

Psychometric Augmentation of an Interactive Genetic Algorithm for Optimizing Cochlear Implant Programs

Sean K. R. Lineaweaver
Cochlear Ltd.
13059 East Peakview Ave.
Centennial, CO 80111
slineaweaver@cochlear.com

Gregory H. Wakefield
Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, MI 48109
ghw@umich.edu

ABSTRACT

As suggested in the *Blind Watchmaker*, human selection can be a remarkable source of information for guiding a genetic algorithm when objective cost functions are unknown [2]. Properly harnessing such input, however, requires an understanding of the “numbers” humans produce as well as the limitations humans face when performing extensive judgment tasks. The Interactive Augmented Genetic Algorithm (IAGA) modifies both the procedural and algorithmic components of Interactive Genetic Algorithms to better match the human-selection process. Experimental results show that cochlear implant recipients are successful in using the IAGA to select processing parameters to improve their perception of music.

Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems—*human information processing, software psychology*

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Genetic Algorithm, IGA, Cochlear Implants, Music, Psychometric methods

1. INTRODUCTION

Interactive Genetic Algorithms (IGAs) explicitly incorporate human feedback into the evolutionary process of the genetic algorithm (GA). In its simplest form, the IGA uses a human-generated goodness of fit in place of some objective function. Population members may either be rank ordered or scaled by the user, based on his or her own subjective criteria. In turn, based on such values, the population is evolved within a standard GA framework.

Since their inception, IGAs have been applied in a variety of design problems for which optimizing human preference is

a paramount, but poorly quantifiable, objective [15]. These include designs where visual and/or auditory qualities determine the utility of the artefact, e.g., visual design systems [1] or hearing aids [3]. Although each of these applications has been deemed a success by their authors, the extent to which each implementation made full use of the human input has not been studied, nor has the degree to which errors or biases in judgement affect the performance of the IGA. Until recently, even the question of whether an IGA search converges to the same member of a population as would be selected using more standard psychometric techniques has been left unanswered [7].

The present paper proposes modifications to the standard components of a GA, which result in a framework that is more consistent with psychometric theory and practice. Specifically, modifications with respect to *representation, generation, selection, crossover, mutation, stopping criteria* and *initialization* are introduced. These modifications define an Interactive Augmented Genetic Algorithm (IAGA). In the remainder of the present section, we present the key components of the IAGA. In Section 2, we introduce the IAGA as a subjective method for improving the perception of sound by the severely hearing-impaired implanted with a cochlear prosthesis and present results from a human study. Finally, in Section 3, we revisit the general components of the IAGA and consider the extent to which several of the proposed modifications were utilized by human participants in the cochlear-prosthetic design problem.

1.1 Psychometric Limitations on IGA Designs

At the core of any IGA is a measure of the human “designer’s” preference. Psychometric techniques for obtaining such preference judgements (e.g., scaling or ranking) impose certain limits on general properties of an IGA. From the standpoint of *memory load*, the number of stimuli a human participant can judge at any one time is typically bounded by the “ 7 ± 1 ” rule. While it is possible to exceed this bound, in practice, the experimenter must provide some means for the participant to compare among options as part of the response procedure. From the standpoint of *task load*, human participants often fatigue after 1-2 hrs. of testing, which implicitly limits the number of generations in a run of the IGA and favors testing procedures that don’t allow the participant the option of reviewing or comparing members of the current generation (see also the discussion in [6]). In the types of audio design problems that the authors are most familiar with, a human participant is likely to complete 2-3 runs of an IGA in the course of one two hour session. When compared with the more standard practice of 1000s of runs

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

GECCO’11, July 12–16, 2011, Dublin, Ireland.

Copyright 2011 ACM 978-1-4503-0557-0/11/07 ...\$10.00.

in GA evaluations, the very limited number of repeated measures afforded in an IGA can impact the quality of the search results.

Finally, from the standpoint of *stimulus variation*, humans are much more likely to handle heavier memory and task loads if the artefacts they are evaluating have a sufficient degree of variability. This less-formalized concept in psychometrics reflects the differences in performance observed, for example, when running fixed-level vs. adaptive psychophysical methods. Fixed-level methods, in which the same stimulus condition is repeated for 50-100 trials, invariably suffer from lags in attention, either because the discrimination or detection task is too difficult (performance is near chance) or because it is too easy (performance is nearly perfect). In contrast, sequential adaptive estimation methods, such as that originally proposed by Levitt [8], are better able to sustain the observer's attention over 50-100 trials by varying the stimulus condition from trial to trial. This desire for a procedure with a sufficient degree of stimulus variation runs counter to the desire for homogeneity within a current generation as an IGA-run evolves. Stimulus variation also is known to be important in "teaching" the participant to discriminate among those properties of the stimulus which are relevant to the task from those which are not. Accordingly, as the population homogenizes over an IGA run, participants are more likely to attend to those stimulus properties that make the scoring task manageable, rather than those that are indicative of potentially better variations. This is particularly a problem when the participant may not really know, at the outset of a run, what stimulus properties they prefer and only learn these from the generated exemplars [7].

Whereas the three factors above reflect the insertion of the human in the "feedback loop" of a GA, psychometric theory also points to the inherent limitations of how data generated by the measurements themselves can be interpreted. In any sensory scaling task, the experimenter must decide whether the data should be treated on a categorical, ordinal, interval or ratio scale [14]. In general, without additional assumptions or more elaborate psychometric procedures, it is recognized that preference scores provided by a human participant are no stronger than ordinal. Should the rules of parent selection assume the figure of merit to be drawn from interval or ratio scales, as is typically the case in GAs, then they must be adjusted to accommodate the weaker ordinal or even categorical nature of the data the participant provides.

The consequences of the psychophysical limitations to the implementation of an IGA are that smaller, as opposed to larger, search spaces are likely to yield valid results, and mechanics of the selection, cross-over, and mutation processes should be scrutinized to ensure that subjects provide reliable data that is uncontaminated by fatigue, inattentiveness, and bias. In the work that follows, we consider IGA designs drawn from a narrow niche of the broader GA search space. We limit our discussion to improved search methods involving relatively small spaces (as might be spanned by 10-14 bit representations) which are robust to highly quantized fidelity criteria. In the latter case, we consider a binary response (accept/don't accept) as the strongest feedback a participant can provide.

1.2 Representation

Given the need for small search spaces over which to search, the mapping of designs to genetic representation takes on greater significance than is typically the case for GA. Binary representations inflate the size of the search space. More efficient representations are M-length strings

$$\alpha \in A_1 \times A_2 \times \dots \times A_M$$

where each A_k is a finite alphabet of size N_k and M typically corresponds to the number of (physical) parameters that define a given design. In our typical application, M is often no greater than 5 and the number of levels of each parameter, N_k , is often no greater than 8.

1.3 Generation

The best psychometric methods are those that make the most from the fewest number of observations. Applying this principle to the IAGA, duplicates within a generation are wasteful observations, despite the fact that increasing homogeneity is a desired outcome of an IGA run. The IAGA modifies the standard form of generational updating through *culling*, *tagging*, and *selective insertion*. The process of *culling* removes duplicates from the next generation. Each member that remains is *tagged* with the number of copies that were removed. In place of the duplicates, unique members are inserted into the generation. The rules for such insertion are variable. In our work, we consider rules that draw members from regions of the search space that haven't yet been explored, while ruling out members that have already been rejected in previous generations (tabu).

1.4 Selection

An often-reported comment from participants in IGA experiments is that at the beginning of a run, it is hard for them to accept any option, but as the run persists, it is even harder for them to evaluate among the very small of a nearly homogeneous population. *Culling* and *selective insertion* are intended to mitigate the task difficulty encountered in later generations of a run, whereas *variable acceptance* is intended to help with task difficulty at the beginning of a run. In standard IGA, a certain number of parents are necessary to avoid pre-mature convergence of the population. In practice, we have required participants to accept half of a generation's members, regardless of whether any or all of them are judged acceptable. The IAGA instructs the participant to accept as many, or as few, of the members of the current generation as appropriate.

1.5 Cross-over and Mutation

At the most general level, the purpose of cross-over is to perpetuate those schema in the search space with positive value and to eliminate all others. Mutation's role is to improve existing schema by introducing new variations in the population. *Culling* and *selective insertion* (to the extent non-visited portions of the search space have very different schema) are counter-productive to schema formation, whereas *variable acceptance* is likely to reinforce larger schema to the detriment of smaller ones during the initial phase of an IAGA run.

A natural way to incorporate binary selections (of arbitrary number) with *tagging* (which preserves the relative dominance of a particular string within the current generation) into the cross-over operator is to generate the set of

all possible children and sample without replacement from the set. Specifically, suppose $A = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ is the set of accepted members from the current generation. We form the parents P from A by augmenting A with each member's collection of duplicates

$$P = \{\alpha_{1,1}, \dots, \alpha_{1,N_1}, \alpha_{2,1}, \dots, \alpha_{2,N_2}, \dots, \alpha_{k,1}, \dots, \alpha_{k,N_k}\}$$

We form C , the population of potential children by crossing, in all possible ways, each pair of parents drawn from P . We then draw, without replacement, the proper number of strings to form the next generation. These strings undergo mutation and then are subject to the *culling*, *tagging*, and *selective insertion* operators.

Although the complexity of the proposed cross-over operation is larger than most cross-over operators in the GA literature, the number of computations required remains relatively small in practice owing to the fact that the size of the search space is small. In our typical application, we employ single-cut crossover for search spaces on the order of 2000, such that the computation time is negligible when compared with the time it takes for the user to make their judgements.

1.6 Initialization and Stopping Criteria

Initialization of IGAs is subject to the same issues encountered with any GA. Whereas standard GAs work around the problems of initialization by repeated measures, IGAs, in practice, cannot rely on more than a handful of runs. The IAGA utilizes *selective insertion* to introduce genetic materials into the population that may not have been encountered as well as an initialization procedure in which values of each parameter are as distinct across the population as possible.

Both *variable acceptance* and the *culling/tagging* operators complicate the application of standard rules for terminating an IGA run. We have found that measures of genetic drift, when applied to the population of potential children during crossover, are useful indicators. In the example application that follows, we utilized a more global stopping criterion based on the proportion of the search space that the user has explored.

1.7 Summary

The IAGA addresses components of an IGA which may be compromised by the psychometric limitations of the assessment procedure. It is designed to efficiently utilize the human participant's time, promote their attentiveness, minimize their fatigue, and work around their bias. The use of *culling*, *tagging*, *selective insertion*, and *variable acceptance* to achieve these design goals require a substantial reworking of cross-over, mutation, initialization, and stopping criteria. What follows is a particular instance of the IAGA approach involving the design of audio processors in a cochlear implants.

2. OPTIMIZING THE PERCEPTION OF MUSIC AND SPEECH BY COCHLEAR IMPLANT USERS

Cochlear implants (CIs) can provide a sense of sound to people with a severe to profound hearing impairment. A sound processor transforms sound signals into patterns of electrical pulse trains that stimulate the auditory nerve via implanted electrodes. The ability of a CI user to understand

speech or recognize other sounds strongly depends on the selection of processing parameters that dictate the processor program's behavior. No single combination of parameters provides the best sound quality for all users. Rather, each recipient has unique hearing that is best served by individually tailored parameters. However, the clinical task of fitting an optimal set of parameters from the thousands of possible combinations is complicated by the parameters themselves because they can exhibit strong nonlinear, non-monotonic, and often unpredictable interactions. Not only are CI programs subject to the preferential judgment of each recipient, different programs perform better with different kinds of sound. To maximize the ability to communicate, most well-fit CI programs aim to best resolve speech sounds. As a result, CI users often achieve good performance in speech recognition, but report that music sounds unnatural and even unrecognizable. To optimize sound processor programs for music listening, the IAGA was used in a study with human CI recipients.

2.1 Methods

The IAGA relies on human feedback to evaluate population fitness and is thus prone to compromised and distorted judgments brought about by fatigue. Each generation of the process involves listening attentively and critically to music through multiple programs. Traditional IGAs compound the mental stress because they require participants to select a fixed number of designs in every generation. This often makes the task more taxing because participants are forced to attend very closely and repeatedly to designs before deciding which bad ones should be selected or which good ones should be rejected just to meet a requisite number of selections. The dynamic nature of the subjective human decision process can exacerbate fatigue further. Until the proper design schema have developed and have been presented in generational populations, bad programs may seem good because they are better in comparison to worse programs. This is especially true in early generations comprised largely of bad or suboptimal designs. Thus, the decision criteria will likely change over the course of an evolution as listeners experience more and potentially better programs available in the search space and learn more about their own preferences and inclinations. As a result, dynamic fitness criteria may impact convergence that is based upon conditions of design homogeneity and may protract the duration of a successful evolution. The IAGA fatigue mitigation strategies include imposing a rigid stopping criterion to limit the duration of an evolution and allowing participants to select as many or as few good designs as they see fit.

IAGA evolutions are terminated after a predefined extent of the search space has been explored, but precautions are taken to avert premature convergence on a design that is not optimal. Based on the behavior of a traditional IGA used in earlier implementations to optimize CI speech programs [16] [9], it was determined that a 5.0% sampling of the search space was sufficient to find several good programs. The IAGA subsequently allows recipients to directly compare these programs against each other in what has been termed *playoff* rounds. Programs that were preferred in earlier stages of the evolutionary process but may have been lost during operations that expanded the divergence of the search are revisited and reassessed within the context of other good programs. The playoffs not only help determine

the best overall program but also ensure that evolutionary termination does not unfairly bias the final program. In this study, *playoff* rounds were seeded with the four programs that were statistically most persistent across evolutionary rounds. Parameters that did not seem to factor into participants' fitness judgments, so called invisible parameters, were not considered in the determination of program persistence. The "top four" programs were compared head to head in three to five *playoff* rounds until a final winner emerged from them.

Mitigating fatigue is a common problem with many IGA implementations, but the GA search space used to optimize CI programs is uniquely challenging. Literally thousands of CI programs can be created by combining the various available processing parameters and setting each parameter to various values. For this study, a search space was constructed with eight different music-relevant processing parameters. Each parameter was limited to two or three different possible values. For instance, the high frequency suppression parameter had possible values of "on" or "off", and the electrical stimulation rate parameter had possible values of 500, 900, or 1200 Hz. As these examples begin to suggest, CI parameters generally cannot be represented by continuous functions or even with simple conventions such as ordinality. A comprehensive discussion of particular parameters is not appropriate in this context but can be found in [4], [12], [10], and [5]. A search space comprised of CI programs generally represents a sparse distribution of designs. For this study, the IAGA was implemented on customized software. The IAGA search space was limited to 972 programs. This was comparable to earlier studies which utilized a traditional IGA to search through an 8-bit space of 256 programs [16] and a 10-bit space of 1024 programs [9] in an effort to optimize speech.

2.2 IAGA Software

Custom software was developed to implement the IAGA, to control CI processors, and to provide a graphical user interface (GUI) for participants to submit feedback. An example screen of the software's GUI is shown in Figure 1. Each screen embodied a single segment of the IAGA session. The term "segment" is used here to denote either an evolutionary generation or a non-evolutionary *playoff* round, both of which appear the same and require the same actions from participants. Throughout the program optimization process, music was played continuously through speakers into the ambient environment of the room. Each screen gave participants access to eight programs. In evolutionary generations, every program is unique. In *playoff* rounds, the "top four" programs were distributed randomly among the eight available locations and therefore presented twice apiece. When each screen was first displayed, the application sequentially switched between each of eight programs for 20 seconds. In this case, switching implies deactivating the current program, loading a new program onto the processor, and then activating that program. The process of switching between programs was almost imperceptibly short, taking under 400 milliseconds to complete. Rather than waiting for the sequential presentation of programs to run its course, participants could at any time switch between programs by using the program activation buttons labeled "1" through "8". The sequential presentation sequence could

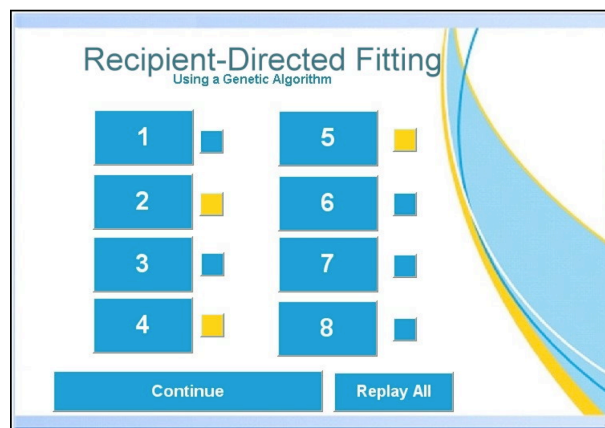


Figure 1: Example view of the customized software used by participants in this study. The screen depicts a single generation. Participants were instructed to listen to music using programs 1 through 8, indicate which of them sounded "good" using the adjacent toggle buttons, and then proceed to the next group of programs (next generation) by using the [Continue] button after all good programs had been selected. In this example, programs 2, 4, and 5 are selected.

also be reinitiated by depressing the button labeled [Replay All].

The toggle buttons adjacent to each program activation button were used by participants to provide subjective feedback. Participants were instructed to indicate which programs processed the ambient music in a manner that sounded good to them. Participants were able to select as many (up to eight) or as few (even zero) programs as was appropriate. The definition of "good" was deliberately vague because the aim of the task was to identify good programs for listening to music, and music appreciation is not universally consistent concept. This study appreciated that different people value different aspect of music above others, whether the beat, the clarity of the instruments or vocals, or some other factor. In addition, there is considerable variability between how different CI recipients hear, whether because of implant placement variations, neurophysiological abilities, and countless other differences. Ultimately, this study aimed to provide a common tool to optimize music programs to each subject's particular liking, despite their differences.

The [Continue] button progressed participants from one screen to the next. Participants were instructed to use this button once all good programs on the current screen had been selected. During the evolutionary process, this initiated the reproductive behavior of the IAGA and produced a new generation of programs based on subject preferences. The new generation was presented as a new screen of eight unique programs. New screens progress as new generations are produced until the IAGA is stopped and the process enters *playoff* rounds. *Playoff* screens were simply a clandestine opportunity for participants to essentially vote for the programs they preferred. Since generation and *playoff* screens looked and behaved similarly, participants did not

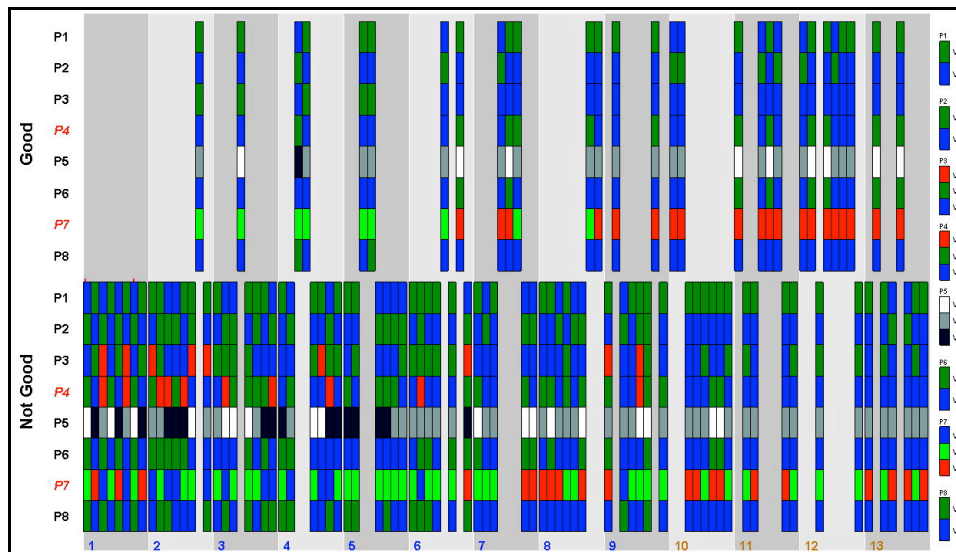


Figure 2: Example graphical representation of an IAGA fitting session.

know whether they were engaging in evolutionary or *playoff* rounds. Ensuring that the optimization task be temporally succinct was an important factor in implementing the IAGA for this study. Although the 5.0% sampling of the search space seemed small, it was quite reasonable considering the required time commitment and the fact that participants would be judging 60 to 80 programs during evolutionary generations and comparing an additional 20 to 30 more during three to five *playoff* rounds.

2.3 Listening Evaluation

This study incorporated both music comprehension and speech perception testing. Six subjects participated in testing using the IAGA programs they had optimized for music and the clinically fit speech programs they used in everyday life. Two participants also had clinically fit everyday music programs in addition to their everyday speech programs. Each program was tested separately. This protocol allowed for comparisons between the quality of the IAGA-fit program relative to clinically fit programs, and thus provided a practical evaluation of the effectiveness of the IAGA fitting process. Music testing utilized the Clinical Assessment of Music Processing (CAMP) test, as described by [11]. One component of the CAMP test battery incorporates a melody identification test using common melodies which have been processed to remove identifying temporal cues. The CAMP battery also utilizes a timbre recognition test in which the listener is asked to identify commonly recognizable musical instruments by their distinctive spectral overtone signatures. Speech testing utilized the AzBio sentence test, as described by [13]. Sentences are natural and conversational in nature, but the test is generally considered more difficult than traditional tests, such as HINT and CUNY sentences. To make the AzBio test more difficult so as to avoid ceiling effects, sentences were presented at 60 dBA in 4-talker babble noise.

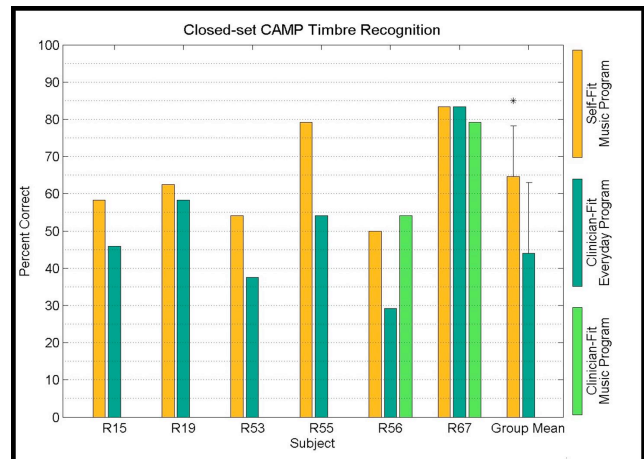


Figure 3: CAMP timbre recognition scores for six participants. Each subject used both GA and everyday programs. Two also used clinically fit music programs. Group mean scores for the GA program were 20.5% higher (statistically significant) than for the everyday program.

2.4 Results and Discussion

2.4.1 Music Program Optimization

The IAGA software was developed to record all data regarding each fitting session. These data include algorithmic information, such as randomly generated values and results of reproduction, and information provided by the subject, such as time-stamped button presses and program selections. A top-level analysis of these data was possible by representing them graphically, as is demonstrated for one session in Figure 2. Each screen encountered while using the IAGA software is represented within a dark or light grey

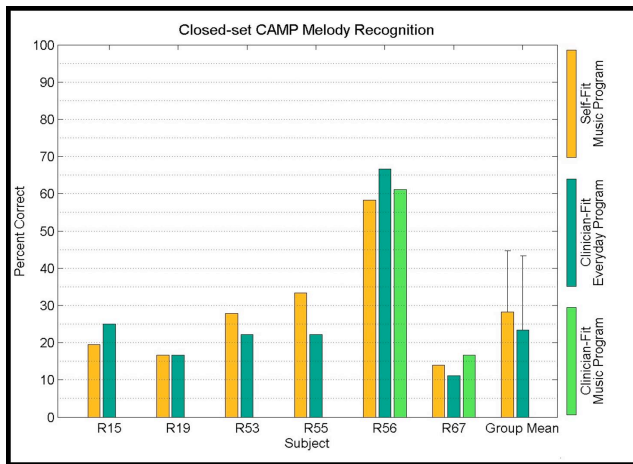


Figure 4: CAMP melody recognition scores for six participants. Each subject used both GA and everyday programs. Two also used clinically fit music programs. The difference between group mean scores were not statistically significant.

vertical stripe. The process progresses from left to right, as indicated by the numbering along the bottom. Blue numbers indicate evolution generations, whereas those in orange indicate *playoff* rounds. The eight different programs encountered on each screen can be found within each stripe. The programs in the top half depict those that sounded good to participants, whereas programs in the bottom half depict those that did not. Each program is represented by a vertically arranged combination of color coded parameters and parameter values. For the purpose of this discussion, parameter names have been replaced by an ordinal numeric identifiers ranging from P1 to P8. Parameters are shown along the left-hand side. Parameter names in red indicate parameters that were considered invisible to the subject in terms of their contribution to fitness. Parameter names have also been removed, but a legend along the right-hand side illustrates the different colors associated with different values (V1, V2, or V3) as well as the number of value levels within each parameter.

2.4.2 Music and Speech Testing

Six participants were tested while wearing either their IAGA music programs or their everyday speech programs. Additionally, two participants were also tested while wearing previously fit clinical programs specifically designed for music listening. The CAMP test battery included tests of both instrumental timbre and musical melody recognition. Results from timbre recognition subtest (Figure 3) show that the IAGA program outperformed clinical speech programs by a statistically significant group mean performance margin. The melody recognition subtest (Figure 4) did not show statistical differences between programs. This was likely due to the small subject sample size, which was definitely the case when comparing clinically fit music programs. Although these results are inconclusive, they are encouraging.

The AzBio test results (Figure 5) are also encouraging. The results were again not statistically significant but demonstrated that the music program fit by the IAGA could also

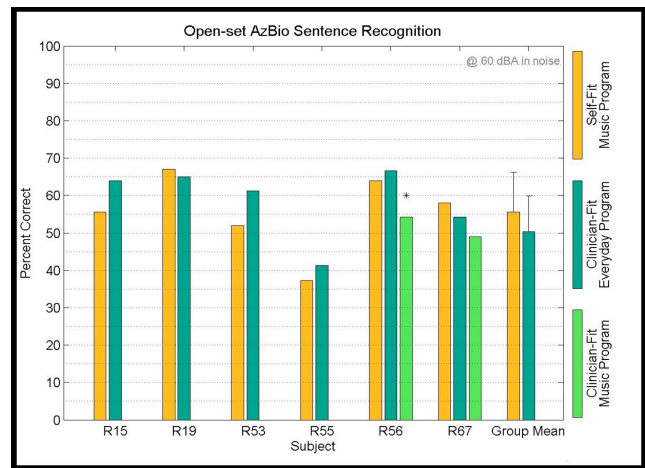


Figure 5: AzBio sentence recognition scores for six participants. Each subject used both GA and everyday programs. Two also used clinically fit music programs. The difference between group mean scores were not statistically significant.

be used for speech perception. This study will continue with an expanded subject pool.

The testing results provide a real world validation of the IAGA. They offer practical proof that the steps taken to augment a traditional IGA produced an algorithm capable of meeting the unique challenge of fitting CI programs. This contention is further buttressed when comparing their respective algorithmic data, particularly in the context of minimizing participant fatigue while expanding the diversity of the search.

3. THE USE OF IAGA FEATURES IN THE COCHLEAR IMPLANT EXPERIMENT

The IAGA modifies several components of the IGA to better reflect the psychometric theory and practice. In the experiment of the preceding section, one factor (proportion of the space searched) was controlled and results for designs selected using IAGA were presented. From the data, it is clear that participants were able to use the IAGA effectively. What is not clear from the data is the extent to which the psychometric goals of the IAGA were met. In the final section of the paper, we consider several of these goals in light of the outcomes in the cochlear implant experiment.

3.1 Fatigue mitigation - reduced time commitment

Participant fatigue is largely a function of the time required by the task at hand. A cursory analysis of the exemplary data in Figure 1 illustrates the temporal characteristics of the IAGA fitting procedure. The algorithm performed nine evolutionary generations before reaching or surpassing the 5.0% search space sampling threshold. During the evolution, the participant experienced 72 programs, 52 of which were unique, and indicated that music sounded good in 15 of those presentations. Following evolution, the subject required four *playoff* rounds to determine which program was ultimately the best for listening to music. Although the

Table 1: Comparing the IAGA used for optimizing music programs to traditional 10-bit IGA used for optimizing speech programs.

| | IAGA | IGA |
|---|-------|------|
| <i>Study</i> | | |
| Participants | 7 | 20 |
| Evolutions | 16 | 104 |
| <i>Search Space</i> | | |
| Unique programs | 972 | 1024 |
| Parameters per program | 8 | 7 |
| <i>Selection</i> | | |
| Programs selected per generation | 2.3 | 4 |
| Proportion of times user selected four programs | 16.5% | 100% |
| <i>Search</i> | | |
| Screens per session | 10.6 | 21.4 |
| Proportion of search space visited per evolution | 5.3% | 6.7% |
| Duplicate programs per generation | 0 | 1.6 |
| New programs per generation | 4.9 | 3.2 |
| <i>Incongruous Operations</i> | | |
| Invisible parameters per evolution | 2.4 | 2.6 |
| Proportion of programs in tabu list per evolution | 8.3 | 0 |

entire process required 13 screens, it took 901.0 seconds (15:01). On average, each comparison screen required only 69.31 seconds (01:09) to complete. The 16 fitting sessions involved in this study required an average of 10.63 screens and all optimized a music listening program in less than 20 minutes. This represents a substantial improvement over the previous IGA implementations. The IGA required an average of 21.40 similarly configured screens for a similar purpose. The IAGA timeframe is consistent with regular clinical programming sessions, and as such, would not impose undue fatigue.

3.2 Fatigue mitigation - relaxed task load

Fatigue can also be a function of mental demand. Active critical listening for extended periods of time can be a demanding burden of concentration for a normal hearing listener, but it is often much more difficult for the hearing impaired listener. A particularly striking aspect of the session shown in Table 1 is how many programs were judged as good, for listening to music and therefore selected for reproduction, in each generation. The previous IGA implementations required that four programs were selected in every generation. The IAGA employed the variable acceptance rule and had no such requirement. Participants were free to select any programs that met their personal fitness criteria. As a result, participants using the IAGA selected an average of 2.28 programs per generation, and there was only a 16.47% chance that they would select four. The distribution of selection count is shown in Figure 6. Whereas the previous IGA studies required four design selections every generation, participants in this study were more likely to choose fewer and even zero programs. Selecting one, two, or three programs per generation was more likely than choosing four. Allowing participants the freedom to select only those

programs as were appropriate, rather than forcing them into deliberations, eased the mental demand of this application.

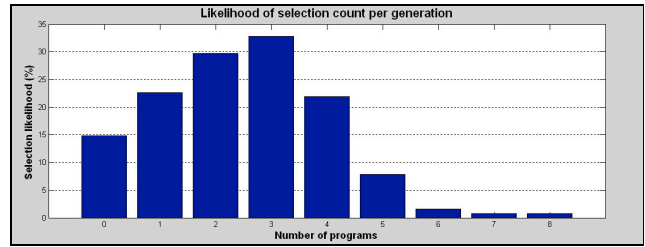


Figure 6: Likelihood breakdown of selecting any-where from zero to eight programs presented in a generation. Participants were instructed to select any or none of the eight. Data were compiled over 16 evolutions comprising 221 generations.

3.3 Search efficiency of the IAGA

Another interesting point of comparison between the IAGA and previous IGA implementations is the occurrence of duplicate programs in a generation. By design, the IAGA eliminates duplicate designs from appearing in any given generation. This enables the search to more broadly sample the search space rather than wasting operations on redundantly presented designs. The information about would-be duplications is retained as tagging data. The traditional IGA has no such provision, and redundancies are common. Therefore, whereas IAGA generations contain no duplicate designs, IGA generations contained 1.65 duplicates in previous studies. To put this within the context of creating a more extensive and diverse search, the IAGA added an average of 4.93 new programs and explored an additional 0.51% of the available search space in every generation. The previous IGA implementations only added an average of 3.19 programs and explored an additional 0.31% of the available search space per generation. These numbers speak to the increased efficiency of the IAGA.

3.4 Participant sensitivity to design parameters

The graphic representation in Figure 1 also informs about the participant's sensitivity to certain parameters. In this example, the subject did not seem to show a strong preference for two parameters (P4 and P7) during this session, and they were classified invisible. Information such as this could be useful during a clinical programming session as it may streamline the process by suggesting which parameters are more important than others. In this study, programming sessions identified 2.38 invisible parameters per evolution on average. It should be emphasized that different sound characteristics may interact unpredictably with different processing parameters, so invisible parameters identified in one setting may not be those identified in another. For comparative purposes, the invisibility calculation was applied to data from IGA sessions. This analysis identified an average of 2.56 invisible parameters per evolution, which was a similar result. Removing invisible design parameters from the design space is likely to further simplify the deliberative burden and improve the efficiency of the search.

3.5 Coupling the IGA with a tabu operator

Incorporating a tabu operation within a GA has the potential of positively affecting the algorithm. However, the IAGA tabu operator was interestingly inconsequential in this study. Tabu programs had to be rejected in two generations and thus considered “not good” for music listening. These programs were subsequently removed from the search space and placed on a tabu list. In theory this approach can reduce the size of the design space and thus improve the GA search. In this study, only an average of 8.25 programs were rejected per evolution. This represented only 0.85% of the search space. Applying the same “2 strikes, you’re out” condition on the earlier IGA evolutions showed the tabu operator to be similarly negligible in terms of pruning the search space. In these cases, the tabu list included an average of 22 programs per evolution. That number, although small, is also inflated because the previous IGA implementations did not remove would-be tabu programs. The common condition among the IAGA and IGA implementations is a small and sparsely populated search space. Further study may be warranted to characterize this properly, but the tabu operator in these experiments provided modest improvement at best.

4. CONCLUSION

Modifications to the general structure of IGAs have been proposed from the standpoint of psychometric theory and practice. The IAGA provides efficiencies in the human-guided search while still achieving evolutionary behavior that is normally associated with GAs. In an example application, cochlear implant recipients demonstrated their ability to use the IAGA to improve their perception of music. Post hoc analysis of the evolutionary behavior of the participants’ searches supports the claims of greater efficiencies when using the IAGA.

5. REFERENCES

- [1] S. B. Cho. Towards creative evolutionary systems with interactive genetic algorithm. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, 16(2):129–138, 2002.
- [2] R. Dawkins. *The Blind Watchmaker*. Longman Scientific and Technical, Harlow, 1986.
- [3] E. A. Durant, G. H. Wakefield, D. J. Van Tasell, and M. E. Rickert. Efficient perceptual tuning of hearing aids with genetic algorithms. *IEEE Transactions on Speech and Audio Processing*, 12(144-155), 2004.
- [4] R. A. Fearn. *Music and Pitch Perception of Cochlear Implant Recipients*. PhD thesis, University of New South Wales, Sydney, Australia, 2001.
- [5] Q. J. Fu. Loudness growth in cochlear implants: effect of stimulation rate and electrode configuration. *Hearing Research*, 202(1-2):56–62, 2005.
- [6] R. Kamalian, Y. Zhang, H. Takagi, and A. M. Agogino. Reduced human fatigue interactive evolutionary computation for micromachine design. In *Proceedings of the 2005 International Conference on Machine Learning and Cybernetics*, volume 9, pages 5666–5671, 2005.
- [7] J. C. Kelly, G. H. Wakefield, and P. Y. Papalambros. Evidence for using interactive genetic algorithms in shape preference assessment. *Int. J. Product Development*, to be published:1–17, 2011.
- [8] H. Levitt. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2):467–477, 1971.
- [9] S. K. Lineaweaver, W. S. Parkinson, P. W. Dawson, Y. Abrahams, G. H. Wakefield, and C. van den Honert. Study of automated fitting software for cochlear implants. page poster, Minneapolis, MN, 2006. *Audiology/NOW*.
- [10] H. J. McDermott. Music perception with cochlear implants: a review. *Trends in Amplification*, 82(2):49–82, 2004.
- [11] G. L. Nimmons, R. S. Kang, W. R. Drennan, J. Longnion, C. Ruffin, T. Worman, B. Yueh, and J. T. Rubinstein. Clinical assessment of music processing in cochlear implant listeners. *Otol. Neurotol.*, 29(2):149–155, 2008.
- [12] K. L. Plant, L. A. Whitford, C. E. Psarros, and A. E. Vandali. Parameter selection and programming recommendations for the ace and cis speech-processing strategies in the nucleus 24 cochlear implant system. *Cochlear Implants Int.*, 3(2):104–125, 2002.
- [13] A. J. Spahr and M. F. Dorman. Performance of subjects fit with the advanced bionics cii and nucleus 3g cochlear implant devices. *Arch. Otolaryngol. Head Neck Surg.*, 130(5):624–8, 2004.
- [14] S. S. Stevens. On the theory of scales of measurement. *Science*, 103(2684):677–680, 1946.
- [15] H. Takagi. Interactive evolutionary computation: fusion of the capabilities of ec optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.
- [16] G. H. Wakefield, C. van den Honert, W. S. Parkinson, and S. K. Lineaweaver. Genetic algorithms for adaptive psychophysical procedures: recipient-directed design of speech processor maps. *Ear and Hearing*, 26(4 Supplement):57S–72S, 2005.