System for Automated Speech and Language Analysis (SALSA)

Kyle Marek-Spartz, Benjamin Knoll, Robert Bill, Thomas Christie, Serguei Pakhomov

University of Minnesota, Minneapolis, USA
mare0132@umn.edu, knol0061@umn.edu, bill0154@umn.edu,
tchristie@umn.edu, pakh0002@umn.edu

Background

SALSA automates administration and scoring of tests used to characterize cognitive impairment resulting from neurodegenerative disease, traumatic brain injury, and drug toxicity.

A number of widely used cognitive test batteries include speech-based tasks such as picture description, picture naming, spontaneous narrative, and verbal fluency. While many of these tests were initially designed to assess aphasia, they have been demonstrated to be useful for assessing other conditions that affect cognition including neurodegenerative disease [1] and neurotoxic medications [2] [3].

Test administration and scoring is manual and requires trained personnel, thus limiting use for routine assessment of large numbers of people for clinical or research purposes. This limitation can be overcome with automation. However, existing neuropsychological test batteries tend to rely on keyboard-based interaction with the person being tested and currently do not include speech-based assessments.

To address these limitations we developed SALSA, a system for collecting (via mobile and telephony platforms) and analyzing spoken responses to cognitive tests (using automatic speech recognition (ASR)) to calculate speech characteristics that may be useful in assessment of cognitive function.

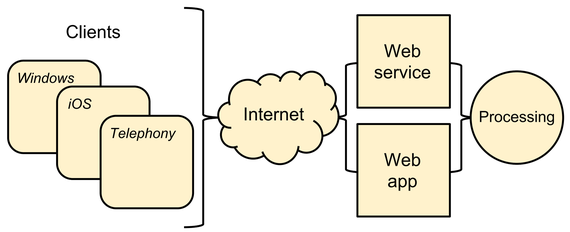

Figure 1: High-level architecture of SALSA.

Assessment clients

- Collect metadata
- Interpret a Test Prototype (defined on the server)
- Recording Subject response
- Upload data to the web service
- iOS, Windows, Telephony

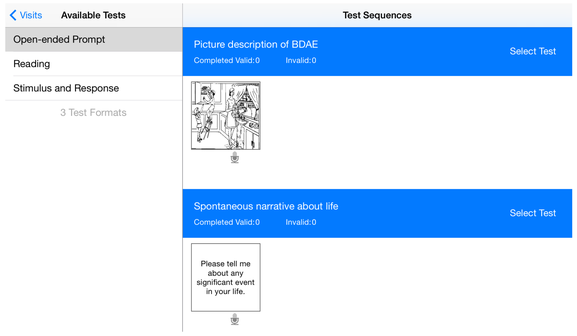

Figure 2: iOS client showing available Test Prototypes for a project.

Server

- Service-oriented Architecture
- Web application:
  - Project administration and transcription interfaces
  - PowerPoint to Test Prototype conversion
  - Starting individual and batch processing
  - Viewing attachments and results
- Web service (RESTful API)
- Database:
  - Most tables are immutable; helps with scale, reliability, consistency
  - Schema follows cognitive assessment paradigm (Projects, Subjects, Visits, Tests)
  - Protocol description (Test Prototypes, Stimuli)
- Server-side:
  - Implemented with Python, Flask, SQLAlchemy
  - Runs on LAMP
  - Supports running with Mac OS X, SQLite
- Client-side:
  - Implemented with Bootstrap, HTML5 Web audio, JavaScript
  - Compatible with recent versions of modern browsers (Firefox, Safari, Chrome)
- Open source license (See: `http://rxinformatics.umn.edu`)

Processing

- Generic, but focused on speech analysis
- Also supports other data, e.g. Mechanical Turk interactions
- Task Queue embedded in web service
- Workers can be distributed to multiple computers
- Fault tolerant

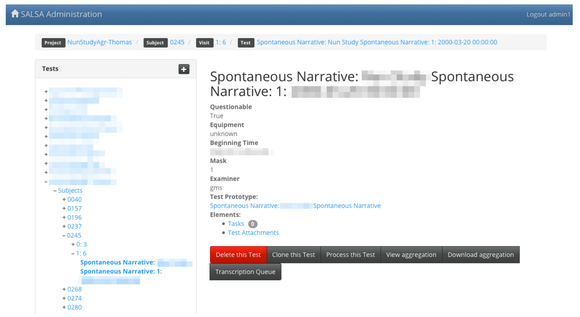

Figure 3: The web application showing metadata about a Test.

Automatic Speech Recognition (ASR)

- KALDI ASR toolkit performs acoustic modeling and decoding [4]
- Acoustic models
  - Speaker independent
  - 88 base phones primarily from the CMU dictionary [5]
  - Other phones: Silence, speech noise, non-speech noise, filled pauses
  - HMMs based on occurrences in Wall Street Journal and TRAINS
- Task-specific language models constructed using SRILM [6]
- Speech is preprocessed into 25ms frames, shifted by 10ms

Speech Analysis Modes

- ASR
  - Phoneme-level language model from biphone probabilities in CMU dictionary [5]
  - ASR estimates utterance boundaries
  - Some extracted features:

utterance count	utterance duration (μ)
utterance intensity (μ, σ)	F0 variability (μ)
ratio of silence to speech	ratio of silence to total duration
silent pause count	silent pause duration (μ)
silent pause densities (μ)	...

- Forced-alignment
  - Manual transcription
  - Deterministic network used to force alignment
  - Similar speech characteristics are calculated, but using word boundaries
- VF-meter
  - Focused on verbal fluency tests
  - Phonemic VF language model constructed using unigram model from CMU dictionary [5]
  - Semantic VF language model constructed using bigram model of words from semantic category, e.g. animals

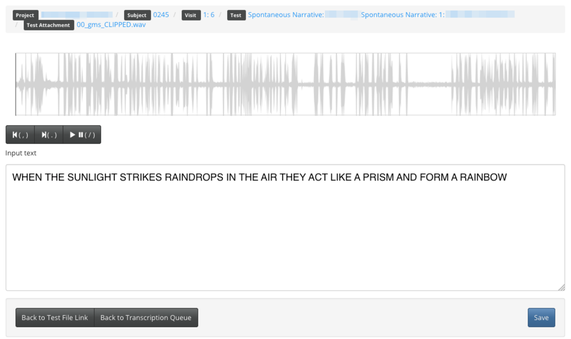

Figure 4: The transcription interface from the web application.

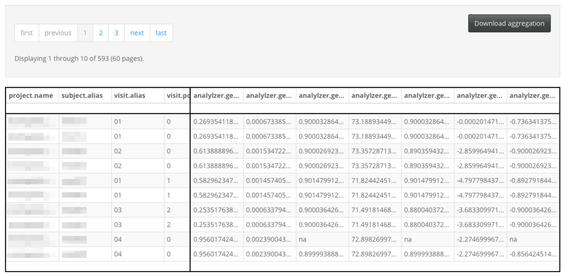

Figure 5: The web application showing analysis results.

References

[1] S. Pakhomov, G. Smith, D. Chacon, Y. Feliciano, N. Graff-Radford, R. Caselli, and D. Knopman, "Computerized analysis of speech and language to identify psycholinguistic correlates of fronto-temporal lobar degeneration," *Cognitive and Behavioral Neurology*, vol. 23, no. 3, pp. 165–177, 2010.

[2] S. Marino, S. Pakhomov, S. Han, K. Anderson, M. Ding, L. Eberly, and et al., "The effect of topiramate plasma concentration on linguistic behavior, verbal recall and working memory," *Epilepsy and Behavior*, vol. 24, no. 3, pp. 365–372, 2012.

[3] S. Pakhomov, S. Marino, and A. Birnbaum, "Quantification of speech dysfluency as a marker of medication-induced cognitive impairment: An application of computerized speech analysis in neuropharmacology," *Computer Speech and Language*, vol. 27, no. 1, pp. 116–134, 2013.

[4] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The Kaldi speech recognition toolkit," in *Proc. ASRU*, 2011, pp. 1–4.

[5] R. L. Weide, "The CMU pronouncing dictionary," 1998. [Online]. Available: http://www.speech.cs.cmu.edu/cgibin/cmudict

[6] A. Stolcke, "SRILM: an extensible language modeling toolkit." in *INTERSPEECH*, 2002.

Acknowledgements

We thank Aaron Free for his contributions to the development of SALSA, and Bartlomiej Plichta for helping with the development of acoustic analysis components. This work was supported in part by the Center for Clinical and Cognitive Neuropharmacology (C^3N) and grants from NIH – NINDS (R01NS076665), Alzheimer's Association (DNCFI-12-242985), Geoffrey Beene Foundation Alzheimer's Disease Challenge, and the University of Minnesota Academic Health Center Faculty Development Grant.