

# Eliciting evolving information structure and audienceless vs. audience oriented prosodies: experimentation on Android tablets

Gina Cook<sup>a</sup> & Siddhartha Kattoju<sup>b</sup>

gina.c.cook@gmail.com

siddharthakattoju@gmail.com

<sup>a</sup>Department of Computer Science and Software Engineering, Concordia University

<sup>b</sup>Department of Electrical and Computer Engineering, Concordia University

## Research Questions

- Is it possible to use a tablet for phonetics experiments?
- How is the data collected in a lab context, different from a real-world context?
- Are prosodic cues (i.e., pitch, duration, loudness, and voice quality Breen et al., 2010) minimal in a stream of consciousness dictation (audienceless) as compared to a podcast dictation (audience-oriented).

## Background

Evaluating speaker's prosody in an experimental setting is a difficult task which requires speaker and listener to truly participate in a conversation (correct and augment each other's background knowledge) yet allow for **quantitative analysis** of the speaker's production of the same words under varying information structure (focus, contrastive focus or givenness) in minimally different utterances.

We present a number of **Open Source application modules** which are designed to aid researchers in conducting structured experiments to discover contextual influences on prosody in settings where users are blind to the experiments underlying their activities. We present one configuration of the modules which allow users to **publish a blog and podcast on the same topic**, as is currently the practice among many popular bloggers. Like comprehension scores, the **user's engagement** with the application can be measured via usage frequency, time actively using the application and quality/quantity of results (blogs/podcasts) produced.

## Methodology

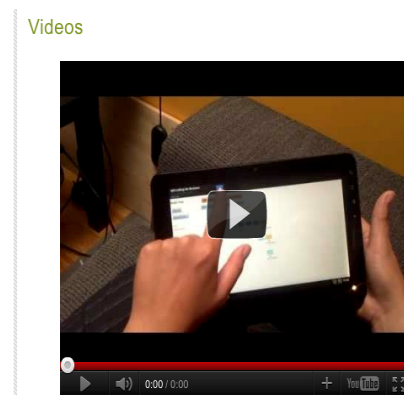
We put the App on the Market to measure user engagement and gather data in a “living room” context. We also **tested the App in a normal laboratory setting**.

Google analytics and the transcription server logs were used to gather data on user engagement.

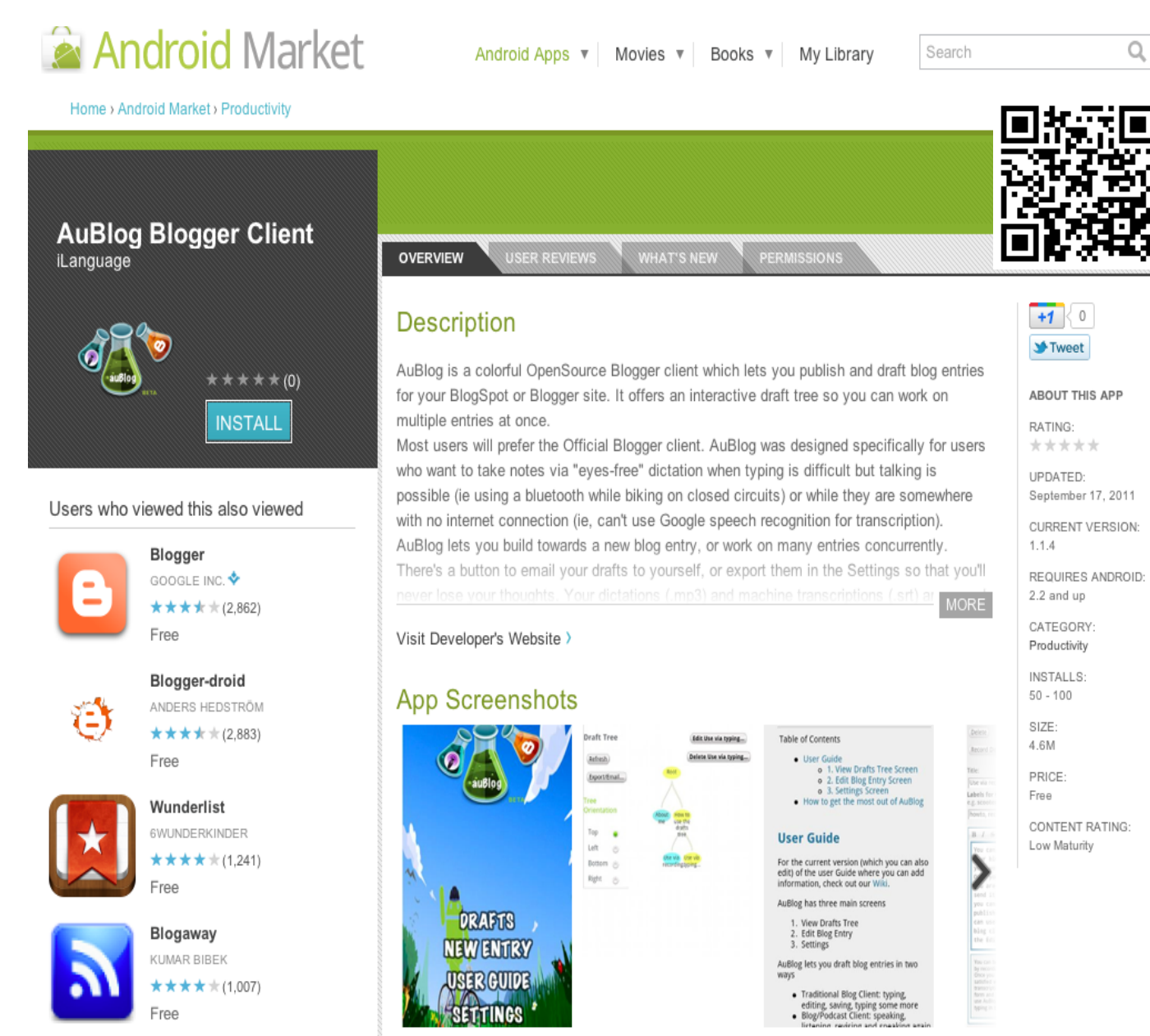
### Lab Procedure (15 minutes total)

Participants were instructed that they were going to create a blog entry, on any subject they would like, in their first language. Materials: a Web Browser, YouTube and a PDF software which can annotate and read PDFs out loud.

- Watch 2 demo videos
  - How to publish a Blog entry (1 min)
  - How to dictate a Blog draft (2 min)



Videos



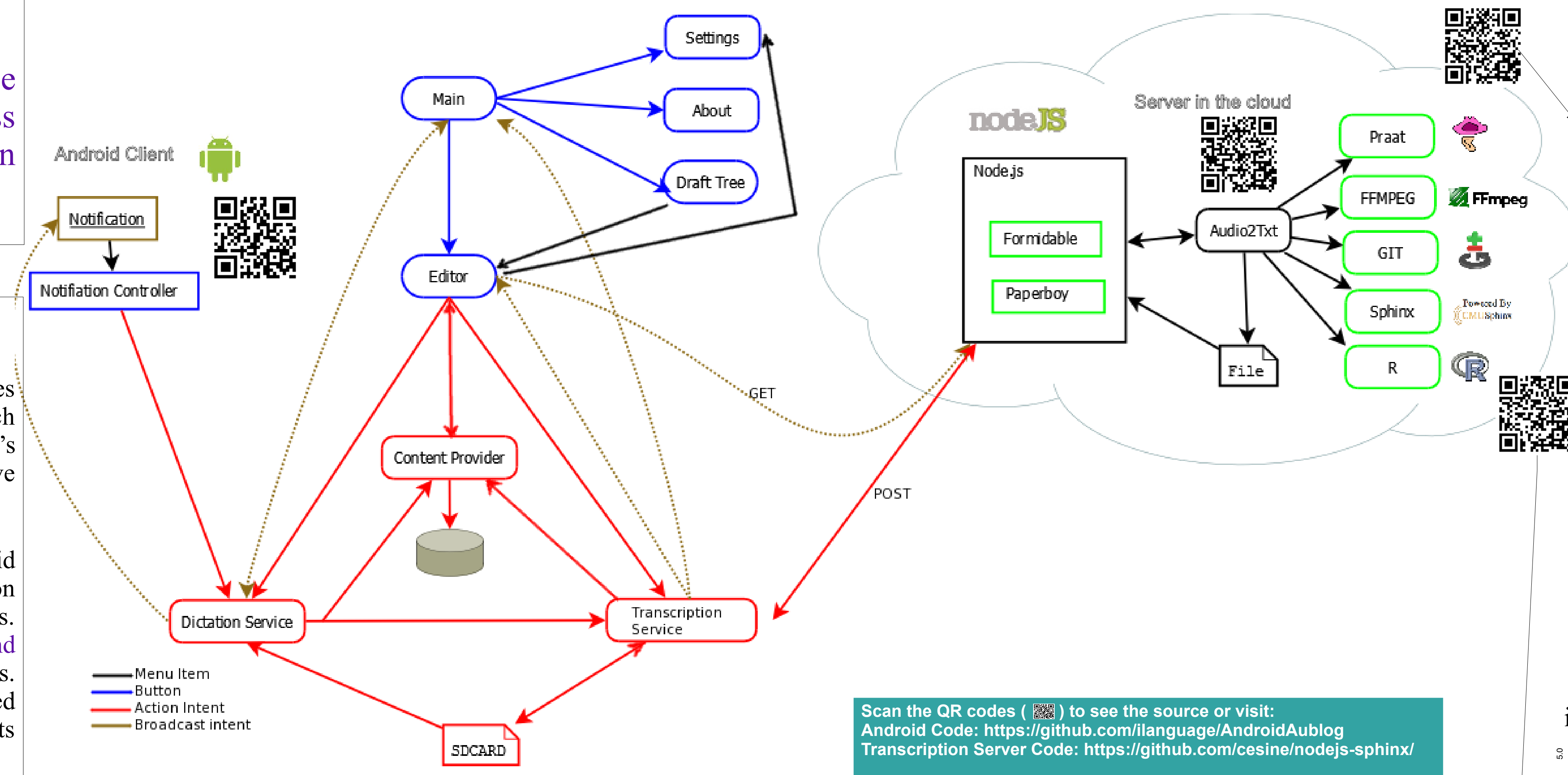
### Non-Lab Procedure

The App was placed on the Android market with colorful screenshots, 2 demo videos and detailed list of features. A website was created which featured the User Guide.

Upon install of the App, the user was presented with a “What's new” dialog which told them fun features of the app, and how to turn off the user tracking feature.

On the main screen of the app the User Guide was prominently featured.

In the Drafts Tree, a sample drafts tree was provided which contained also instructions on how to use the App.

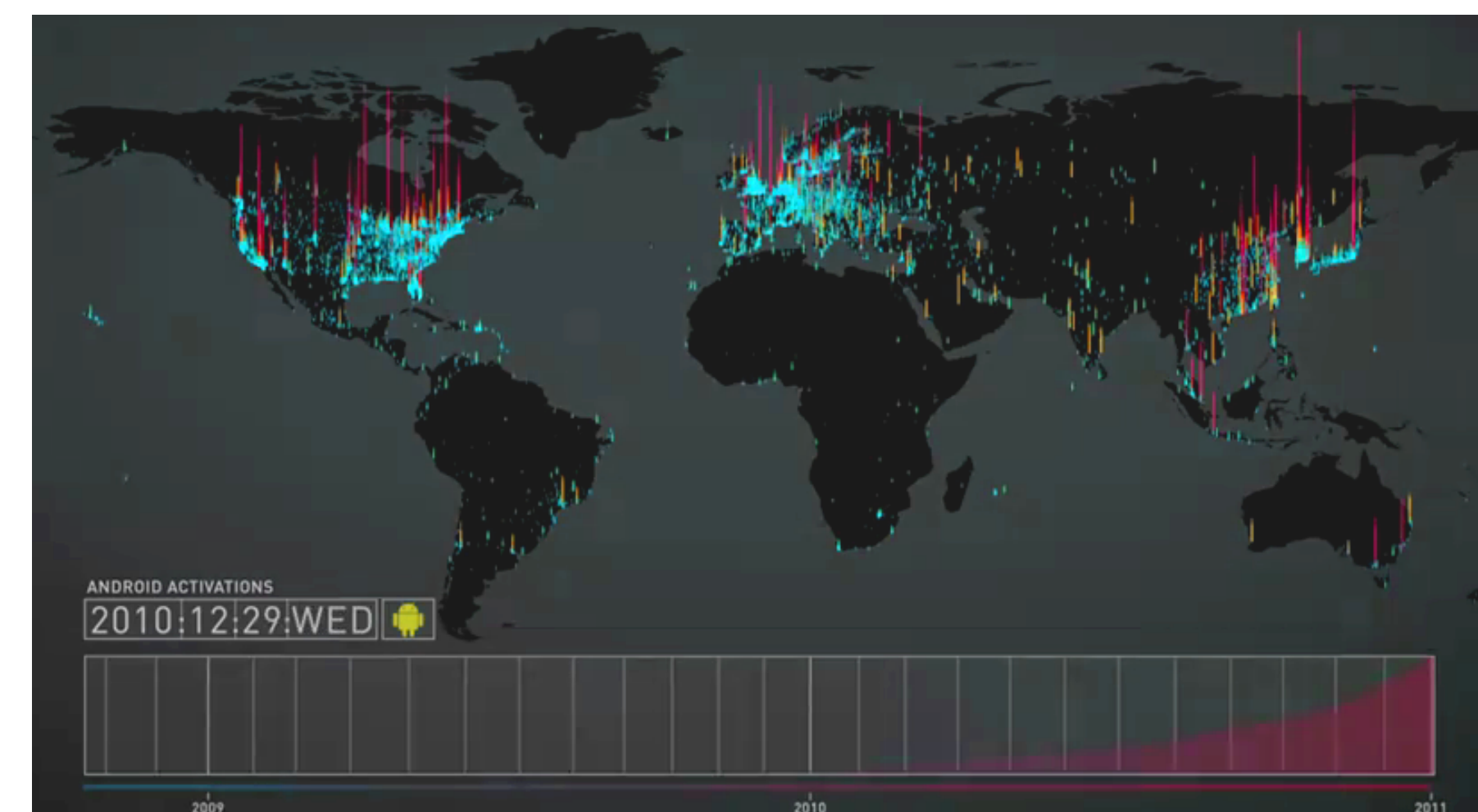


### Lab Participants

113 LING 201 Students were told they could participate in “a linguistics experiment on a tablet”

48 Students signed up to be tested

12 Students came to be tested



### Non-Lab Participants

10+ million users with a significant number in nearly every country and unprecedented access to minority language speakers

55 users downloaded the App

22 users still have the App installed  
 8 Malay/Indonesian  
 7 English  
 1 Polish  
 1 Swedish  
 1 Hindi  
 1 French (France)  
 1 German (Austrian)

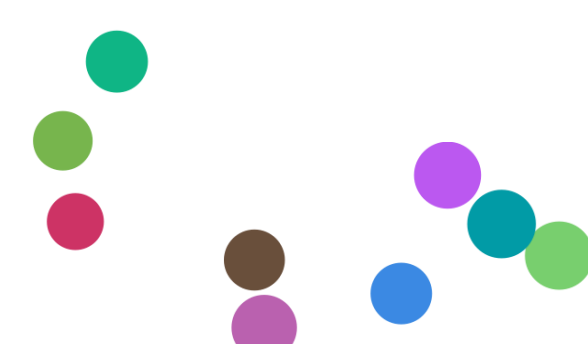
5 Students completed a Blog entry

3 English  
 1 Romanian  
 1 Mauritian Creole

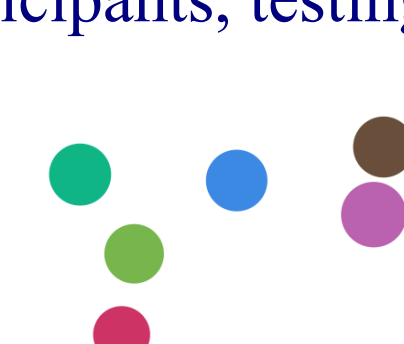
It took 2 weeks to recruit participants, and test them

0 users used the app to create a Blog entry

It takes an unknown amount of time to recruit participants, testing them takes no time.



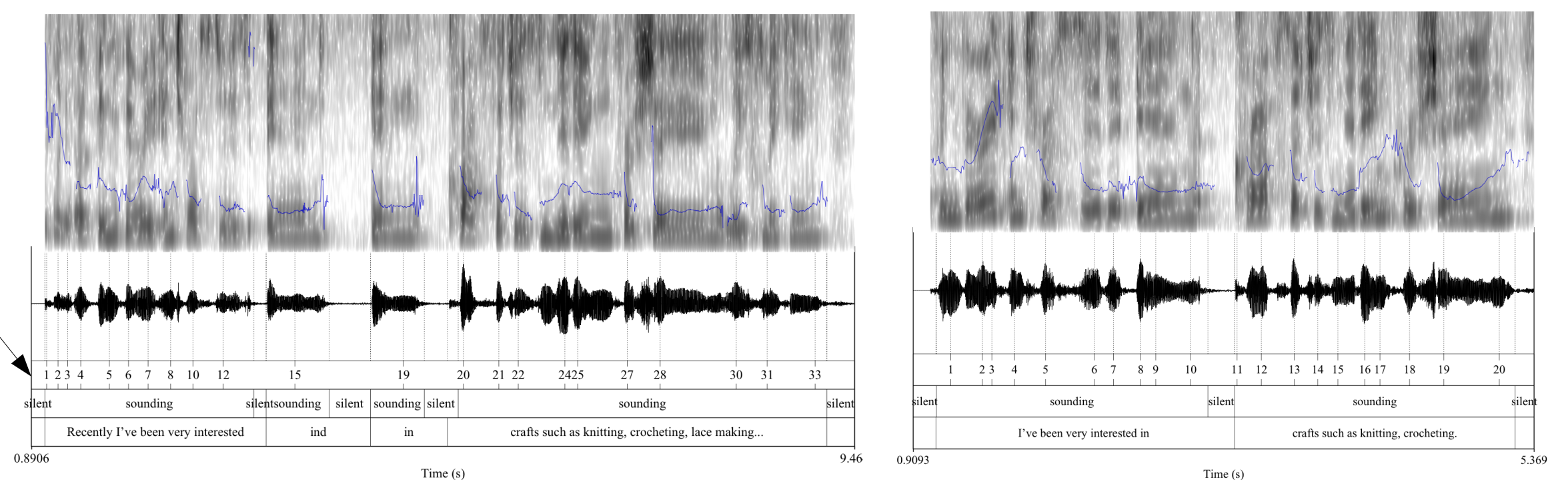
To Email authors:



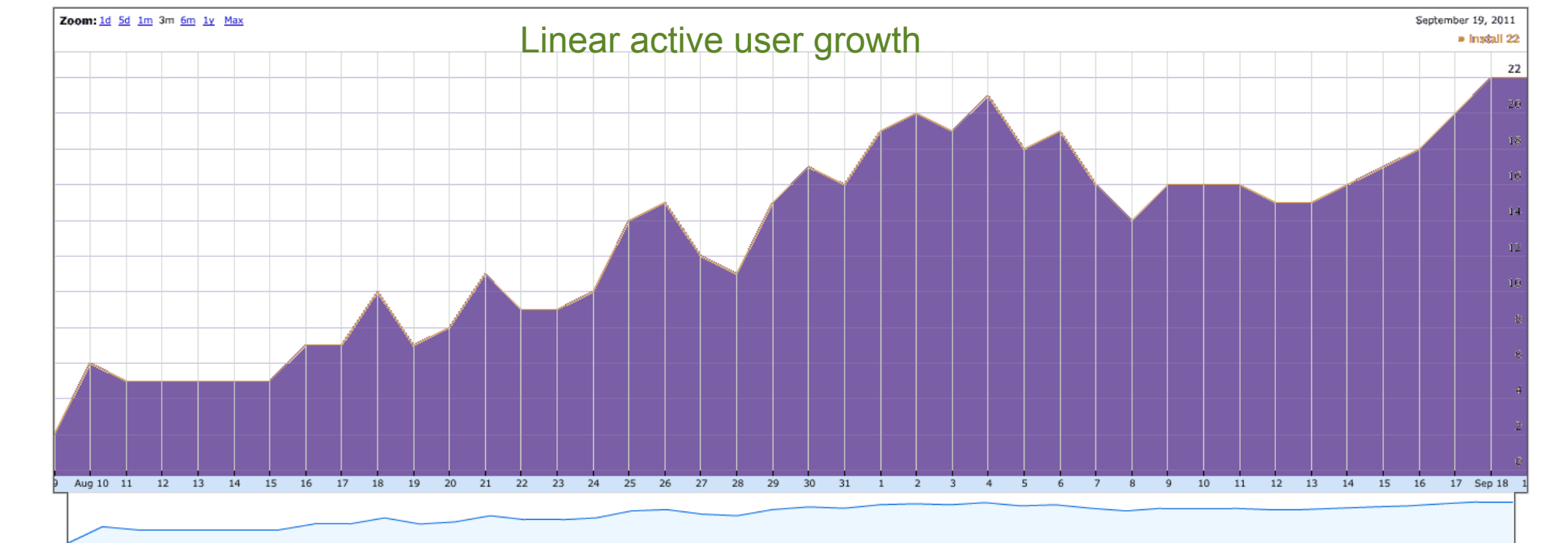
## Results

A) Yes it the acoustic quality is good enough for phonetic analysis.\*

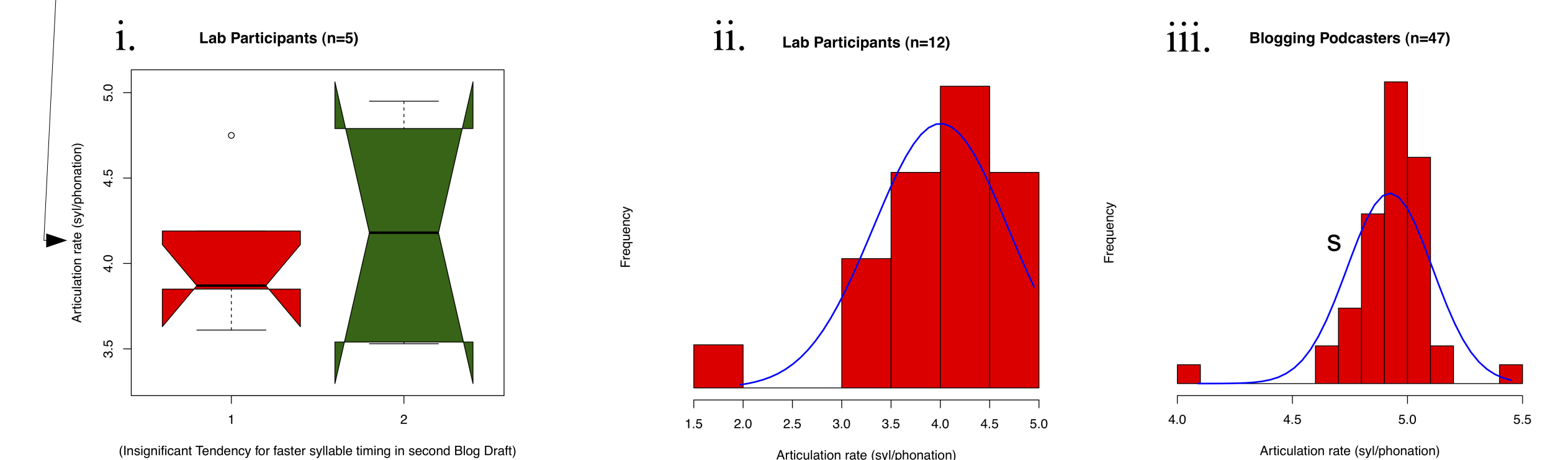
\*as good as a laptop and mass-market microphone



B) Unknown as of yet, we expect more Market data in the next 6-9 months.



- Three out of five participants showed faster speech rate in the second draft.
- Lab participants (n=12) speech rate was significantly different from
- professional blogging podcasters (n=47) (Welch t-test p<.0001)



## Discussion

By investigating both the evolution of information structure in a text, along with two sorts of prosody, there is a potential to tease out more factors involved in prosody which must be controlled in experiments in a laboratory setting, as well as factors which lead to confounds in over-predictability in a laboratory setting.

All lab participants displayed audience-oriented prosodies, in some cases even said “Hello [Experimenter], I'm going to tell you about...” thus we were unable to test the prosodic differences of stream of consciousness prosody vs audience oriented prosody.

We ran some syllable rate analysis and found a general inclination for a faster syllable timing in the second blog drafts. However the sample size was very small (n=5) and very diverse and unbalanced on many factors (3 different languages, 4 female, 1 male).

Our main goal was to pave the path so that researchers can use these OpenSource components in their labs or data collection procedures. It is the hope of this project that by collecting open ended data in a structured way, we will create a strong community of linguists and software developers to better tackle the complexity of investigating prosody and other phenomena at the phonetics-semantics interface.

## Acknowledgements

We would like to thank the users who participated in our experiments both in the lab and on the market, Laura Spinu for volunteering her Ling 201 students, Drs. Sabine Bergler and Leila Kosseim for their support and guidance in their respective Natural Language Interfaces and Statistical Natural Language Processing courses, without which we would have had a much more difficult time deciphering how to customize and train the Sphinx Language Models. We would also like to thank the members of the Clac Lab for their stimulating conversations, the audience at CUSEC Demo Camp 2011 for their feedback and questions about Android speech recognition and text to speech, and finally the dozens of OpenSource developers for creating the software we used in our project and deciding to keep it open and free.