Computerized Language and Speech Analysis to Measure Effects of Anti-Epileptic Drugs on Cognition

Serguei VS Pakhomov, Kyle MV Marek-Spartz, Robert W Bill, Benjamin C Knoll, S Thomas Christie, Angela K Birnbaum, Ilo E Leppik, Susan E Marino

Center for Clinical and Cognitive Neuropharmacology, University of Minnesota, Minneapolis, USA

Abstract

Adverse effects of anti-epileptic drugs (AEDs) on cognitive function frequently manifest themselves in altered speech and language patterns that are different in different individuals. Currently, these effects are captured through patients' self-report or informal observation of their speech, resulting in coarse-grained and subjective assessments. Our goal is to develop a novel paradigm for computerized capture and assessment of naturalistic spontaneous speech to produce objective and ecologically valid measures of global effects of AEDs on cognitive function. To this end, we have designed the System for Automated Language and Speech Analysis (SALSA) and implemented it using a Client-Server architecture. The Client consists of a computer program for presenting pictures, videos, or audio description stimuli to patients and capturing their spoken responses. The audio-recorded responses are encrypted and securely transferred via a wireless network to the Server for automated processing. The Client can currently operate on an iPad or Windows-based tablets, as well as over the telephone for tasks not requiring visual stimuli. The Server consists of a database for storing and cataloguing the samples, a web-based administrative interface for viewing and manipulating the database, and a speech-processing engine based on automatic speech recognition technology. The processing engine extracts a set of measurements from the audio reflecting several speech and language characteristics including the number of utterances, speaking rate, density and length of hesitations (silent and filled pauses, word fragments, repetitions), and variability in voice pitch. We are currently using SALSA to study effects of topiramate on cognition in a sample of healthy volunteers to test the hypothesis that spontaneous speech and language characteristics are sensitive to plasma concentrations of topiramate.

System description

We developed a system to process audio recordings collected during cognitive testing. The high-level architecture is shown in Figure 1. SALSA consists of the following components:

- A set of distributed, multi-platform data collection instruments (Clients)
- A RESTful web service used to store and organize data
- A speech analysis system that interacts with the web service
- A web application for controlling the system and viewing results

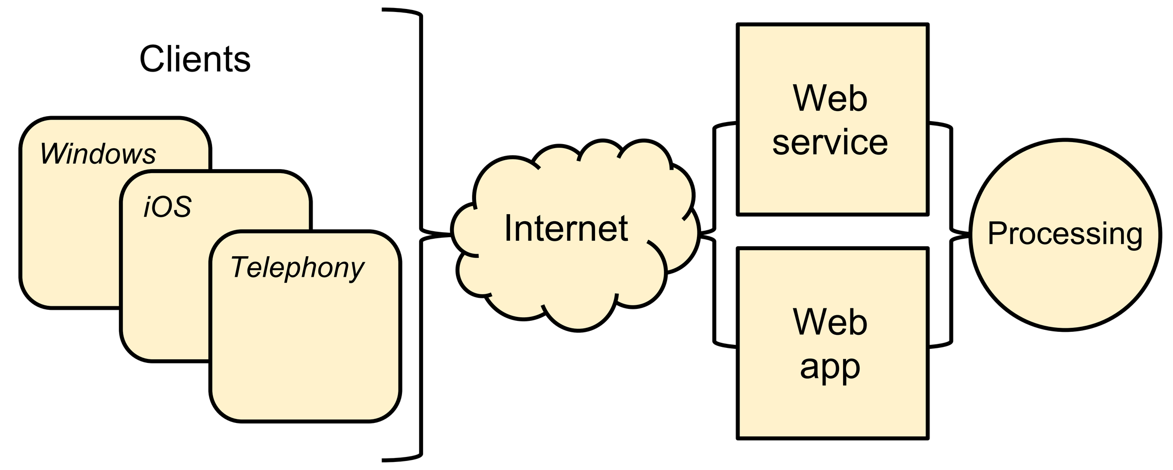

Figure 1: High-level architecture.

Assessment instruments

The data collection instruments are used by the person conducting cognitive testing. These instruments communicate with the web service over the Internet by sending audio data and receiving project and subject related information. We have implemented assessment instruments for multiple computer platforms, including:

- iOS 6 and 7 (See Figure 2)
- Windows 8
- A secondary web service providing a telephony interface

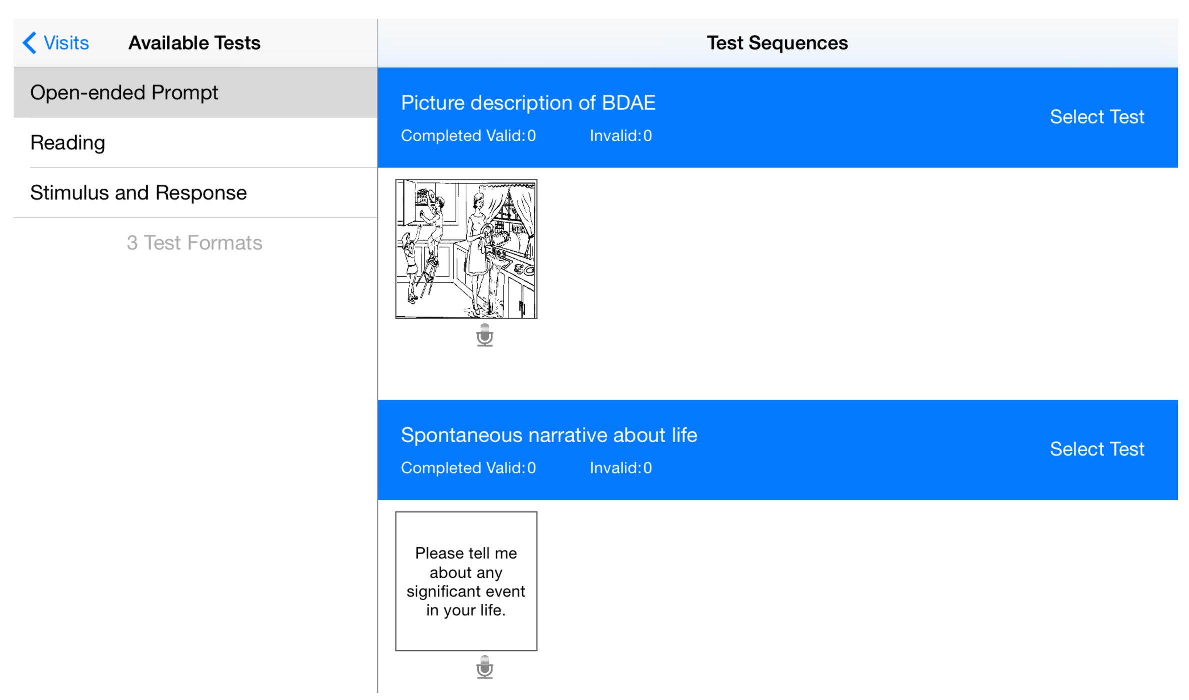

Figure 2: The iOS client showing available test prototypes for a project.

Web service

The web service component provides a distributed way of interacting with the SALSA database. We implemented a RESTful API (Fielding and Taylor, 2002) in Python using Flask web services technology. Flask uses JSON as the serialization protocol (Crockford, 2006). With the exception of direct file interactions, all request and response bodies are JSON. For security, we use HTTPS combined with HTTP Basic authentication. According to the access controls predefined in the SALSA database, queries are accepted, denied or results are filtered based on authentication credentials provided by the user.

Speech analysis

To analyze audio recordings, the speech analysis component relies on open source software including KALDI (Povey et al., 2011) automatic speech recognition toolkit, SRILM (Stolcke, 2002) language modeling toolkit, Sox audio utilities package, R statistics package, and Praat (See Figure 3) acoustic analysis package. After processing is completed, the processing component reports results back to the web service component so that they can be accessed by the end users. The processing component generates a number of variables that correspond to various prosodic and temporal characteristics of the speech being analyzed. Below is a select set of speech variables produced by SALSA:

- Total utterance count
- Mean utterance duration
- Mean utterance intensity standard deviation
- Mean F0 - variability in voice pitch
- Density of short (100 ms) silent pauses
- Density of medium (1000 ms) silent pauses
- Density of long (2000 ms) silent pauses
- Mean silent pause duration
- Mean word duration
- Ratio of silence to speech
- Ratio of silence to total duration
- Hesitation count (sum of raw filled pause count, raw fragment count, and single repetition count)
- Hesitation rate (ratio of hesitations to total utterance count)
- Speaking rate (phones per second)

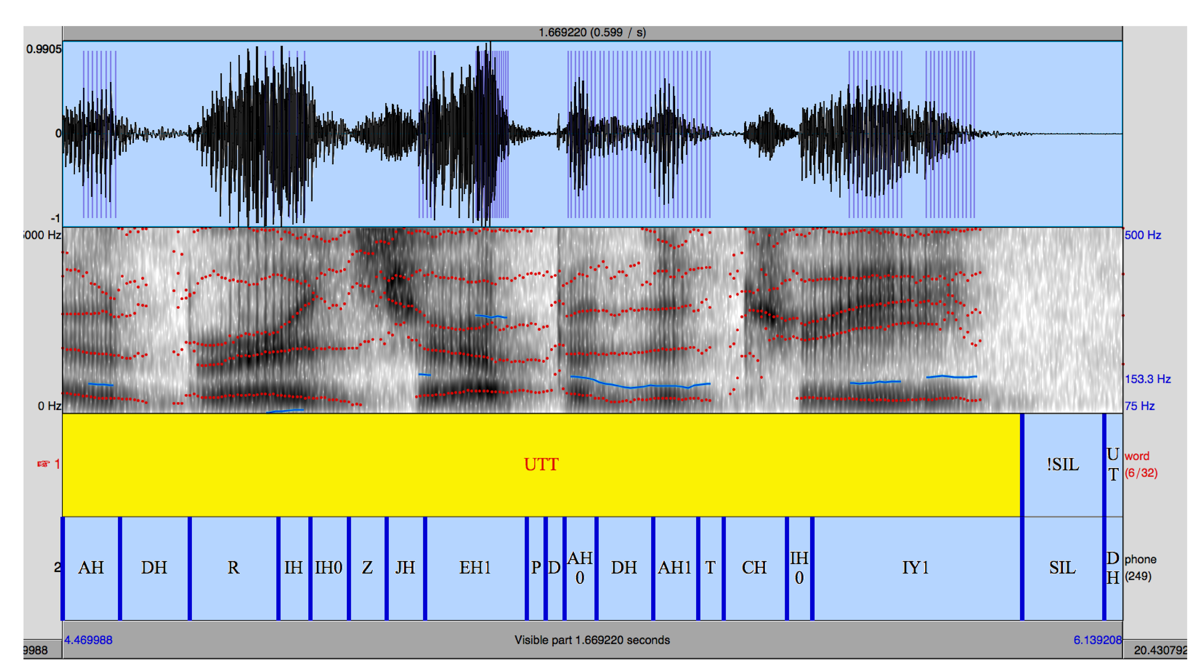

Figure 3: Example of speech recognizer output time-aligned with audio. This alignment is used to calculate SALSA variables.

Web application

The web application component (See Figure 4) is used for project administration and transcription. It provides for the following capabilities:

- Adding and removing projects, subjects and tests
- Querying to view projects, participants and test recordings
- Defining tests (including stimuli and timing parameters)
- Transcribing audio recordings
- Initiating batch analyses
- Viewing speech analysis results

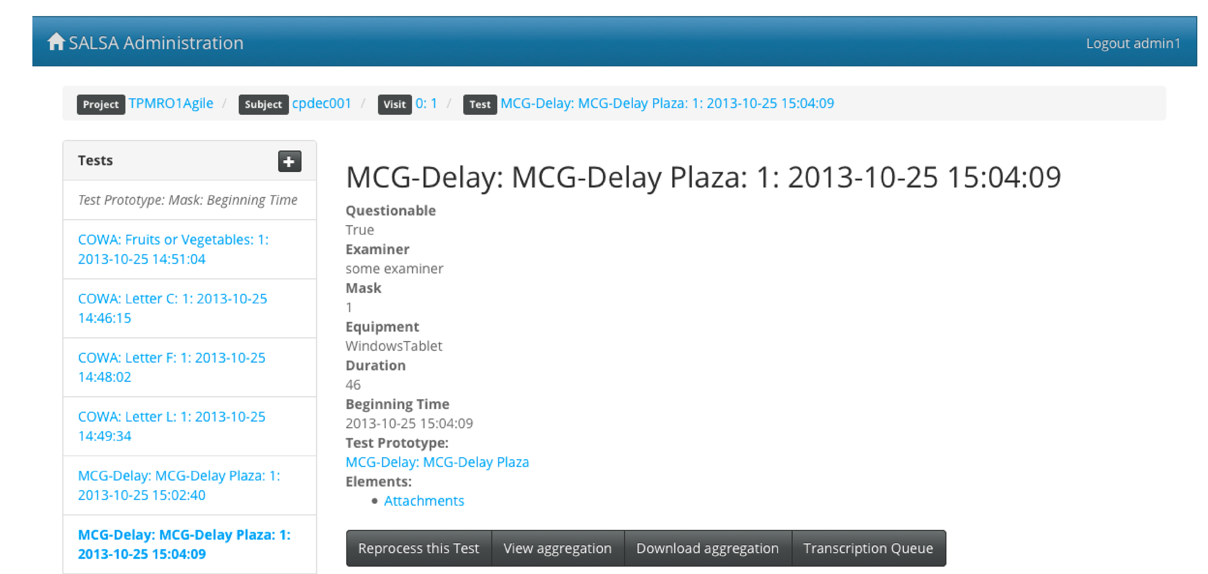

Figure 4: The web application showing information about a test.

References

Crockford, D. (2006). The application/json media type for JavaScript object notation (JSON). RFC 4627, IETF.

Fielding, R. T. and Taylor, R. N. (2002). Principled design of the modern web architecture. *ACM Transactions on Internet Technology (TOIT)*, 2(2):115–150.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., et al. (2011). The Kaldi speech recognition toolkit. In *Proc. ASRU*, pages 1–4.

Stolcke, A. (2002). SRILM: an extensible language modeling toolkit. In *INTERSPEECH*.

Acknowledgements

This work was supported in part by 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.