AUTOMATIC SPEECH RECOGNITION
FOR THE DIAGNOSIS OF
ALZHEIMER'S DISEASE

Luke Zhou
Supervised by Frank Rudzicz & Katie Fraser
June 23, 2015
DEMENTIA & ALZHEIMER’S DISEASE

747,000 Canadians living with dementias or cognitive impairment in 2011 (Alzheimer Society, 2012)
• That’s 14.9% of Canadians 65 and older.

Alzheimer’s Disease: most common form of dementia
• Symptoms:  
  • memory loss  
  • difficulties with thinking  
  • problem-solving  
  • language  
  • mood & behaviour

• Most AD patients tested for language function are found to have some degree of language impairment (Kirshner, 2012)
ALZHEIMER’S & LANGUAGE

“Cookie theft” elicitation task
AD speech:

- Decrease in informational content & density (Giles et al., 1996; Tomoeda et al., 1996; Croisile et al., 1996; Forbes-McKay and Venneri, 2005; Ahmed et al., 2013)

- Repetition of ideas (Tomoeda et al., 1996)
  
  all the bad things
  sink’s overflowing
  the stool’s going over
  and the cookie jar
  I guess the little girl she’s saying
  give me shh
  and the sink’s overflowing
  I might not be very observant but I don’t see anything else

- Increased production of vague words & pronouns with no referent
  
  ‘He’ — but who’s ‘he’?
ALZHEIMER’S & LANGUAGE (CTUD)

Clinician interview
Transcription
Analysis

Time-consuming process.
What if we could automate it?

Open-source toolkit for automated speech recognition.

I see cookies
AUTOMATIC SPEECH RECOGNITION SYSTEMS

WSJ/TIMIT Corpuses
(Healthy speakers) + Manual transcripts

DementiaBank Corpus
(Alzheimer’s & control speakers) + Manual transcripts

KALDI TRAINING

Language Model
- Lexicon
- Word sequences

Acoustic Model
- ‘Phones’ (sounds)

Models trained on healthy speakers, on “domain-general” content.

KALDI ADAPTATION

Readjustment of models:
- Older/impaired voices
- Reduced vocabulary
AUTOMATIC SPEECH RECOGNITION SYSTEMS (CTUD)

Adapted language & acoustic models

I see cookies

DECODING

What's the most likely word sequence $X'$ corresponding to a given acoustic sequence $Y$?

$$X' = \arg\max_X P(Y|X)P(X)$$

Classifier

Alzheimers? Healthy?
CLASSIFICATION FEATURES

TEXTUAL

i. SYNTACTIC

ii. PSYCHOLINGUISTIC:
   Familiarity, frequency, imageability, age of acquisition, …

iii. FLUENCY OF SPEECH:
   # of um, ah, pauses, …

ACOUSTIC

From the ASR-generated transcripts (& the original sound files), we can measure the following:

Various properties of the physical signal:
Kurtosis, skew, pitch, formant frequencies, pause duration, total duration, phonation rate, jitter, shimmer, …
RESEARCH QUESTIONS

Experiment 1: Adaptation of language & acoustic models
• What adaptation weighting parameters yield the best ASR accuracies?

Experiment 2: ASR errors
• What will be the effects of differing (i) types and (ii) rates of ASR errors on diagnosis accuracy?
• What classification features will remain statistically significant between the manual and ASR transcripts?

Experiment 3: Noise reduction
• What acoustic noise reduction techniques will improve diagnosis accuracies the most?
REFERENCES


QUESTIONS & COMMENTS?

Thank you for listening!