prior rankings are in error, new solutions are ranked correctly relative to the members at the top of the prior estimated rank order, which is where the hypergeometric distribution is most sensitive. In an extreme case, random inputs would drive the system in a random direction, but upon switching to rational decision

making, the incorrectly (randomly) ranked solutions would be pushed towards lower ranks as soon as any newly generated solution was judged superior to the randomly selected top ranked solution. This would eventually lead to removal of the random solutions and ranking information from the population.

The efficiency of our approach comes at the cost of eliminating redundancy. As discussed above, the steady-state GA mitigates this cost, but as the user error rate increases (whether due to simple input error or to difficulty in discerning between similar solutions), the redundancy may in fact be needed. Redundancy could be incorporated into the algorithm such that old and important decisions are occasionally verified, but we do not do so in the current work.

9.1 Background

Runkle, Blommer and Wakefield [55] also used paired comparisons to incorporate user input into a GA. In particular, they used an insertion sort for inferencing. We shall see that the present method requires slightly fewer comparisons and uses heuristic rules to ensure that the most meaningful comparisons are made first, so that truncating the comparison sequence at any point yields the most useful information possible.

A related problem is discussed in the operational research (see [23] for an overview and [2, 32, 35, 47, 51] for extensions to the basic problem) and mathematics applications literature [60]. The basic problem addressed by this literature is the “multicriteria decision problem” in which each alternative is scored on several objective criteria on scales that have monotonic relationships to desirability. Expert decision makers (DMs) then rank exemplary alternatives on a one dimensional scale. These rankings are used to determine a set of weights that may be applied to the criterion scores to arrive at a composite score consistent with the DM’s ranking. The various papers address extensions to this basic model that deal with missing information [60], the use of stochastic dominance¹, and uncertainty [35] in both the probability of events

¹One solution dominates another if it is at least as desirable on all criteria. Stochastic dominance extends this idea to the case in which the properties of the solution specify probability distributions for desirability instead of directly specifying desirability. One probability distribution has first order stochastic dominance over another if the cumulative value of the former is greater than that of

and outcomes due to those events.

Although the operational research literature does not directly apply to our situation — a situation in which the user is trying to optimize subjective qualities, perhaps through implicit tradeoffs that are not reducible to an additive linear model — it does suggest improvements to our approach. For example, Greco *et al.* [23], among others, note that DMs make inconsistent decisions when given the opportunity. The authors note that these inconsistent decisions are not merely errors or noise. Indeed, a perceptual decision model such as the one suggested by Torgerson [61] explains the apparently inconsistent decisions well, but requires a much larger sampling of paired comparisons in order to yield useful results.

Runarsson and Yao [53] considered a ranking procedure based on the bubble sort for use with a genetic algorithm. Their procedure was an objective method based on tradeoffs between searching for the optimum and enforcing design space constraints. Our approach, on the other hand, is to build design space constraints into the genetic encoding method. Because of this, and of their focus on explicit objective functions, their work does not apply to ours.

9.2 Notation

The matrix and graph notation we use is demonstrated in Figure 9.1. The graph is usually the most comfortable representation for people, while the matrix is efficient for implementing the algorithm.

9.3 Algorithm Components

Our algorithm has four components. This first component chooses which pair to present to the user. Then, the second component draws all possible inferences based on the user’s response and previous information. The third component determines whether another comparison is required. If another comparison is not required, the fourth component is invoked and generates ranks for all members of the population.

the latter throughout the domain. Rietveld and Ouwersloot [51] provide a concise introduction to stochastic dominance.

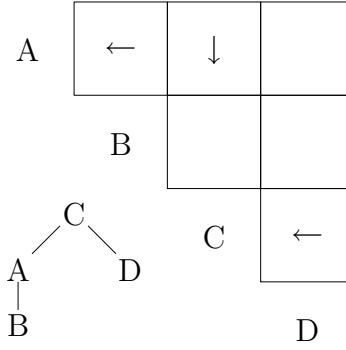


Figure 9.1: Partial dominance matrix and corresponding graph. Each arrow points to the superior node and corresponds to a link in the graph. From either the matrix or the graph, we can infer that C is superior to B. One of the functions of our algorithm is to make such inferences. Since there are $N = 4$ members of the population, there are $\frac{N(N-1)}{2} = 6$ cells in the upper triangular portion of the matrix. In this example, no members are known to be equivalent (although D and A or B might be); equivalence is represented by a 0 in the matrix and by putting symbols in the same node of the graph.

9.3.1 Choosing the Next Pair

Two heuristics are used to choose the next pair to present to the user. First, the graph must be connected to allow generation of complete rank information. The first heuristic addresses this by looking for isolated nodes. If it finds two such nodes, they are presented as a pair to the user. If it only finds one such node, that node is passed to the second heuristic.

Second, many rank-based fitness functions (in particular the hypergeometric distribution, which we use) are more sensitive to relative order among the best solutions than among the worst solutions. So, the second heuristic gives preference to pairs near the top of the graph. It first calculates a score for each node by subtracting the number of arrows pointing away from it from the number of arrows pointing toward it. If the first heuristic did not produce any nodes, the node with the highest score among those with at least one empty cell in the matrix becomes the first node. The second node is the one having the highest score in the set of nodes whose relation to the first is unknown. This heuristic effectively closes (gives complete ordering, up to equivalence) the tree from the top down.

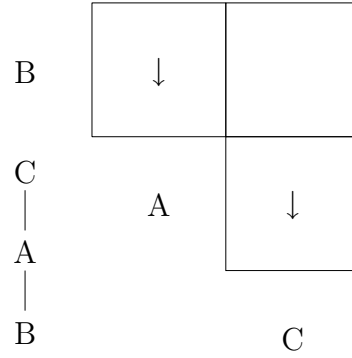


Figure 9.2: Partial dominance submatrix and corresponding subgraph resulting from dropping D and swapping A and B from Figure 9.1. When presented this way, it is clear from the matrix that $C \succ B$ and we may put an arrow in the empty cell pointing to C.

9.3.2 Inferences in the Dominance Matrix

Referring to the graph in Figure 9.1, C is obviously superior to B ($C \succ A \wedge A \succ B \therefore C \succ B$), but this is not plain from the matrix. Our algorithm considers each empty cell (such as BC) and pairs it with all other nodes (A and D in this case). Each of these nodes is shifted between the pair, flipping and moving arrows to maintain the relationships. The resulting matrix and graph for A are shown in Figure 9.2. If the arrows point the same way (or are both 0, giving an equivalence, which is transitive in our algorithm), we place a third arrow of the same type in the empty cell. This comparison is accomplished efficiently in software by negating signs depending on the position of the third node relative to the two nodes related to the cell in question.

The matrix is processed in this manner until a pass is completed with no new inferences. This is necessary since all inferences may not be found in a single pass. For example, if we have only the three cells filled from directly reading $A \succ B \succ C \succ D$, the test for $A \succ D$ fails if it is considered before we infer either $A \succ C$ or $B \succ D$.

9.3.3 Determining Completion

If the dominance matrix contains enough information to generate a complete ranking, with the exception that we do not require the order of the two least fit members, we move on to the next component; otherwise we go back to the first component to

gather more information. After making all inferences as discussed previously, we have a complete order if there are no empty cells in the dominance matrix. If there are two or more empty cells in the matrix, there is not enough information to generate a ranking. If there is exactly one empty cell, we do not have a complete ranking, but there may be enough data to rank all but the two least fit members. This is the case if and only if all arrows in the dominance matrix point away from the empty cell.

9.3.4 Ranks from the Partial Dominance Matrix

Our rank generation algorithm gives ranks from 1 (best) to N (worst). Ranks may be averaged in two cases. The first case is when two solutions have equivalent quality. For example, if there is a two-way tie for second place, both solutions receive the rank $(2 + 3)/2 = 2.5$. The other instance is for unranked solutions at the bottom of the hierarchy.

Our algorithm destructively modifies the dominance matrix, so it may not work with the original matrix if we want to maintain that data.² The algorithm begins by putting zeros, indicating equivalence, in all empty cells of its copy of the dominance matrix. There should be no more than one empty cell since this part of the algorithm is only run when the previous component determines there is enough data to generate ranks. We then initialize three variables:

- u – unassigned set, initially contains the entire population
- b – being assigned set, initially empty
- r – rank counter, initially 1

Then, the following steps are taken repeatedly until u is empty:

- For each element of u that has no arrows in the matrix pointing away from it, move the element to b .
- $m = r + (\|b\| - 1)/2$; m is the pooled rank for members of the set b

²And we do wish to maintain that data. Alternatively, the dominance matrix could be reconstructed from the rank data, with the exception that, if the bottom two members receive the same rank, we do not know whether they were from an equivalence class, or were just unranked solutions at the bottom of the hierarchy.

- $r = r + \|b\|$
- For each element of b , assign the rank m and set the associated rows and columns of the matrix to 0.
- $b = \{\}$

9.4 Comparison with Efficient Sorting Algorithms

The method presented in this chapter is a special instance of a sorting algorithm. It is explicitly designed to never make a redundant comparison and it sometimes neglects sorting the 2 least fit items in the interest of efficiency. Keeping these goals in mind, we compare it with well known efficient sorting algorithms.

Figure 9.3 shows that our method is comparable to quicksort and superior to Shell’s method and heapsort. In this comparison, our method is penalized one full comparison (the worst case) since it sometimes saves a comparison by not determining the relation between the 2 worst stimuli. Also, our method becomes more efficient when equivalences are allowed, but this is not considered in the figure. All trials were repeated 50 times with random input data. For large stimulus sets (much larger than used in our work, and probably too large for most perceptual applications of the GA), our method degrades to a modified insertion sort, performing worse than the efficient algorithms considered here.

Between population updates, our method carries forward knowledge of the order of the surviving elements, but the basic sort algorithms cannot readily take advantage of this information. Some even have their worst case performance ($O(N^2)$) for pre-sorted input. The worst case performance of our algorithm after a population update is $O(N)$, specifically $N - 1$. This occurs when the children are at or very near the bottom of the new order.

Runkle, Blommer and Wakefield [55] used an insertion sort for inferencing from paired comparisons. Quicksort uses insertion sort on its subpopulations, so for populations smaller than the subpopulation size, the efficiency of quicksort is equivalent to that of insertion sort. We use the threshold $M = 7$ as suggested by an authoritative reference on sorting algorithms [48].

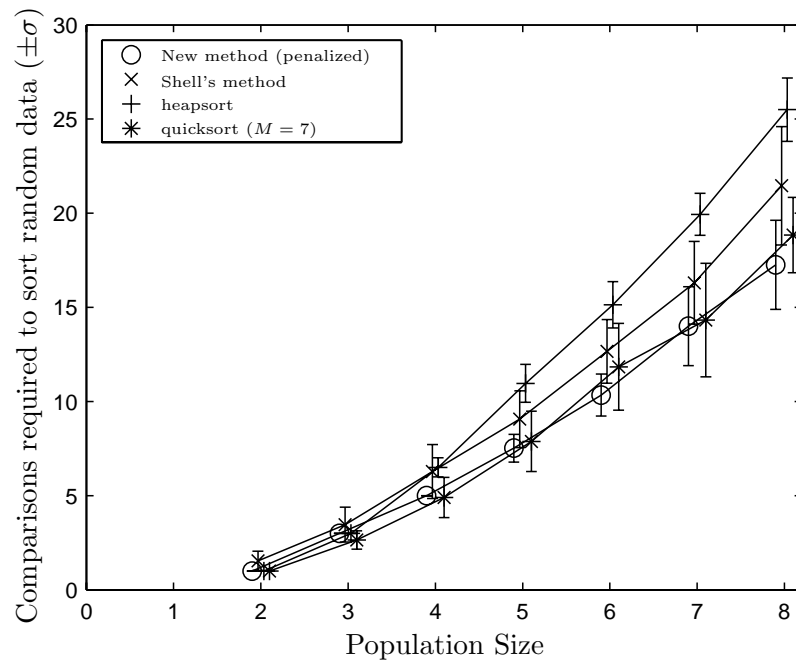


Figure 9.3: Efficiency of sorting algorithm relative to various well known efficient algorithms. Our method has been conservatively penalized by one whole comparison since it does not always sort the 2 least fit members of the population.

CHAPTER 10

Non-binary Operators

In this chapter, we develop crossover and mutation operators to be used with our parametric encoding method [§3.1]. We investigate the shortcomings of using the standard binary crossover and mutation operators on bitstring representations of parametric values and then discuss the operators we use in light of existing research in this area.

When used on parametric values, binary crossover would defeat most of the intent of crossover, which is to combine useful information from two solutions. A binary crossover operator would combine information in the bit space, which often has a tenuous relationship to the parameter space. For example, crossover operators often select a child that is between the two parents in the parameter space, but it is not possible to accomplish this with binary crossover. Specifying that binary crossover of parents having adjacent parametric values must generate a child having one of the two parent values results in the requirement that a Gray code [22] be used. However, a Gray code has the property that values that differ by 2 have encodings that differ in exactly 2 bits. Since two bits differ, selecting the crossover point between the differing bits results in two children distinct from the parents. There is only one encoding between the parents in the parameter space, so it is impossible to choose an encoding such that the goal of all crossovers resulting in children between the parents in the parameter space is achieved.

Binary mutation is ill suited to working with parametric encodings. Due to its structure, it produces children with, to the first order,¹ mutation distance distribu-

¹By first order, we mean mutations of zero or one bits per parameter.

tions with sparse and heavy tails. As we will explain shortly, the mutation strength changes the weight of the tails relative to the probability of no mutation, but does not mitigate the sparseness. Thus, a weak binary mutation operator would have a high probability of causing large parameter jumps, biasing it towards global and effectively random search. This severely limits its usefulness as a local perturbation operator. Hollstien [28] found that a Gray code representation worked slightly better than direct binary encoding for numerical optimization. However, a bit-based operator still seems poorly suited to parametric optimization.

10.1 First Order Analysis of the Binary Mutation Operator

Binary mutation results in binomially distributed probabilities. The binomial distribution is given by $P_k = \binom{n}{k} p^k (1-p)^{n-k}$, where n is the number of trials (bits in a parameter), p is the probability of success (probability of bit change, which is half the mutation strength), and k is the number of successes (bit changes). The zeroth order mutation (no mutation) probability is $P_0 = (1-p)^n$. The first order mutation (1-bit mutation) probability is $P_1 = np(1-p)^{n-1}$. With direct encoding, the 1-bit mutations cause jumps of 2^m in the parameter space, where $m \in [0, n-1]$. Figure 10.1 shows that this distribution is sparse in the parameter space.

In the manner that the exponential distribution is the continuous analog of the geometric distribution, we consider the density function analog of the mass function of parameter jumps resulting from single bit mutations. This density function has the property $f(2x) = f(x)/2$ throughout its domain. This is good in that smaller mutations are more probable than larger ones, but the sparseness of the discrete distribution is nonetheless undesirable.

Neglecting the higher order mutations is acceptable for this analysis. To demonstrate this, consider the probability that any given mutation is not of higher than first order. This is $P = P_0 + P_1 = (1-p)^{n-1}(1-p+np)$ and is given for some common values of n and p in Table 10.1.

We also find evidence for neglecting the multibit changes by looking at the prob-

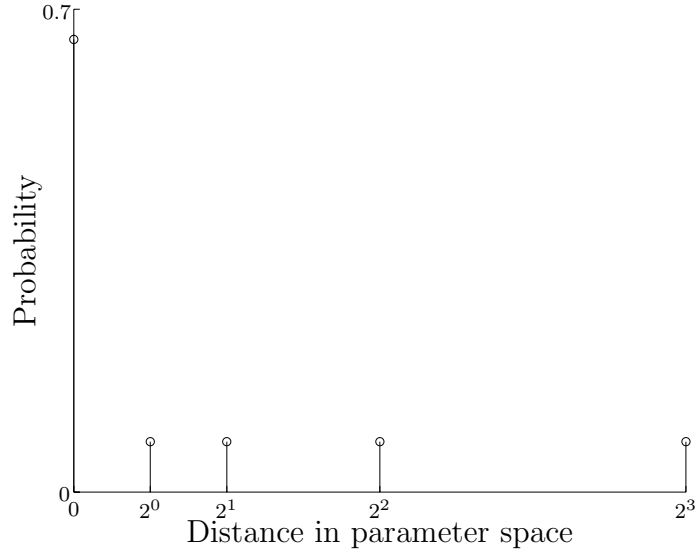


Figure 10.1: Zeroth and first order binary mutation distance probabilities for a gene length (n) of 4 and a bit mutation probability (p) of 0.1. The first order mutations are sparse and have a constant probability. These are undesirable attributes for mutation in an ordered parameter space, so we will design a more appropriate mutation operator.

Table 10.1: Probability that a binary mutation is of order 0 or 1 for various gene lengths (n) and mutation probabilities (p). Since this probability is quite high over the normal GA operating range, it justifies neglecting higher order mutations. For example, [15] used $n \in \{4, 5, 6\}$ and $p \in [0.018, 0.025]$.

$p \setminus n$	3	4	5	6
0.01	0.9997	0.9994	0.9990	0.9985
0.02	0.9988	0.9977	0.9962	0.9943
0.05	0.9928	0.9860	0.9774	0.9672
0.1	0.9720	0.9477	0.9185	0.8857
0.2	0.8960	0.8192	0.7373	0.6554
0.5	0.5000	0.3125	0.1875	0.1094

Table 10.2: Probability that any given mutation is of order 1 for various gene lengths (n) and mutation probabilities (p). As in the previous table, the high probabilities support neglecting the higher order mutations in the analysis.

$p \setminus n$	3	4	5	6
0.01	0.9900	0.9850	0.9800	0.9750
0.02	0.9799	0.9699	0.9600	0.9502
0.05	0.9492	0.9244	0.9001	0.8763
0.1	0.8967	0.8479	0.8011	0.7561
0.2	0.7869	0.6938	0.6092	0.5329
0.5	0.4286	0.2667	0.1613	0.0952

ability that when a mutation occurs, it is of order 1. The probability of this is $\frac{P_1}{1-P_0} = \frac{np(1-p)^{n-1}}{1-(1-p)^n}$ and is given in Table 10.2 for the same common values of n and p . These probabilities are very high over the range of interest, indicating that the 1-bit mutations account for most of the behavior of the binary mutation operator

10.2 New Operators

Our new operators work on parameter indexes. A parameter index may take on any integer value within some range. Operating on either parameters or indexes into parameter sets was suggested by Davis [12] and other researchers.

Michalewicz *et al.* [39] list three kinds of crossover operators commonly used in parametric optimization problems and proposes a fourth:

1. uniform crossover — each child parameter value is randomly taken from one of the parents; the sibling takes its values from the opposite parents; uniform crossover is the generalization of 1-point, 2-point, and multi-point crossover
2. arithmetical crossover — each child lies on the line segment connecting the parents in the parameter space; the children are equidistant from the point halfway between the parents
3. heuristic crossover — inspired by hill climbing, a child is on the line connecting the parents in the parameter space, but is not between the parents; the child is closer to the parent having greater fitness

4. geometric crossover — children are between the parents in a manifold of lower dimension than the parameter space; this manifold need not be linear (for example, the manifold could be an ellipsoid and the space could be a 3-D linear vector space); often, the manifold is the feasible region of the parameter space

We note that in many cases the complexity of geometric crossover can be eliminated by judicious representation of the parameter space. For example, assume a 3-D linear vector space in which only the surface of some ellipsoid is feasible. An azimuth and elevation relative to a given point inside the ellipsoid form a 2-D representation of a single point on the ellipsoid. Further, any azimuth and elevation specify a feasible point. Contrast this to the natural 3-D representation in which each coordinate constrains, and is constrained by, all the other coordinates. In this spirit, we used a simplified parameter representation in Chapter 7, in which the 2-D feasible region of a 3-D gain space was represented by 2 parameters. Ultimately, these 2 parameters were mapped to 3 hearing aid parameters, but, for crossover and mutation purposes, there were only 2 parameters.

Our crossover operator, demonstrated in Figure 10.2, is a variation on the 1-point crossover operator from which Michalewicz's uniform crossover operator is generalized. Our operator chooses a parameter position using a uniform distribution. The first child inherits the parameters of the first parent to the left of this position and the parameters of the second parent to the right of it. The second child inherits the opposite parameters. The values at the crossover position are chosen from the range of the value in the parent genes. A geometric distribution is used to determine how far the child values are from the parent values. The rate of the geometric distribution is a system parameter multiplied by the number of levels that the particular parameter may take. This distribution is truncated at the difference of the parent parameters to prevent the child values from going outside the parent parameter range.

Our mutation operator, demonstrated in Figure 10.3, perturbs each parameter by the value of a normal random variable with standard deviation equal to a system parameter multiplied by the number of levels for that parameter, subject to range constraints.

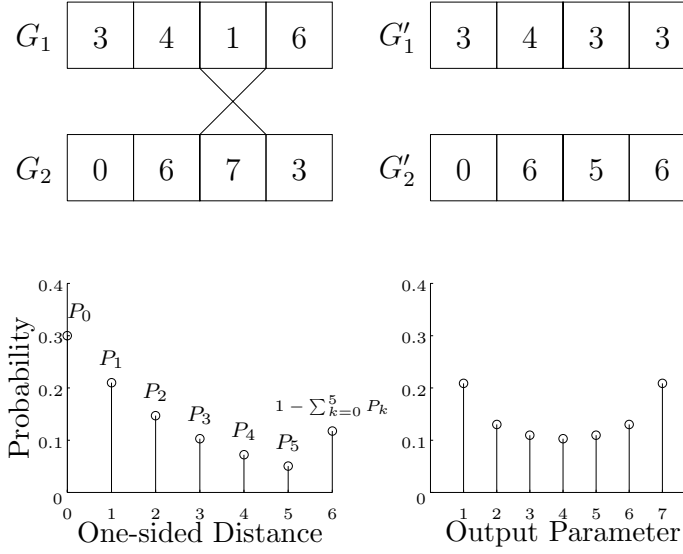


Figure 10.2: New crossover operator. G_1 and G_2 are the parent genes. The third position is chosen for crossover from a uniform distribution ($x = 3$). The child genes, G'_1 and G'_2 , get the parameters of their corresponding parent to the left of x , and the parameters of the opposite parent to the right of x . At x , a geometric distribution is constructed (lower left). The distribution is truncated, giving the maximum distance all remaining probability. Then the distribution is symmetrized and shifted to the parameter range (lower right). A parameter is chosen from this distribution for the first gene ($G_{1,x} = 3$ here). The symmetric value is chosen for the parameter in the other gene ($G'_{2,x} = G_{1,x} + G_{2,x} - G'_{1,x} = 1 + 7 - 3 = 5$).

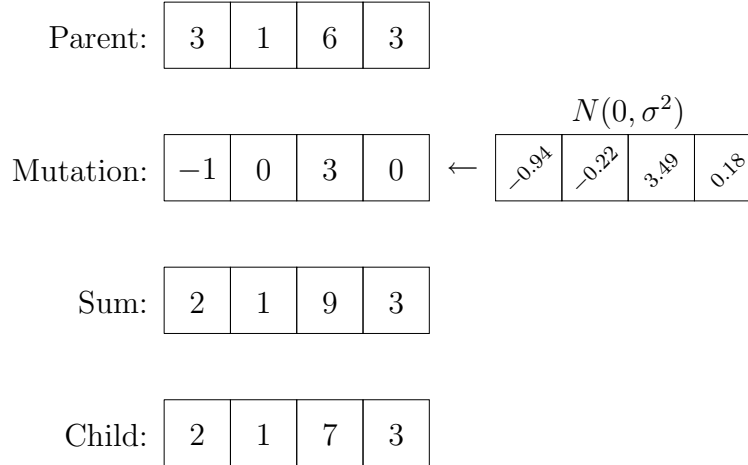


Figure 10.3: New mutation operator. The standard deviation of each element of the mutation is the product of a system constant ($s = 0.2$) and the number of levels in a codeword (8). Here we have $\sigma = 0.2 \cdot 8 = 1.6$. The rounded mutation is added to the parent gene and then the sum is range limited to produce a valid child.

10.3 Summary

In this chapter, we considered the intended functions of mutation and crossover operators in light of our parametric encoding method. We found that the commonly used operators, which manipulate genes at the bit level, do not perform these functions well when used with parametric values. After reviewing the literature on genetic operators, we designed crossover and mutation operators that take advantage of parametric structure. These operators were used successfully in the experiments presented earlier [Chapters 4, 5, and 7].

CHAPTER 11

Summary and Conclusions

This thesis brings together the fields of hearing aid fitting and genetic algorithm research. We addressed issues necessary to adapt the GA for hearing aid fitting, including a simple method of gathering input from subjects, measuring convergence, and evolving hearing aid parameters. We then conducted experiments in which we applied our method to fitting two sets of hearing aid parameters.

In the first experiment, we showed that subjects can use our GA fitting system to find good, repeatable acoustic feedback canceller settings. This conclusion is supported by a combination of ensemble performance data, gain margin measurements, and follow-up repeatability comparisons. We also found that subjects had a wide variety of preferred settings. This confirms that, in addition to objective performance measures, idiosyncratic preferences are a necessary component of evaluating canceller performance.

In the second experiment, we showed that subjects can use the GA system to successfully fit multiband dynamic range expansion parameters for suppression of unwanted low level sounds. The data for this experiment did not speak as strongly as those for the feedback canceller experiment. This was due to an increase in complexity (six parameters in contrast to three for feedback canceller fitting) and oversampling of the parameter space. Nonetheless, the combination of the survey results and the repeatability analysis speak to the success of the method.

Also, we presented several aspects of our fitting system in detail. We proposed a convergence metric appropriate for populations of solutions made up of parametric values and determined its lower bound for the two experiments conducted. We then

presented our efficient paired comparison inferencing method, which enables dominance judgments from subjects to be used by the GA. Finally, we reviewed the purpose of genetic operators and constructed crossover and mutation operators appropriate for parametric optimization.

The simulated results of varying the mutation strength highlighted an important tradeoff that is made in any search algorithm — tuning the search to provide an improvement for one class of problems will yield a decrease in performance for another class of problems [69]. So, as we did by choosing a parametric representation instead of a bitstring representation, we must be careful that any changes made to increase performance have more benefits than costs in the domain of interest, which is, in our case, hearing aid fitting.

The experimental results suggest some areas of future work. First, an on-line method of adapting the coarseness of the parameter tilings in response to observed discrimination ability could lessen the waste of user input that we observed in the latter rounds of our expansion experiment. Also, enhancing our inferencing method to use some redundancy might give better performance for the more error-prone subjects.

Our experimental focus has been on *in-field* optimization of modern DSP-based hearing aid fitting in a variety of environments. This work also has great potential as a tool for clinical use by the audiologist for refining a prescription for an individual's preferences and making efficient and effective use of new capabilities that may not have been part of the original rationale, but that interact with its prescribed parameters.

More generally, our goal has been to improve the arsenal used to attack perceptual tuning problems. GAs have successfully addressed many aspects of such problems in a variety of applications, but the great majority of successes until now have been in objective domains. Thus, our theoretical focus has been adapting the GA to perceptual tuning problems by using knowledge of perception and person-machine interfaces.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] P. J. Antsaklis and A. N. Michel, *Linear Systems*, McGraw-Hill series in electrical and computer engineering. Control theory, McGraw-Hill, New York, 1997.
- [2] S. H. Azondékon and J.-M. Martel, “‘Value’ of additional information in multi-criterion analysis under uncertainty,” *European Journal of Operational Research*, vol. 117, no. 1, pp. 45–62, 1999.
- [3] I. Borg, *Modern multidimensional scaling: theory and applications*, Springer, New York, 1997.
- [4] G. E. P. Box and K. B. Wilson, “On the experimental attainment of optimum conditions,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 13, no. 1, pp. 1–45, 1951.
- [5] D. Byrne and G. Walker, “The effects of multichannel compression and expansion on perceived quality of speech,” *Aust. J. Audiol.*, vol. 4, pp. 1–8, 1982.
- [6] D. Byrne, “Hearing aid selection for the 1990s: where to?,” *Journal of the American Academy of Audiology (JAAA)*, vol. 7, no. 6, pp. 377–395, 1996.
- [7] D. Byrne, “Now that hearing aids can do almost anything, what should they do to be really helpful?,” in *Hearing Aid Amplification for the New Millennium*, Sydney, Australia, 1999.
- [8] D. Byrne and H. Dillon, “The National Acoustic Laboratories’ (NAL) new procedure for selecting the gain and frequency response of a hearing aid,” *Ear and Hearing*, vol. 7, no. 4, pp. 257–265, 1986.
- [9] D. Byrne, *et al.*, “An international comparison of long-term average speech spectra,” *Journal of the Acoustical Society of America (JASA)*, vol. 96, no. 4, pp. 2108–2120, 1994.
- [10] D. Byrne and F. K. Kuk, “Paired comparison procedures in hearing aid evaluations,” *Ear and Hearing*, vol. 15, no. 6, pp. 476–479, 1994.
- [11] C. H. Coombs, R. M. Dawes, and A. Tversky, *Mathematical Psychology: An Elementary Introduction*, Prentice-Hall, Englewood Cliffs, N.J., 1970.

- [12] L. Davis, *Handbook of genetic algorithms*, Van Nostrand Reinhold, New York, 1991.
- [13] H. Dillon, "Tutorial: Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times?," *Ear and Hearing*, vol. 17, no. 4, pp. 287–307, 1996.
- [14] H. Dillon, "NAL-NL1: A new procedure for fitting non-linear hearing aids," *Hearing Journal*, vol. 52, no. 4, pp. 10–16, 1999.
- [15] E. A. Durant and G. H. Wakefield, "Efficient model fitting using a genetic algorithm: Pole-zero approximations of HRTFs," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 1, pp. 18–27, 2002.
- [16] L. S. Eisenberg, D. D. Dirks, S. Takayanagi, and A. S. Martinez, "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 2, pp. 327–339, 1998.
- [17] D. A. Fabry and D. J. VanTasell, "Evaluation of an articulation-index based model for predicting the effects of adaptive frequency response hearing aids," *Journal of Speech and Hearing Research*, vol. 33, pp. 676–89, 1990.
- [18] D. B. Fogel, "An introduction to simulated evolutionary optimization," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 3–14, 1994.
- [19] D. B. Fogel, "What is evolutionary computation?," *IEEE Spectrum*, vol. 37, no. 2, pp. 26–32, February 2000.
- [20] M. Gen and R. Cheng, *Genetic algorithms and engineering design*, Wiley, New York, 1997.
- [21] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
- [22] F. Gray, "Pulse Code Communication," U.S. Patent no. 2632058, Mar. 1953.
- [23] S. Greco, B. Matarazzo, and R. Slowinski, "Rough sets theory for multicriteria decision analysis," *European Journal of Operational Research*, vol. 129, no. 1, pp. 1–47, 2001.
- [24] R. L. Haupt and S. E. Haupt, *Practical genetic algorithms*, Wiley, New York, 1998.
- [25] R. L. Haupt and S. E. Haupt, "The creative use of genetic algorithms: Computers evolve into the artistic realm," *IEEE Potentials*, vol. 19, no. 2 (supplement), pp. 26–29, April/May 2000.

- [26] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, University of Michigan Press, Ann Arbor, 1975.
- [27] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, pp. 66–72, July 1992.
- [28] R. B. Hollstein, *Artificial genetic adaptation in computer control systems*, PhD thesis, The University of Michigan, 1971.
- [29] H. A. L. Joon, F. Asano, Y. Suzuki, and T. Sone, "Adaptive feedback cancellation with frequency compression for hearing aids," *Journal of the Acoustical Society of America (JASA)*, vol. 94, no. 6, pp. 3248–3254, 1993.
- [30] J. M. Kates, "Feedback cancellation in hearing aids: results from a computer simulation," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 553–562, 1991.
- [31] M. Kawamata, J. Imakubo, and T. Higuchi, "Optimal design method of 2-D IIR digital filters based on a simple genetic algorithm," in *1994 IEEE International Conference on Image Processing (ICIP-94)*, volume 1, pp. 780–784, Department of Syst. Inf. Sci., Tohoku University, Sendai, Japan, 1994, IEEE.
- [32] C. W. Kirkwood and R. K. Sarin, "Ranking with partial information: A method and an application," *Operations Research*, vol. 33, no. 1, pp. 38–48, 1985.
- [33] F. K. Kuk and N. M. C. Pape, "The reliability of a modified simplex procedure in hearing aid frequency-response selection," *Journal of Speech and Hearing Research (JSHR)*, vol. 35, no. 2, pp. 418–429, 1992.
- [34] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *Society of Industrial and Applied Mathematics (SIAM) Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [35] A. Langewisch and F. Choobineh, "Stochastic dominance tests for ranking alternatives under ambiguity," *European Journal of Operational Research*, vol. 95, no. 1, pp. 139–154, 1996.
- [36] K.-S. Leung, Q.-H. Duan, Z.-B. Xu, and C. K. Wong, "A new model of simulated evolutionary computation— convergence analysis and specifications," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 1, pp. 3–16, 2001.
- [37] J. A. Maxwell and P. M. Zurek, "Reducing acoustic feedback in hearing aids," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 304–313, 1995.

- [38] Z. Michalewicz, T. D. Logan, and S. Swaminathan, "Evolutionary operators for continuous convex parameter spaces," in *3rd Annual Conference on Evolutionary Programming*, A. V. Sebald and L. J. Fogel, editors, pp. 84–97. World Scientific Publishing, 1994.
- [39] Z. Michalewicz, G. Nazhiyath, and M. Michalewicz, "A note on usefulness of geometrical crossover for numerical optimization problems," in *5th Annual Conference on Evolutionary Programming*, pp. 305–312, San Diego, CA, 1996, MIT Press.
- [40] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin, second edition, 1994.
- [41] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA, 1996.
- [42] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 1997.
- [43] T. Nagao, "Homogeneous coding for genetic algorithm based parameter optimization," in *IEEE International Conference on Evolutionary Computation*, volume 1, pp. 426–430, Nagoya, Japan, 1996, IEEE.
- [44] A. C. Neuman, M. H. Bakke, S. Hellman, and H. Levitt, "Effect of compression ratio in a slow-acting compression hearing aid: Paired-comparison judgments of quality," *Journal of the Acoustical Society of America (JASA)*, vol. 96, no. 3, pp. 1471–1478, 1994.
- [45] A. C. Neuman, M. H. Bakke, C. Mackersie, S. Hellman, and H. Levitt, "The effect of compression ratio and release time on the categorical rating of sound quality," *Journal of the Acoustical Society of America (JASA)*, vol. 103, no. 5, pp. 2273–2281, 1998.
- [46] A. C. Neuman, H. Levitt, R. Mills, and T. Schwander, "An evaluation of three adaptive hearing aid selection strategies," *Journal of the Acoustical Society of America (JASA)*, vol. 82, pp. 1967–1976, 1987.
- [47] K. S. Park and S. H. Kim, "Tools for interactive multiattribute decisionmaking with incompletely identified information," *European Journal of Operational Research*, vol. 98, no. 1, pp. 111–1123, 1997.
- [48] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, second edition, 1992.
- [49] J. Radecki, J. Konrad, and E. Dubois, "Design of multidimensional finite-wordlength FIR and IIR filters by simulated annealing," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 42, no. 6, pp. 424–431, 1995.

- [50] B. Rafaely and M. Roccasalva-Firenze, "Control of feedback in hearing aids—A robust filter design approach," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 754–756, 2000.
- [51] P. Rietveld and H. Ouwersloot, "Ordinal data in multicriteria decision making, a stochastic dominance approach to siting nuclear power plants," *European Journal of Operational Research*, vol. 56, no. 2, pp. 249–262, 1992.
- [52] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 96–101, 1994.
- [53] T. P. Runarsson and X. Yao, "Stochastic ranking for constrained evolutionary optimization," *IEEE Transactions on Evolutionary Computation*, vol. 4, no. 3, pp. 284–94, 2000.
- [54] P. R. Runkle, *Optimization of Head-Related Transfer Functions Using Subjective Criteria*, PhD thesis, The University of Michigan, 2000.
- [55] P. R. Runkle, M. A. Blommer, and G. H. Wakefield, "Comparison of head related transfer function interpolation methods," *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, p. 4, 1995.
- [56] M. G. Siqueira, *et al.*, "Subband adaptive filtering applied to acoustic feedback reduction in hearing aids," in *Thirtieth Asilomar Conference on Signals, Systems and Computers*, volume 1, pp. 788–792, Pacific Grove, CA, USA, 1996, IEEE.
- [57] K. E. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Transactions on Automatic Control*, vol. 10, pp. 461–464, 1965.
- [58] A. D. Sterian, P. R. Runkle, and G. H. Wakefield, "Active sensory tuning of wind-noise using a genetic algorithm," in *Proceedings of the 1995 20th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pp. 2967–2970, Detroit, MI, USA, 1995, IEEE.
- [59] H. Takagi and M. Ohsaki, "IEC-based hearing aid fitting," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC '99)*, volume 3, pp. 657–662, IEEE, 1999.
- [60] E. Takeda and J. Satoh, "A data envelopment analysis approach to multicriteria decision problems with incomplete information," *Computers and Mathematics with Applications*, vol. 39, no. 9–10, pp. 81–90, 2000.
- [61] W. S. Torgerson, *Theory and methods of scaling*, Wiley, New York, 1958.
- [62] C. W. Turner and R. R. Hurtig, "Proportional frequency compression of speech for listeners with sensorineural hearing loss," *Journal of the Acoustical Society of America (JASA)*, vol. 106, no. 2, pp. 877–886, 1999.

- [63] R. A. van Buuren, J. M. Festen, and T. Houtgast, "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," *Journal of the Acoustical Society of America (JASA)*, vol. 105, no. 5, pp. 2903–2913, 1999.
- [64] F. Vavak and T. C. Fogarty, "Comparison of steady state and generational genetic algorithms for use in nonstationary environments," in *IEEE International Conference on Evolutionary Computation (ICEC)*, pp. 192–195, Nagoya, Japan, 1996, IEEE, New York, NY, USA.
- [65] M. Velmans and M. Marcuson, "The acceptability of spectrum-preserving and spectrum-destroying transposition to severely hearing-impaired listeners," *British Journal of Audiology (BJA)*, vol. 17, pp. 17–26, 1983.
- [66] G. H. Wakefield, A. D. Sterian, and D. Everstine, "Psychophysics of windnoise and sound quality engineering," *Acoustique Canadienne (Canadian Acoustics)*, vol. 25, no. 3, pp. 31, 1997.
- [67] G. Walker, D. Byrne, and H. Dillon, "The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise," *Journal of the Acoustical Society of America (JASA)*, vol. 76, no. 3, pp. 746–57, 1984.
- [68] R. Wang and R. Harjani, "Acoustic feedback cancellation in hearing aids," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-93)*, volume 1, pp. 137–140, Minneapolis, MN, USA, 1993, IEEE.
- [69] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [70] A. M. S. Zalzala and P. J. Fleming, *Genetic algorithms in engineering systems*, Institution of Electrical Engineers, London, 1997.

Index

- adaptive methods, 4
 - simplex, 11, 90
- AI, *see* articulation index
- articulation index, 27
- audiogram, 8
- automatic volume control, 10
- ball, 84
- bias, decision, 42, 73
- binomial distribution, 104
- Blommer, Michael Alan, 21, 96, 101
- Byrne, Denis, 9, 10, 61
- child, 13, 103
- compact set, 84
- competition, 13
- compression, 5
 - illustration, 6
 - kneepoints, 62
 - patient preferences, 61
 - rationales, 2
 - representing thresholds, 62
 - uses, 9
- connected set, 84
- convergence, 77, *see* genetic algorithm,
 - convergence
- crossover, *see* operator, crossover
- Davis, Lawrence, 22
- decode, 13
- density, 84
- Dillon, Harvey, 9, 61
- disjoint, 84
- diversity, 77, *see* variation
- encoding, 14, 15, 20, 23
 - direct, 13
 - homogeneous, 23
- equalizer, 20
- expansion, 60, 61
- feedback
 - acoustic, 5
 - cancellation, 32, 44
 - triggers, 33
 - subjective, 1
- fitness, 14
 - from paired comparisons, 95
 - function, 14
- fitting, 8
- frequency transposition, 2
- gain margin, 27, 41, 44
- gene, 14
- generational, *see* update, generational
- genetic algorithm
 - convergence, 26

- overview, 12
- performance statistics, 28
- previous applications, 1, 12, 15–18, 20, 22, 23, 77, 78, 96, 101, 104–106, 111
- genotype, 14
- Gray code, 103, 104
- head-related transfer function, 21
- hearing aids, 4
 - gain, 4
 - maximum output, 5
- Holland, John H., 12, 15
- HRTF, *see* head-related transfer function
- hypergeometric distribution, 16, 17, 24, 95, 98
- JND, *see* just noticeable difference
- Kawamata, M., 12
- loudness recruitment, 5
 - compensation, 10
- Macready, William G., 22
- MCL, *see* most comfortable levels
- Michalewicz, Zbigniew, 12, 106
- Minkowski norms, 15, 85–88
 - multiple dimensions, 87
 - theorems, 86
- modulation depth, 61
- most comfortable levels, 8
- mutation, *see* operator, mutation
- Nagao, Tomoharu, 22, 23, 78
- NAL-NL1, 10
- Neuman, Arlene C., 11
- observer models, 28
- operator
 - conceptual view, 20
 - crossover, 13
 - examples, 19
 - mutation, 14, 103
 - binary, 104
 - strength, 46, 104
 - reproduction, 14, 18
- paired comparisons, 21, 23, 24, 58, 96, 101
 - inferencing from, 95
 - reliability, 95
 - sound quality, 61
- parameters
 - controlling GA, 46
 - examples, 5
 - expansion fitting, 62
 - values, 63
 - feedback cancellation, 32
 - preferred, 54
 - values, 33
 - model based on, 22
 - scaling, 29
 - vs. graph theory, 15
- parent, 14
- phenotype, 14
- population, 14
 - initial, 24, 34, 42, 66

- updating, 18
- prescriptive parameters, 4
- presentation order
 - 5 alternative optimal, 57
 - and decision bias, 42
 - of fixed alternatives, 55
- rank order, *see* sorting
- rationales, 1, 4
 - adaptive, 11
 - conflicting, 2
 - prescriptive, 8
 - compressive, 9
 - linear, 9
- reproduction, 14
- response surface, 11, 22
- Runkle, Paul R., 16, 20, 21, 96, 101
- scaling
 - parameters, 29
 - perceptual, 21, 23, 62
 - similarity, 55
- selection probability, 14
- set, 85
 - compact, 84
 - connected, 84
 - disjoint, 84
- SII, *see* speech intelligibility index
- simulated annealing, 13
- sorting, 21, 96, 101
- sound pressure level, 5
- speech intelligibility, 27
- speech intelligibility index, 27
- SPL, *see* sound pressure level
- Starkey, iii, 26, 32
- steady-state, *see* update, steady-state
- Sterian, Andrew, 20
- stochastic dominance, 96
- subjects
 - consistency, 57
 - gain margin, 44
 - hearing aid settings, 62
 - instructions, 25, 33, 35, 55
 - repeatability, 73
 - simulated, 28
 - survey, 65
- Takagi, Hideyuki, 21
- Tetris problem, 85
 - generalized, 84
- UCL, *see* uncomfortable levels
- uncomfortable levels, 8
- update
 - generational, 14, 18, 79
 - steady-state, 14, 18, 21, 79, 95
- variation, 77
 - bitstrings, 79
 - multilevel, 82
 - stopping criteria, 25
- Wakefield, Gregory H., 20, 21, 96, 101
- windnoise, 20
- Wolpert, David H., 22