

Improving Ability to Verify Audio CAPTCHAs to Serve Visually-Impaired

IPro 316 Project Plan
Advisor: Matt Bauer

Spring 2010

Contents

- 1 Abstract** **2**
- 2 Background** **2**
- 3 Objectives** **3**
- 4 Work Breakdown Structure** **3**
 - 4.1 Phase One 3
 - 4.2 Phase Two 3
 - 4.3 Phase Three 3
- 5 Expect Results** **4**
- 6 Project Budget** **4**
- 7 Team Structure and Assignments** **4**
 - 7.1 Phase One 5
 - 7.2 Phase Two 5
 - 7.3 Phase Three 5
- 8 Team Members' Background and Expectations** **7**
 - 8.1 Team Members' Background 7
 - 8.2 Team Members' Expectations 8

1 Abstract

CAPTCHAs (Completely Automated Public Turing Test to Tell Computers and Humans Apart) are used to prevent automated access to sensitive information online. In its usual format, users are presented with distorted text and asked to enter the displayed text in an answer box. If successful, humans, but not computers, will be able to interpret the distorted text. Another format of CAPTCHAs asks users to identify audio information (usually a string of digits or phrase of words) that has been distorted or placed against a background of noise (“white” noise, reversed speech, etc.). Users type the words they hear into an answer box. The audio format is intended to be accessible to blind and low-vision users who cannot use the visually-based format. Unfortunately, audio CAPTCHAs are difficult for humans to use (Bigham and Cavendar 2009) but relatively easy for computers to solve (Tam et al. 2008), which is exactly the opposite outcome desired. To take two extreme examples, in one recent study (Sauer et al. 2008), users were able to solve only 46 percent of audio CAPTCHAs, while in another study (Burztein and Bethard 2009), a computer program was able to break 75 percent of audio CAPTCHAs. At issue is whether audio CAPTCHAs can be designed so that users can easily solve them but computers cannot. Yan and Ahmad (2008) propose testing different kinds of background noise to determine which is the most effective at blocking computers but admitting humans. To this end, Tam et al. (2008) suggest using other human voices as background noise (to thwart computers) but familiar phrases as the string to decode (to aid listeners). The current project will focus on selecting from a set of potential solutions to test. These solutions include:

1. Using two concurrent undistorted, high-quality streams of speech, both of which can serve as the target signal or background noise.
2. Manipulating talker characteristics, such as speaking rate and intonation.
3. Using audio puzzles, such as a math problem, sound question, or an instruction.
4. Use portions of music and ask the user to input what lyrics they have heard.
5. Changing the interface to something that would be easier for low-vision users to use but perhaps harder for computers.

2 Background

This IPRO continues the basic work of IPRO 343 F08 and S09 [reabeled as IPRO 316 for Spring 2010] in examining acoustic and cognitive factors that contribute to understanding speech for public and commercial purposes. In fall 2008, the project focused on auditory factors that may improve accuracy of taking customer orders in a simulated fast food drive-thru environment. In spring 2009, the project identified and proposed benchmarks for speaking rate and pitch of synthetic speech designed for public announcements. The goal of the current project is to improve user ability to solve audio CAPTCHAs (Completely Automated Public Turing Test to Tell Computers and Humans Apart) while preventing computers from doing so.

The basic problem the current IPRO team will address is how to make an audio CAPTCHA highly intelligible to human users while impossible to solve by computers. Much of the current work aimed at improving audio CAPTCHAs ignores characteristics of the signal for users to interpret, such as speaking rate (syllables per second), fidelity of the signal, degree of hyperarticulation (“enunciation” of speech sounds), etc. The current project will add to existing work by testing these and other factors known to affect a speech signal’s intelligibility. In terms of contemporary significance, audio CAPTCHAs are currently used mostly by people with blindness or low vision, but Soupionis et al. (2009) note that audio CAPTCHAs may be useful in Voice over IP contexts to combat spam over Internet telephony (e.g. bots making calls for political or product advertisement). Therefore, the quality of audio CAPTCHAs has widespread significance for the general population.

3 Objectives

This IPRO's goal is to study and recommend one or more of various potential means of decreasing computers ability to break audio CAPTCHAs while at the same time improving human ability to break them. To that purpose:

- I. The IPRO team will study speech and speech recognition in terms of its acoustical properties in order to better understand the problem and potential solutions.
- II. IPRO sub teams will explore possible solutions and their pros and cons in order to select which options should be pursued.
- III. The IPRO team shall devise and conduct an experiment to determine how effective the chosen solutions are at improving audio CAPTCHAs.
- IV. The IPRO team shall devise a recommendation as to how to improve audio CAPTCHAs based on an analysis of the data obtained in the experiment.

4 Work Breakdown Structure

4.1 Phase One

Task	Description	Deadline
Learn Acoustic Foundations of Speech	The team will learn the fundamentals of acoustics and how this transposes into speech.	01/28/10
Project Plan/IRB Form	Revise and submit the project plan and the IRB form.	02/05/10
Budget Proposal	Revise and submit the proposed budget.	02/05/10
Ethics Training	Complete the web training on research ethics.	02/10/10
Evaluate Existing CAPTCHAs	A team will evaluate existing CAPTCHA systems for further analysis.	02/17/10
Devise Solutions	A team will devise possible solutions and applicable experiments to test those solutions.	02/17/10
Midterm Presentation	A team will compile the data acquired and present.	02/23/10

4.2 Phase Two

Task	Description	Deadline
Recruitment	A team will recruit IIT students to be our test subjects.	04/01/10
Design Stimuli	A team will devise the stimuli necessary for the experiment.	04/13/10
Design Measurement Tools	A team will formulate accurate tools to measure results achieved from experiment.	04/13/10
Administer the Experiment	The team will administer the experiment compile the results.	04/13/10
Plan of Analysis	A team will construct a plan to analyze data obtained from experiment.	04/13/10

4.3 Phase Three

Task	Description	Deadline
Analyze Results	A team will analyze results from experiment.	04/15/10
Final Report	A team will craft the final report including the analysis of results and further recommendations.	04/20/10
Final Presentation	The team will present the findings from the IPRO.	04/23/10

5 Expect Results

We anticipate that our results will allow us to provide a recommendation as to how to improve audio CAPTCHAs by increasing computer failure rate and human success rate. Also, we expect our findings to illustrate why our recommendation makes sense as opposed to other potential solutions as well as approximate the magnitude of its advantage.

6 Project Budget

Experimental Expenses	Spring 2009 Actual Expenses			Spring 2010 Projected Budget		
	Days	Price Per Day	Total	Days	Price Per Day	Total
Participant Incentive/Support - Pizza	4	\$125.00	\$500.00	4	\$125.00	\$500.00
IPRO Day Expenses	Amount	Price Per Unit	Total	Amount	Price Per Unit	Total
Team Polo Shirts	8	\$24.25	\$194.00	8	\$24.25	\$194.00
Exhibit Materials	-	\$90.00	\$90.00	-	\$90.00	\$90.00
Other Expenses	Amount	Price Per Unit	Total	Amount	Price Per Unit	Total
Travel Expenses	-	\$20.00	\$20.00	-	-	-
Equipment - Microphone	-	-	-	1	\$150.00	\$150.00
Equipment - Software	-	-	-	1	\$150.00	\$150.00
TOTAL EXPENSES			\$804.00			\$1084.00

7 Team Structure and Assignments

To better facilitate the completion of the project's objectives, the team has been divided into groups and roles have been assigned as follows:

IPRO 316 Team Leader: Maxwell Kaim

Final Report Leader: Gabriel Klansky

Ethics Training Leader: Michael Fabian

Experiment Organizer: Daniel Kipp

Minute Taker: Adam Ciarkowski

Agenda/Time Keeper: Sean Wallace

The Groups are as follows:

7.1 Phase One

Group	Members	Description
Learn Acoustic Foundations of Speech	All	We will learn the Phonetic alphabet and the acoustical properties of speech to improve the tools to be used in the task at hand.
Project Plan	Sean Wallace, Daniel Kipp	This group is in charge of constructing this document (the project plan).
Ethics Training & IRB	All	These people will become certified to administer an experiment of the nature required and handle the submission of the budget and the Institutional Review Board form (IRB).
Evaluate Existing CAPTCHAs	Michael Fabian, Erick Schneider, Maxwell Kaim	This group will collect data on existing CAPTCHAs relating to their format and evaluate what works and what should be changed.
Devise Solutions	All	The IPRO group will come up with potential solutions to the problem that can be explored by way of brainstorming.
Midterm Presentation	Gabriel Klansky, Michael Fabian, Adam Ciarkowski	This group shall handle the midterm presentation in terms of both fabrication and the actual presenting.

7.2 Phase Two

Group	Members	Description
Recruitment	Gabriel Klansky	The IPRO group will run experiments to determine the effectiveness of a chosen solution.
Design Stimuli	Erick Schneider, Maxwell Kaim, Daniel Kipp	These people will be in charge of designing the stimuli to be used in the experiments.
Design Measurement Tools	Sean Wallace, Adam Ciarkowski	This group will design the measuring tools for the collection of experimental data.
Administer the Experiment	Erick Schneider, Adam Ciarkowski	These people will be in charge of any administrative aspects of the experimental process.
Plan of Analysis	Sean Wallace, Adam Ciarkowski	This group will plan how to analyze the data gathered in the experimental process.

7.3 Phase Three

Group	Members	Description
Analyze Results	Sean Wallace, Daniel Kipp	These people will perform the analysis of the experimental data.
Final Report	All	The IPRO group will build a report of our findings and recommendations.
Final Presentation	Gabriel Klansky, Michael Fabian, Adam Ciarkowski, Sean Wallace	This group will present our findings in the final presentation.
Oral Presentation	Erick Schneider, Daniel Kipp	This group will perform the IPRO oral presentation.
IPRO Booth	All	The team will present findings to all interested at IPRO day.

8 Team Members' Background and Expectations

8.1 Team Members' Background

Name	Major	Year	Team(s)	Skills	Interests
Gabriel Klansky	Humanities	3rd	Final Report Leader, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Midterm Presentation, Recruitment, Final Presentation, Final Report, IPRO Booth	MS Office, Praat, semi-fluent in French	Music, Art, Photography, Writing, Food
Erick Schneider	Computer Science	3rd	Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Evaluate Existing CAPTCHAs, Design Stimuli, Administer the Experiment, Oral Presentation, Final Report, IPRO Booth	MS Office, Java, Visual Basic, Scala, GIMP	Computers, Gaming, Books
Maxwell Kaim	Computer Science / Psychology	4th	Team Leader, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Evaluate Existing CAPTCHAs, Design Stimuli, Final Report, IPRO Booth	Java, C, C++, Python, Open Office, Open Canvas, Photoshop, Psych Testing Experience	Language / Semantics, Natural Language Processing, AI Programing, Drawing, Creative Writing
Adam Ciarkowski	Computer Science	3rd	Minute Taker, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Midterm Presentation, Design Measurement Tools, Administer the Experiment, Plan of Analysis, Final Presentation, Final Report, IPRO Booth	MS Office, Java, C, Sound Mixing	Music, Games, Computers, Bike-riding
Michael Fabian	Computer Science	3rd	Ethics Training Leader, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Evaluate Existing CAPTCHAs, Midterm Presentation, Final Presentation, Final Report, IPRO Booth	C, C++, Perl, Java, HTML, MS Office, Visual Basic, SQL	Technology, Movies, Cognitive Science
Sean Wallace	Computer Science	3rd	Agenda/Time Keeper, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Project Plan, Design Measurement Tools, Plan of Analysis, Analyze Results, Final Presentation, Final Report, IPRO Booth	C, Objective-C, Java, PHP, SQL, Adobe Programs, MS Office, JavaScript, HTML	Technology, Wakeboarding, Movies

Name	Major	Year	Team(s)	Skills	Interests
Daniel Kipp	Mathematics	3rd	Experiment Organizer, Explore Multiple Voices and Music Lyrics, Learn acoustic foundations of speech, Devise Solutions - Design Experiment, Ethics Training, Generate Project Plan, Running Experiments, Designing Stimuli, Analyze Results, Final Report, Final Presentation Booth	OpenOffice, MS Office, Audacity, Java, C, Praat, Fourier Analysis, Probability & Statistics, Matlab, Derive, Python	Music, Reading, Math, Video Games, Puzzles, Climbing Things, Skiing, Board Games

8.2 Team Members' Expectations

Name	Short Term Goals	Long Term Goals
Gabriel Klansky	I hope that I will be able to express the difficulties and intricacies of this project in the presentations and reports. Also I look forward to running an experiment and seeing through to completion.	I hope to further my knowledge and ability in phonetics and language processing. I hope that through the project I will improve my skills in writing in a academic setting.
Erick Schneider	To bring my abilities and experience to the group to facilitate the development of a new type of CAPTCHA that will be both easy for humans to recognize and understand, yet hard for an intelligent agent to reason out.	To do the best possible in order to place high in IPRO day and to lay out a foundation for the next group of individuals who will work on this project.
Maxwell Kaim	Keep the team on track and together in terms of our deliverables, benchmarks, goals, et cetera. To keep my own contributions on time and their content up to group expectations.	To learn something if not useful than at least interesting about language and language analysis.
Adam Ciarkowski	Explore several different background sounds and being able to successful incorporate them with spoken words, Prepare presentations, Discuss various options for solutions.	Work together as a group to build a solid final presentation, Learn how the experiment process works when testing others, Further improve my knowledge about sound and speech.
Michael Fabian	To successfully help in recruiting participants for our experiments, work with the group as a whole to devise the next step in research in thus upcoming technology, and to help in providing a supportive environment through ideas of each team member may be freely expressed.	To provide valid experimental results for future research to build upon through strong teamwork and dedication to the problem at hand.
Sean Wallace	Better round my knowledge and experience with technology by utilizing previous skills to accomplish new things.	To learn more about language as it pertains to technology and hopefully help focus future study on this subject more precisely.
Daniel Kipp	Within my groups I hope to complete my tasks well and on time. I hope to learn things about speech and acoustics and I expect things to run reasonably smoothly.	I hope to have a successful project that is also interesting.