# Early Stage Detection of Speech Recognition Errors(FINAL DRAFT)

by

**Stephen Choularton**

This thesis entitled:

**Early Stage Detection of Speech Recognition Errors(FINAL DRAFT)**

written by **Stephen Choularton**

has been approved for the Division of Information and Communication Sciences
Department of Computing

Robert Dale

Dr. Steve Cassidy

Date ____________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Choularton, Stephen (Ph.D., COMPSC)

**Early Stage Detection of Speech Recognition Errors(FINAL DRAFT)**

Thesis directed by Professor Robert Dale, Dr. Steve Cassidy.

Machines mishear human utterances. This mishearing represents a gateway problem for speech applications, introducing errors into dialogues, or giving rise to clarification sub-dialogues that often cause as many problems as they solve.

Misrecognition is so ubiquitous that commercial speech recognizers make use of a confidence metric to deliver a numerical assessment of the probability that an utterance has been correctly heard. Each recognition is then classified as either correct or false depending on whether its confidence exceeds a pre-determined threshold. Even with an optimally chosen threshold value, this classification decision is still incorrect between 10 and 25% of the time.

The aim of this thesis is to improve upon this capability of the machine to know when it has misheard. We do this by first exploring techniques for assessing the likelihood of error separately in the acoustic domain and the language domain, and then combining these methods in a unified classification mechanism.

In the acoustic domain, we establish that individual speaker characteristics are a major factor in determining whether or not a speech recognizer will mishear an utterance, and that, using logistic regression, we can train a model to identify such speakers. Working with data that we know to be free of errors in the language domain, we show that we can identify the utterances of problematic speakers with 85% accuracy.

In the language domain, we explore a range of techniques for identifying out-of-language utterances, and determine the circumstances under which these different approaches are most appropriate. Working on data that we know to be free of acoustic errors, we show that a domain-independent technique can identify out-of-language errors with an accuracy of 82%.

Finally, we combine these techniques to provide a unified mechanism for predicting the recognizability of an utterance. Using a dialogue state where the user confirms the systems understanding as an example, we demonstrate that, while the highest accuracy achievable via an existing confidence measure is 91%, we can achieve an accuracy of over 95%.

# Contents

**Chapter**

# Tables

## Table

# Figures

**Figure**