IPRO 316: SPEECH RECOGNITION & HYPERARTICULATION

PROBLEM

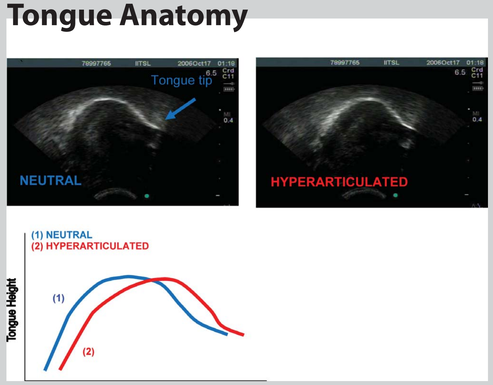
After error prompts users exaggerate their speech which worsens recognition rates

OBJECTIVE

To determine the factors that cause speech exaggeration after error prompts

BACKGROUND INFO.

Tongue Anatomy



Existing Error Prompts

Existing prompts were collected and analyzed. We looked at speaker characteristics of these prompts such as inflection, intonation, intensity, etc.

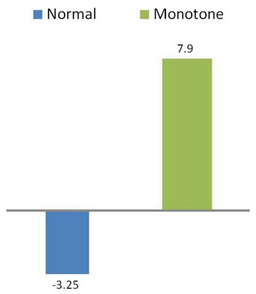
HYPOTHESIS

-Speakers will match their speech characteristics to those of the prompt
-Speakers are less likely to exaggerate their speech if the prompt says that it made a mistake

RESULTS

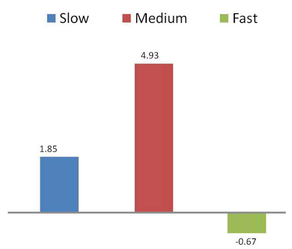
Intonation

(% recognized change pre-error vs. post-error)

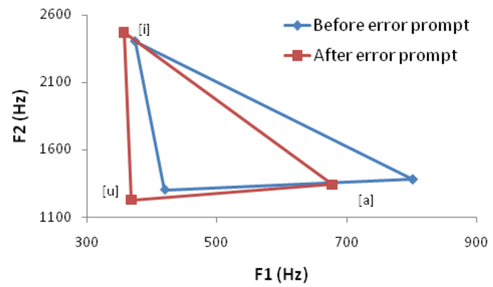


Rate of Speech

(% recognized change pre-error vs. post-error)



Hyperarticulation Analysis



The three vowels are [a] as in economies, [i] as in three, [u] as in two. This chart shows that the [a] sound was less clearly articulated while the [i] and [u] sounds were hyperarticulated slightly when responding to the error prompt. This can be seen because less articulated vowel sounds are closer to the center of the chart, while hyperarticulated vowels are closer to the edges or axises of the charts.

Statistics

	F	P
Intonation	F(1,73) = 4.7	< 0.05
Speaking Rate	F(2,73) = 0.39	0.68
Interaction	F(2,73) = 0.91	0.41

ANALYSIS & CONCLUSIONS

Overall, results show that recognition rates of responses following an error prompt improve significantly when the intonation contour of the prompt is natural-like, but nearly monotone. Results for speaking rate of the prompt suggest a medium rate leads to slightly improved recognition of response compared responses to slower or faster spoken prompts, but differences did not reach significance.

APPLICATIONS

There are many companies at the forefront of speech recognition software design. They all face the problem of error prompt enduced hyperarticulation. These findings can be utilized by their developers to better the user experience of their product.

at&t

CISCO

PROCEDURE

Step 1.
Participants were placed in a sound proof box for high audio fidelity. A micrphone and headphones were placed on their head and they were given a script. In order to simulate voice recognition software, the subjects could not see or interact with the testers.

Step 2.
While the participants were listening to the stimuli and r responding from the script, the testers were recording the subject using a solid-state recorder. At the same time, the testers were listening playing the stimuli to simulate voice recognition software.

Step 3.
The experiment took approximately 3-5 minutes to complete for each participant. Participants were all college-aged students of IIT or Shimer College. Among them, many were non-native speakers of English. Once they were done and were taken out of the box, they were given their incentives of pizza and a raffle ticket.

TYPICAL REPSONSE

Normal Response

Error Prompt

Hyperarticulated Response

PATH OF INFORMATION