Real-Time Captioning by Groups of Non-Experts

Walter S. Lasecki, Christopher D. Miller, Adam Sadilek, Andrew Abumoussa, Donato Borrello, Raja Kushalnagar, and Jeffrey P. Bigham

(UIST 2012)

Presentation by Felicia Ng for CS mini 2017
TECHNOLOGY NEWS 18 July 2012

Crowdsourcing serves up the subtitles to your life

By Jacob Aron

GIZMODO

Crowdsourcing Could Help Deaf People Subtitle Their Everyday Life
Real-time crowdsourcing

LEGION: SCRIBE

Walter Lasecki (PhD student)  UIST 2012

Mobile tool for (nearly) real-time answers to visual questions by multiple individual crowdworkers
-introduced quikTurkit

VizWiz

Jeff Bigham (Advisor)  UIST 2010

Mobile tool for (actual) real-time captions to audio streams by groups of crowdworkers
-uses quikTurkit

Wobbrock was on my PhD committee!
Scribe
System Overview

Speech Source
we have a crystal that has a two-fold axis ...

Flash Media Server

Caption Stream

Merging Server

Output
we have a crystal that has a two-fold axis

Crowd Corrections
Real-time captioning interface

Crowdworker interface
- “Locks in” words after 800 milliseconds
- Submits phrase after 2 second delay
- Varies volume to encourage coverage

User interface
- Live text stream
- Highlights conflicted words or spellings
- User and workers can edit directly
Real-time input combiner

<table>
<thead>
<tr>
<th>Graph</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>open</td>
<td>open</td>
<td>open</td>
</tr>
<tr>
<td></td>
<td>java</td>
<td>java</td>
<td>java</td>
</tr>
<tr>
<td></td>
<td>the</td>
<td>the</td>
<td>the</td>
</tr>
</tbody>
</table>

- Algorithm combines timing info, observed overlap between partial captions, and models of natural language to inform final output.
- Augments multiple sequence alignment (MSA) model with keyboard- and context-based features, and graph building process for speed and scalability.
Evaluation

- More workers → More coverage
- Adjusting volume → More coverage

- SCRIBE covered more audio than ASR or average individual worker
- Theoretically, could beat CART too
• SCRIBE beat ASR and average individual worker on word error rate and precision
• Also beat trained stenographers in latency (average 2.89 vs. 4.38 sec)
Contributions to technical HCI

- LEGION:SCRIBE, new end-to-end system
  - real-time captioning interface
  - real-time input combiner algorithm

- The idea of using non-expert crowds to caption audio in real-time

- The idea of automatically merging real-time inputs of dynamic groups of workers to outperform individuals on human performance tasks
Discussion

● What might be some of the **limitations** of Scribe? In what contexts do you think it could use more technical development to do even more/better?

● Besides captioning for the deaf and transcribing lectures for students, what **other applications** could you see benefiting from this real-time on-demand crowdsourcing approach?

● What do you think is going to be the **relationship between crowd-powered systems and AI/ML** in the future? Will they eventually be used to replace one another, complement one another, or just used for different purposes?