CHI 2020

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

VoiceCoach: Interactive Evidence-based Training for Voice Modulation Skills in Public Speaking


Xingbo Wang


Haipeng Zeng


Yong Wang


Aoyu Wu


Zhida Sun


Xiaojuan Ma


Huamin Qu

Introduction

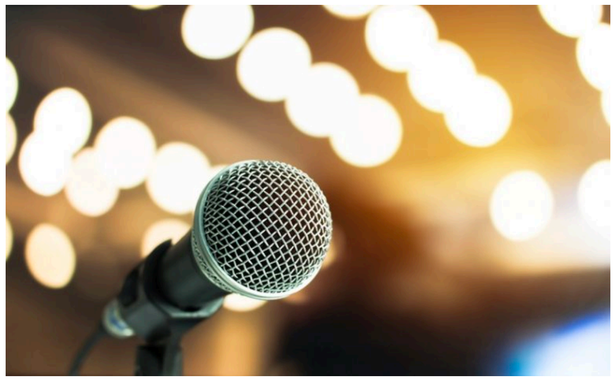
Voice Modulation

Pitch

Volume

Speed

Pause



Public Speaking

audience engagement

convey key ideas

Introduction

Voice Modulation

Pitch

Volume

Speed

Pause

- **Higher** *volume/pitch* → vocal emphasis

audience engagement

- **Increasing** speed → excitement

convey key ideas

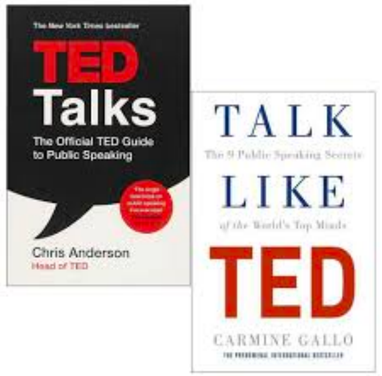
- **Slowing** down and **pauses** → content reflection → personal connections

Introduction

Voice Modulation

It is challenging for novice speakers to master various voice modulation skills

Soln1: Follow the guidelines from the books



Problem: No timely feedback

Introduction

Voice Modulation

It is challenging for novice speakers to master various voice modulation skills

SoIn2: Join the public speaking training programs

Problems: (feedback)

1. No quantitative evaluation
2. Subjected to personal preferences

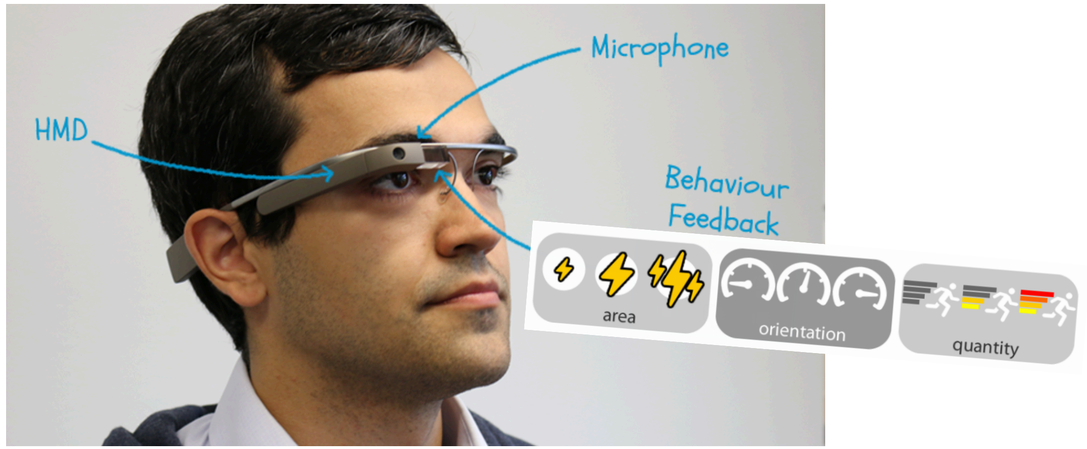
Introduction

Training Systems for Voice Modulation

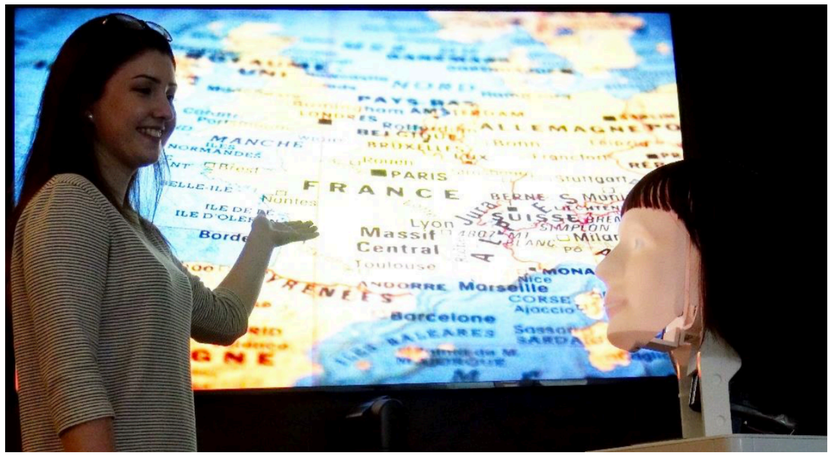
To facilitate the training process:

- **Automatic feedback**
- **Effective feedback**

on speech quality
(pitch, speech rate, loudness)



Logue (Damian et al., 2015)



RoboCOP (Trinh et al., 2017)

→ Pre-defined thresholds regardless of sentences and contexts

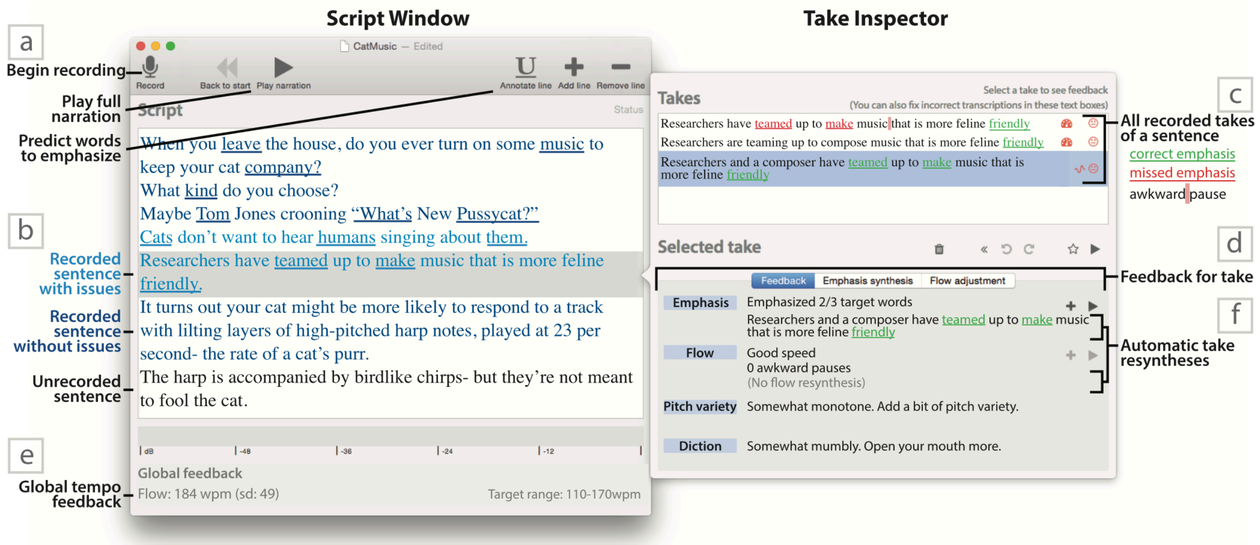
Introduction

Training Systems for Voice Modulation

To facilitate the training process:

- **Automatic feedback**
- **Effective feedback**

on speech quality
(pitch, speech rate, loudness)



NarrationCoach (Rubin et al., 2015)

Check whether speakers satisfy specific requirements for words in sentences

→ Difficult for novice speakers to specify requirements

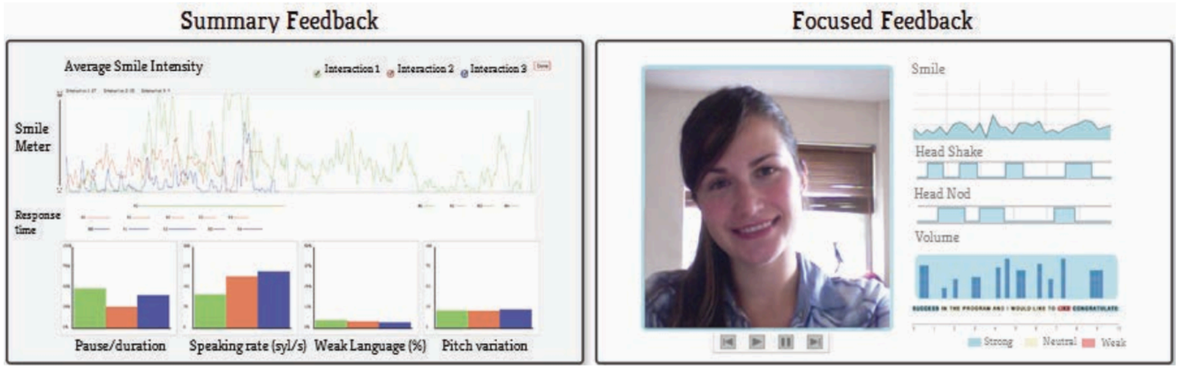
Introduction

Training Systems for Voice Modulation

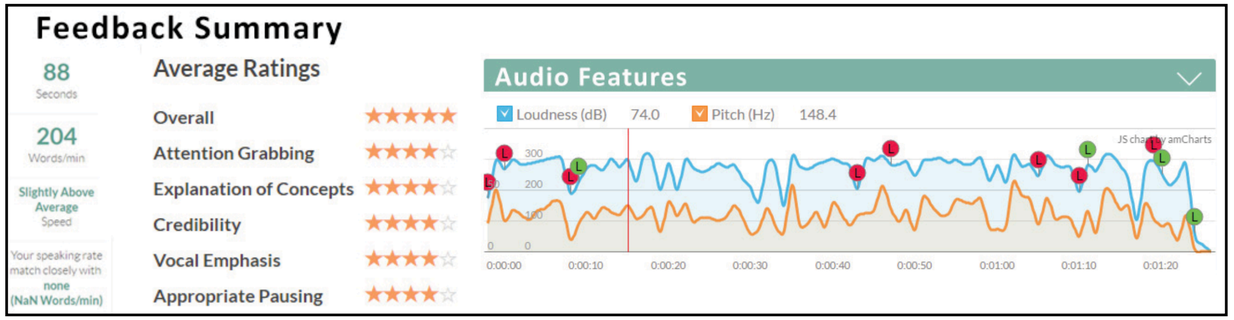
To facilitate the training process:

- Automatic feedback
- Effective feedback

for self-reflection & practice



MACH (Hoque et al., 2013)



ROC Speak (Fung et al., 2015)

Simple charts with limited interaction support

→ insufficient to **compare** their performance and **practice deliberately**

Design Process

To help novice speakers improve voice modulation

- Iterative development (8 months)
- Worked with **4 professional communication coaches**
  - Well-experienced (>= 6 years) in training of public speaking

    → gain deeper understanding of voice modulation skills
  - “Proxies” to novice speakers

    → better aware of difficulties that they may encounter

    → deep insights into limitations of traditional methods

System Requirements

- R1: **Inform** speakers of their voice modulation
- R2: Provide **hints and evidence** to guide potential improvements
- R3: Illustrate the evidence with **concrete examples**
- R4: Enable **on-the-fly** feedback on speakers' vocal performance
- R5: Promote **deliberate** and **iterative** practice

VoiceCoach

Our Approaches

- For personalized guidelines *(R2, R3)*
  - Data-driven approach to retrieve **GOOD** learning examples

    (sentences that have similar **structures** and **meaning**)
  - **TED Talks** as benchmark

- For self-awareness (feedback) & self-adjustment (practice) *(R1, R4, R5)*
  - **Interactive Visualization**

VoiceCoach

Our Approaches

User Input

Sentences & words

Voice modulation

Recommender

Speech content

Sentence structure

TED Talk Videos

Sentences & words

Voice modulation

User Interface

Visualization

Exploration & comparison

On-the-fly feedback

VoiceCoach

Our Approaches

User Input

Sentences & words

Voice modulation

Audio

Transcribe

Text

Voice Modulation

volume, pitch,
pause, speed

1) … **?**

2) **If your actions, …** **.**

slower

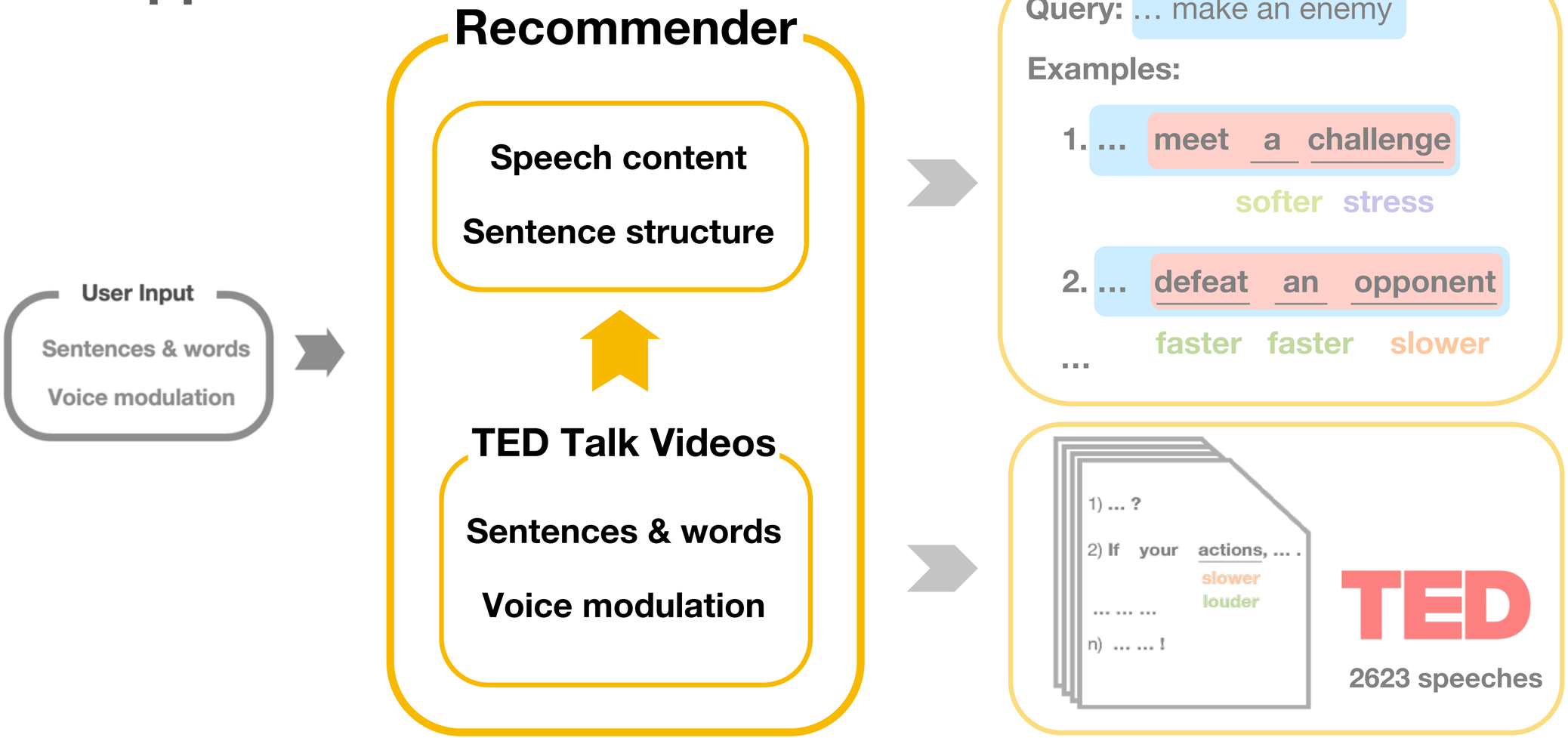
louder

… … …

n) **… …** **!**

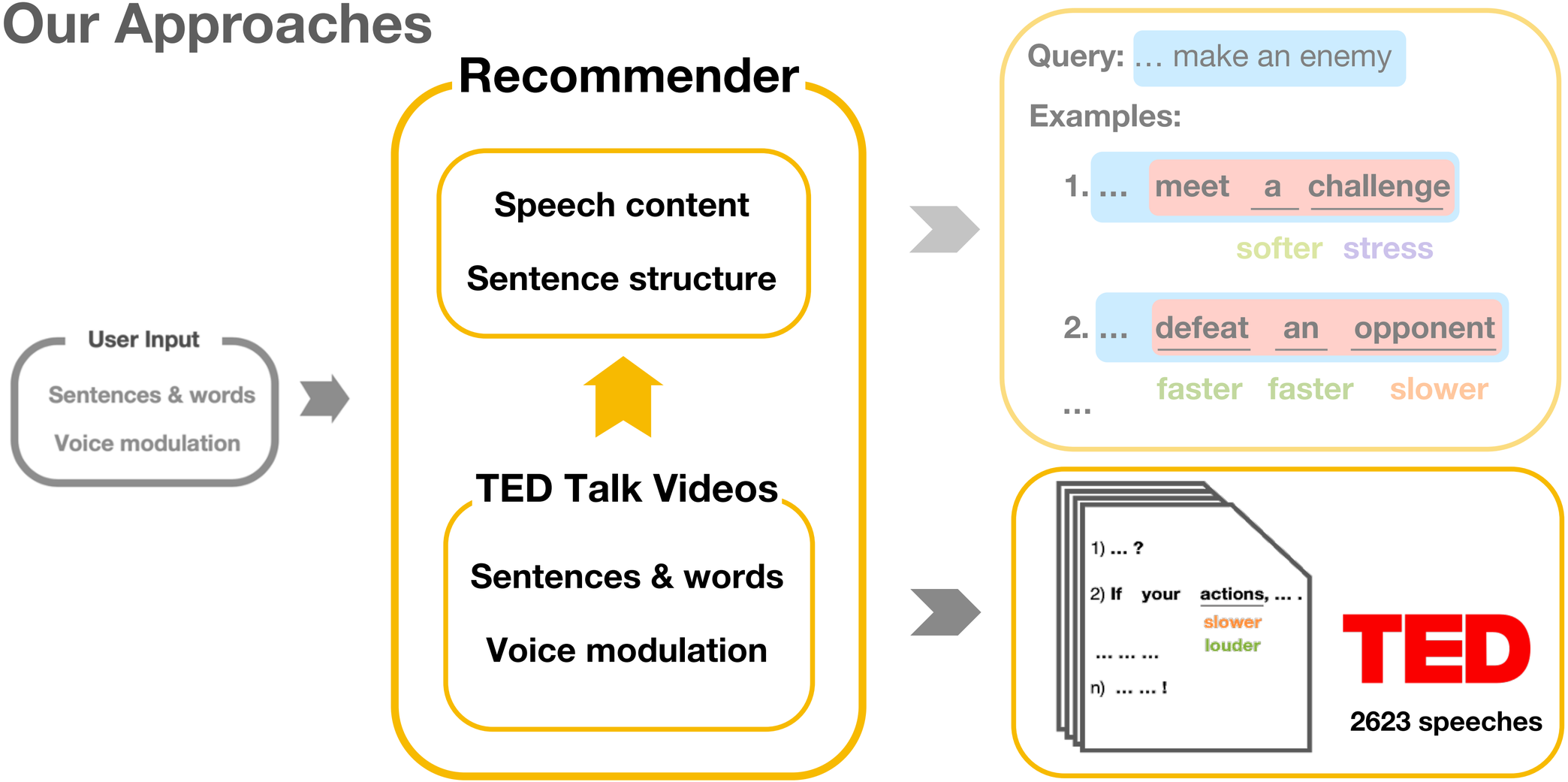
VoiceCoach

Our Approaches



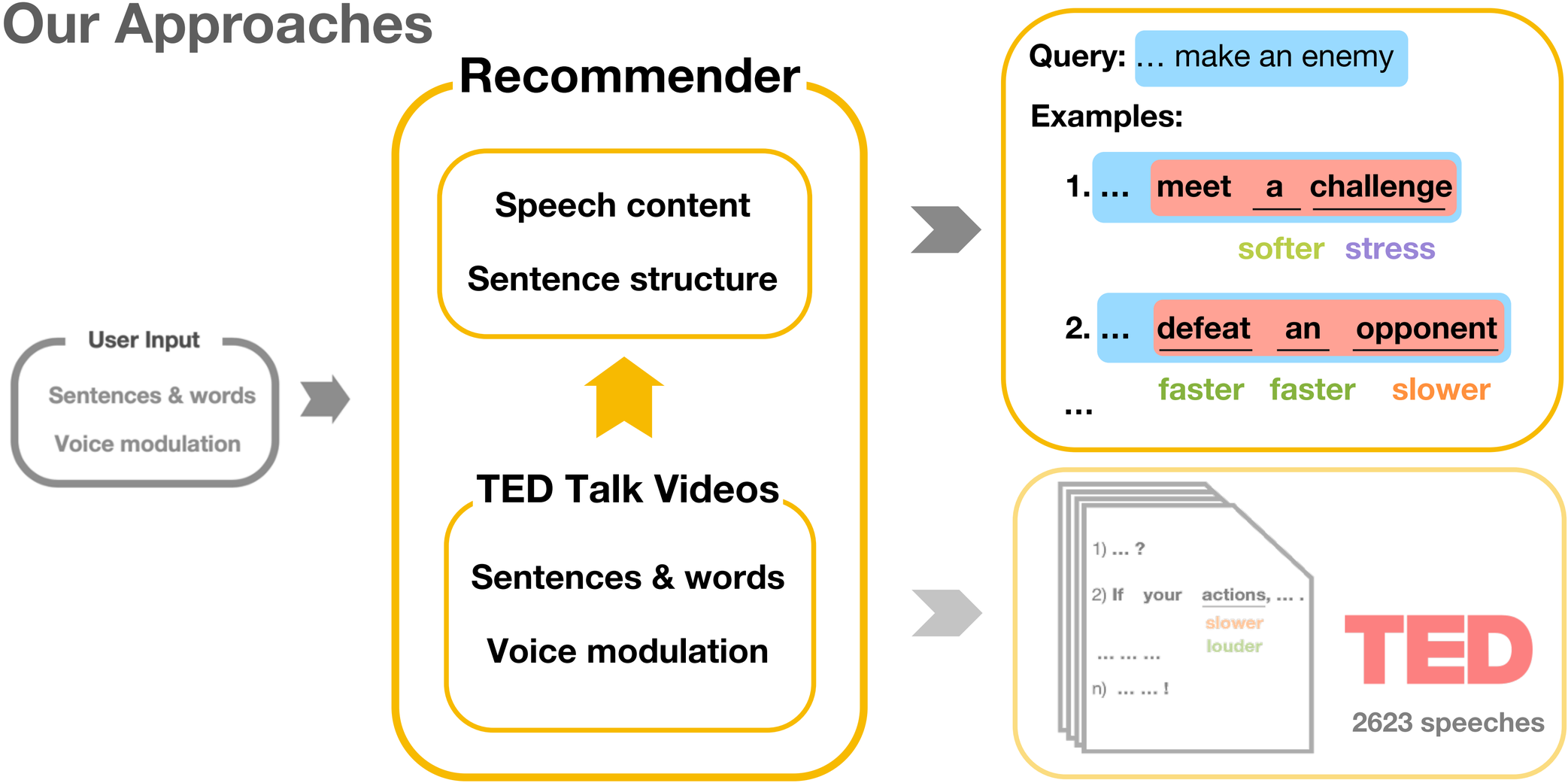
VoiceCoach

Our Approaches



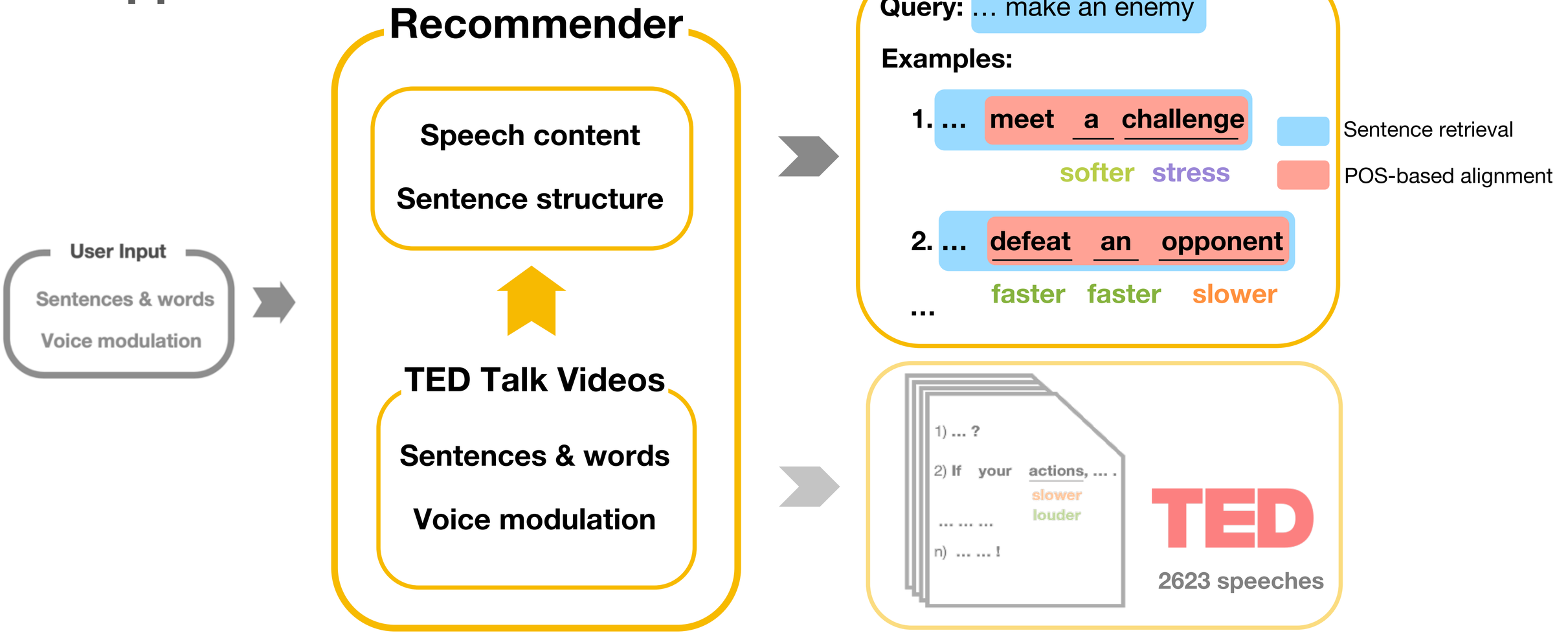
VoiceCoach

Our Approaches



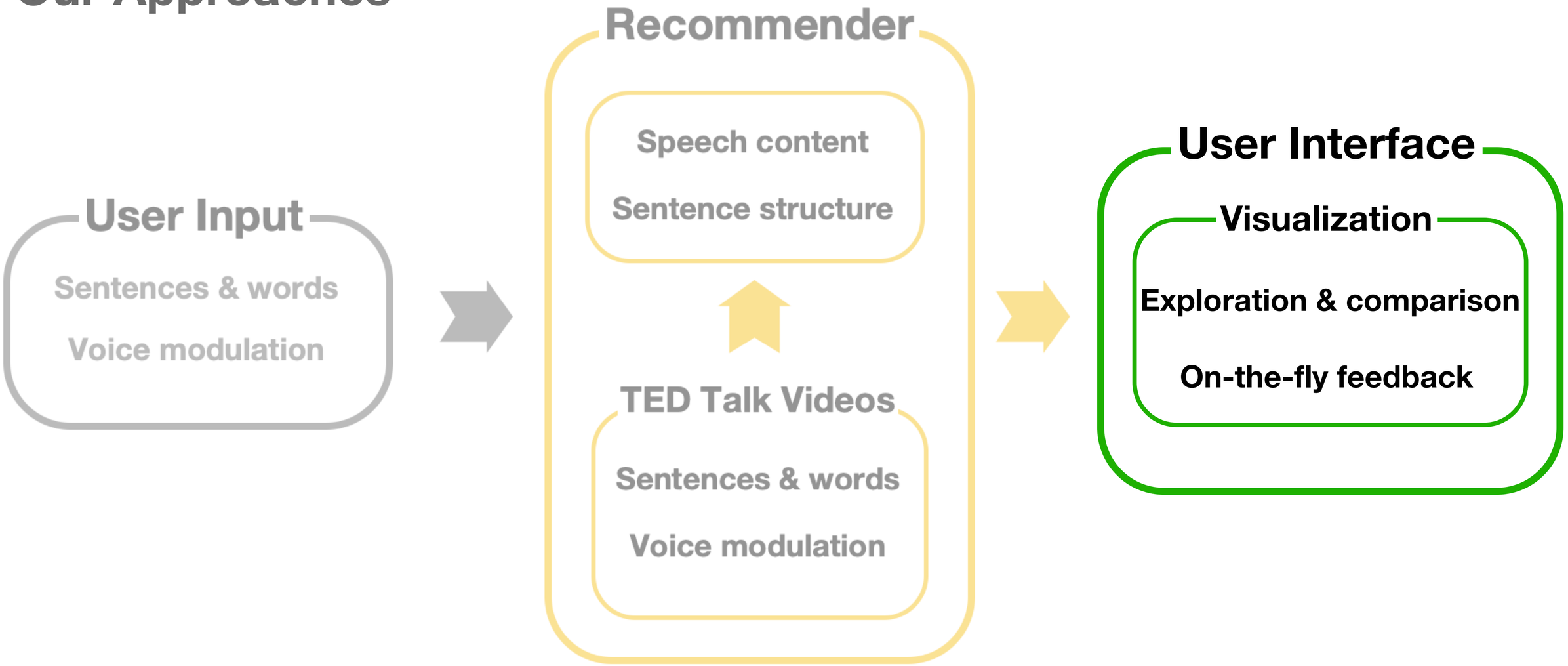
VoiceCoach

Our Approaches



VoiceCoach

Our Approaches

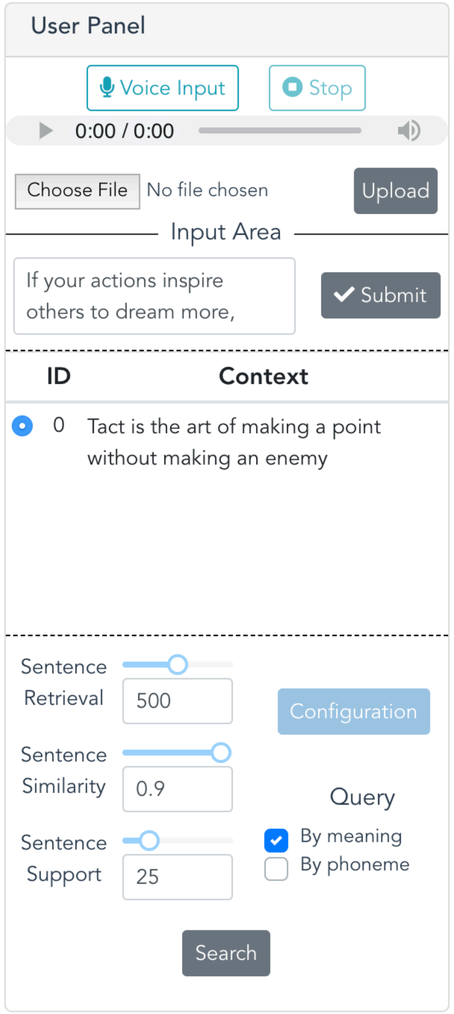


VoiceCoach

User Interface

VoiceCoach

User Interface - User Panel



A: Audio streaming

B: Speech file uploading

C: Text input

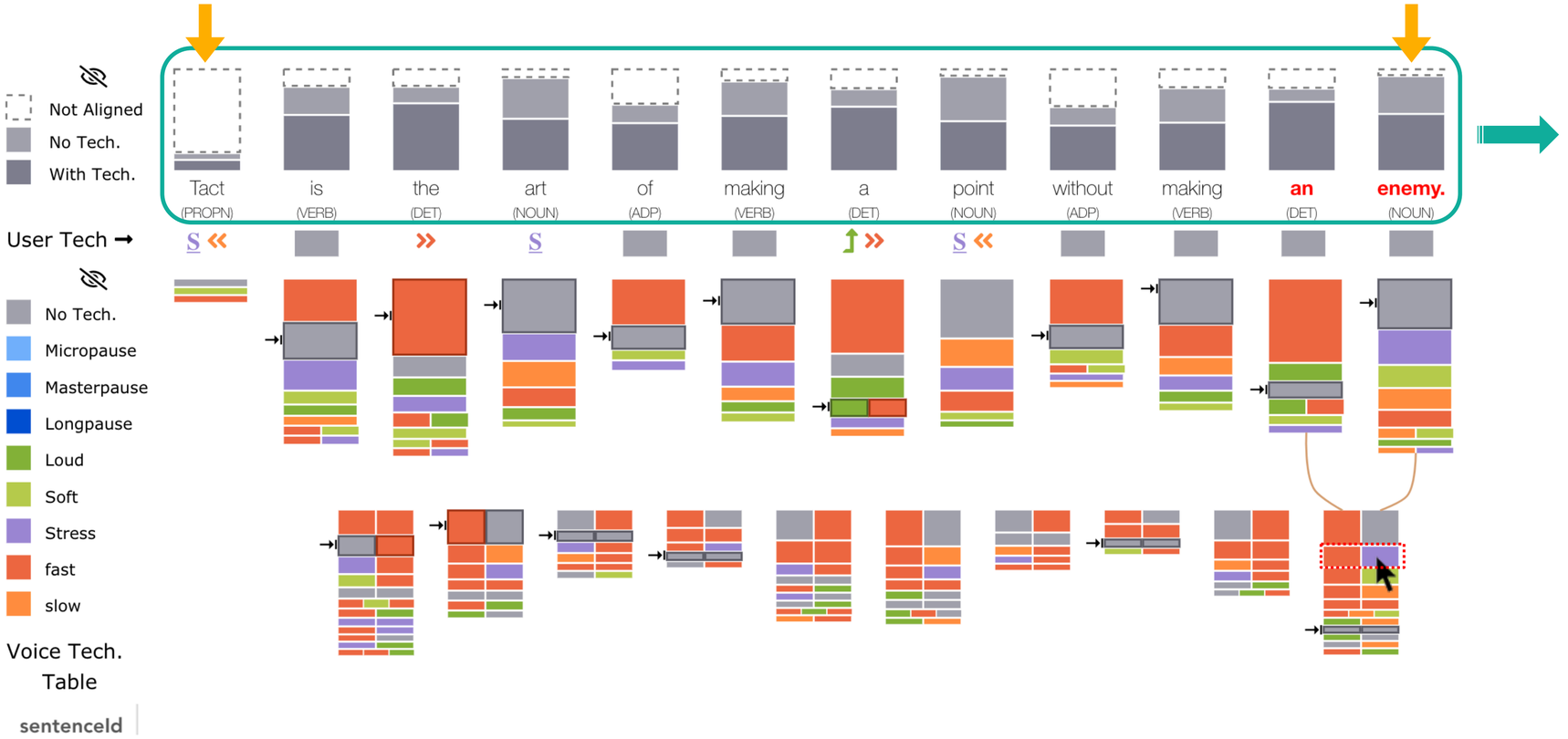
speech input

Parsing the speech input into sentences

Configuration of example retrieval

VoiceCoach

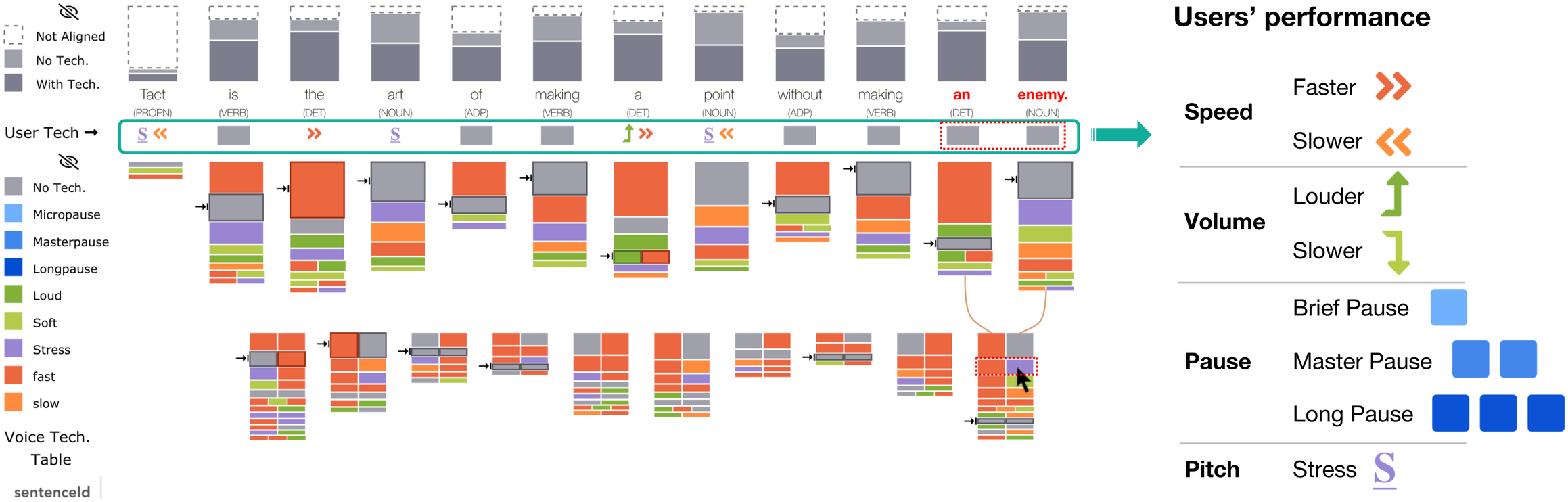
User Interface - Recommendation View



General summary of retrieval results

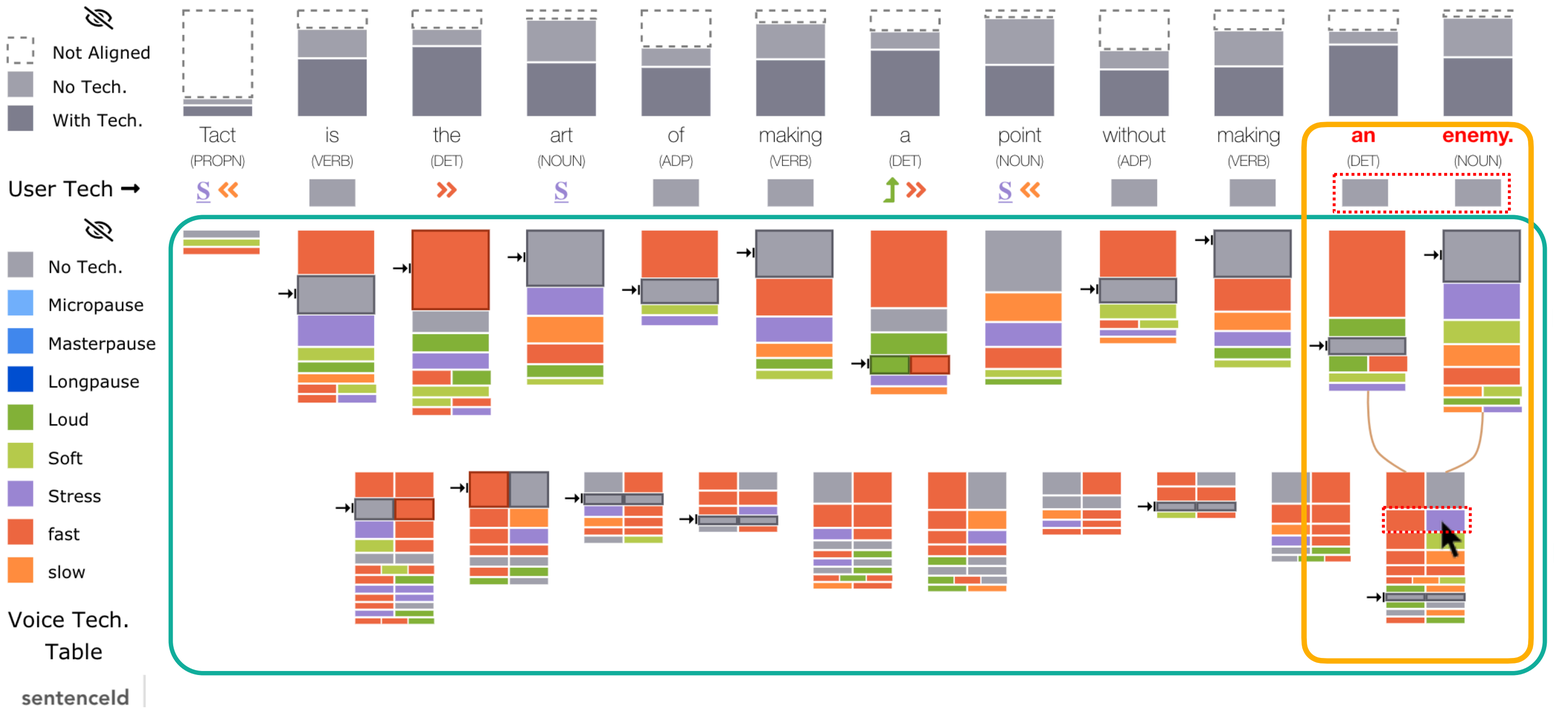
VoiceCoach

User Interface - Recommendation View



VoiceCoach

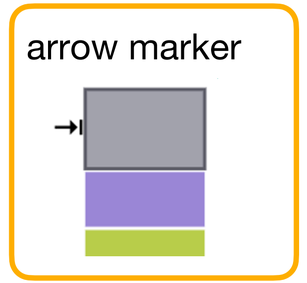
User Interface - Recommendation View



Word Combinations

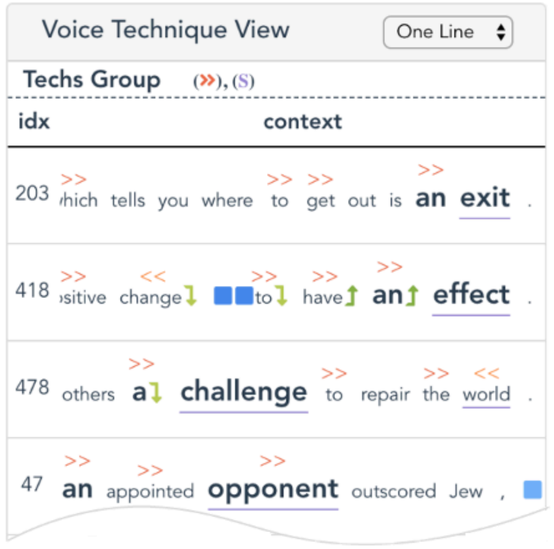
1-gram summary

2-gram summary



VoiceCoach

User Interface - Voice Technique View

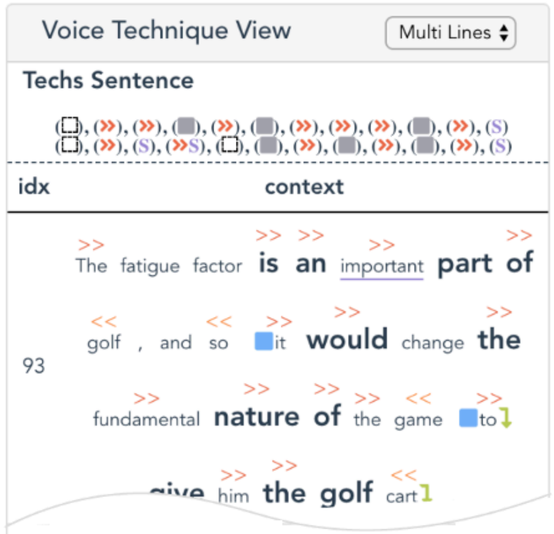


Query: Tact is … making **an enemy**

One-line mode

Phrase candidates for modulation of interest

- Combinations of modulation: (Faster », Stress S)
- Contexts of focused modulations



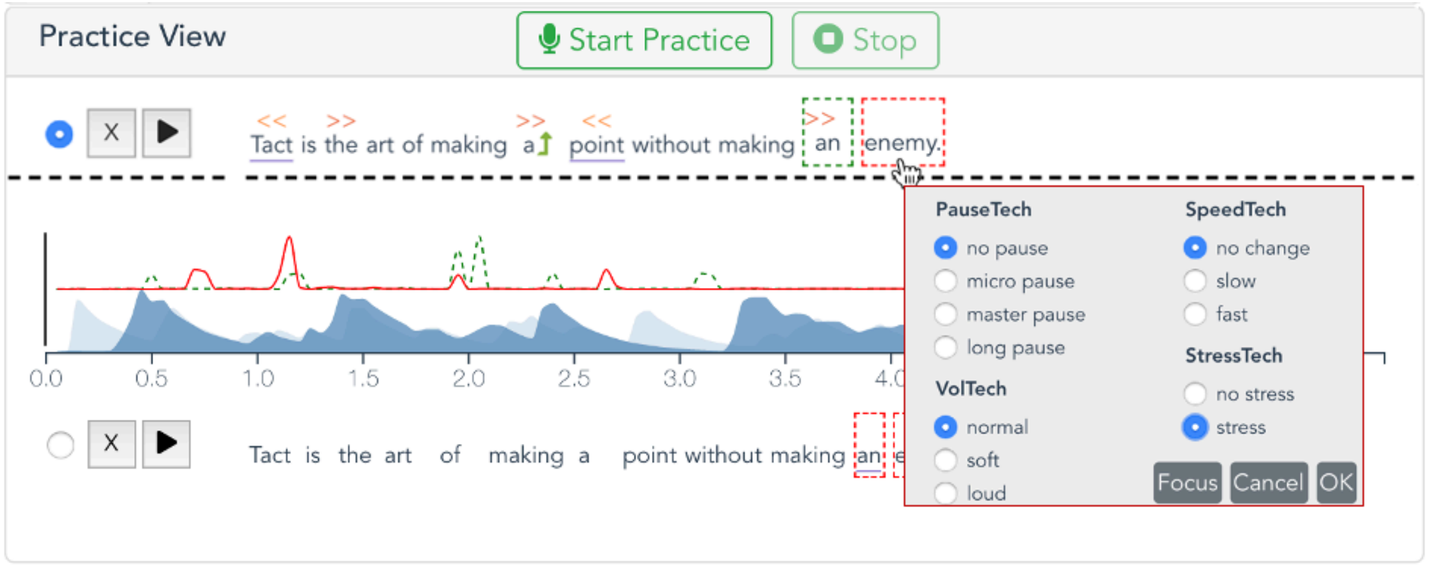
Multi-line mode

Corresponding sentence for modulation of interest

- Summary of modulation
- Contexts of focused sentences

VoiceCoach

User Interface - Practice View

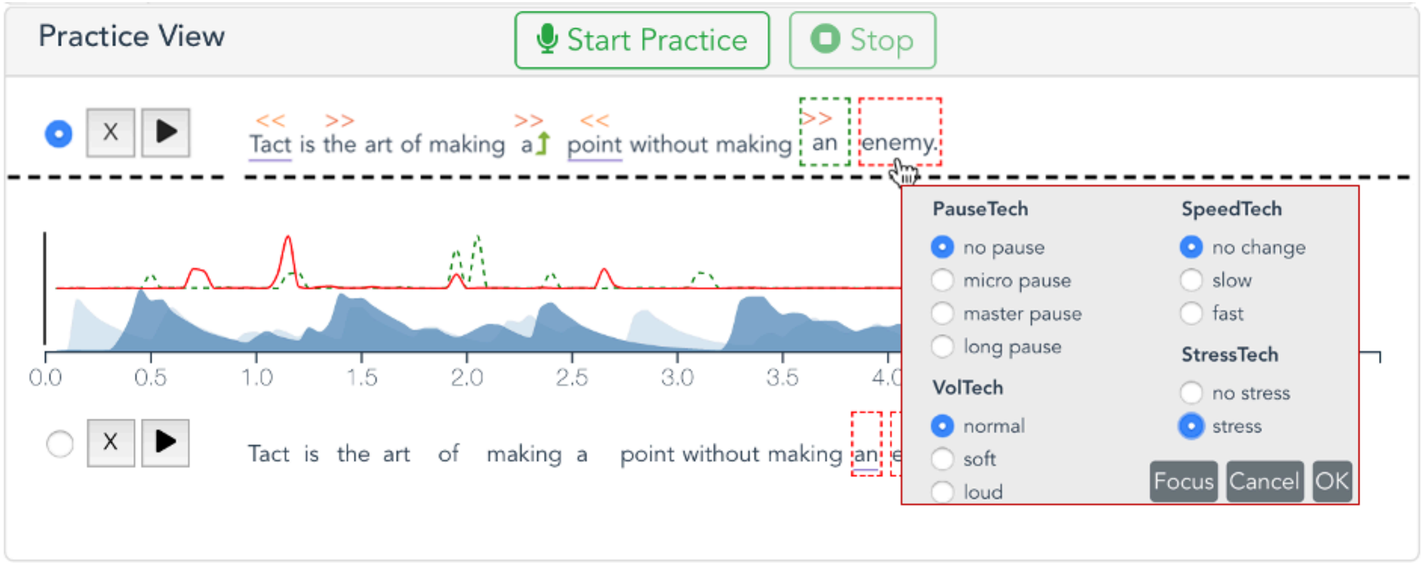


Original sentence and modulation

Configuration: an enemy (Faster », Stress S)

VoiceCoach

User Interface - Practice View



Realtime quantitative feedback curve & iterative practice

Last performance | Current performance

Line chart **(Pitch)**

Area chart **(Volume)**

Highlighting of differences

Pitch: line

Volume: area

Pause: segments that have volume = 0

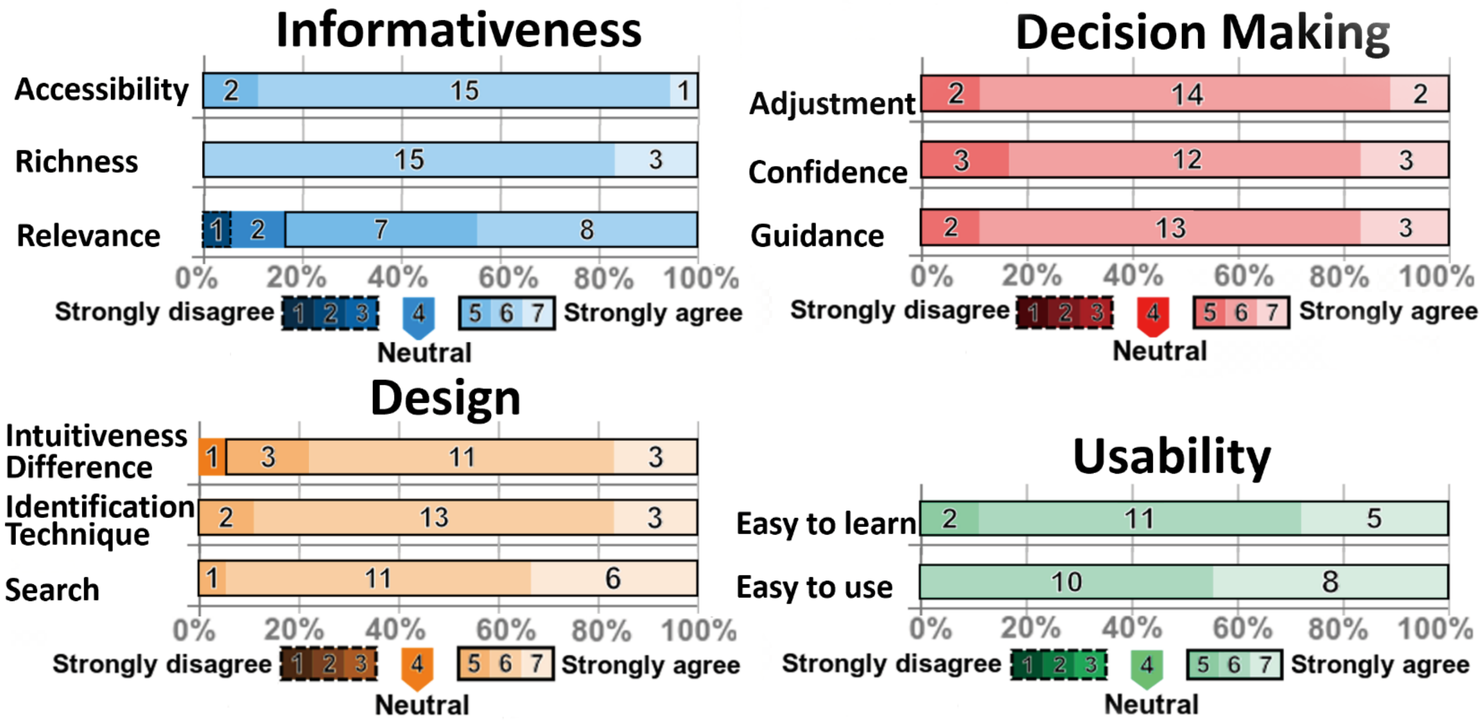
Speed: speed of curve

Evaluation

- **Expert Interviews**
- **User Study**
  - Recommendation helpfulness (recommendation view, voice technique view)
  - Effectiveness of immediate feedback (practice view)
  - Overall usability & effectiveness

User Study

Evaluation on recommendation (Recommendation View, Voice Technique View)



Questionnaire results of recommendation in 4 aspects

Recommended examples

(Voice Technique View)

- 4.21 (SD = 1.21) of top 5 satisfied the participants' needs.
- The relevance rate was 89%.

Overall feedback

- Positive, e.s.p. decision making & usability

User Study

Evaluation on Practice View

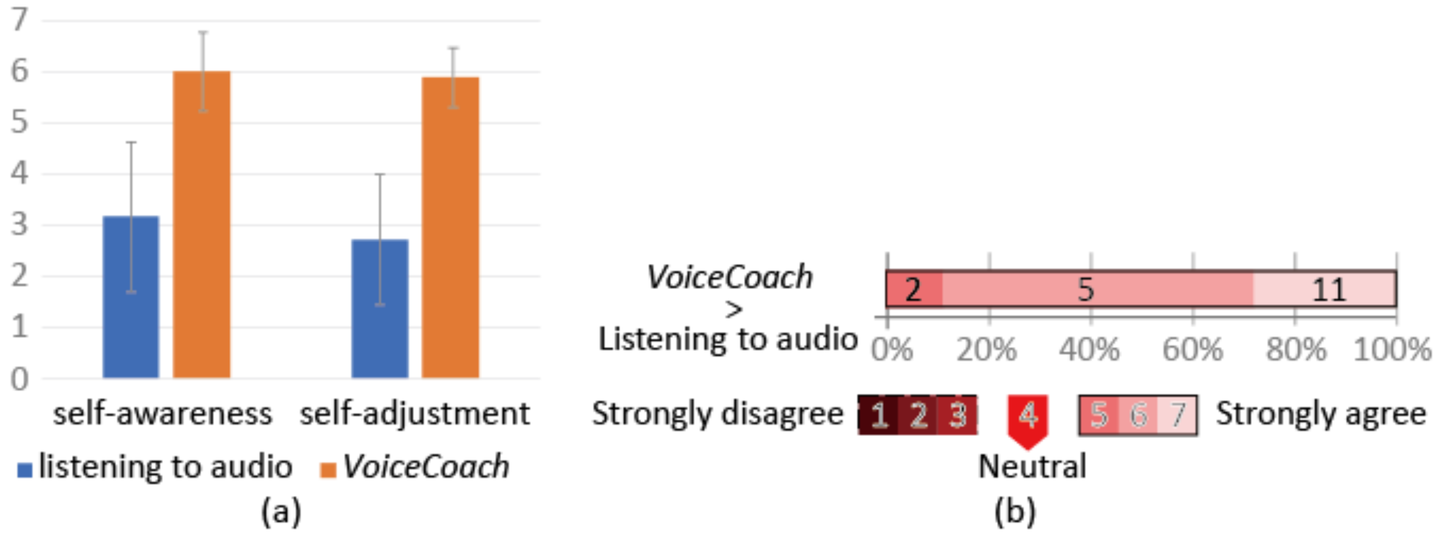


Fig. Results of questionnaire about user experience of practice:

VoiceCoach V.S. Listening to audio

Wilcoxon signed-rank test

- Self-awareness ($P<0.001$, $Z=-3.75$)
- Self-adjustment ($P<0.001$, $Z=-3.70$)

Overall feedback

- All participants prefer VoiceCoach

User Study

Overall Evaluation

1. Participants' feedback on VoiceCoach

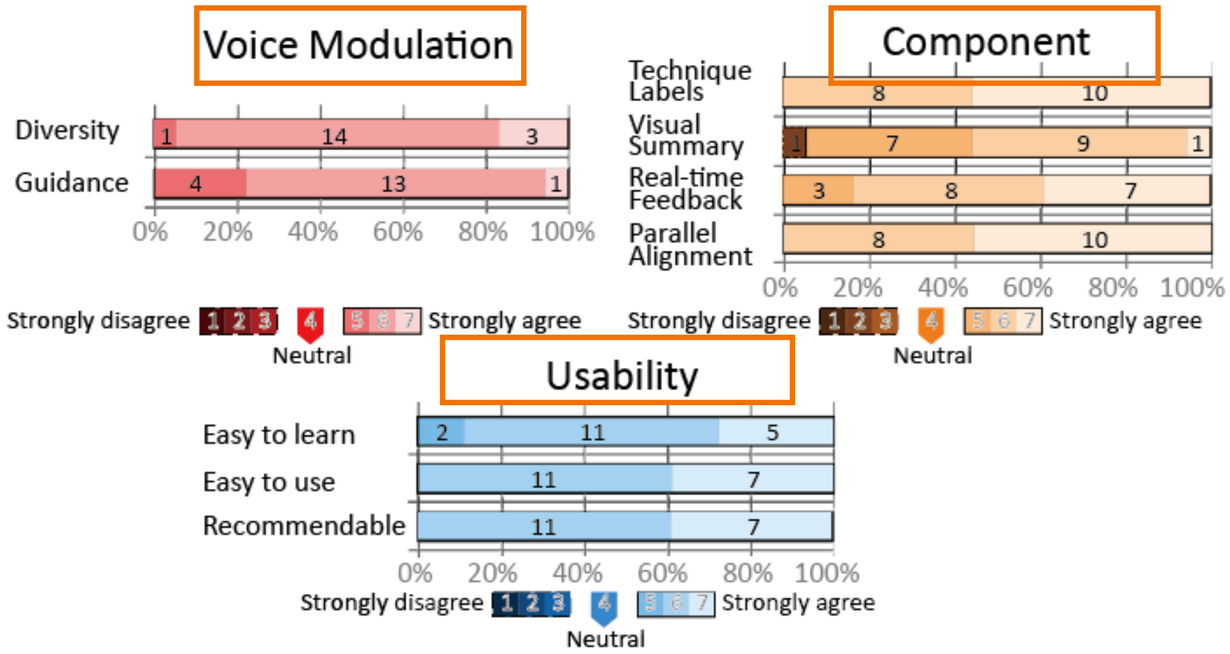


Fig. Results of questionnaire about system.

2. Scoring by coaches on participants' performance

Wilcoxon signed-rank test (**with/without** VoiceCoach)

- Performance (P=0.03, Z=-2.15)
  - VoiceCoach (5.08 +/- 1.08)
  - Without VoiceCoach (4.17 +/- 1.03)

Future Work

- **Extend** current dataset (e.g., academic talks)
- Learning from “**BAD**” examples of voice modulation
- **Improve** recommendation
  - Users’ preferences (e.g., examples with more pauses)

END

VoiceCoach

Thank you!

Xingbo Wang

Ph.D. Student @VISLab, HKUST

Email: xwangeg@cse.ust.hk

Home Page: http://www.cse.ust.hk/~xwangeg/