

CS492F Pitch Presentation

SCOOBY



will pull an all-nighter helping you deliver the “clearest” pitch ever

THE SHY team

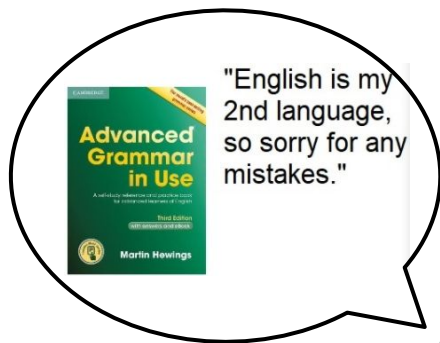
Shyngys Aitkazinov | Hyunsung Cho | Yewon Kim

SCOOBY

01

Problem

You are a non-native English speaker and...



You are invited as a
presenter of TED talk!

Steps to prepare the presentation



Write a **script**,
memorize & **practice**
before the speech!

Non-native English speakers would think...

"Am I pronouncing this correctly?"
"Will people understand what I speak?"



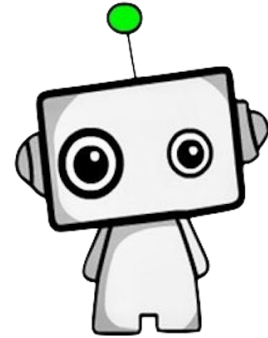
You would look for...



An agent who will help you practice pronunciation with scripts!

Problem Statement

There is no **AI-interactive** platform for **non-native** English speakers to **evaluate** and **practice** English pronunciation skills with **personalized scripts**.



SCOOBY

02

Demo

chris.kaist.ac.kr:5432

KVPN required for people outside the campus

Prototype Demo

Scooby

enter your script to practice

upload your recording

Choose File No file chosen

Practice

Grade:

Why this result?

What does the grade mean?

Why am I getting this result?

Get more evaluation

Prototype Demo

Scooby

enter your script to practice

1. User types the script here

upload your recording

Choose File No file chosen

3.



2. User uploads the recorded audio*

Grade:

Why this result?

What does the grade mean?

Why am I getting this result?

Get more evaluation

Prototype Demo

Scooby

Simply marking them as wrong and giving no explanation would not help. As a solution, we suggest two approaches. First, listening to their own recorded voice and correct pronunciation from AI, and comparing them would help their learning.

upload your recording

Choose File Hyungsung_3.m4a

Practice

Grade: A

Word-level analysis of the speech

Simply marking them as wrong and giving no explanation would not help. As a solution, we suggest two approaches. First, listening to their own recorded voice and correct pronunciation from AI, and comparing them would help their learning.

Why this result?

AI gives the score

Detailed speech analysis

What does the grade mean?

From your pronunciation of a sentence, word, syllable, or phoneme, the accuracy and intelligibility of the overall pronunciation is measured on a scale of 0 to 100.

- A: Excellent. Native or native-like
- B: Very good and clearly intelligible
- C: Good. Intelligible but with one or two evident mistakes.
- D: Fair. Possibly not intelligible with several evident mistakes.
- F: Poor and must be reattempted.

Why am I getting this result?

Get more evaluation

ntation

Prototype Demo

Simply marking them as wrong and giving no **explanation** would not help. As a solution, **we suggest** two approaches. First, listening to their own recorded voice and correct **pronunciation** from AI, **and** comparing them would help their learning.

Why this result?

sihmpli maarkihng dhehm aez raong aend gihvihng now ehksplahneyshahn wuhd naat hhehlp ehz ah sahluwshahn wiy
Correct: sahghjehst tuw ahprohchihz ferst lihsnihng tuw dhehr own rihkaordihd voys aend kerehkt prawnahnsiyeishahn frahm eyay aend
kahmpehrihng dhehm wuhd hhehlp dhehr lernihng
sihmpli maarkihng dhehm aez raong aend gihvihng now ehksplahneyzhahn wuhd naat hhehlp ehz ah sahluwshahn ahiy
Yours: lahchjehst tuw ahprohchihz ferst lihsnihng tuw dhehr own rihkaordihd voys aend kerehkt pwnahnseyshahn frahm eyay aen
kahmpehrihng dhehm wuhd hhehlp dhehr lernihng



If Scooby heard your speech without the script, he might understand it as...

"simply marking them as long and giving Louis cremation would not help it. The reason we have a chance to practice first listening to their own recorded voice and correct pronunciation from in comparing gun without their their name"

So what?

Ignore this result (Dumb Scooby!)

1. Phoneme-level analysis

2. Listening to my and correct audio

3. If output is completely wrong, ignore option provided

Prototype Demo

What does the grade mean?

Why am I getting this result?

We count the number of correctly spoken words and syllables to grade your speech. Among 38 words, you pronounced 35 words right. Among 61 syllables, you pronounced 56 words right.

Get more evaluation

Your estimated IELTS speaking score is: 8.5. For more information, please visit: [IELTS Score Description](#)

Your estimated PTE speech score is: 83. For more information, please visit: [PTE Score Description](#)

1. More explanation on evaluation: syllable and word level

2. Estimation of IELTS, and PTE scores, evaluating your pronunciation

SCOOBY

03

Tasks & Solutions

Task 1: Let users know where they were wrong

Visualize the output & color wrong parts

Grade: B

Due to the coronavirus outbreak, people face lots of restrictions in day-to-day lives.

Why this result?

Wrongly spoken words are colored red

If Scooby heard your speech without the script, he might understand it as...

"due to the coronavirus outbreak people face lots and restrictions and they said they'd like"

Task 2 : Let users effectively **practice** by **interacting** with AI

Let users **compare** their own speech and correct answer.

&

Provide more sophisticated **analysis** from AI.

Correct: duw tah dhah kerownahvayrahs awtbreyk piypahl feys laats ahv riystrihkshahnz ihh deytahdey layvz

Yours: dih tah dhah keraetahvayrihz awtbreyk piypahl feys laats ahv riystrihkshinyz ihh deytahdey laytt



Correct
Pronunciation



My
Pronunciation

Phoneme-level analysis

Replay the record &
compare with the **answer**

Task 3 : Let users know their progress

By **scoring** the speeches!

Grade: B

Due to the **coronavirus** outbreak, people face lots of **restrictions** in day-to-day **lives**.

Why this result?

What does the grade mean?

Meaning of the letter grades.

B: Very good and clearly intelligible

Why am I getting this result?

Briefly explain how the grade is counted;
provide the number of words and
syllables that are spoken right

Get more evaluation

Provide users AI-estimated
IELTS and PTE scores

Task 4 : Let users be aware of model uncertainty

Give users an option to **ignore** the model.

Ignore this result (Dumb Scooby!)

chris.kaist.ac.kr:5432 says

Dumb scooby reported! Sorry for the inconvenience. Please try other recordings.

OK

SCOOBY

04

Evaluation & Results

User Study: Method

Part 1. Scooby Experience Study

Goal: *How* does Scooby *help users practice* English pronunciation?

8 participants

1. Record a sentence
2. Use Scooby
3. After exploring Scooby's explanations, re-record the same sentence
4. Answer our Google Form survey (5-point Likert scale & open-ended questions)

Part 2. Blind Test

Goal: Is using Scooby *effective* in improving pronunciation?

9 participants (not overlapping with Part 1)

1. Listen to recordings from Part 1, before & after using Scooby (without knowing which one is before/after)
2. Select which one has a better pronunciation

Part 1. Scooby Experience Study

Interpretability of AI-generated 'Grade'

(1: Strongly Disagree, 5: Strongly Agree)

*The 'grade' **reliably** represents my pronunciation level.*

Average: **4.25**; Median: 4.5; STD: 0.88

*The 'grade' **accurately** represents my pronunciation level.*

Average: **4.125**; Median: 4.5; STD: 0.99

Reasons

"reliably because **two times the same grade**, and accurately because I am **confident** in my English skills and it **gave me A**"

"For the second recording, I focused on the syllables that I got wrong the first time and Scooby **reflected those efforts.**"

Part 1. Scooby Experience Study

Explainability of the three AI-based explanations

A. Word-level highlight

Without numerical visualization, users would feel that the AI model is ambiguous and lose interests.

Why this result?

B. Phoneme-level highlight

Correct: wih**th**awt nuwmehrihk**ahl** vihzh**uh**ah**lah**zeyshahn yuwzerz wu**hd** fiyl dhaet dhi**y** ay maadahl ihz
aembihgyuwahs aen**d** luwz ihnterihsts

Yours: wih**da**ot yuwmehrih**kr** vihzh**ah**lay**z**heyshahn yuwzerz d**hd** fiyl dhae**ay** diy y maadahl ihz
aembih**ai**uwahs aen luwz ihnterihsts

C. Listening to audio



Correct
Pronunciation



My
Pronunciation

Part 1. Scooby Experience Study

Explainability of the three AI-based explanations

	A. Word-level highlight	B. Phoneme-level highlight	C. Listening to audio
It helps me understand why I got the grade.	Avg: 3.37 Median: 3.5 Std: 1.40	Avg: 4.62 Median: 5 Std: 0.51	Avg: 4.37 Median: 5 Std: 0.91
It helps me plan how to improve my pronunciation.	Avg: 3.37 Median: 3.5 Std: 1.40	Avg: 3.75 Median: 5 Std: 1.03	Avg: 4.37 Median: 5 Std: 1.06

Part 1. Scooby Experience Study

Explainability of the three AI-based explanations

- A. Word-level highlight *explains why* I got the grade, but it is *too vague* and less useful.
- B. Phoneme-level highlight is very accurate and *helpful in explaining why* I got the grade, but *hard to practice* with it.
- C. Listening to audio *helps plan* for *further improvement* in pronunciation.
- Combined: "*Listening to audio after observing the syllables I got wrong* were quite helpful in improving my pronunciation."

Discussion for Part 1

Participants thought the grade is reliable and accurate.

→ Users trust AI when the results & explanations are consistent with their expectation.

Each explanation contributed differently to interpreting the AI-graded result and planning to improve pronunciation.

→ It is important to consider not only **individual benefits of different explanations** but also how different explanations **complement** each other and create a **synergy** effect towards interpretability and explainability.

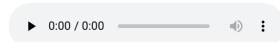
Part 2: Blind Test

Q. Which audio has a better American English pronunciation?

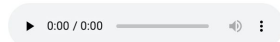
Blind test

Participant 1

(A)

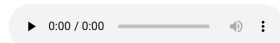


(B)

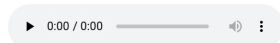


Participant 2

(A)

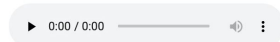


(B)

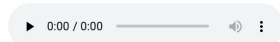


Participant 3

(A)



(B)



Participant 4

(A)

(B)

Participant 1 *

- ☐ A
- ☐ B
- ☐ Same

Participant 2 *

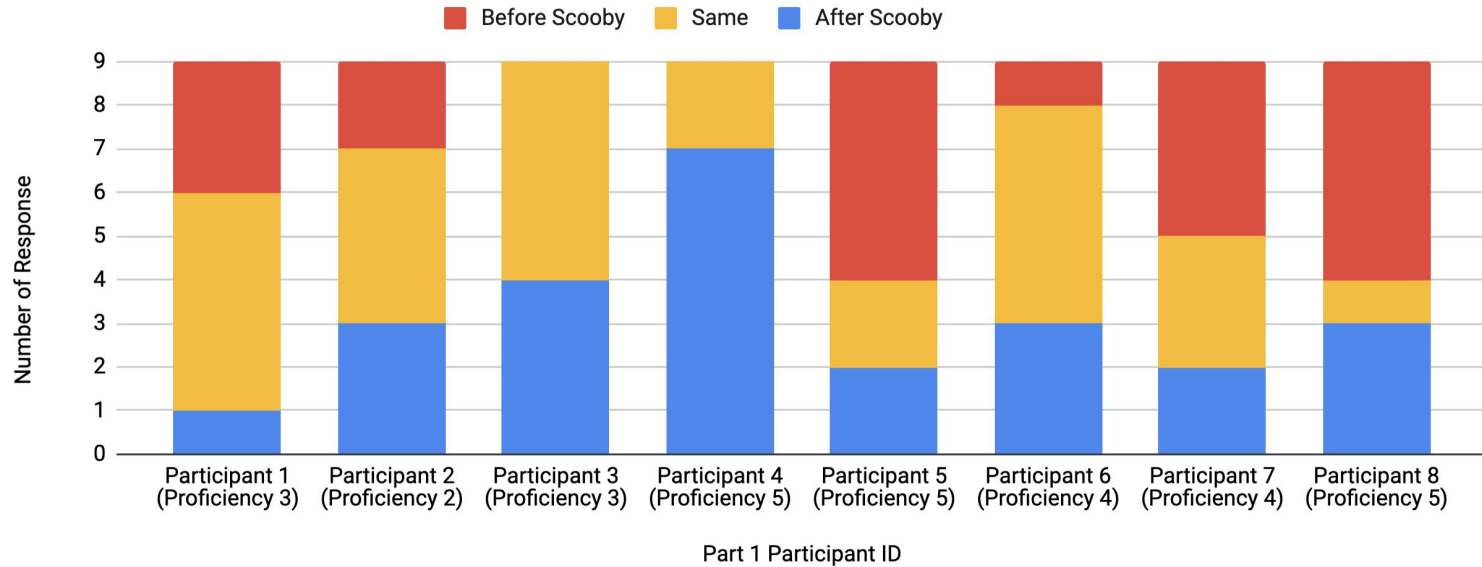
- ☐ A
- ☐ B
- ☐ Same

Participant 3 *

- ☐ A
- ☐ B
- ☐ Same

Part 2: Blind Test

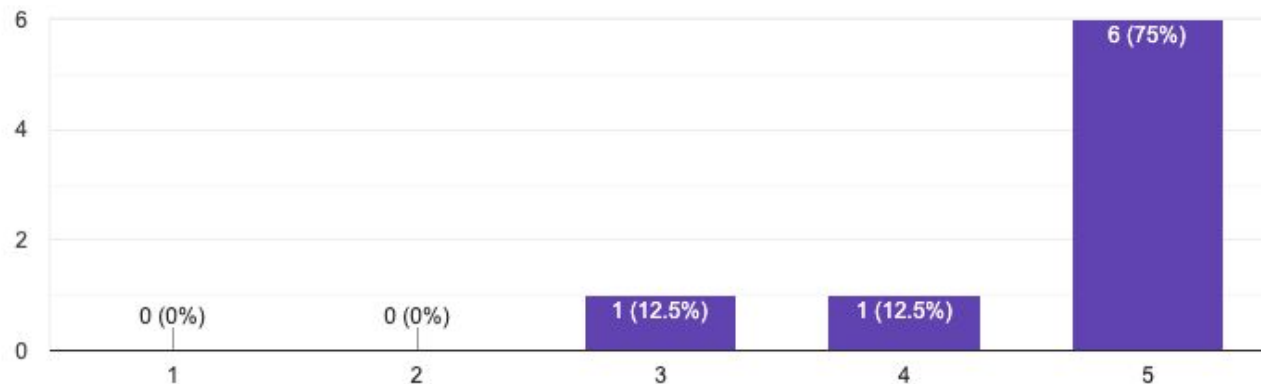
Blind Test Results



Part 2: Blind Test

2-a. My pronunciation improved in the second recording (compared to the first trial) after using Scooby.

8 responses



Discussion for Part 2

Audio quality varied across participants.

One-time use created a *subtle* difference that was recognizable to the owners of recordings but not to others.

Future work

Support for more explanations

Correct pronunciation of each word

Diverse accents & voice for correct pronunciation

Visualize pitches, accents, and pauses

Improve evaluation method

Long-term effects of using Scooby

Greater user pool, by groups with different English proficiency levels

Better metrics to measure pronunciation; Native speaking blind testers (Our testers: avg. 2.8 in 0-5 Language Proficiency Levels)

Consistent audio quality for blind testing

Try out Scooby 🐾 <http://chris.kaist.ac.kr:5432>

(kvpn required if you're outside campus)

CS492F Pitch Presentation

SCOOBY



will pull an all-nighter helping you deliver the “clearest” pitch ever

THE SHY team

Shyngys Aitkazinov | Hyunsung Cho | Yewon Kim
