

BeParrot: Efficient Interface for Transcribing Unclear Speech via Respeaking



Riku Arakawa[†]

Carnegie Mellon University



Hiromu Yakura[†]

University of Tsukuba

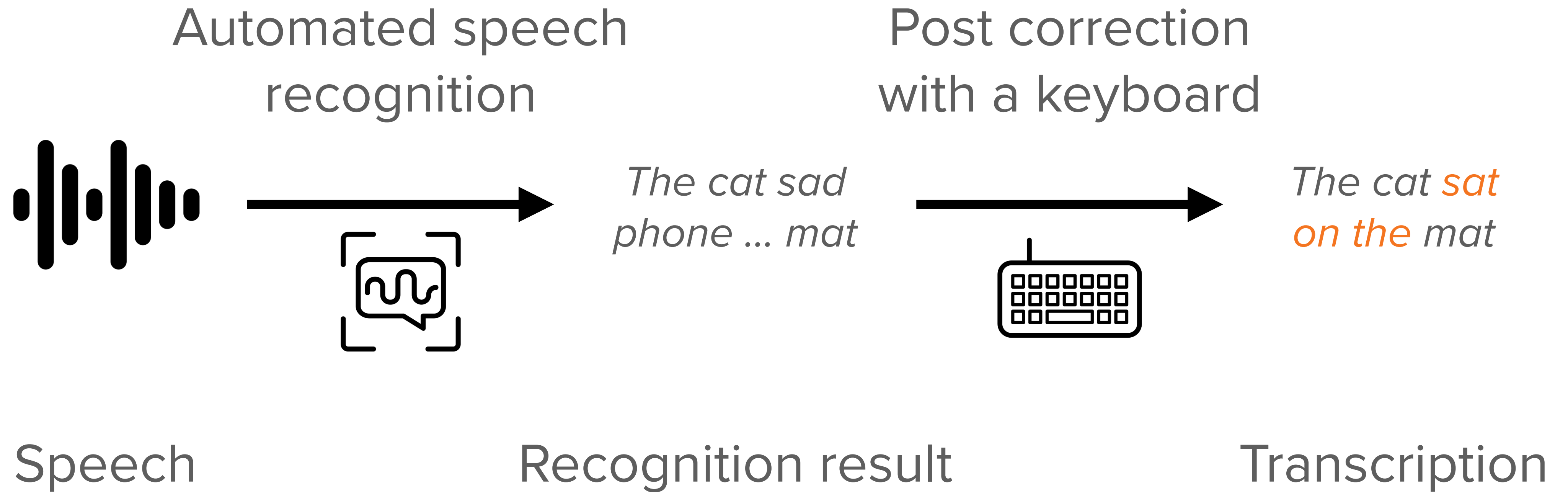


Masataka Goto

National Institute of Advanced Industrial
Science and Technology (AIST)

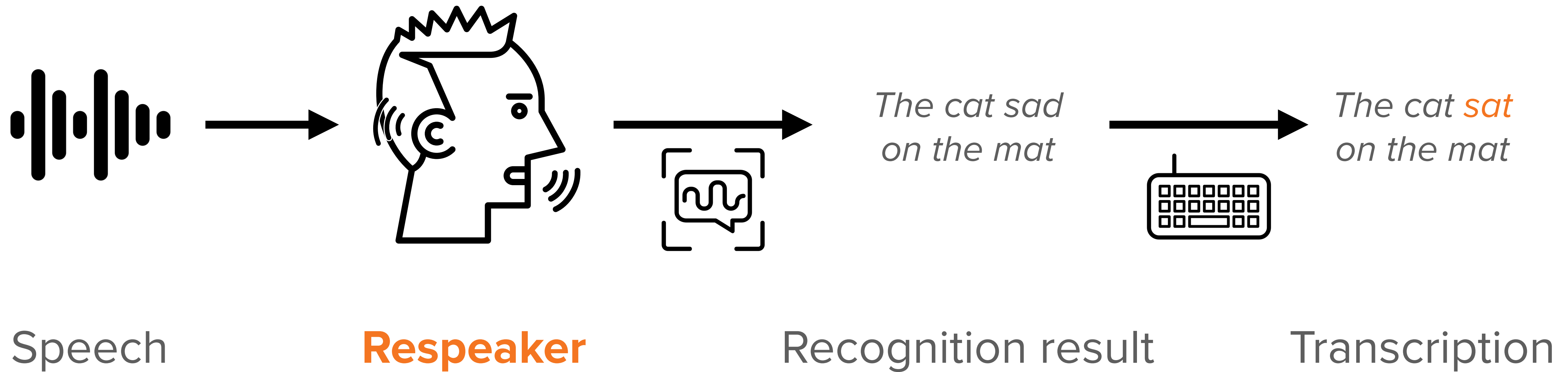
[†] Equal contribution

How to transcribe speech with recognition models



What if the speech is unclear?

Respeaking



Respeaking can reduce the number of error corrections

Respeaking in Practice: Live Captioning

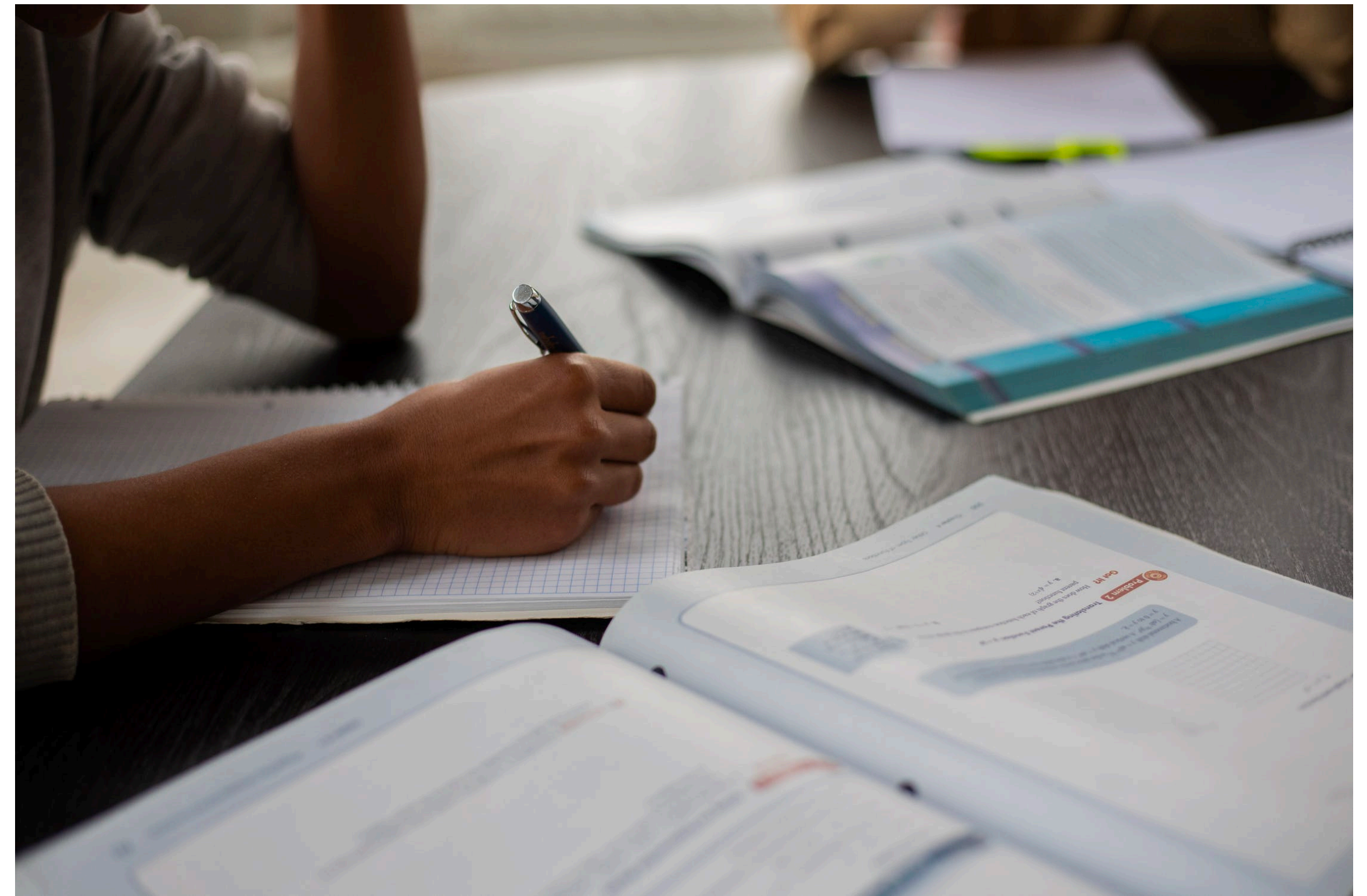


Source: <https://www.nhk.or.jp/stri/publica/rd/182/3.html>

Hurdle of respeaking

- Respeakers should be able to repeat the provided speech clearly without stuttering or stammering
- Respeakers are required not only to repeat the speech but also to memorize the speech content

To be a professional respeaker ...



There is a training program of 75 hours

Proposed System: BeParrot

0:50 / 1:03

Go to the next

Retry this segment

Playback speed x0.9

Segment length 7.0 sec

Pre-recognized content (for reference)

Prev: we will be able to speed up at 10

Curr: but I will

Next: be able to bring in the words of the old Negro spiritual free

Recognition result

but I will

Final transcription

I have a dream my poor little children one one day live in the nation where they will not be judged by the color of their skin but by the content of a character I have a dream today will be able to speed up at the

Words misrecognized

live: labor (1 time)

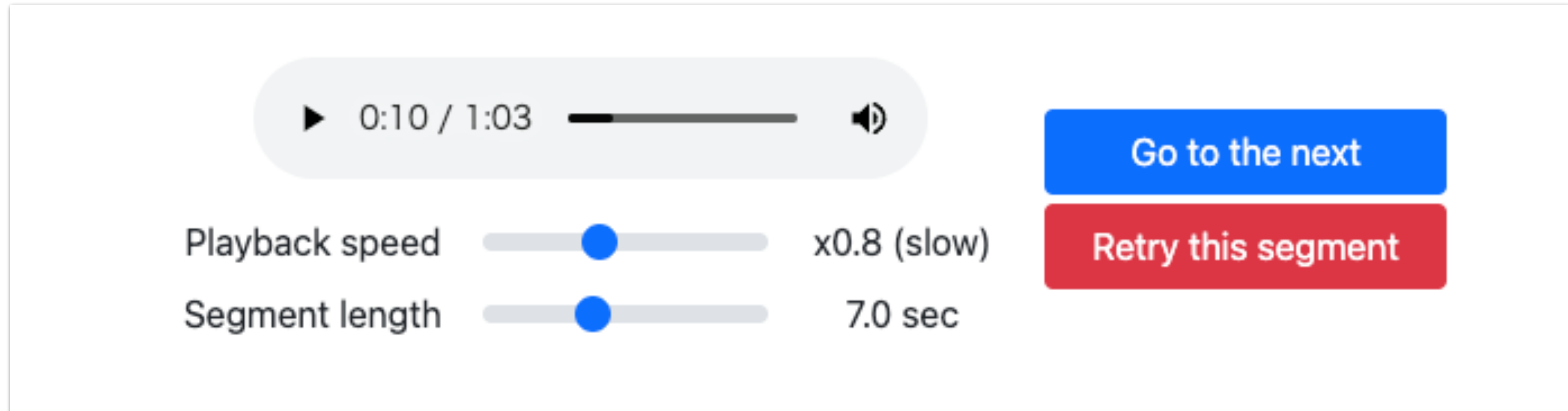
the: 10 (1 time)

Having trouble?

Key features:

- **Parameter adjustment**
- **Pronunciation feedback**

System Feature: Parameter Adjustment



Slower playback speed → Help avoiding stuttering or stammering

Shorter segment length → Reduce the demand of memorizing content

System Feature: Pronunciation Feedback

Words not recognized

at: 1 time

Words misrecognized

the: 10, day (2 times)

live: labor (1 time)

bring: ring (1 time)

Automatically calculated from the correction history

0:23 / 3:00

次へ

再生速度 x1.0
読み上げ長さ 10.0秒

やり直す

自動認識結果

前: 現在地

今: 現在永遠の地下鉄の文学のことではの古典の

次: 世界みたいなのが入った

読み上げ結果

[Redacted]

確定済み

ただいまご紹介に預かりました大変良いところで羨ましい環境だと思います今日は文学の現在 というテーマでお話するわけですが

This is a demo where a Japanese user tries to transcribe a historical speech.

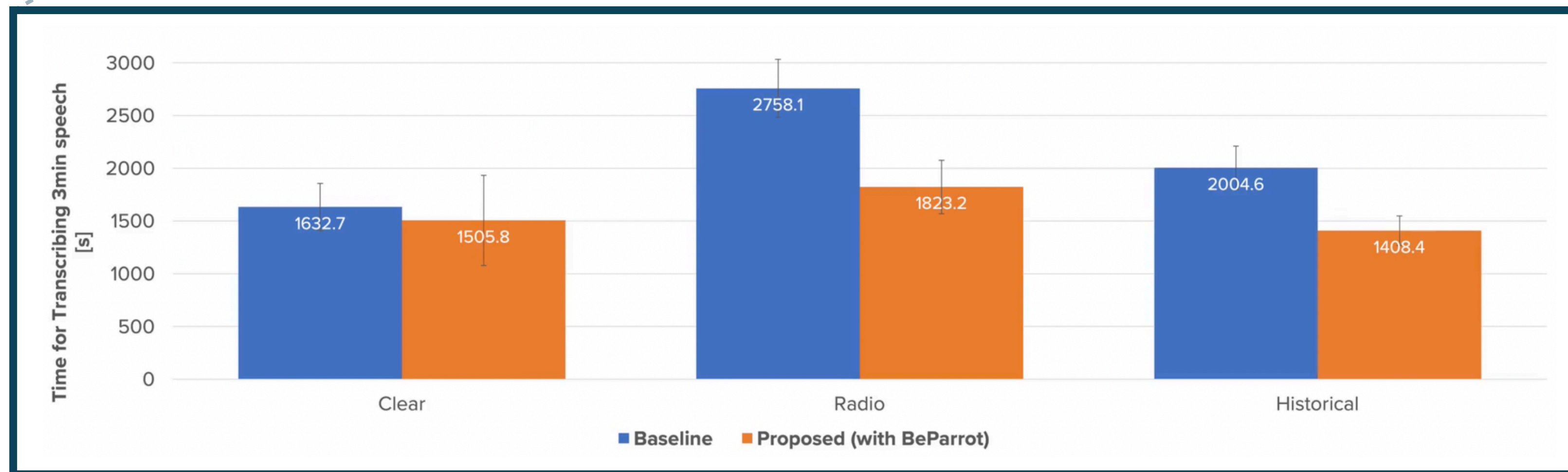
Study Results: Design

- 60 crowd workers
 - 30 workers: transcribe with BeParrot (*proposed*)
 - 30 workers: transcribe with conventional post-correction interface (*baseline*)
- 3 types of speech: *clear*, *radio* and *historical*
- 2 measures: *time* and *character error rate (CER)*

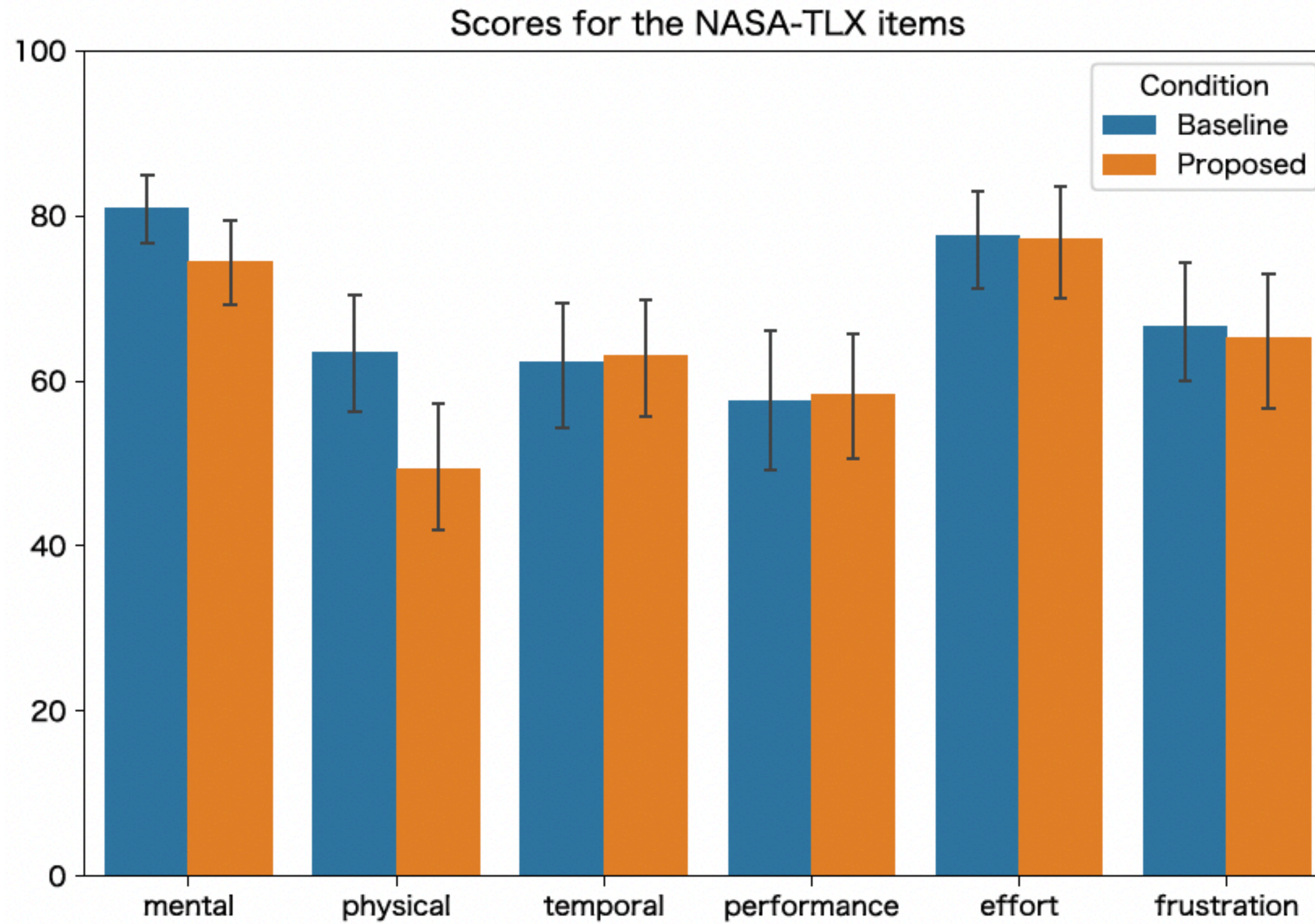
Study Results: Time

Speech type	Time			CER (%)		
	Baseline (s)	Proposed (s)	Reduction (%)	ASR	Baseline	Proposed
Clear	1632.7 (± 224.6)	1505.8 (± 426.8)	7.8	6.16	3.73 (± 0.57)	5.75 (± 1.02)
Radio	2758.1 (± 273.6)	1823.2 (± 253.4)	33.9	30.72	19.05 (± 1.56)	24.81 (± 2.69)
Historical	2004.6 (± 205.0)	1408.4 (± 140.4)	29.7	48.18	29.25 (± 5.16)	29.10 (± 2.17)

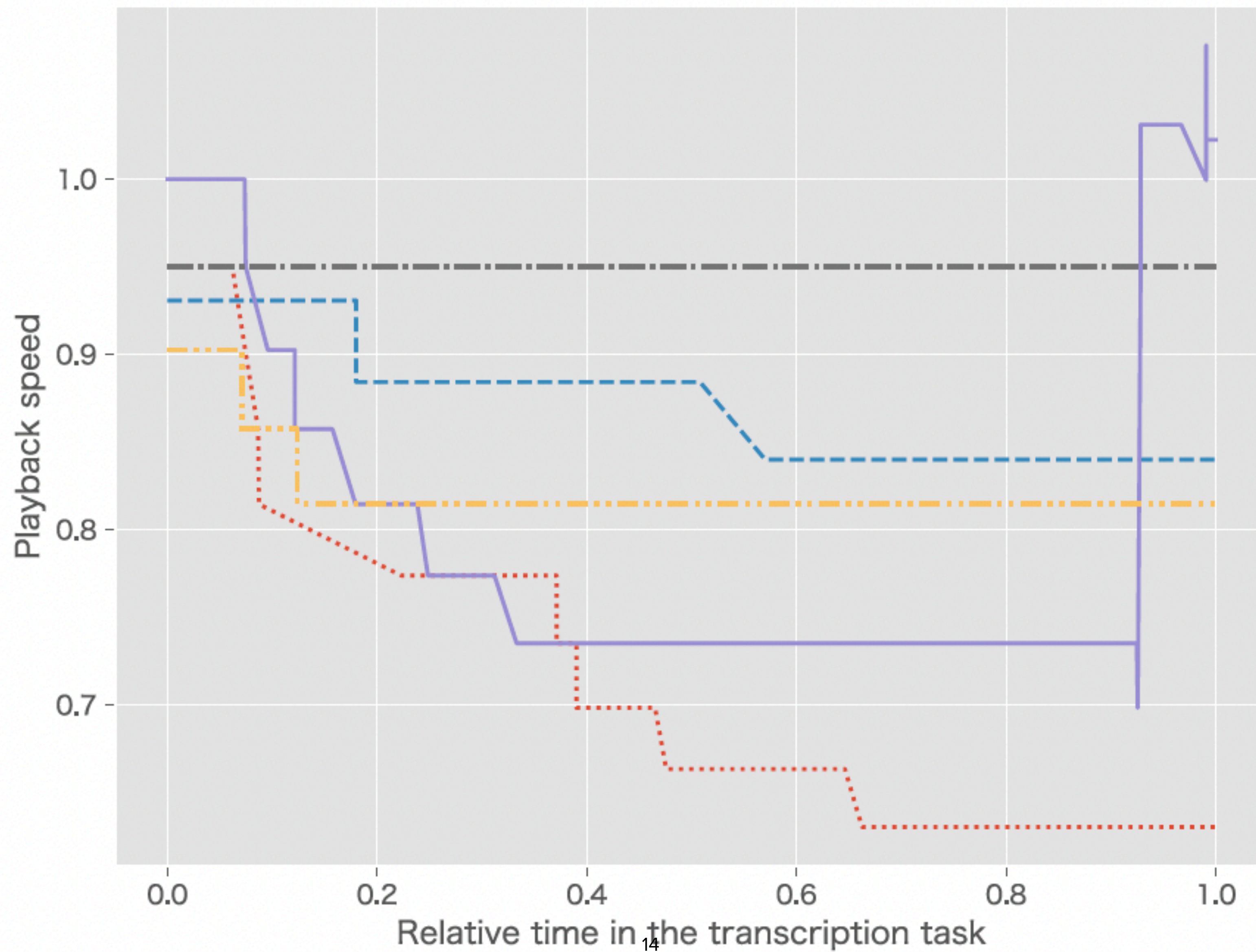
32.1 % time reduction for transcribing unclear speech (radio + historical speech)



Study Results: NASA-TLX



Study Results: Speed History



Study Results: User Comments

- Overall preference and willingness to use BeParrot
- Examples highlighting the effectiveness of the two key features

For more details, please refer to our paper.

Limitation and Future Work

- Evaluate BeParrot with various audios in different languages
- Investigate longitudinal leaning effect
- Introduce online adaptation of the streaming ASR model

Conclusion

- We developed BeParrot as an efficient interface for transcribing unclear speech via respeaking.
- BeParrot has two key features, **parameter adjustment** and **pronunciation feedback**, that enable novice users to conduct respeaking without extensive training.
- Our user study with 60 crowd workers showed that BeParrot makes transcription tasks of unclear speech **32.2% faster**.