

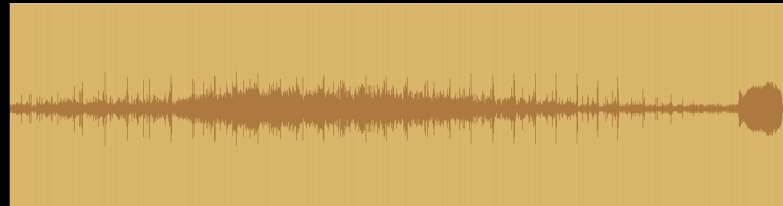
An Iterative Dual Pathway Structure for Speech-to-Text Transcription

Beatrice Liem

Haoqi Zhang

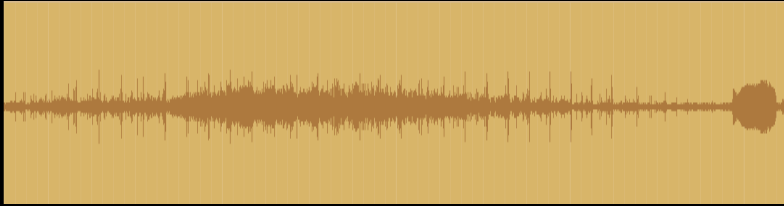
Yiling Chen

audio transcription



"I have a dream that ..."

audio transcription



"I have a dream that ..."

1. Human transcription

- Very accurate
- High cost

2. Computer transcription

- Less accurate
- Lower cost

Crowdsourced transcription



Crowdsourced transcription

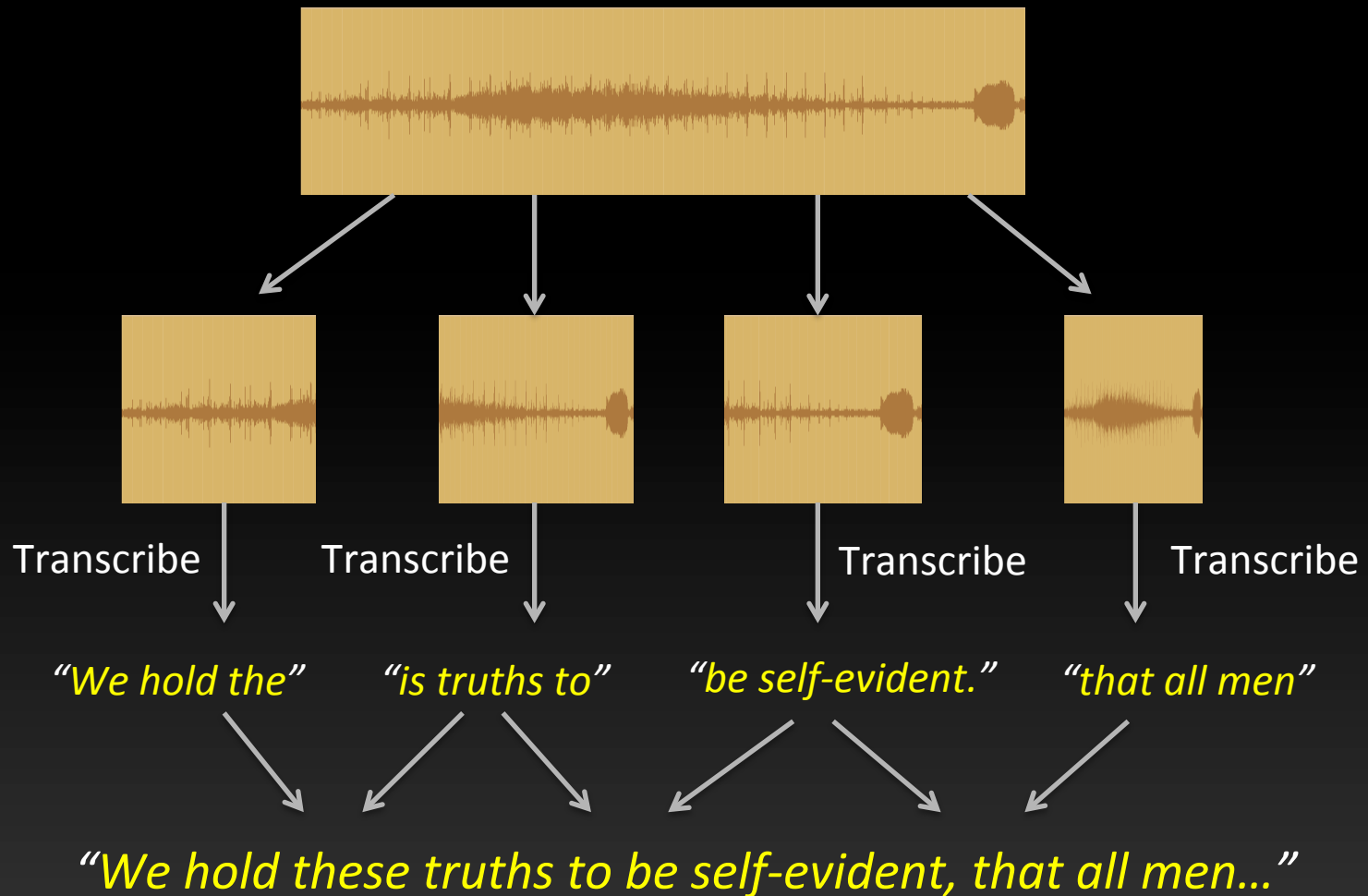


- Recruit Mturk workers to:
 - **transcribe** clips of audio
 - **improve** existing transcripts
 - **grade** transcripts
- Quality control
 - Grading the grader, multiple graders
 - **It works! But, some wasted effort.**

This work: remove checking process

- Iterative dual pathway structure incentivizes people to enter good transcriptions
- Embed iteration in a MapReduce framework
(Dean and Ghemawat '04; Kittur, Smus, and Kraut '11)

MapReduce for audio transcription



Iterative tasks

(Little et al, '10)

task $\longrightarrow A_1 \longrightarrow A_2 \longrightarrow \dots$

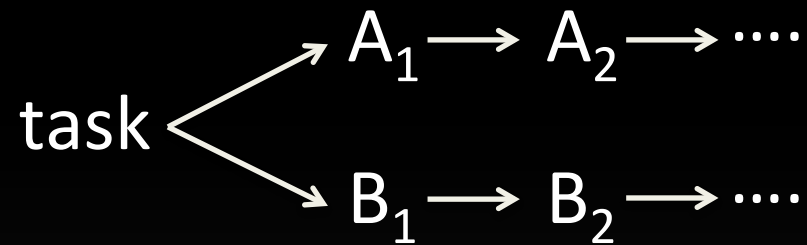
Iterative tasks

(Little et al, '10)

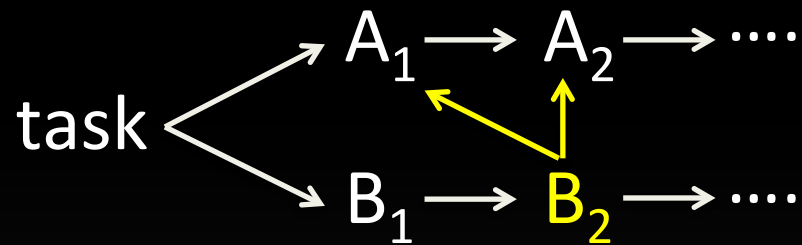
task $\longrightarrow A_1 \longrightarrow A_2 \longrightarrow \dots$

task $\longrightarrow \dots \longrightarrow A_2 \longrightarrow \text{vote}(A_1, A_2) \longrightarrow A_3$

Iterative dual pathway structure



Iterative dual pathway structure



score by similarity
to recent answers
on other path

Iterative dual pathway structure

- Combines iteration with output agreement
- Natural stopping condition by convergence
- Can implement in a game or as a Mturk task

BEATRICE'S SCORE: 0

ROUND 1, CLIP 1

[home](#)

[play game](#)

[tutorial](#)

[about](#)

[login/register](#)



Listen to the clip and write down what you hear.

You have been assigned to Team 2 for this clip. You will get points based on how well you match the transcript produced by the other team. You may use any transcripts generated by your team members (displayed below if applicable) to help you.



[Edit This >](#)

The real story behind the sheep is that Doctor Wilmoth created the prototype for bioindustrial design. He's the Henry Ford of the biotech century.

[Clear Text Box](#)

[Clear Text Box](#)

[Submit](#)

[No Words!](#)

[Skip Clip](#)

Study

- 147 Harvard undergrads
- Iterative version
 - 549 transcripts
- Parallel version (baseline)
 - 308 transcripts

Results

- Parallel: 93.6% word accuracy
- Iterative: 96.6% word accuracy

Example

Iteration 1: red, red, red! what should i do?

Iteration 2: red, red, red! Dear God, where should I go, what should i do?

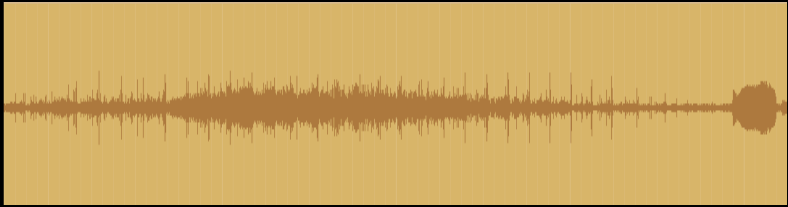
Iteration 3: Fred, Fred, Fred! Dear God, where shall I go, what should i do?

Iteration 4: Rhett, Rhett, Rhett! Dear God, where shall I go, what shall I do? (Correct)

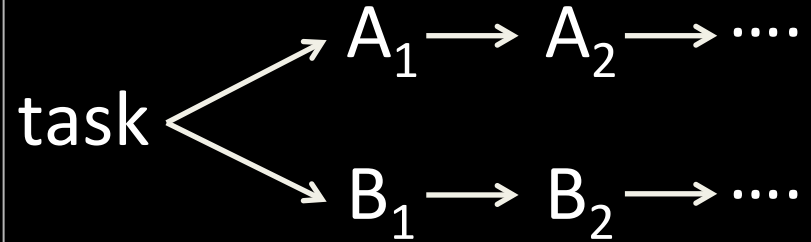
Feedback

- Most subjects claim to give high effort
- Iterative version was more enjoyable

Conclusion



"I have a dream that ..."



Iterative tasks

+

Output agreement games

Thank you

Smell stew hq@eecs.harvard.edu

Smiles to hq@eecs.harvard.edu

Send mails to hq@eecs.harvard.edu