



SRI LinguaKey Active Authentication

Aaron Lawson

SRI International

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA).

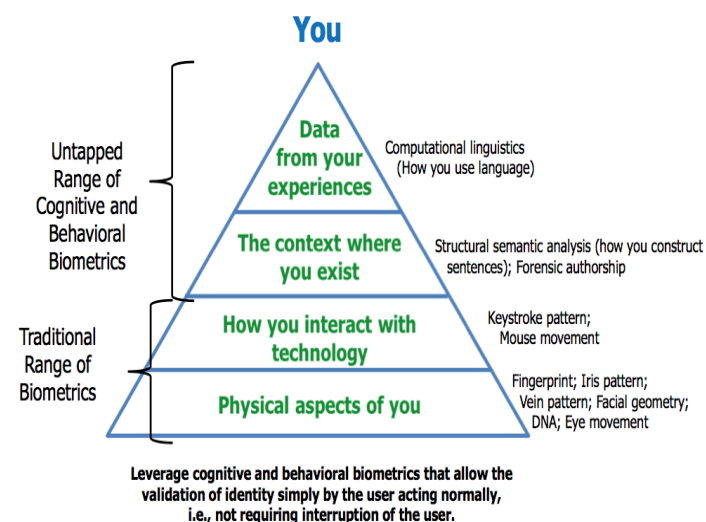
Disclaimer: The views, opinions, and/or findings contained in this article/presentation are those of the author(s)/ presenter(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

“Distribution Statement A – Approved for Public Release, Distribution Unlimited”

Active Authentication Motivation

- Program has the goal of providing continuous authentication of DoD users on government owned computers and mobile devices.
 - Currently users log in once, generally with a CAC Card, and if they leave their computer anyone else can come by and use it.
 - Passwords have also become ridiculously long and impossible to remember, which is a security risk.
- This program will develop alternate means of authenticating users based on their behavioral, cognitive and physical attributes
 - Primary focus is on behavioral and cognitive
 - Research on biometric modalities that will not interrupt a mobile user's normal operations and can be validated by empirical testing,

Finweb = Jane123
DTS = 123Jane
PKI = JaneA123
DiskCrypt = Jane123A
Gmail = Jane123A

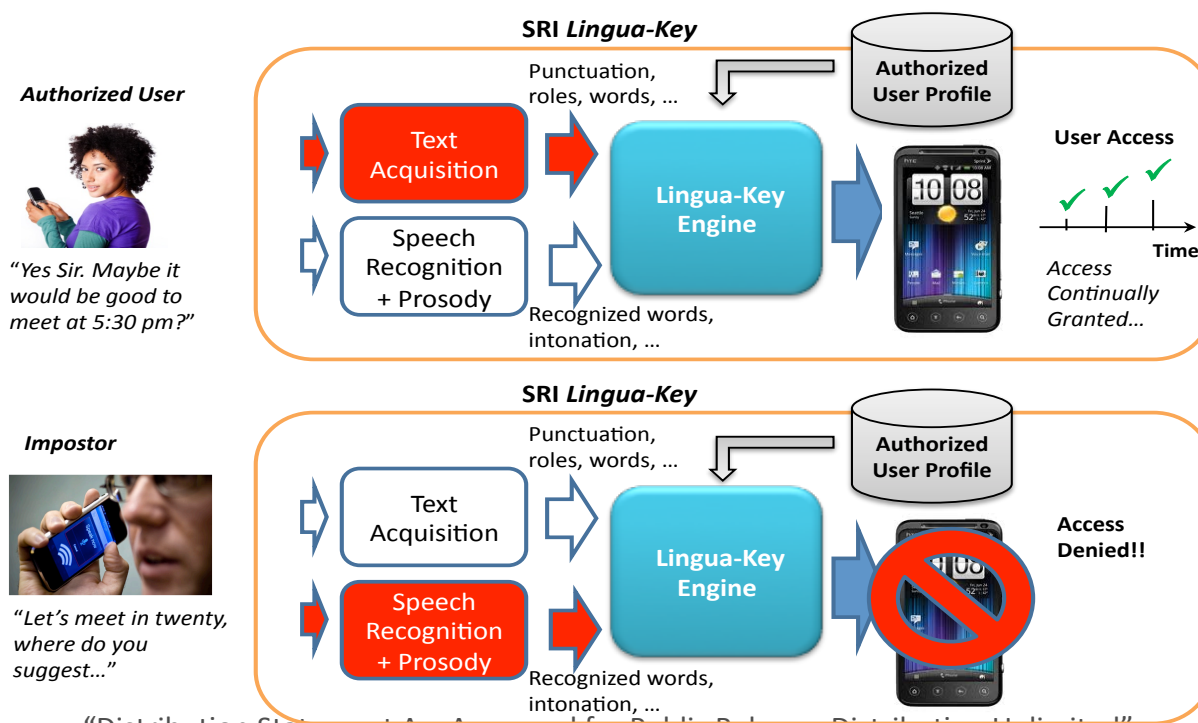


Continuous authentication of mobile device users through their language usage

- Extract key socio-linguistic features shared by both spoken and written language use that provide the capability to characterize a user's personality and background for purposes of user authentication; develop a capability to predict traits of one modality based on behavior observed in the other.
- Identify common “speech acts” (apologies, agreement, questions, etc.) as frames within which user behavior, cognition and style of interaction will be more effectively modeled.
- Model speaker distinctive spoken language behavior, with emphasis on prosody, filled pauses/backchannels, word level stress, intonation contours and spectral features.
- Develop modeling techniques to generate continuous, joint log-likelihood ratios for state of the art user verification.
- Integration of software to acquire speech and text on mobile platforms and software to perform continuous authentication from these inputs
- Data collection of spontaneous written and spoken language use on mobile platforms for system validation and research.

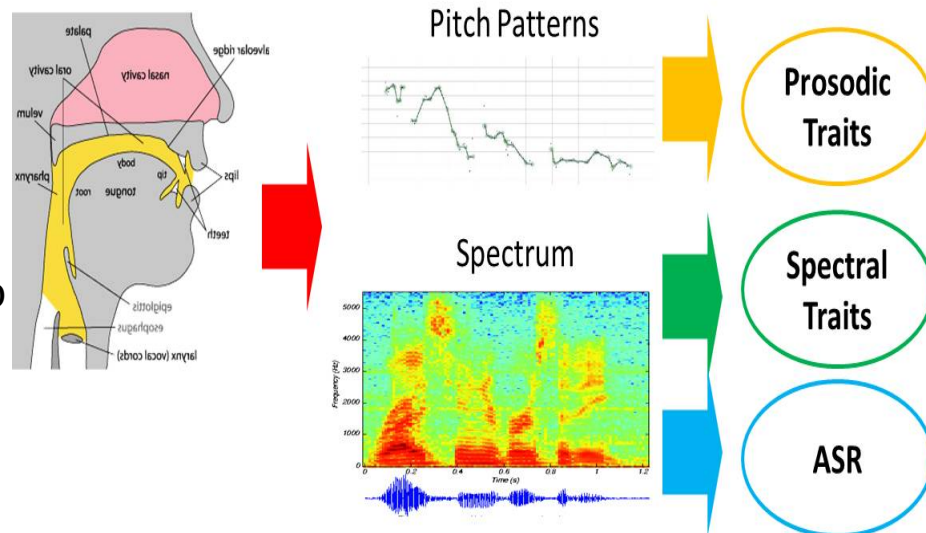
SRI's LinguaKey will provide authentication from spoken and written language usage

- Goal of LinguaKey is continuous authentication of mobile device users through their language usage, through a software package that authenticates users based on their spoken and written inputs.
- What we propose goes beyond speech and text stylistic biometrics to tap into core cognitive and behavioral traits in language usage.



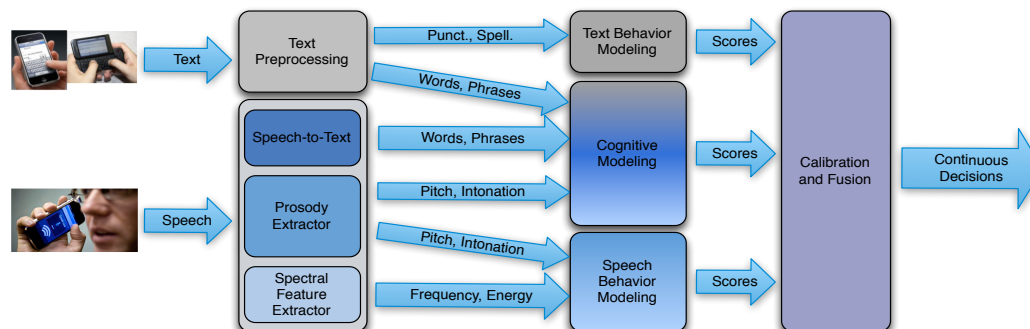
Overview of SRI Effort

- Research
 - Speaker ID
 - Spoken language features
 - Text stylometrics
 - Socio-linguistics traits that transcend speech and text
- Integration/Software development
 - Tools to capture speech and text on mobile device
 - Integration of research algorithms into biometric server that continuously authenticate user on mobile device.
- Data collection
 - Spoken and texted conversations
 - Spoken “PDA” interaction with the device
 - Business related texting
 - Business related emails



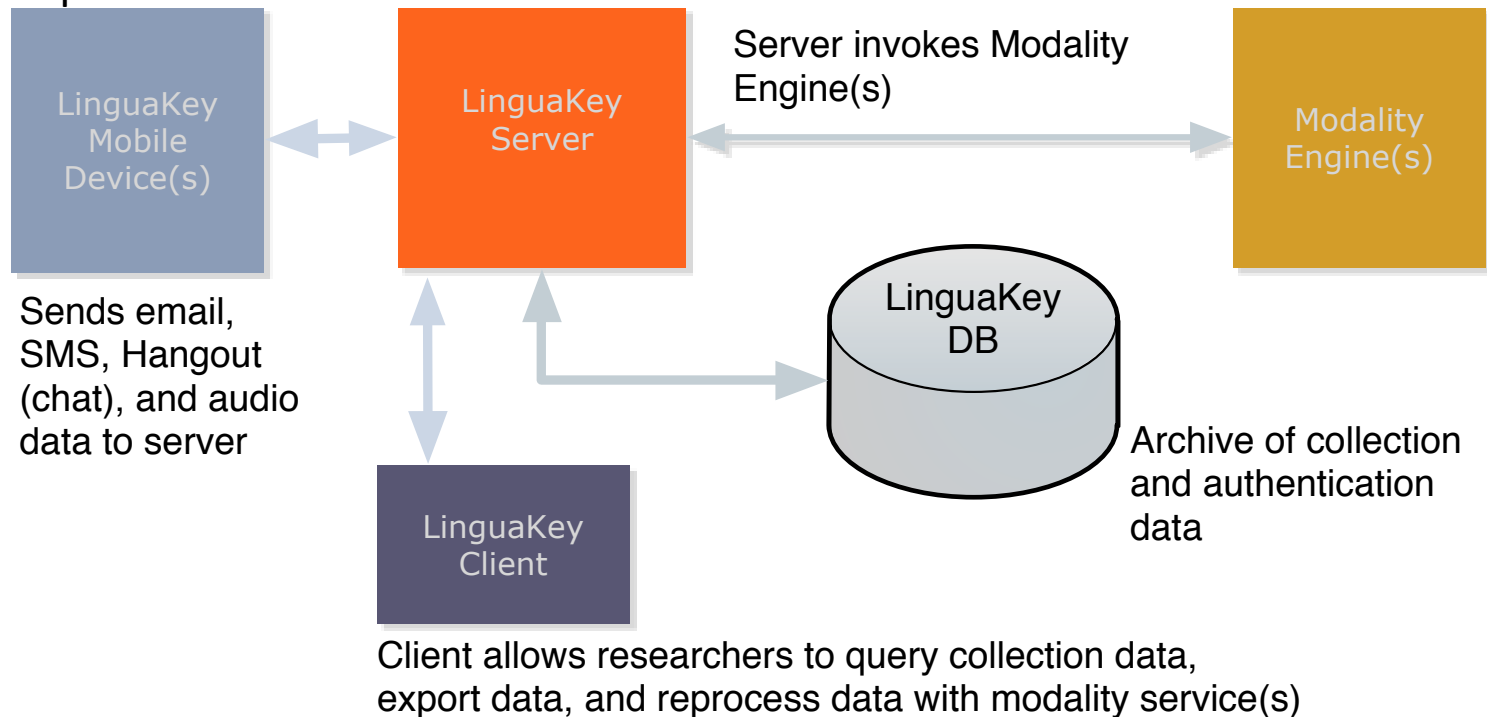
Behavioral and Cognitive Traits

- SRI will focus on behavioral and cognitive traits, in keeping with DARPA's goals.
- Behavioral dimensions are the idiosyncrasies of how users actually produce spoken or written language.
 - In speech this includes factors such as *intonation*, *speech rate*, and *speech energy/frequency* and traditional speaker identification (SID).
 - In writing, behavioral features will target *spelling errors*, *punctuation style*, use of *abbreviations*, *sentence* and *word length*, among other factors.
- Cognitive features are from sociolinguistic activity providing clues about the individual user's background, personality and approach to interpersonal relations.
- Language use is highly colored by social background, with influences from gender, age, ethnicity, native language, region, level of education, social class, dialect and others.
- We further leverage the rich information found in frames, based on how users accomplish intents.

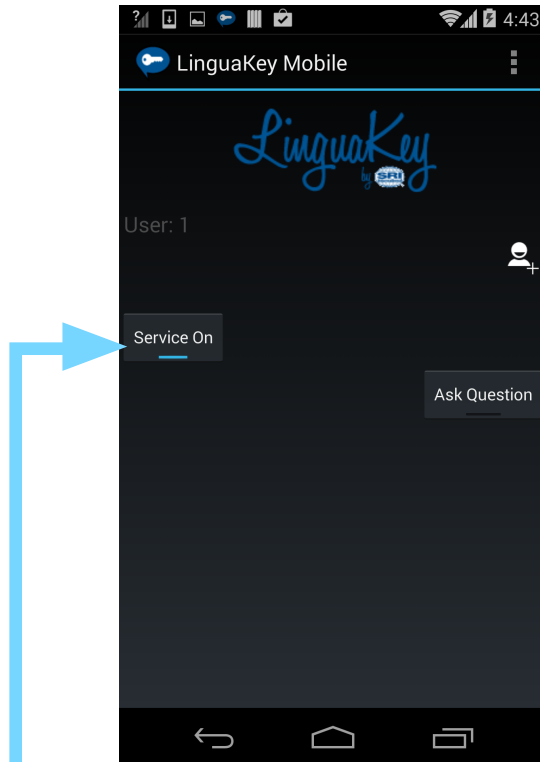


Software Overview and Architecture

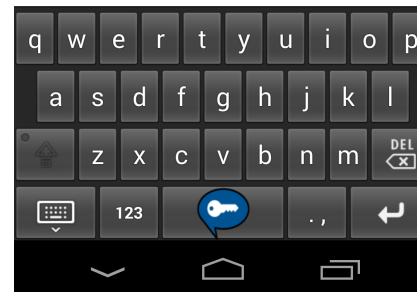
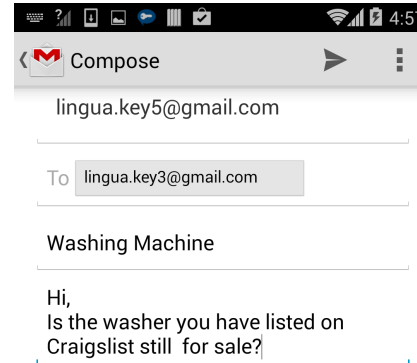
- Our approach to integration follows the opposite timeline to most research projects: we get the pipeline set up first and ready to demo, and then use this to develop and validate research concepts.
 - This approach required the development of custom LinguaKey software that was used to collect and store data, provide user authentication, and by researchers to improve authentication results.



Mobile Interface



Turn the LinguaKey
"Service On" to enable text
and audio data collection



Data collection is non-intrusive to
user. Custom LinguaKey keyboard
shown entering an email message

Feature Research

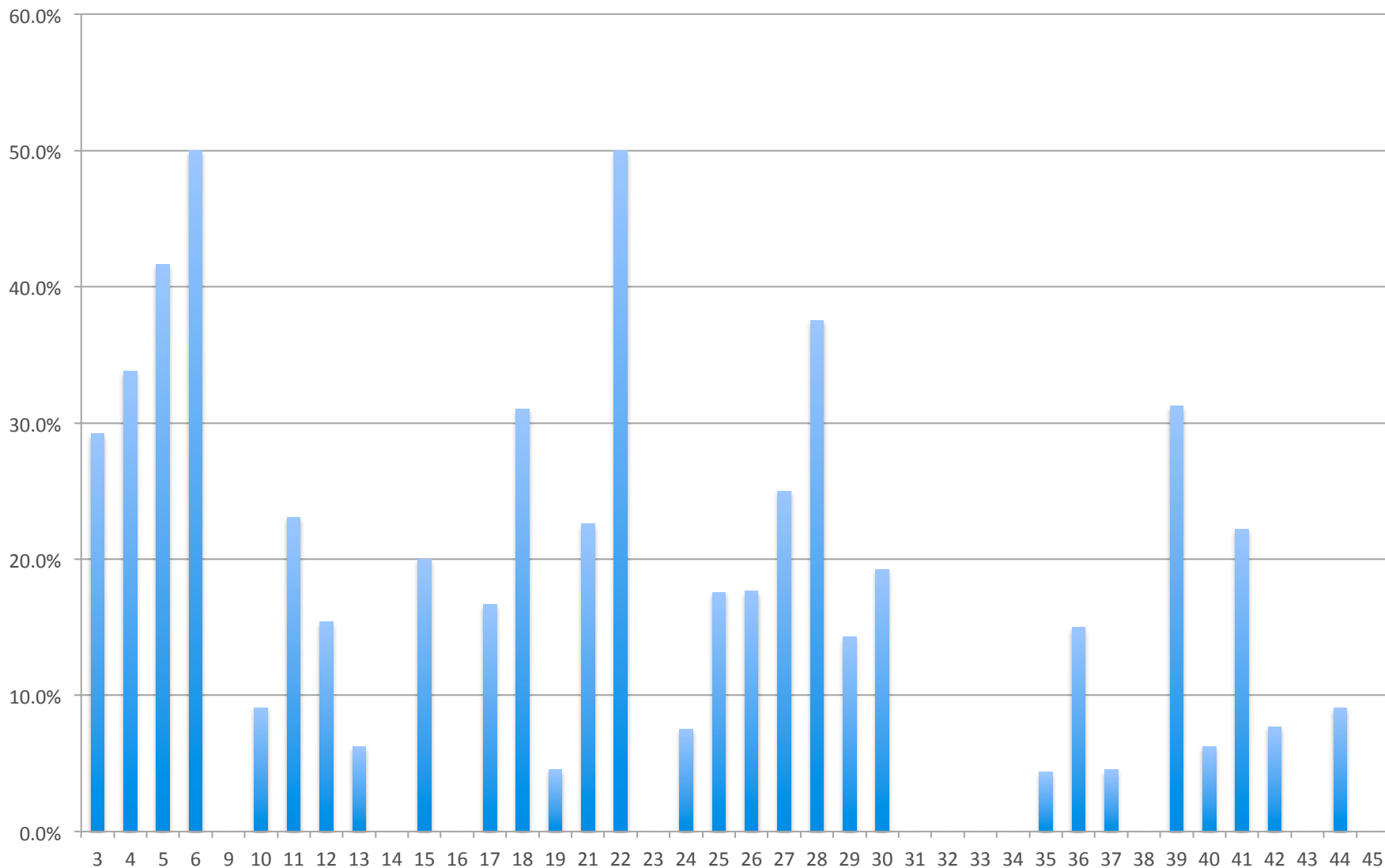
- We have implemented a large set of features covering everything from basic behavioral traits (punctuation, capitalization, word length, etc.) to higher level features focused on word semantics, emotional content, and language frames.
 - Speech specific: pitch, intonation, speech rate, cepstral
 - Text specific: typographic, orthographic, corrections
 - Language general: words, speech act frames, semantics, socio-linguistic trends
- Further goals are to understand and take advantage of entailment –the fact that features are correlated across individuals because they are associated with the same set of personality and background traits

Some Features Close-up

- Part of our research program is to look at how features interact to predict higher level aspects of a person's cognitive and behavioral traits, traits that effectively characterize users but about which they have little control
- In the next few slides we'll look at a few simple features and then see how they fit in to our framework, interacting and correlating with other features
- Our goal is to combine our 100 or so features into groups that correlate with higher level personality, cognition or demographic factors

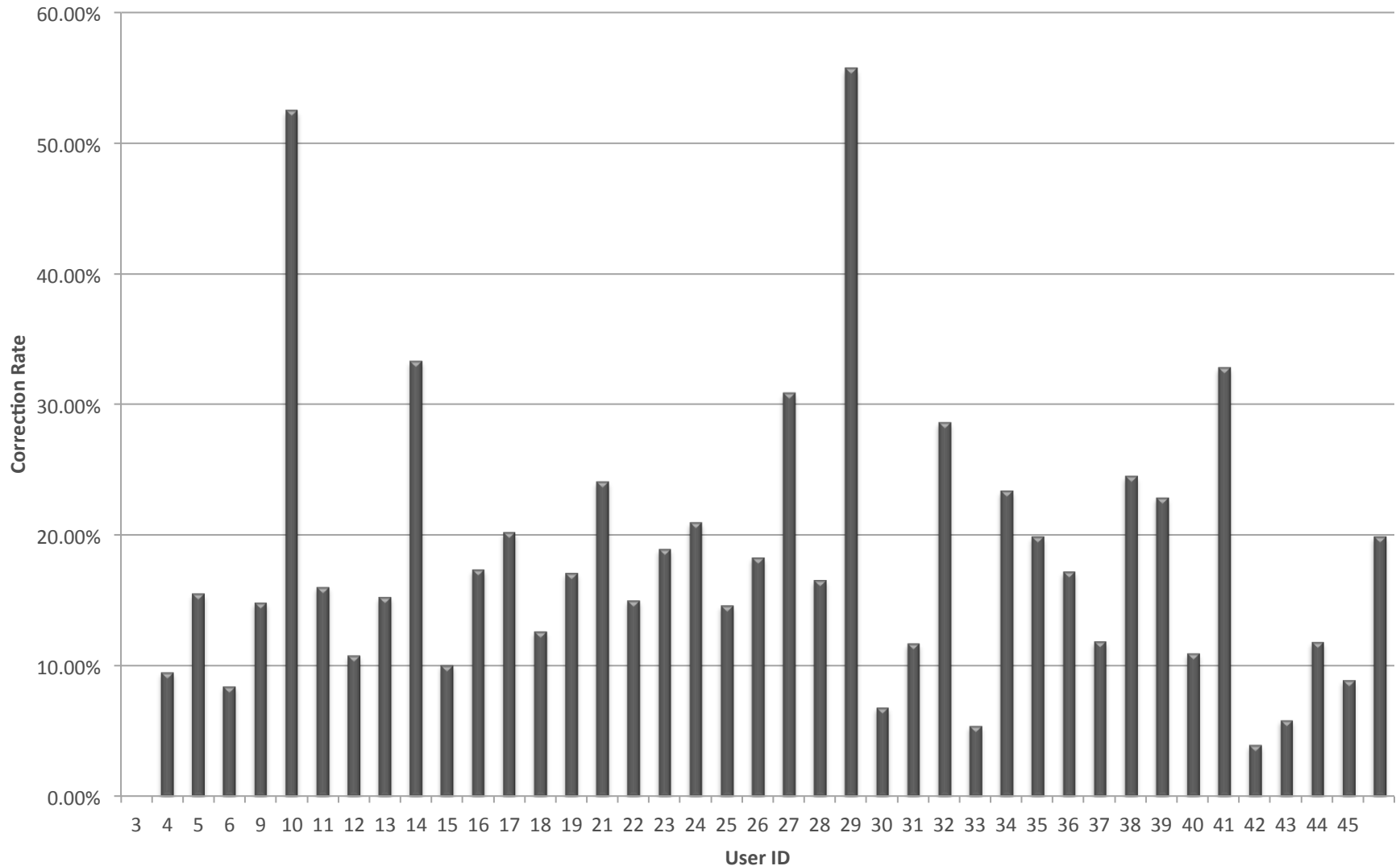
High Variability in Starting a Sentence with a Lower-case Letter

Probability of Lowercase Sentence Start

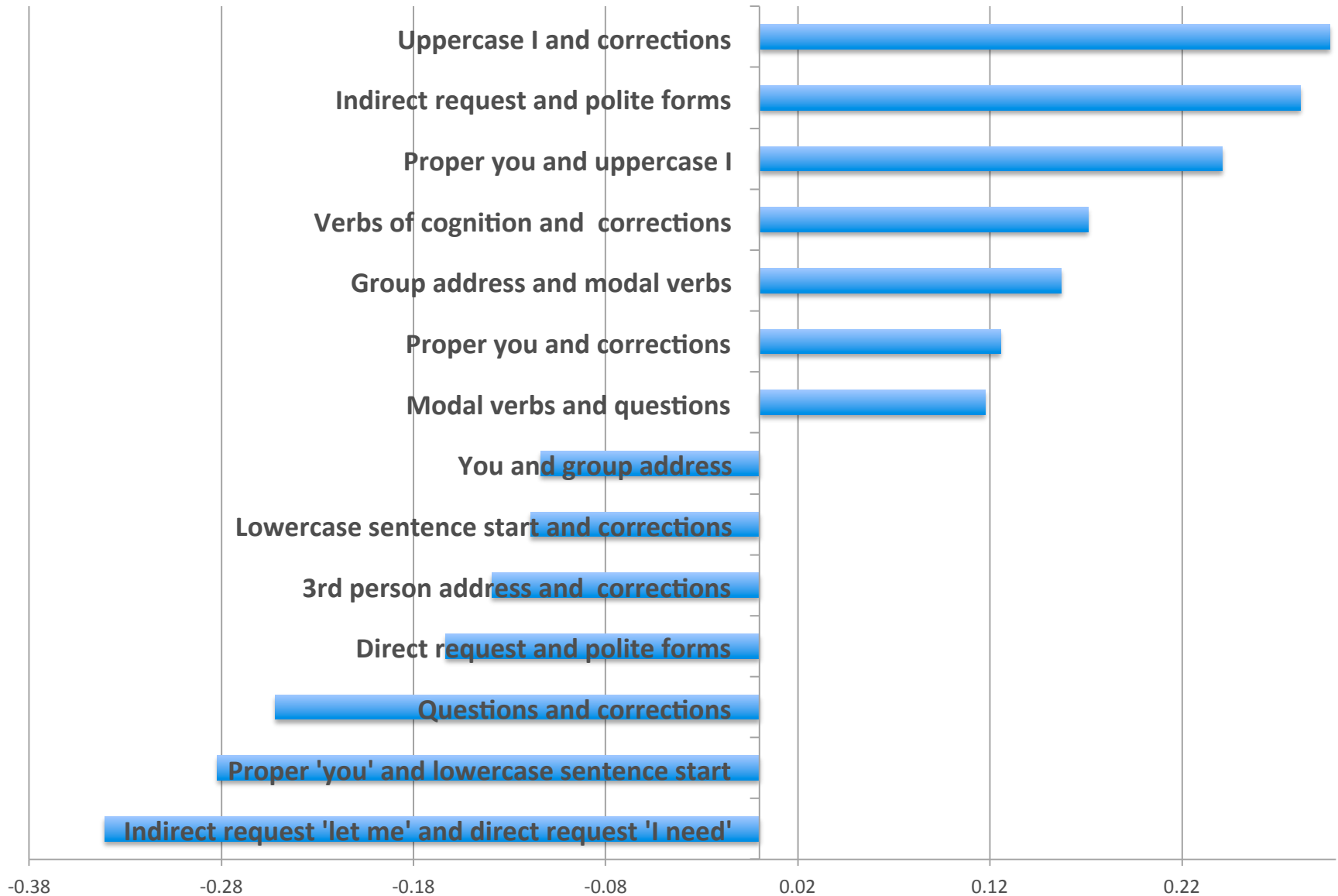


Very High Variability in Correction Rate for Text Input

Variation in Correction Rate for First 40 Enrollees



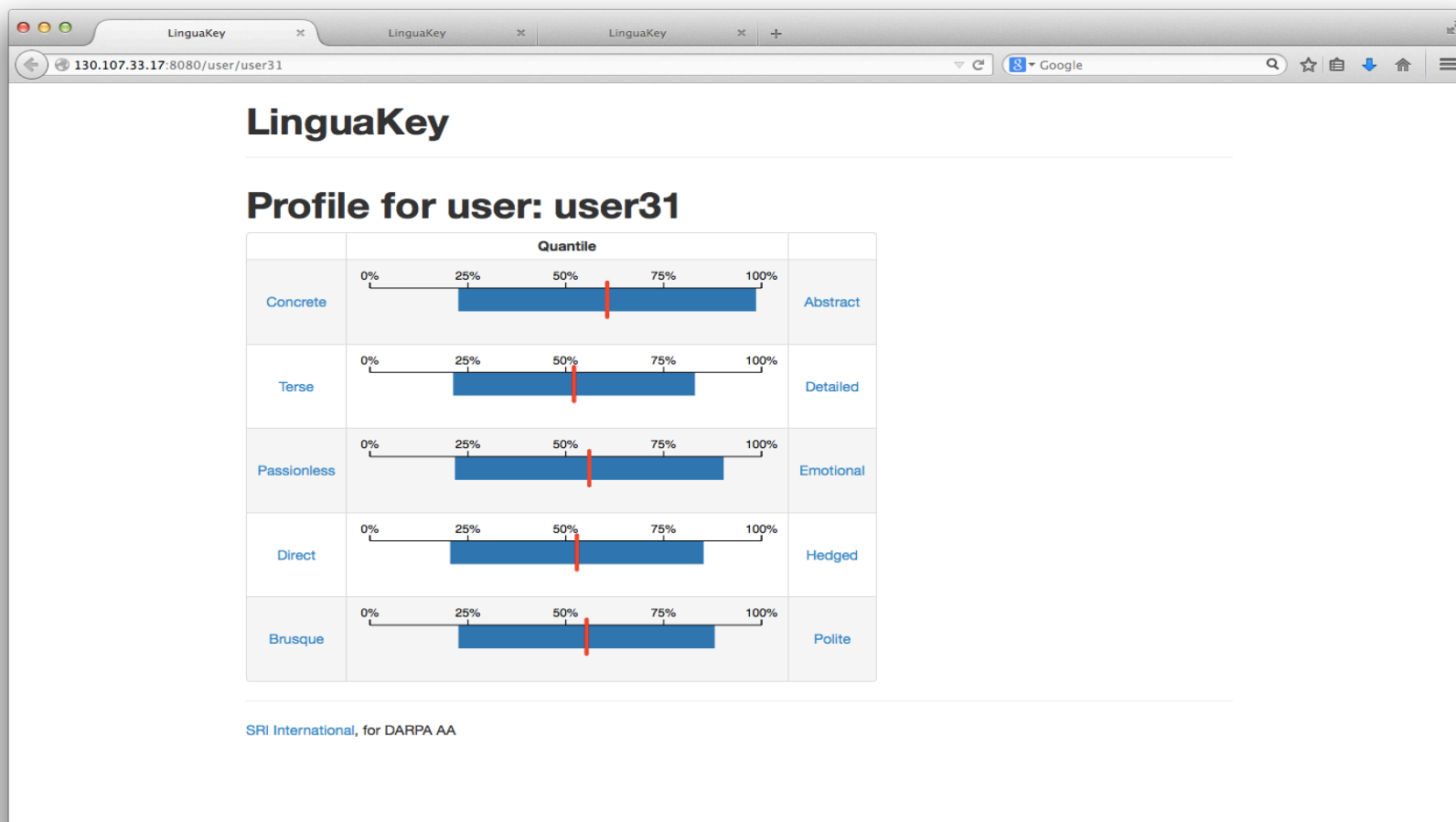
Implication Relationships between Linguistic Features



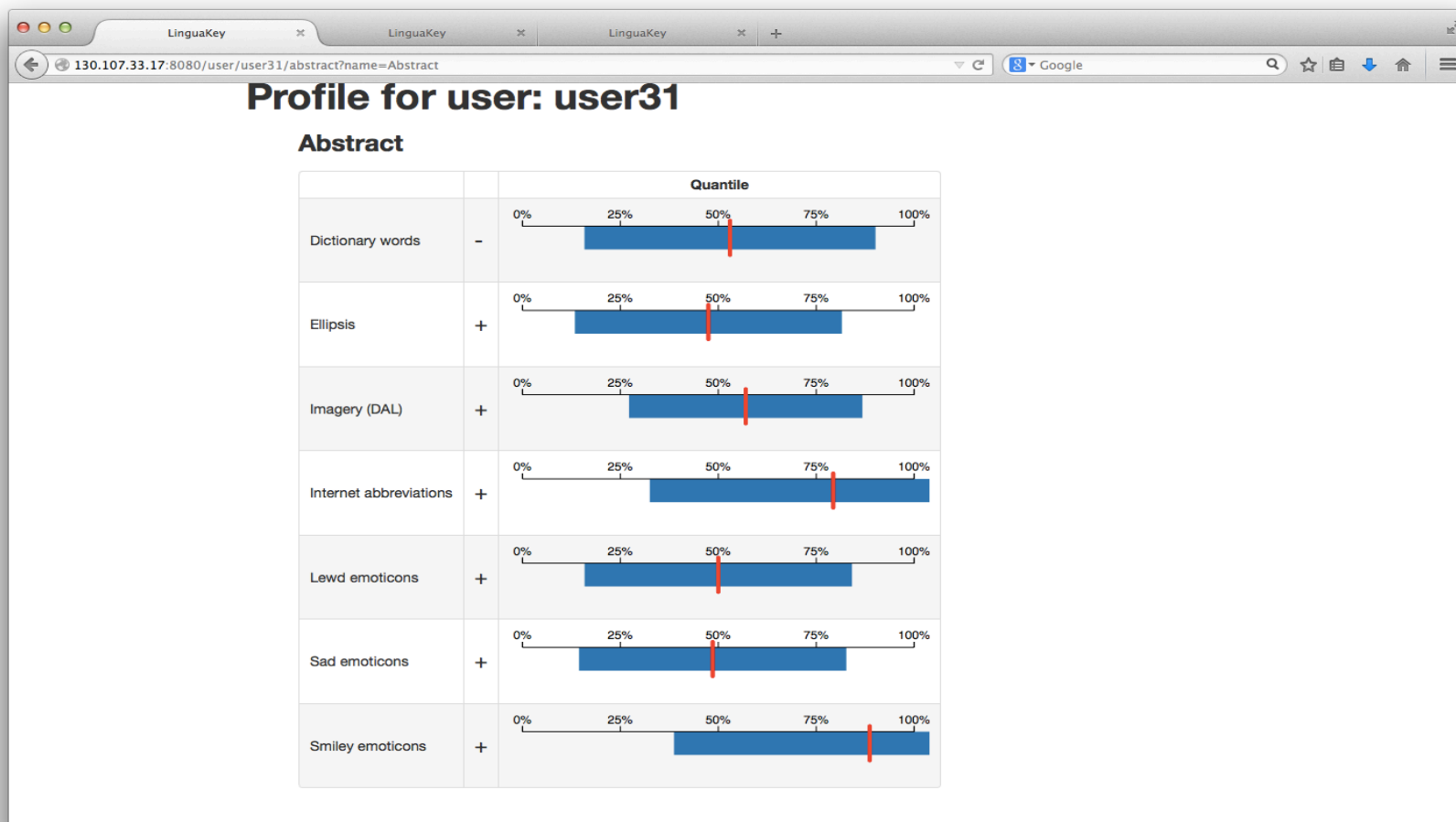
Voice Authentication

- The SID system used for verification in LinguaKey is optimized for short duration test snippets in noisy background conditions.
- Need to determine the proper test segment size, to balance continuous authentication with low error.
- Using one conversation (2.5 minutes) from each SOMLUC participant to enroll and one to test gives an equal error rate of 0%.
- Reducing test data to a single SIRI query (2-3 seconds) increases the false reject rate considerably (5-7%).
- At just 10 seconds of test data authentication is much more reliable.
- We will show a quick demonstration of the voice authentication process to give a sense of what works.

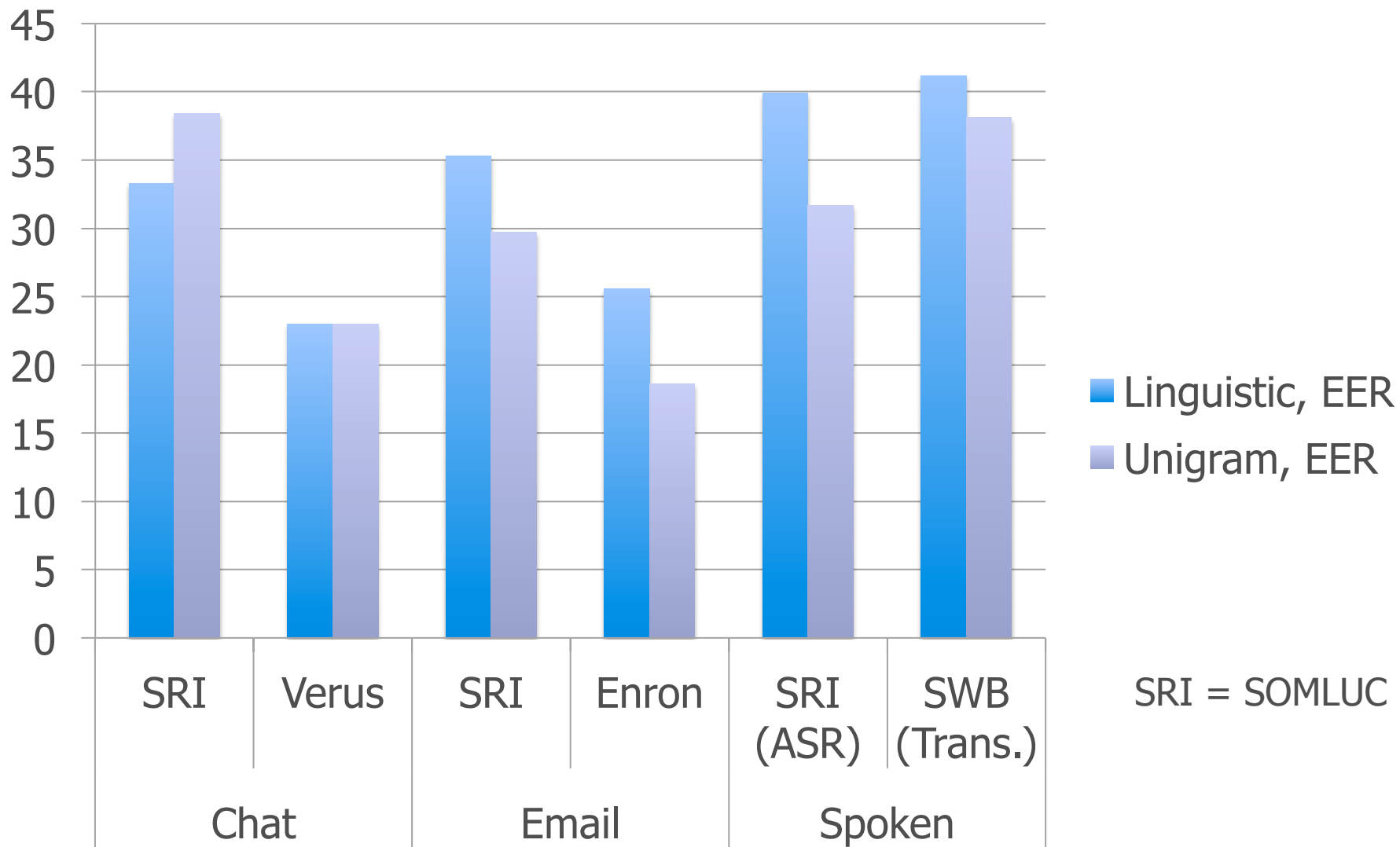
User 31 Profile



Abstract Features –Profile user 31



Modeling and Scoring-some results



Results and Conclusions

Modality	Data Type	Duration	Error (n=100)
Voice	Conversational	≈60 seconds	0% EER
	Conversational	≈10 seconds	3% EER
	“SIRI” queries	≈4 seconds	5% EER
Written	Email	≈60 seconds	18% EER
	Chat	≈10 seconds	25% EER

- Text/Word scores have much more error than voice biometrics on very short time frames
 - Text and words requires much longer time frame to generate the same amount of user-discriminative information
 - Also, there has been 25 years of continuously active research on voice biometrics
- There is still much more to do in this program on modeling, including fusion of different features, additional features, and fusion with voice

Next Steps

- There is still much more to do in this program on modeling, including fusion of different features, and fusion with voice
 - Combine cepstral speaker identification scores with linguistic and prosodic scores into a single set of scores for spoken.
 - Fuse linguistic and lexical features
- Linguistic features out perform lexical for chat, but not for email
 - May be 'red herring' due to sigs, but there is also some good information we are missing