Measuring sociolinguistic perception in real time

Martha Austen & Kathryn Campbell-Kibler

The Ohio State University
Cues over time
Capturing change

Watson & Clark (2014)
Hesson & Shellgren (2015)
Jones (2016)
Montgomery & Moore (2018)
Methods

- Core idea: Test stimuli where we **already know** what effects to expect
  
  - large shifts in enthusiasm, fluency ("gestalt" changes)
  - (ING) tokens
Study Questions

● How do listeners use continuous evaluation tools?

● How can we analyze continuous evaluation data?

● How do after-the-fact ratings relate to in-the-moment reactions?
Methods: Participants
Methods: Gestalt Stimuli

- 4 American English speakers reading texts out loud
  - children's story
  - college radio station announcement
  - Jabberwocky
  - podcast for ESL learners

- **fluency**: hard-to-read font

- **enthusiasm**: asked to read normal, enthusiastic, bored
You're listening to WXYZ 98.7 college radio. This is the classical music hour, run entirely by students in the music department. You just heard Bach's Brandenburg Concerto No. 4 in G Major performed by the Academy of St. Martin in the Fields. Before that was Yo-Yo Ma with the Boston Symphony Orchestra performing Dvorak's Slavonic Dance number 2 in E Minor, Opus 72.

In our last half hour, you'll be hearing works by Mozart, Haydn, and Tchaikovsky. Before that, though, a couple announcements. The university symphony orchestra will be performing a free concert next Thursday night in Smith Hall at 8pm, featuring guest violinist Jamie Jones. I hope to see you all there!
Methods: Gestalt Stimuli

bored

enthusiastic
Methods: Gestalt Stimuli

bored

enthusiastic
Methods: Gestalt Stimuli

(bored)

(bored)

(bored)

(enthusiastic)

(enthus)

(bored)

(listen)
Methods: Procedure

Tested:
buttons vs slider
mouse vs arrow keys vs tablet
“in that moment” vs “overall impression”

Afterwards: Evaluate on 1-10 scale
Results: Enthusiasm

(50 participants)

Time (seconds)

Rating (Enthusiastic)

bored

enthusiastic

bored
Results: Enthusiasm

(50 participants)

![Graph showing changes in enthusiasm over time with ratings on the y-axis and time in seconds on the x-axis. The graph shows a peak of enthusiasm around the middle of the time period, with periods of boredom at the start and end.]
Results: Enthusiasm

(50 participants)
Results: Enthusiasm

It works! But poor time granularity...

1.4 to 7.2 seconds of lag time
Results: Enthusiasm

1.4 – 7.2 seconds
Results: Enthusiasm

1.4 – 7.2 seconds
Results: Enthusiasm

1.4 – 7.2 seconds

16 words, 64 phones
Results: Fluency
Results: Fluency

![Graph showing fluency ratings over time. The graph includes three distinct sections labeled: hard-to-read font at 0-10, enthusiastic at 20-30, and hard-to-read font at 40-60. The y-axis represents rating (articulate), ranging from 0 to 100, and the x-axis represents time (seconds), ranging from 0 to 60.](image-url)
Results: Fluency

It works, too! But time granularity is similarly poor...
Study 2: (ING)

Stimuli from Campbell-Kibler (2006) - “Robert”

<table>
<thead>
<tr>
<th>Description</th>
<th>Num. Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>all -ing</td>
<td>24</td>
</tr>
<tr>
<td>all -in’</td>
<td>25</td>
</tr>
<tr>
<td>alternating ing / in’ / ing / in’</td>
<td>48</td>
</tr>
<tr>
<td>alternating in’ / ing / in’ / ing</td>
<td>51</td>
</tr>
<tr>
<td>alternating ing / in’ / ing / in’, but excerpts in reverse order</td>
<td>77</td>
</tr>
<tr>
<td>alternating in’ / ing / in’ / ing, but excerpts in reverse order</td>
<td>79</td>
</tr>
</tbody>
</table>
### Study 2: (ING)

Stimuli from Campbell-Kibler (2006) - “Robert”

<table>
<thead>
<tr>
<th>Description</th>
<th>Num. Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>all -ing</td>
<td>24</td>
</tr>
<tr>
<td>all -in’</td>
<td>25</td>
</tr>
<tr>
<td>alternating ing / in' / ing / in'</td>
<td>48</td>
</tr>
<tr>
<td>alternating in' / ing / in' / ing</td>
<td>51</td>
</tr>
<tr>
<td>alternating ing / in' / ing / in', but excerpts in reverse order</td>
<td>77</td>
</tr>
<tr>
<td>alternating in' / ing / in' / ing, but excerpts in reverse order</td>
<td>79</td>
</tr>
</tbody>
</table>
Results: (ING)

Mean ratings, single-variant stimuli

- all ING
- all IN
- (ING) token

Rating (Educated)

Time (seconds)
Results: (ING)

Mean ratings, single-variant stimuli

Rating (Educated)

Time (seconds)
But after-the-fact ratings show the expected pattern!
Results: (ING)

Content and Southern accent > (ING)

Mean ratings, single-variant stimuli

Rating (Educated)

Time (seconds)
How do you analyze this data?

- GAMs
- changepoint analysis
- various forms of cluster analysis
- rolling windows of rating slopes
- staring really hard at the plots
- looking at all the individual plots
- ........etc......
How do you analyze this data?

“Token-to-token slopes”
How do you analyze this data?

“Token-to-token slopes”
How do you analyze this data?

“Token-to-token slopes”

B – A for each participant, then compare across guises
Question 3: How do after-the-fact ratings relate to in-the-moment ones?
Question 3: How do after-the-fact ratings relate to in-the-moment ones?
Question 3: How do after-the-fact ratings relate to in-the-moment ones?
Question 3: How do after-the-fact ratings relate to in-the-moment ones?

[Diagram showing the relationship between ratings and time, with key points marked as max, min, mean, median, and final.]
Question 3: How do after-the-fact ratings relate to in-the-moment ones?

+ demographic variables
Question 3: How do after-the-fact ratings relate to in-the-moment ones?

Best predictor: **mean**
Question 3: How do after-the-fact ratings relate to in-the-moment ones?

Also matters: final
Question 3: How do after-the-fact ratings relate to in-the-moment ones?

After-the-fact ratings are a reasonable approximation of real-time impressions.

Listeners maintain distinct “current impressions” and “overall impressions”.

Some methodological conclusions...

High variability means:

• better with larger effects
• use matched guise, not verbal
• try online data collection
... and some theoretical ones

Variables are integrated into the overall impression.

Sociolinguistic perception is not like sticking your hand in a bucket of ice water.
Thanks!
Three buttons

Push a button every time you notice something about the way the person speaks that either CHANGES or CONFIRMS how BORED/ENTHUSIASTIC they sound.

- CHANGE: more bored
- CONFIRMS what I thought
- CHANGE: more enthusiastic
Three buttons

- enthusiastic
- normal
- bored

Time (seconds)