Improving Voice Recognition Prompts for Users in Various Application Environments

IPRO 316 Project Plan
Advisor: Matt Bauer

Fall 2010

1 Abstract

Users of speech recognition technology often hyperarticulate (i.e., exaggerate) their speech in response to recognition failures and subsequent requests to repeat (e.g., "I'm sorry, I didn't understand, please repeat the input."). Hyperarticulation usually leads to further recognition failure. The goal of the current project is to develop a protocol for testing different talker characteristics of voice prompts in speech recognizers with an aim towards minimizing hyperarticulated speech from users. This IPRO is equally suited to students interested in the more technical aspects of acoustic phonetics and voice recognition as well as the cognitive aspects of predicting user behavior in technology-mediated environments.

2 Background

This IPRO continues the basic work of IPRO 343 F08 and S09 and IPRO 316 S10 in examining acoustic and cognitive factors that contribute to understanding speech for public and commercial purposes.

Hyperarticulated speech is exaggerated or more extremely produced speech (Lindblom 1990). Speakers will hyperarticulate their speech to overcome noisy work environments (Tufts and Frank 2003), to address children (Kuhl 1997), to address hard-of-hearing listeners (Picheny, Durlach, and Braida 1985), to address pets (Burnham, Kitamura, and Vollmer-Conna 2002), to accent words (Cho 2005), to convey fussiness (Eckert 2005), to indicate salient points within a sentence (Cho 2005), and to express frustration, sadness, excitement and other emotions (Lee et al 2005, Litman and Forbes-Riley 2006, Ververidis and Kotropoulos 2006).

Hyperarticulation involves enhancement of the acoustic signal and modification of the normal movement of the vocal organs. In particular, hyperarticulated speech is louder and higher pitched. Speech segments are longer, and the acoustic vowel space is larger. Jaw displacement from rest position is more extreme, and tongue body movement is more exaggerated, such that articulations requiring the tongue body to be high and front in the vocal tract are sometimes higher and more forward in the mouth (Lindblom and Moon 1994, De Jong 1995, Johnson et al. 1993, Smiljanic and Bradlow 2005).

Several studies have shown that when speech recognizers fail to identify a string of speech and then ask users to repeat the input, users will hyperarticulate their responses (Oviatt, MacEachern and Levow 1998, Swerts, Litman, and Hirschberg 2000, Goldberg, Ostendorf, and Kirchhoff 2003, Hirchberg, Litman, and Swerts 2001). Interestingly, as a result of such hyperarticulation, once users are issued such failure-to-understand prompts, recognition rates fall significantly as hyperarticulation increasingly distorts the speech string (Swerts, Litman, and Hirschberg 2000). Thus, an ability to correctly predict how exactly speakers will hyperarticulate speech in failure-to-understand situations is a present challenge for speech researchers (Oviatt, MacEachern and Levow 1998).

One factor related to hyperarticulation in failure-to-understand responses is user emotion. A significant body of literature has shown how emotions of speakers affect their speech (Williams and Stevens 1972, Goldberg, Ostendorf, and Kirchoff 2003, Linnankoski et al 2005, Nordstarnd et al 2004, Lee et al 2005, Litman and Forbes-Riley 2006, and see Ververidis & Kotropoulos 2006 for

a bibliography of several dozen other papers). In human-computer interactions, hyperarticulation from frustration is frequently exhibited but can be minimized if the wording of the error message is apologetic, rather than direct (e.g. "I'm sorry, I didn't understand. Please say the sentence again," vs. "Say the sentence again.") (Goldberg, Ostendorf, and Kirchoff 2003). Another factor related to hyperarticulation in failure-to-understand responses is user desire to be intelligible. Lindblom and Moon (1994) observe that speakers instructed to "speak clearly" will usually hyperarticulate their speech, even if doing so undermines intelligibility of speech.

At issue is whether other talker characteristics of the voice prompt, such as its speaking rate, pitch, intonation, and its own degree of hyperarticulation, influence users' speech in predictable ways and can further minimize recognition failure.

3 Objectives

The goal of the IPRO is to develop a protocol for testing different talker characteristics of voice prompts in speech recognizers with an aim towards minimizing hyperarticulated speech from users and improving recognition success rates.

I The IPRO team will learn about the acoustic properties of normal and hyperarticulated speech in order to better understand the problem and potential solutions.

II IPRO subteams will identify relevant factors in the quality of voice prompts to be tested during the experiments.

III The IPRO team will devise and condut experiments to test the effect of varying the properties of the voice prompt's speech.

IV The IPRO team will summarize recommendations for improving voice prompts in voice recognition systems so as to reduce the amount of hyperarticulated speech from users.

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 affects the way speech is interpreted by humans and computers.	9/9/10
Project Plan	Revise and Submit the project plan.	9/12/10
Budjet Proposal	Revise and Submit the proposed budget.	9/12/10
Ethics Training	Complete web training on research ethics.	9/28/10
Evaluate Existing Voice Prompts	A team will collect recordings of existing voice prompts for further analysis.	9/16/10
Devise Solutions	The team will devise solutions and experiments to test those solutions.	9/30/10
Midterm Presentation	A team will compile the data acquired and give a presentation on the current state of the project.	10/14/10

4.2 Phase Two

Task	Description	Deadline
Recruitment	A team will recuit IIT students to be our test subjects.	10/19/10
Design Stimuli	A team will devise the stimuli necessary for the experiments.	10/21/10
Design Measurement Tools	A team will design tools needed to gather data during the experiments.	10/21/10
Administer the Experiments	The team will administer the experiments on test subjects and compile the results.	11/10/10
Plan of Analysis	A team will construct a plan to analyze data obtained from the experiments.	11/16/10

4.3 Phase Three

Task	Description	Deadline
Analyze Results	The team will analyze the results of the experiments.	11/25/10
Final Report	A team will write up the final report, including the analysis of the results and further recommendations.	12/2/10
Final Presentation	A team will present the findings from the IPRO.	12/3/10

5 Expected Results

We expect that by the end of the semester the IPRO team will have established which talker characteristics of voice prompts elicit the most successfully recognized speech, and will be able to make recommendations leading to more successful voice recognition systems.

6 Project Budget

Experimental Expenses	**Days**	**Price Per Day**	**Total**
Participant Incentive/Support - Pizza	4	$125.00	$500.00
IPRO Day Expenses	**-**	**Price**	**Total**
Exhibit Materials	-	$90.00	$90.00
Other Expenses	**Amount**	**Price Per Unit**	**Total**
Audio Equipment	-	$20.00	$20.00
TOTAL EXPENSES			$610.00

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: Naomi Peterson

Final Report Leader: Nithin Winston
Ethics Training Leader: Shashank Gopal
Experiment Organizer: Andrew Bossemeyer
Minute Taker: Alexander Webster
Agenda/Time Keeper: Robert Millonzi

7.1 Phase One

Group	Members	Description
Learn Acoustic Foundation of Speech	All	We will learn some IPA and the acoustic properties of speech in order to determine how best to improve voice prompts in recognition systems.
Project Plan	Ruth Morrison	This group will write the project plan (this document).
Ethics Training	All	We will become certified to administer the necessary experiments.
Evaluate Existing Voice Prompts	Alexander Webster, Vincent Echavarria	This group will collect recordings of existing voice prompts and evaluate their merits.
Devise Solutions	All	We will come up with possible solutions to the problems with existing voice prompts.
Midterm Presentation	Nithin Winston, Andrew Bossemeyer, Gabriel Klansky	This group will create the slides for and give the Midterm Presentation.

7.2 Phase Two

Group	Members	Description
Recruitment	Robert Millonzi, Andrew Bossemeyer, Shashank Gopal	This group will recruit IIT students to participate in the experiments.
Design Stimuli	Ruth Morrison, Nithin Winston, Gabriel Klansky	This group will decide on voice quality variables to test during the experiments.
Design Measurement Tools	Alexander Webster, Andrew Bossemeyer	This group will design measurement tools used in the experiments.
Administer the Experiments	All	We will administer the experiments and record the data collected.
Plan of Analysis	Alexander Webster, Andrew Bossemeyer	This group will plan how to analyze the data gathered during the experiments.

7.3 Phase Three

Group	Members	Description
Analyze Results	All	We will analyze the data collected in the experiments.
Final Report	Nithin Winston	This group will write up the final report containing the findings from the experiments and our recommendations.
Final Presentation	Andrew Bossemeyer, Robert Millonzi, Naomi Peterson	This group will give the final presentation.
IPRO Booth	All	We will present the findings to all interested at IPRO day.

8 Team Members' Background and Expectations

8.1 Team Members' Background

Name	Major	Year	Teams	Skills	Interests
Alexander Webster	Electrical Engineering/ Computer Engineering	3rd	Minute Taker, Learn Acoustic Foundations of Speech, Ethics Training, Evauluate Existing Voice Prompts, Devise Solutions, Design Measurement Tools, Administer the Experiments, Plan of Analysis, Analyze Results, IPRO Booth	Java, C, Open Office, Breadboarding, MS Paint, Circuit Design, Fourier Analysis	Music, Games, Computers, Gadgeteering
Nithin Winston	Biomedical Engineering	4th	Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Design Stimuli, Administer the Experiments, Analyze Results, Final Report, IPRO Booth	MS Paint, MATLAB, MS Office, AutoCAD, Organizational Skills	Books, Television, Music
Vincent Echavarria	Computer Science	3rd	Learn Acoustic Foundations of Speech, Ethics Training, Evaluate Existing Voice Prompts, Devise Solutions, Administer the Experiments, Analyze Results, IPRO Booth	Java, C++, C, MS Office, OpenOffice, LaTeX	Reading, Games, Computers, Movies
Robert Millonzi	Architecture	5th	Agenda/Time Keeper, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Recruitment Administer the Experiments, Analyze Results, Final Presentation, IPRO Booth	Photoshop, Illustrator, In Design, and other design software	Architecture, Music, and various other arts

Name	Major	Year	Teams	Skills	Interests
Andrew Bossemeyer	Architecture	5th	Experiment Organizer, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Midterm Presentation, Recruitment, Design Measurement Tools, Administer the Experiment, Plan of Analysis, Analyze Results, Final Presentation, IPRO Booth	Graphic Design, Leadership	Baseball, Volleyball, Photography, Sketching
Ruth Morrison	Computer Information Systems	5th	Learn Acoustic Foundations of Speech, Project Plan, Devise Solutions, Design Stimuli, Administer the Experiment, Analyze Results, IPRO Booth	C/C++, Java, Word Processors and LaTeX, Familiarity with IPA and Linguistics	Language, Computers, Programming, Reading
Shashank Gopal	Computer Science and Computer Engineering	4th	Ethics Training Leader, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Recruitment, Administer the Experiments, Analyze Results, IPRO Booth	Communication, Effective Teamwork, Organization	Music, Reading, Coding
Gabriel Klansky	Humanities	4th	Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Midterm Presentation, Design Stimuli, Administer the Experiments, Analyze Results, IPRO Booth	Writing, Presenting, Photography, Linguistics background	Semiotics, Photography, Communication, Philosophy
Naomi Peterson	Computer Science	4th	Project Leader, Learn Acoustic Foundations of Speech, Ethics Training, Devise Solutions, Administer the Experiments, Analyze Results, Final Presentation, IPRO Booth	Java, MS Office, Leadership, Communication	Speech accents, Music, Computers, Reading

8.2 Team Members' Expectations

Name	Short Term Goals	Long Term Goals
Alexander Webster	To create working systems that suit the needs of the experiments and, hence, further research into voice-recognition technology.	To gain valuable experience working with a development team towards furthering a research end.
Nithin Winston	I would like to partake in research that will benefit and promote the field of voice-recognition technology.	I would like to have more experience working with a team on a research project.
Vincent Echavarria	I want to help improve voice recognition prompts.	I would like to learn more details about voice recognition technology because it looks to be a major part of everyday life in the future.
Robert Millonzi	I want to see this group provide meaningful research into the development of voice-recognition software.	To work in a team scenario with various disciplinary backgrounds to achieve a common goal.
Andrew Bossemeyer	Develop a command prompt that decreases hyper-articulated responses	
Ruth Morrison	I'd like to learn more about the auditory properties of speech, and how other people react to them.	I hope to gain experience with working as part of a team and conducting experiments in order to further research.
Shashank Gopal	I would like to learn to use Praat. I would like to understand linguistics. I would like to use ultrasound to understand tongue movement.	I would like to help improve voice recognition prompts.
Gabriel Klansky	I hope to run an experiment and analyze the results. I also hope to learn how to analyze speech.	My long term goals are to learn how to be a team player and work in a group effectively. In tandem with that, I hope to learn to subdue my aggressiveness for others.
Naomi Peterson	I would like to understand people better, specifically what causes their spoken response to audio directions to change and what changes are caused.	I hope to gain valuable experience in learning new things quickly in a team environment so I can jump into helping with problem-solving almost immediately.