

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



Xingbo Wang



Haipeng Zeng



Yong Wang



Aoyu Wu



Zhida Sun



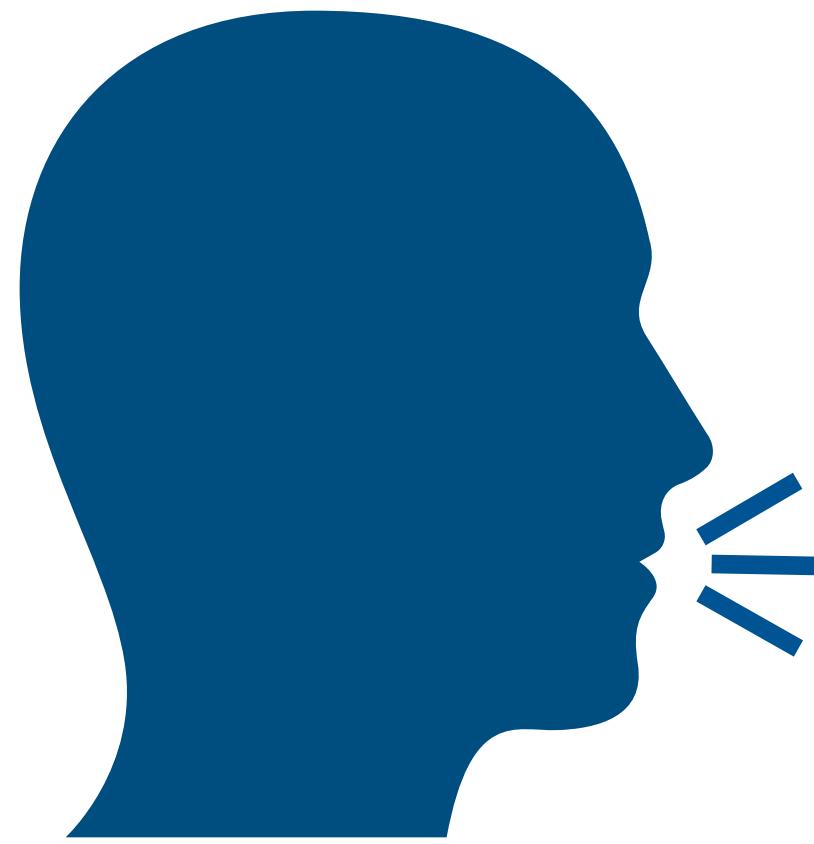
Xiaojuan Ma



Huamin Qu

Introduction

Voice Modulation



Pitch

Volume

Speed

Pause



Public Speaking



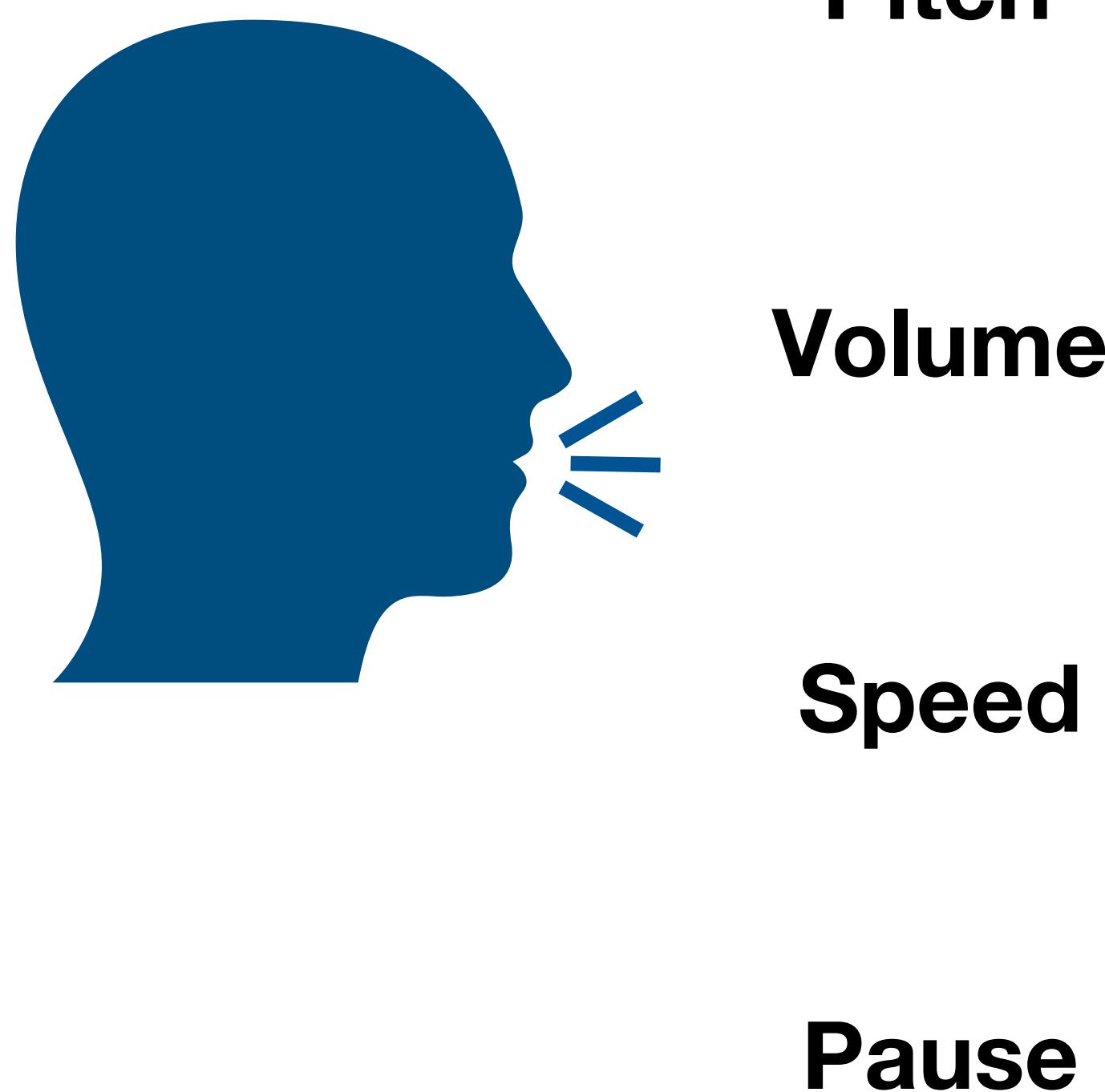
audience engagement



convey key ideas

Introduction

Voice Modulation



- **Higher volume/pitch** → vocal emphasis
- **Increasing** speed → excitement
- **Slowing** down and **pauses** → content reflection
→ personal connections



convey key ideas

Introduction

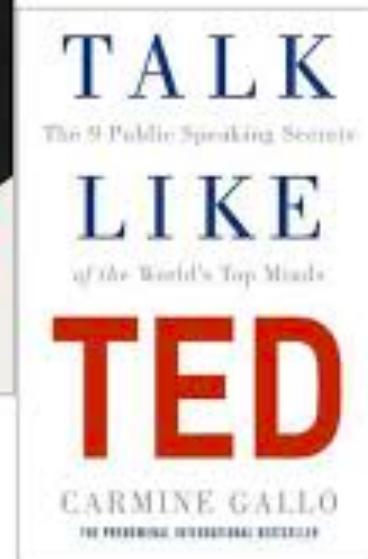
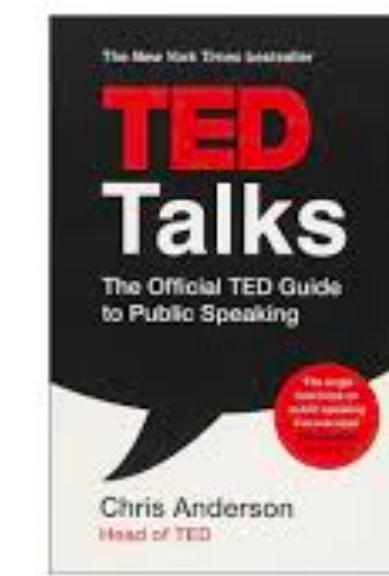
Voice Modulation



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

Soln2: Join the public speaking training programs

Problems: (feedback)

1. No quantitative evaluation
2. Subjected to personal preferences



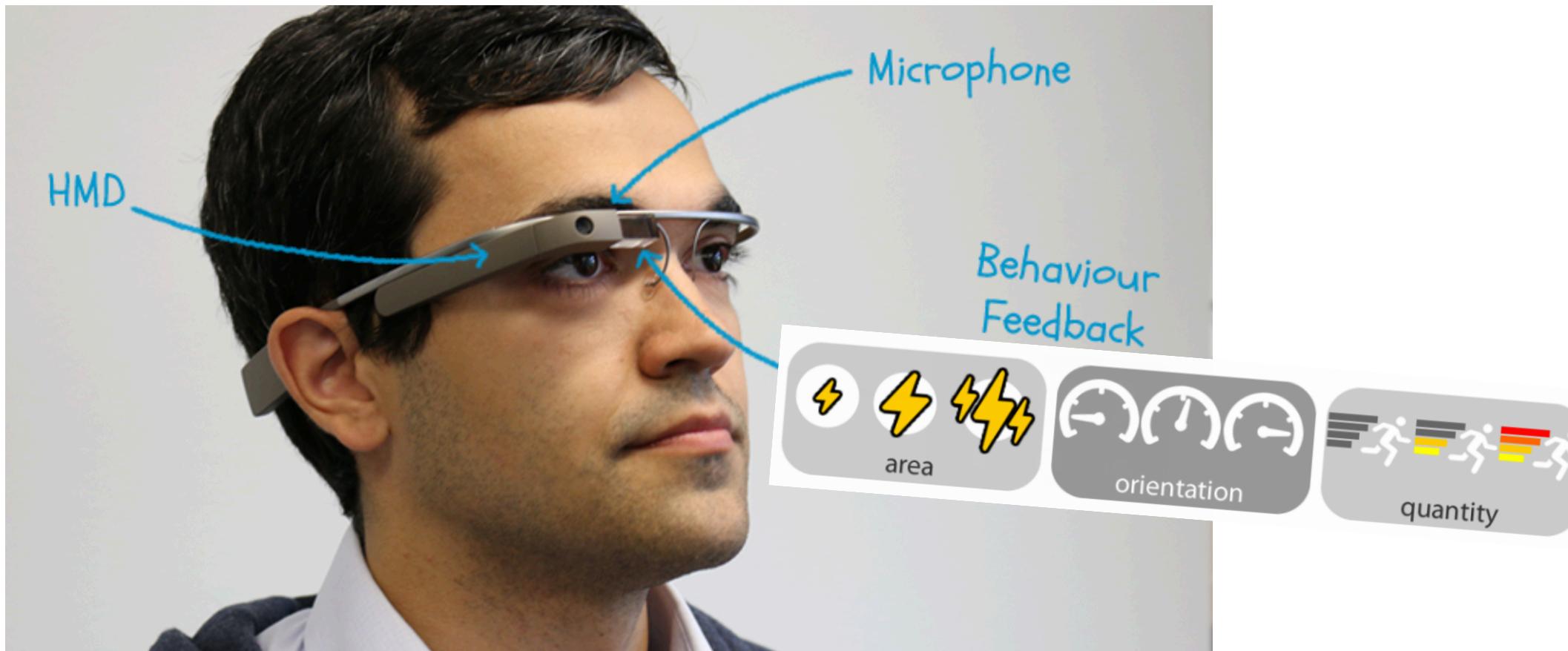
Introduction

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**

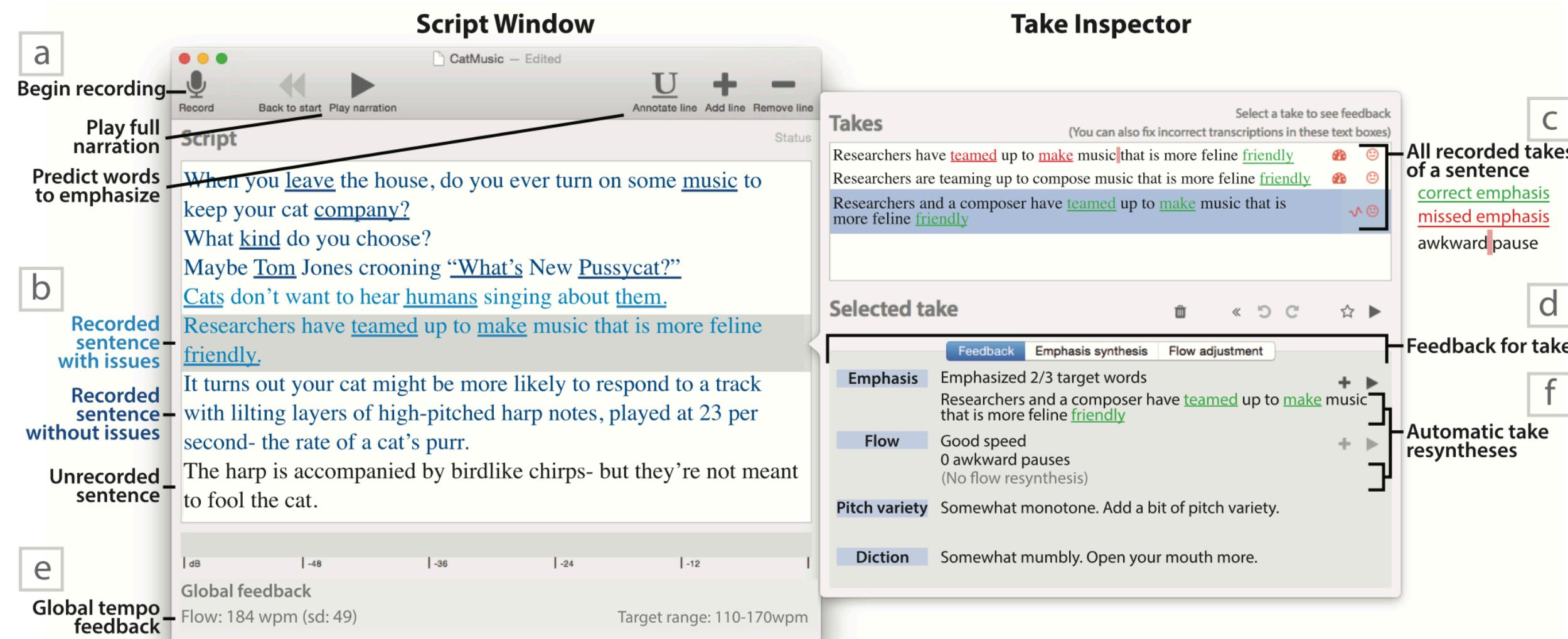
Introduction

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)

→ Difficult for novice speakers to **specify requirements**

Introduction

Training Systems for Voice Modulation

To facilitate the training process:

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

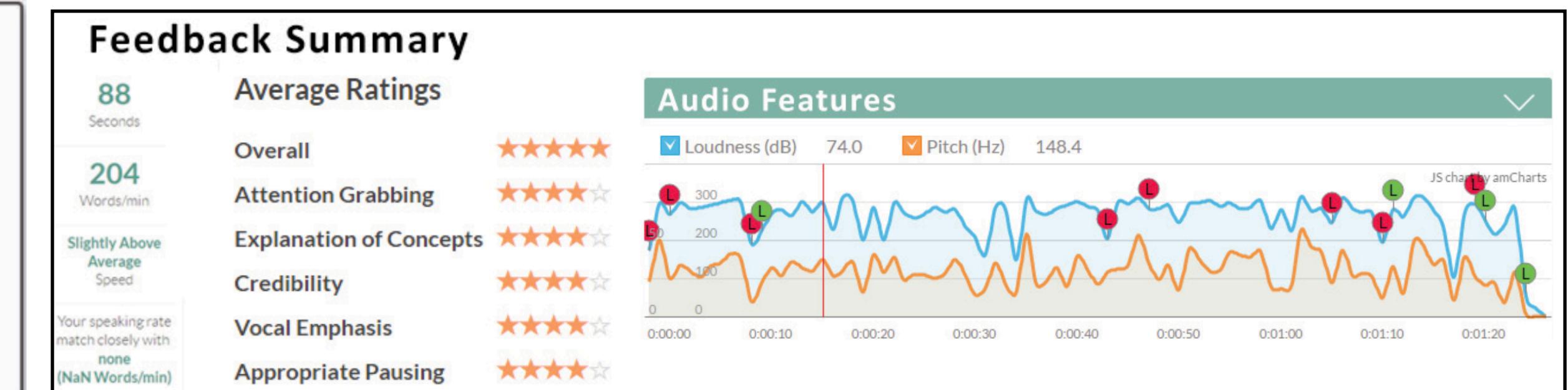
for self-reflection & practice



MACH (Hoque et al., 2013)

Simple charts with limited interaction support

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



ROC Speak (Fung et al., 2015)

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

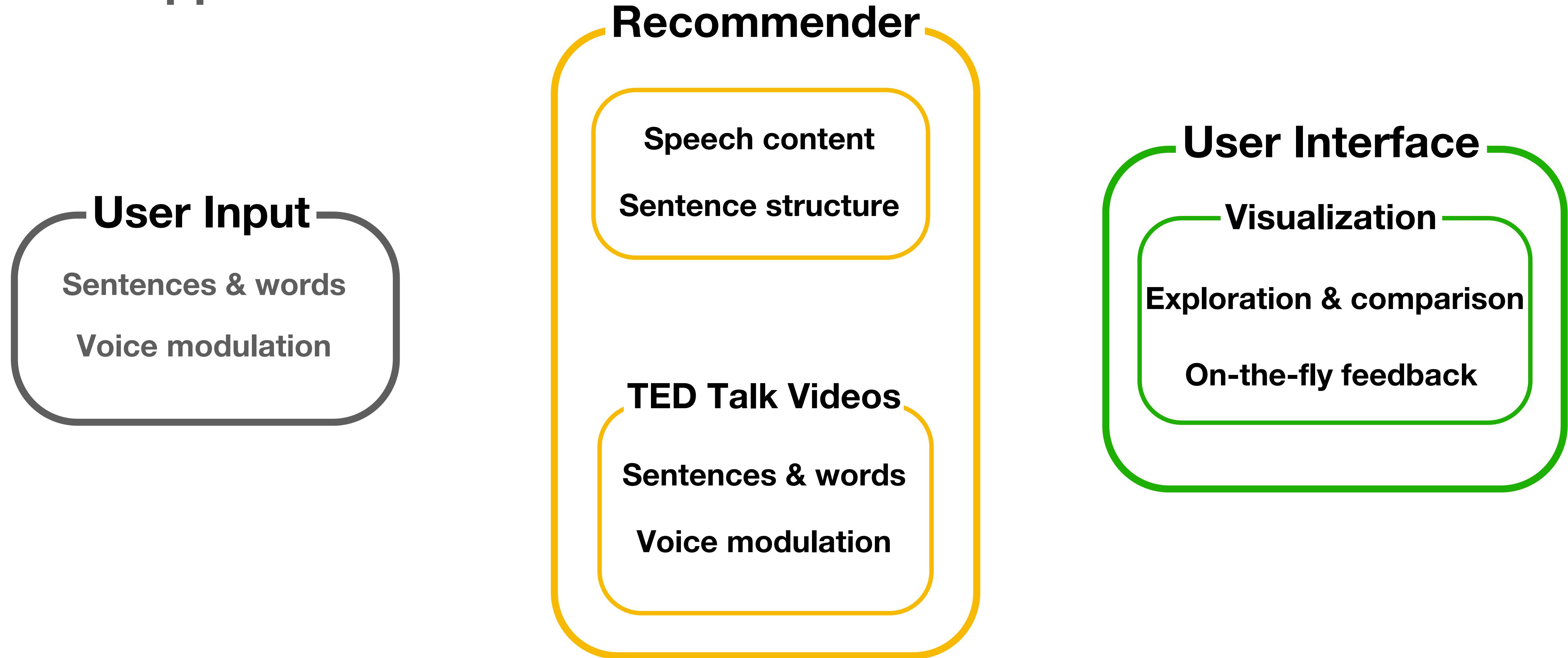
VoiceCoach

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**

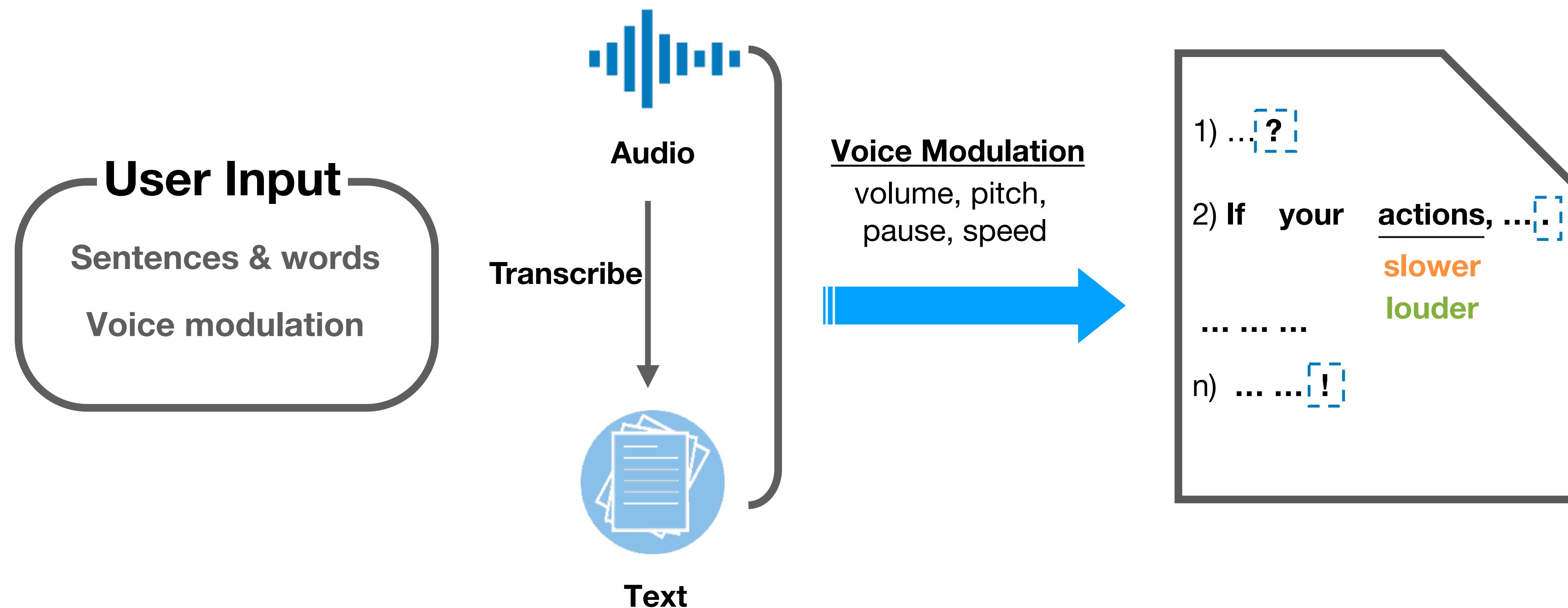
VoiceCoach

Our Approaches



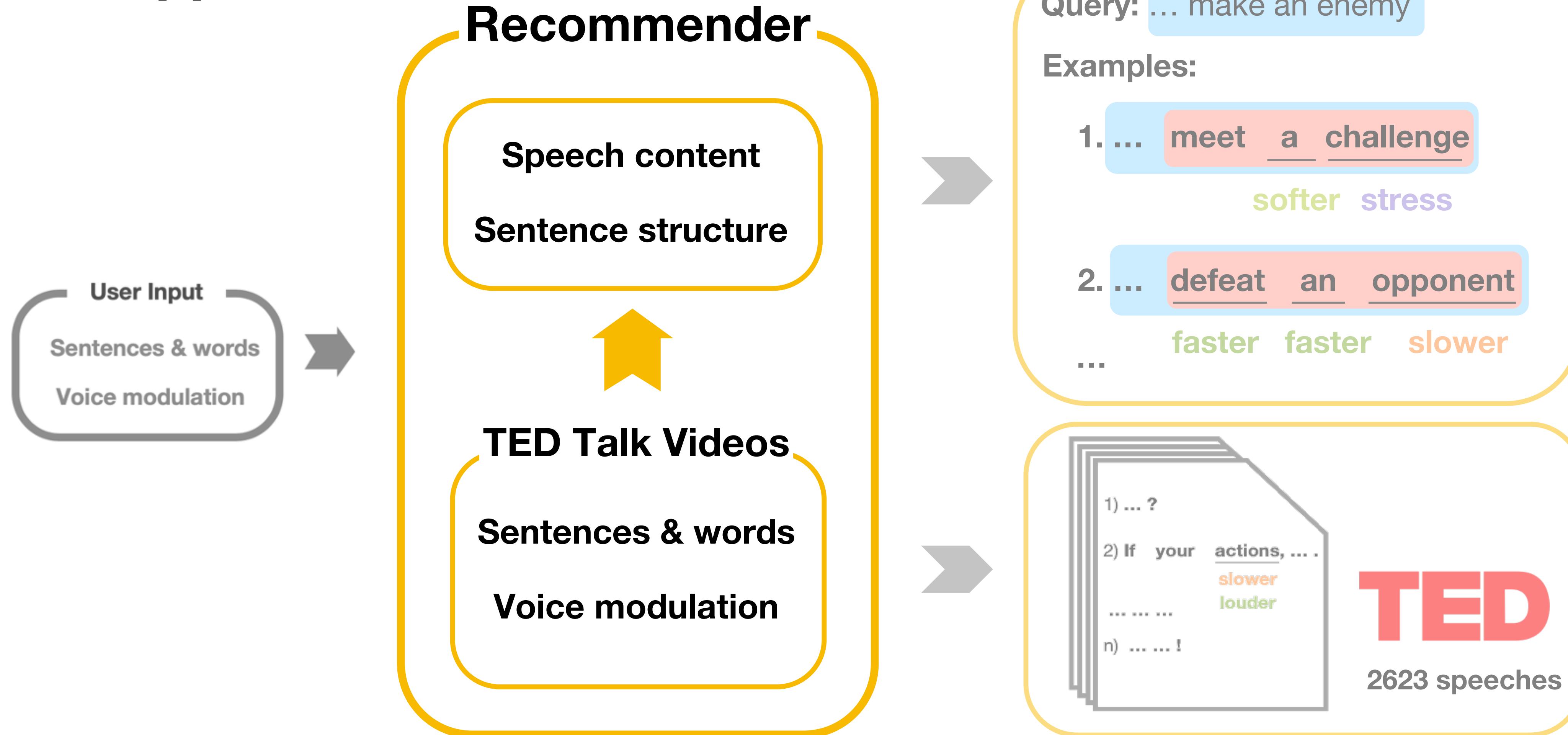
VoiceCoach

Our Approaches



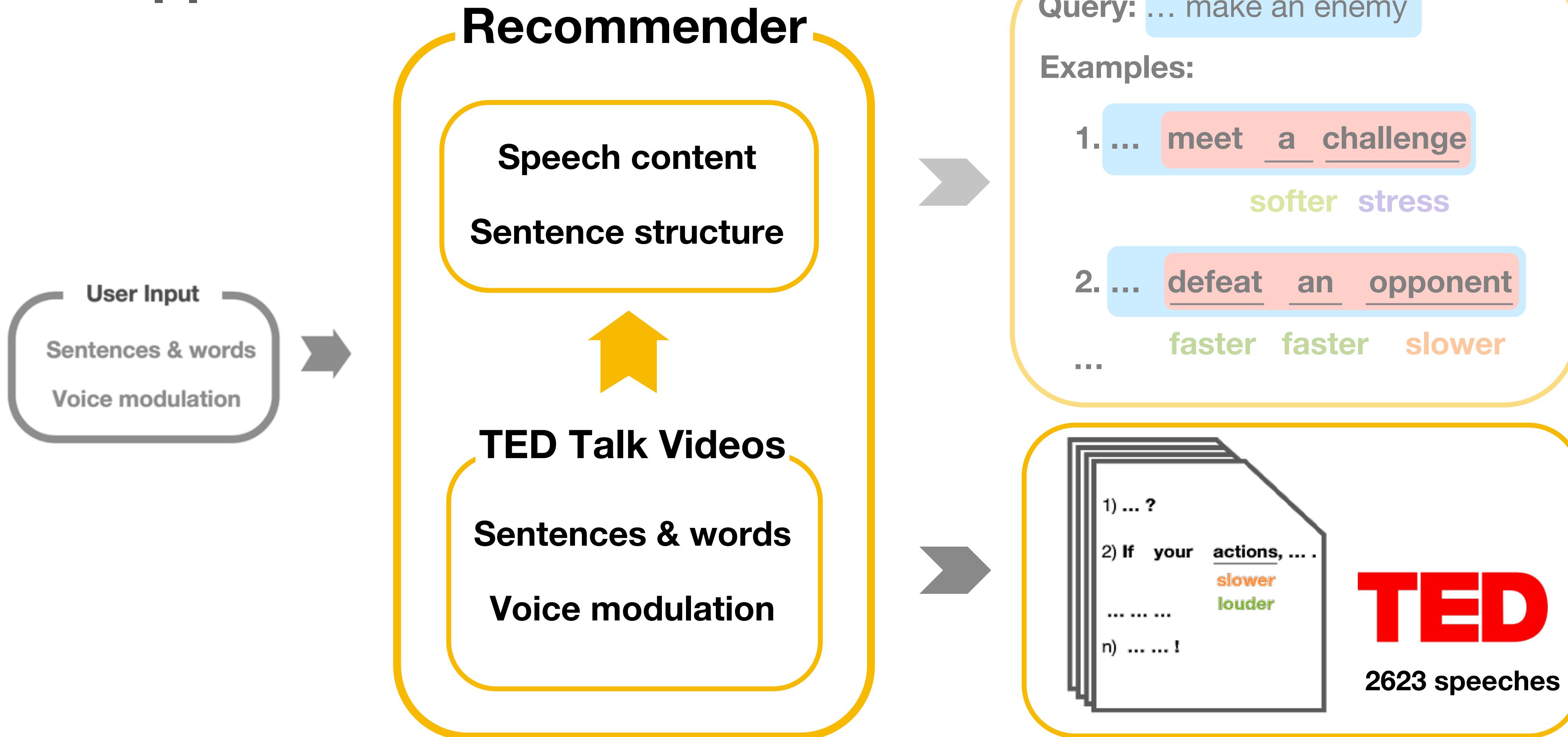
VoiceCoach

Our Approaches



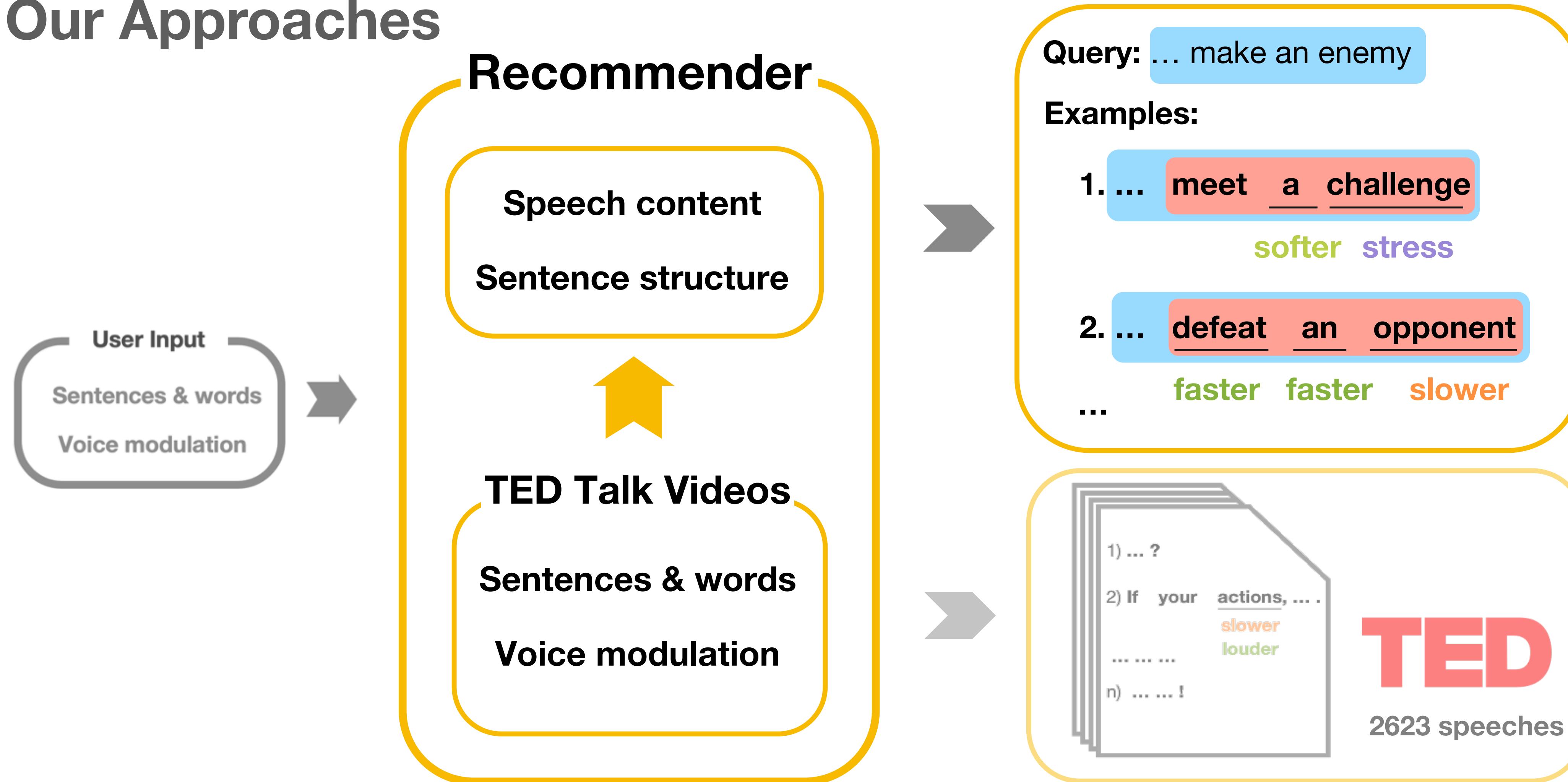
VoiceCoach

Our Approaches



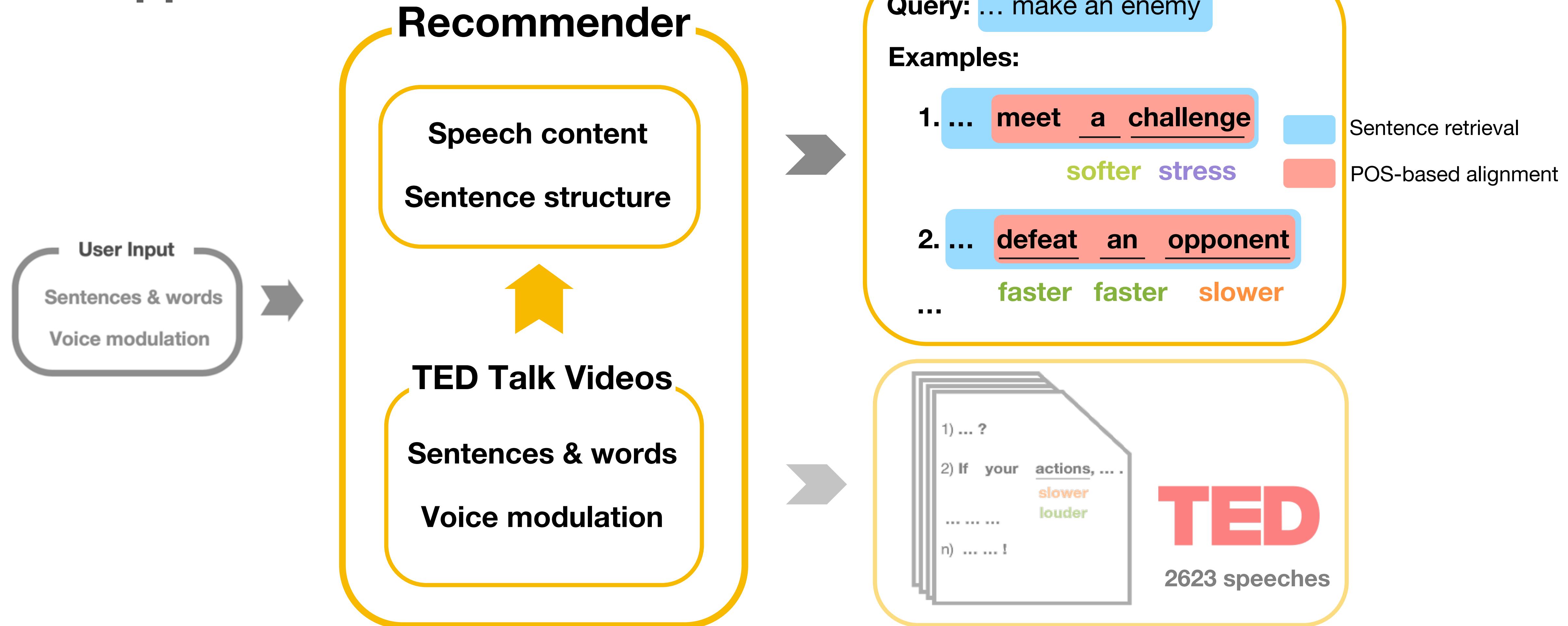
VoiceCoach

Our Approaches



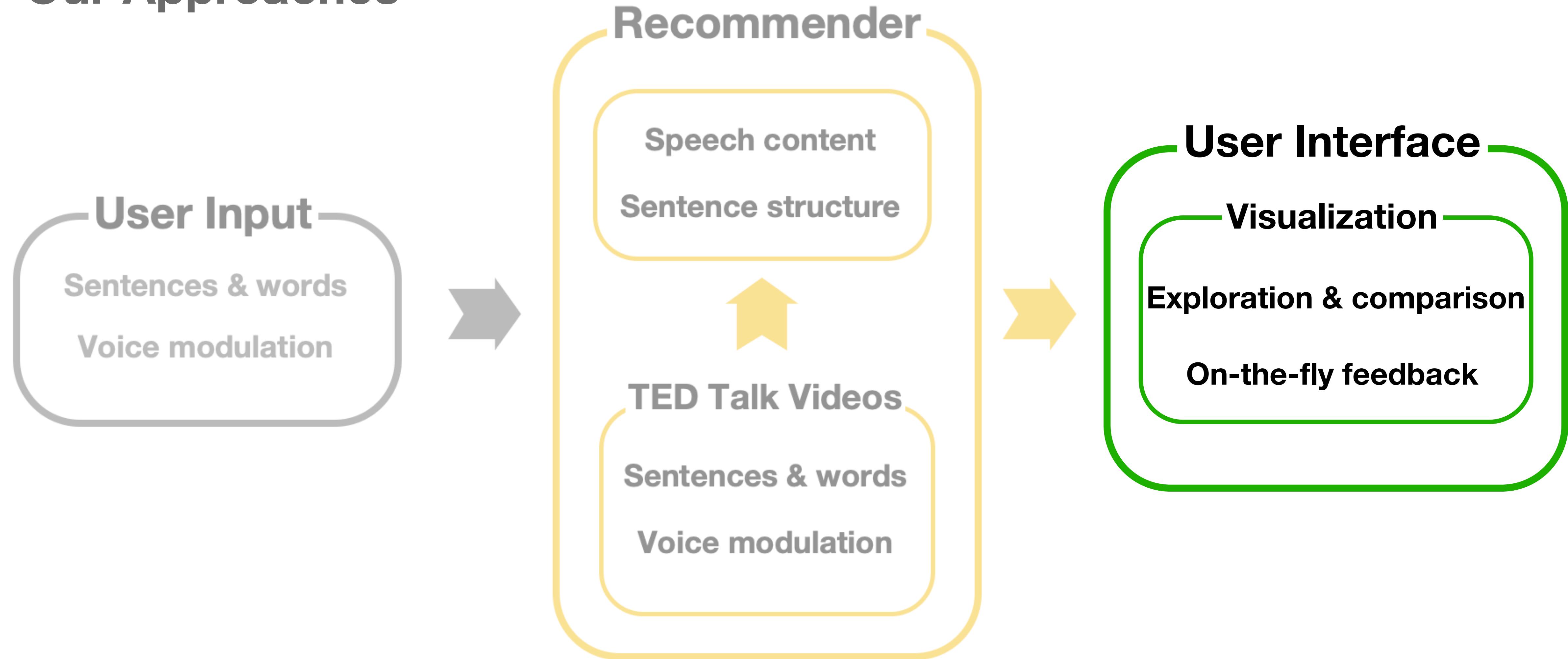
VoiceCoach

Our Approaches



VoiceCoach

Our Approaches



VoiceCoach

User Interface

User Panel

Voice Input Stop
0:05 / 0:05

Choose File No file chosen Upload

If your actions inspire others to dream more, Submit

Sentence Retrieval 500 Configuration

Sentence Similarity 0.9 Query

Sentence Support 25 By meaning By phoneme

Search

Recommendation View

Not Aligned, No Tech., With Tech.

Tact (PROPN) is (VERB) the (DET) art (NOUN) of (ADP) making (VERB) a (DET) point (NOUN) without (ADP) making (VERB) an (DET) enemy. (NOUN)

User Tech →

- No Tech.
- Micropause
- Masterpause
- Longpause
- Loud
- Soft
- Stress
- fast
- slow

Voice Tech. Table

sentenceld

Practice View

Start Practice Stop

<< >> Tact is the art of making a ↑ point without making an enemy.

PauseTech: no pause, micro pause, master pause, long pause

SpeedTech: no change, slow, fast

StressTech: no stress, stress

VolTech: normal, soft, loud

Focus Cancel OK

Voice Technique View One Line

Techs Group (», (S)

idx	context
203	>> which tells you where to get out is <u>an exit</u> .
418	>> positive change ↴ to ↴ have ↑ <u>an ↑ effect</u> .
478	others a ↴ <u>challenge</u> to repair the world.
47	>> an appointed <u>opponent</u> outscored Jew , ↴

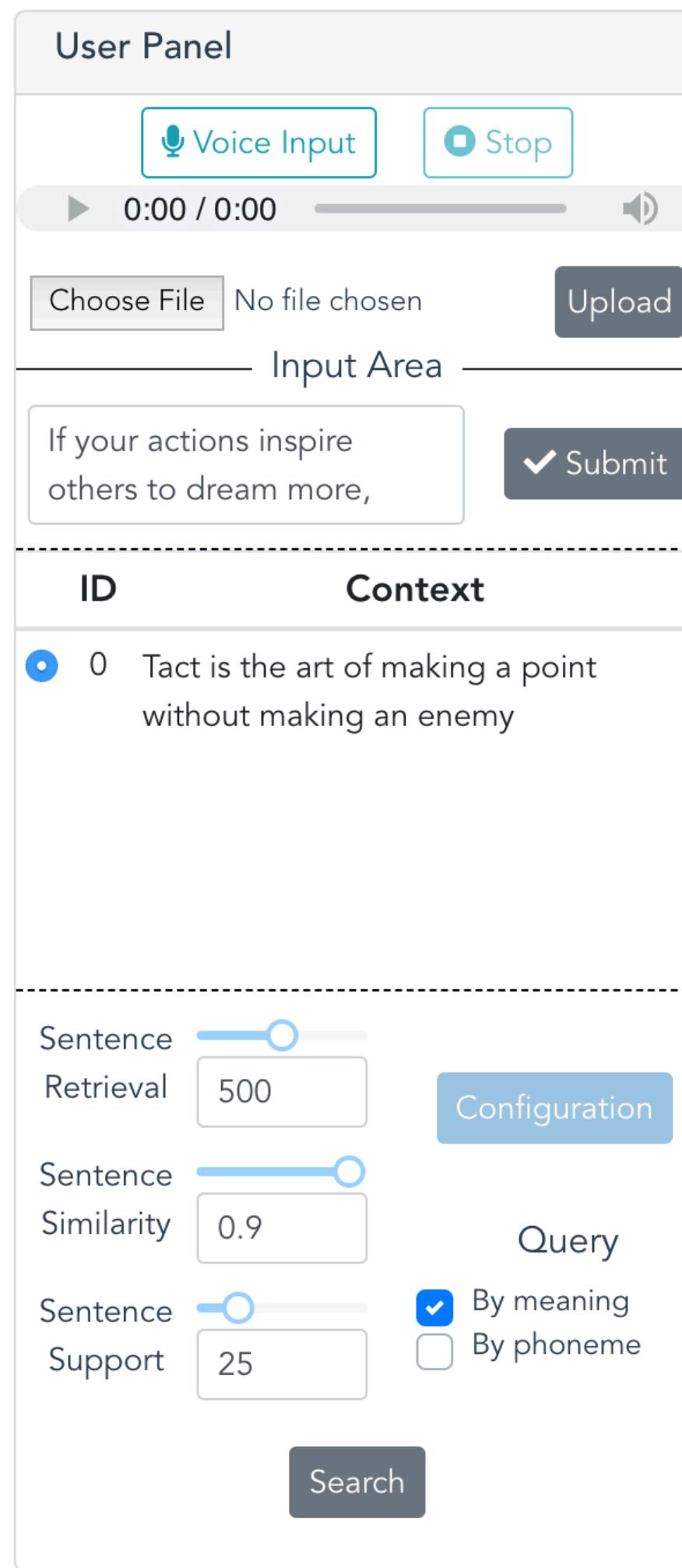
Voice Technique View Multi Lines

Techs Sentence

idx	context
93	>> The fatigue factor <u>is an important part of</u> golf , and so ↴ it <u>would change the fundamental nature of</u> the game ↴ to ↴ give him <u>the golf</u> ↴

VoiceCoach

User Interface - User Panel



- A: Audio streaming
- B: Speech file uploading
- C: Text input

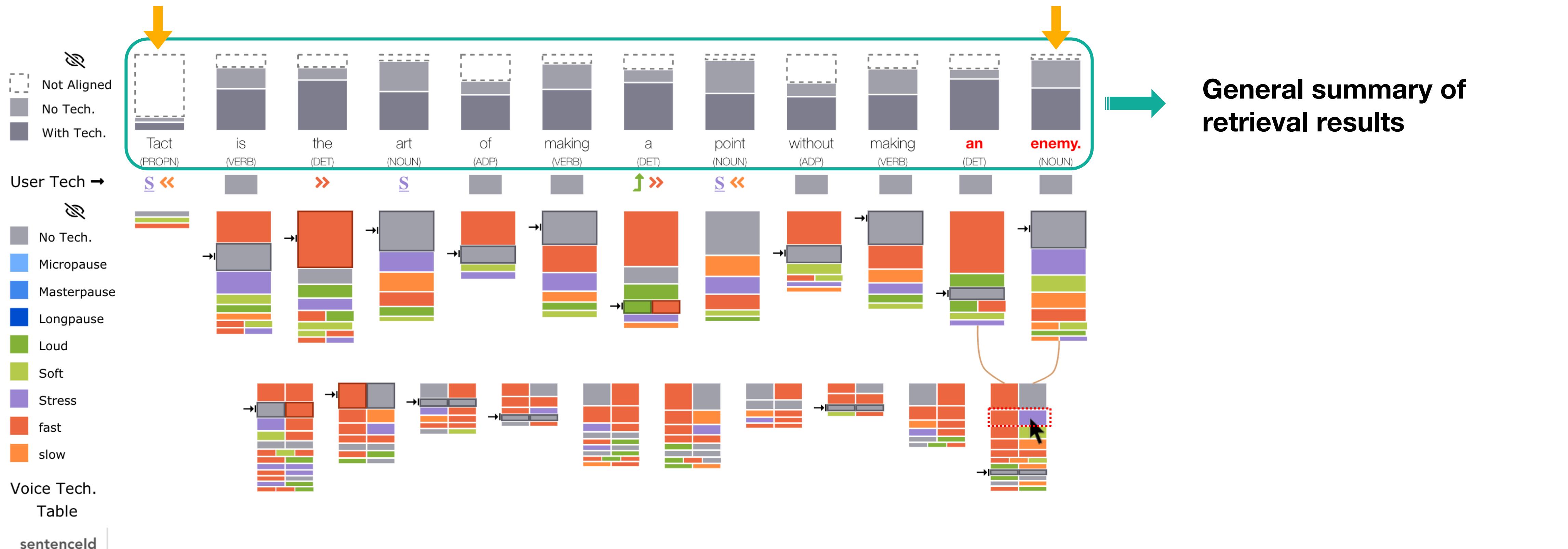
speech input

■ Parsing the speech input into sentences

■ Configuration of example retrieval

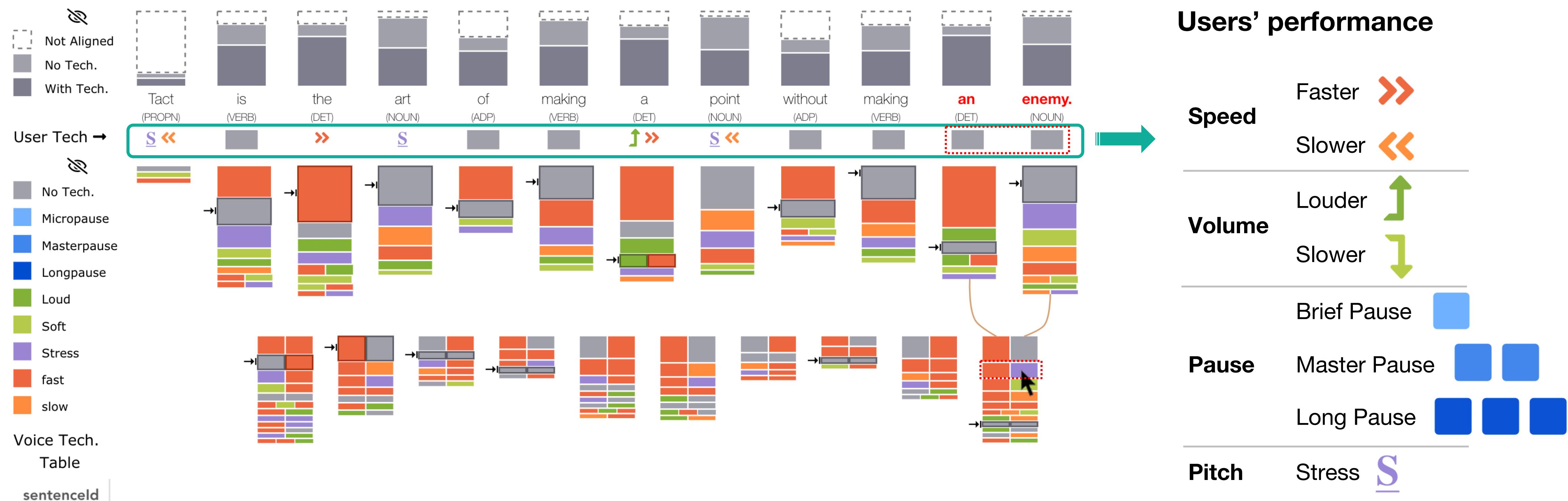
VoiceCoach

User Interface - Recommendation View



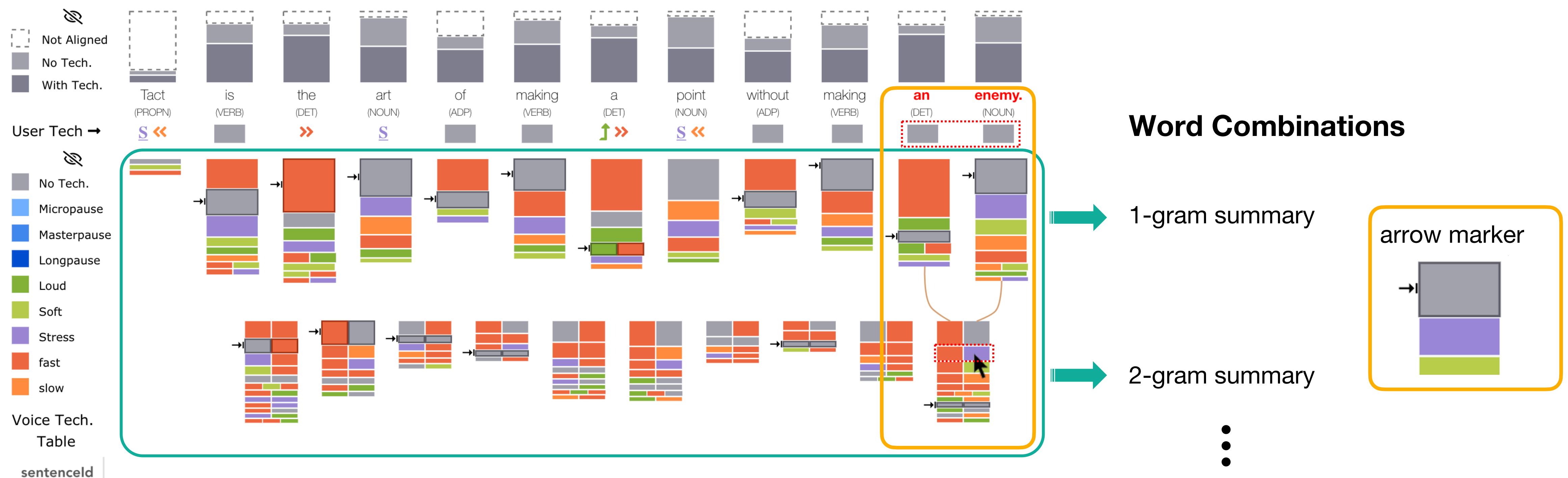
VoiceCoach

User Interface - Recommendation View



VoiceCoach

User Interface - Recommendation View



VoiceCoach

User Interface - Voice Technique View

Voice Technique View	
Techs Group (», (S)	
idx	context
203	>> which tells you where to get out is <u>an exit</u> .
418	>> << positive change ↓ ■■ to ↓ have ↑ <u>an ↑ effect</u> .
478	>> others a↓ <u>challenge</u> >> to repair >> <<
47	>> >> >> an appointed <u>opponent</u> outscored Jew , ■

Voice Technique View	
Techs Sentence	
idx	context
	(□), (»), (»), (□), (»), (□), (»), (»), (»), (»), (»), (»), (S) (□), (»), (S), (»S), (□), (»), (»), (»), (»), (»), (»), (»), (»), (S)
93	>> The fatigue factor <u>is an important part of</u> << golf , and so ■ it <u>would change the</u> >> >> >> << >> fundamental <u>nature of</u> the game ■ to ↓ give him >> >> the golf << 1

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

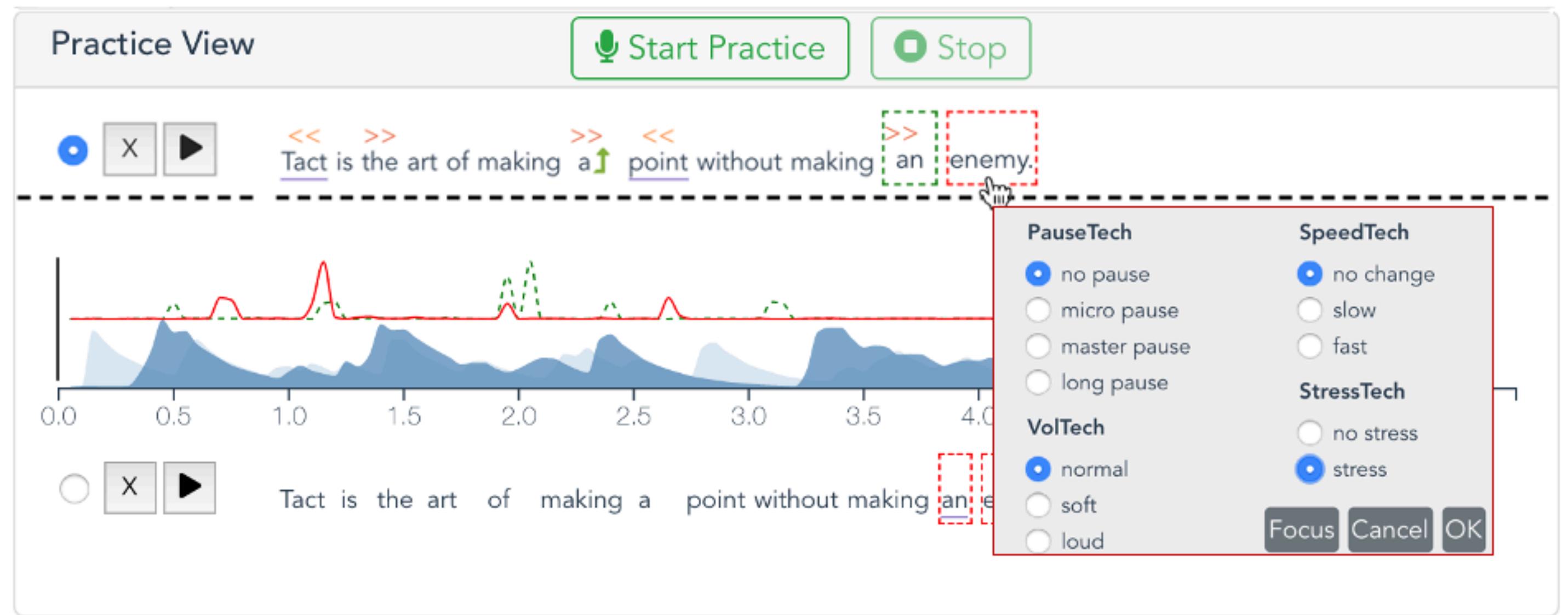
Multi-line mode

Corresponding sentence for modulation of interest

- Summary of modulation
- Contexts of focused sentences

VoiceCoach

User Interface - Practice View

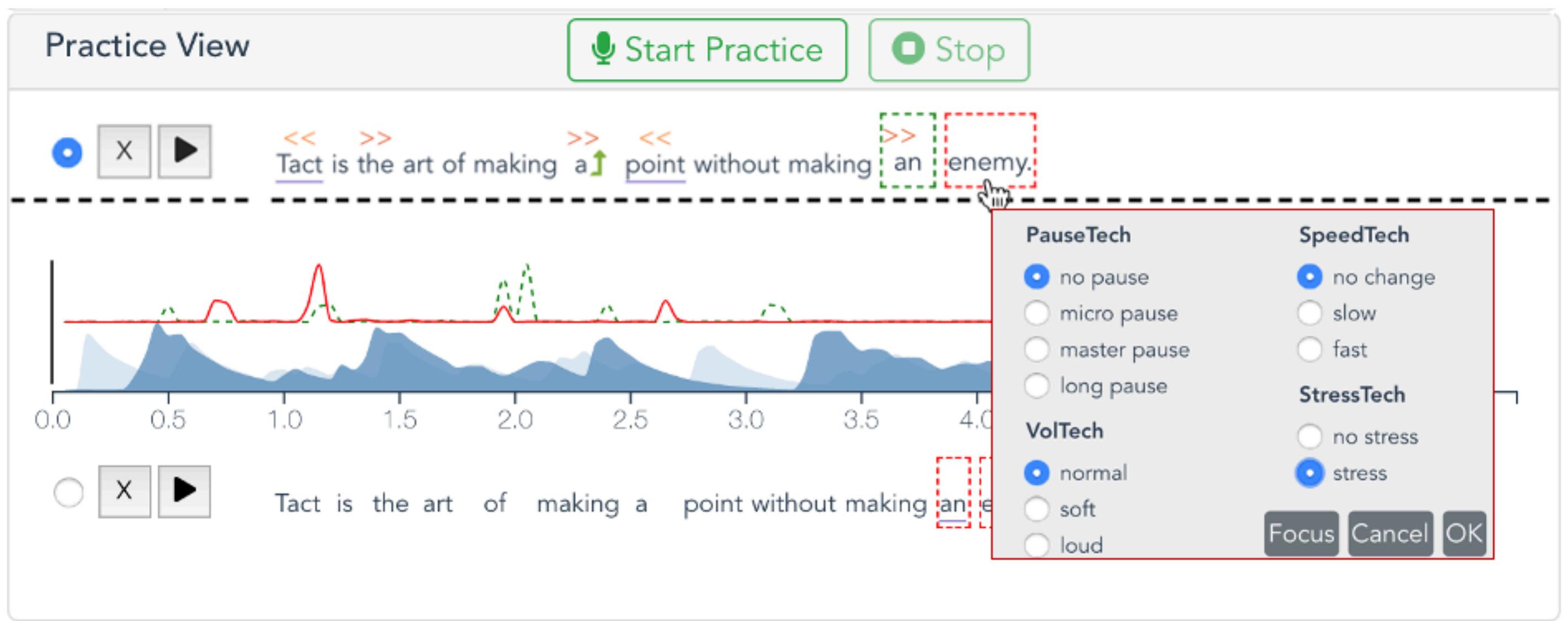


Original sentence and modulation

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

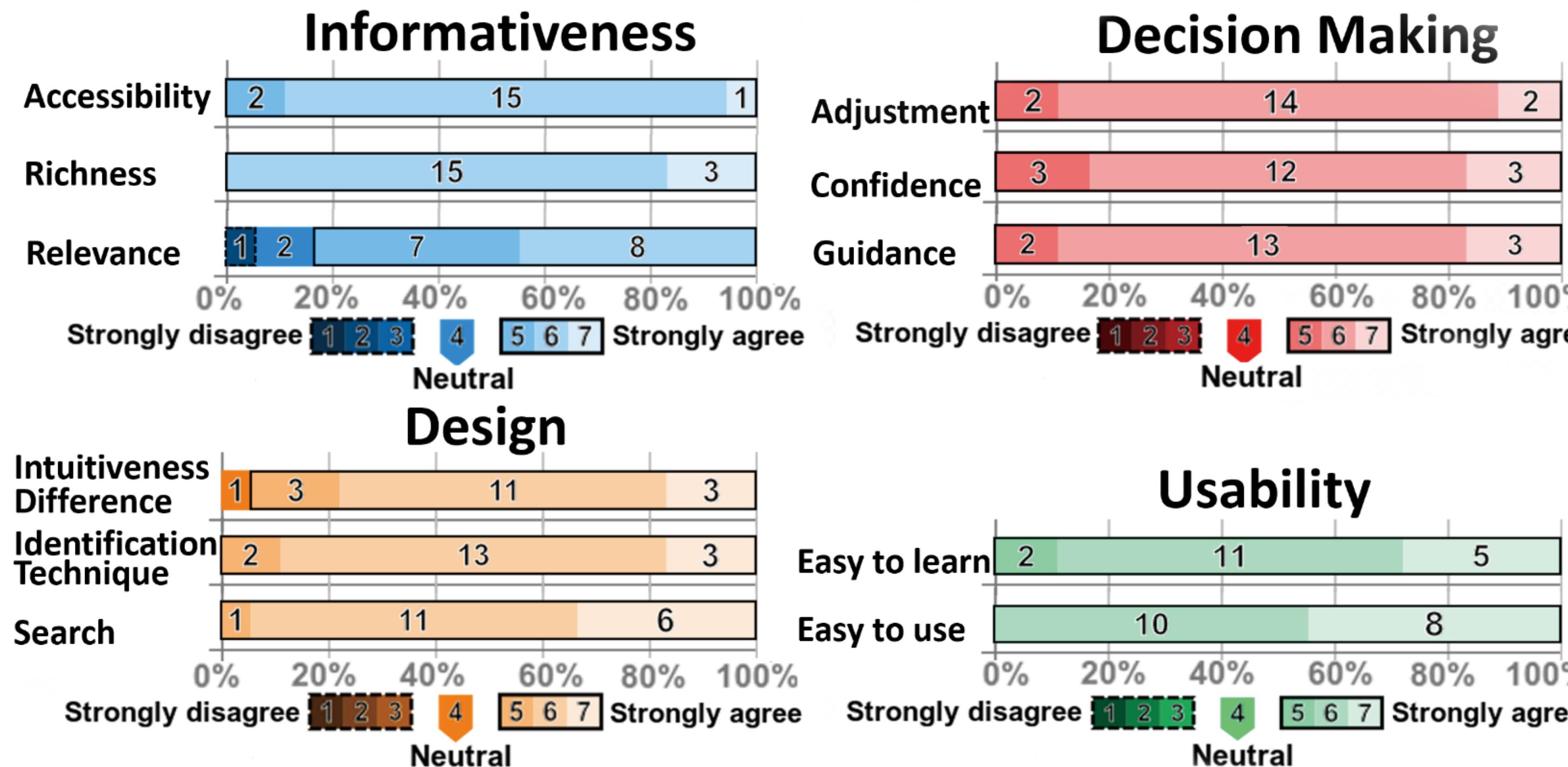
Speed: speed of curve

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

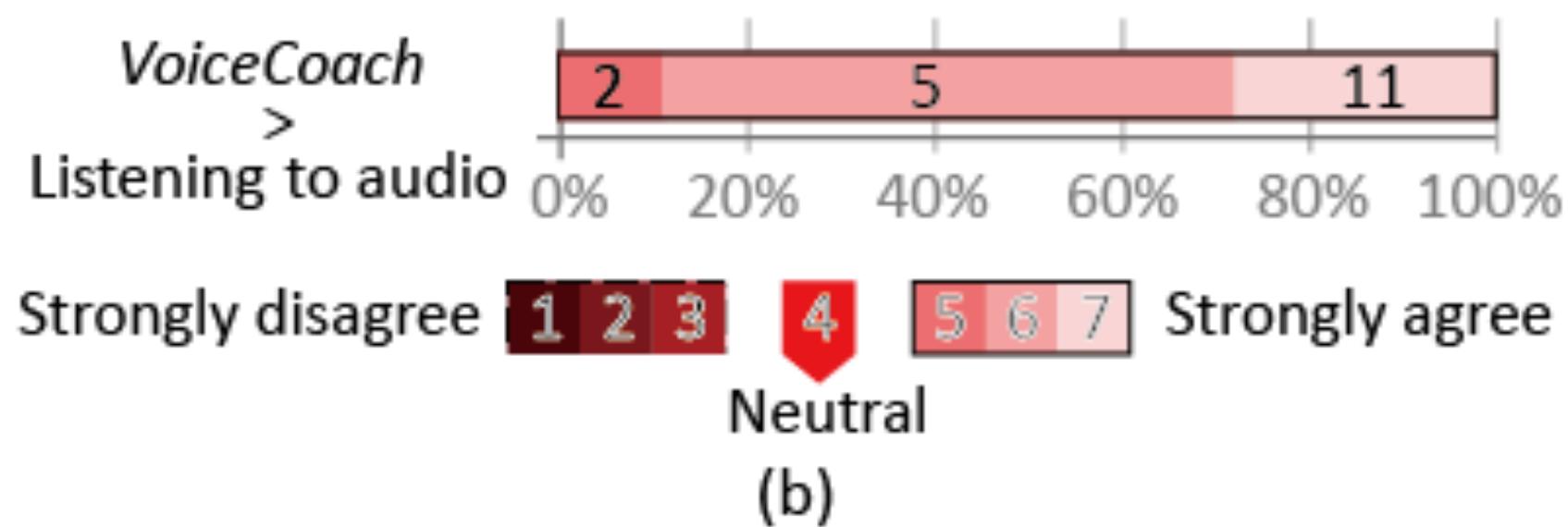
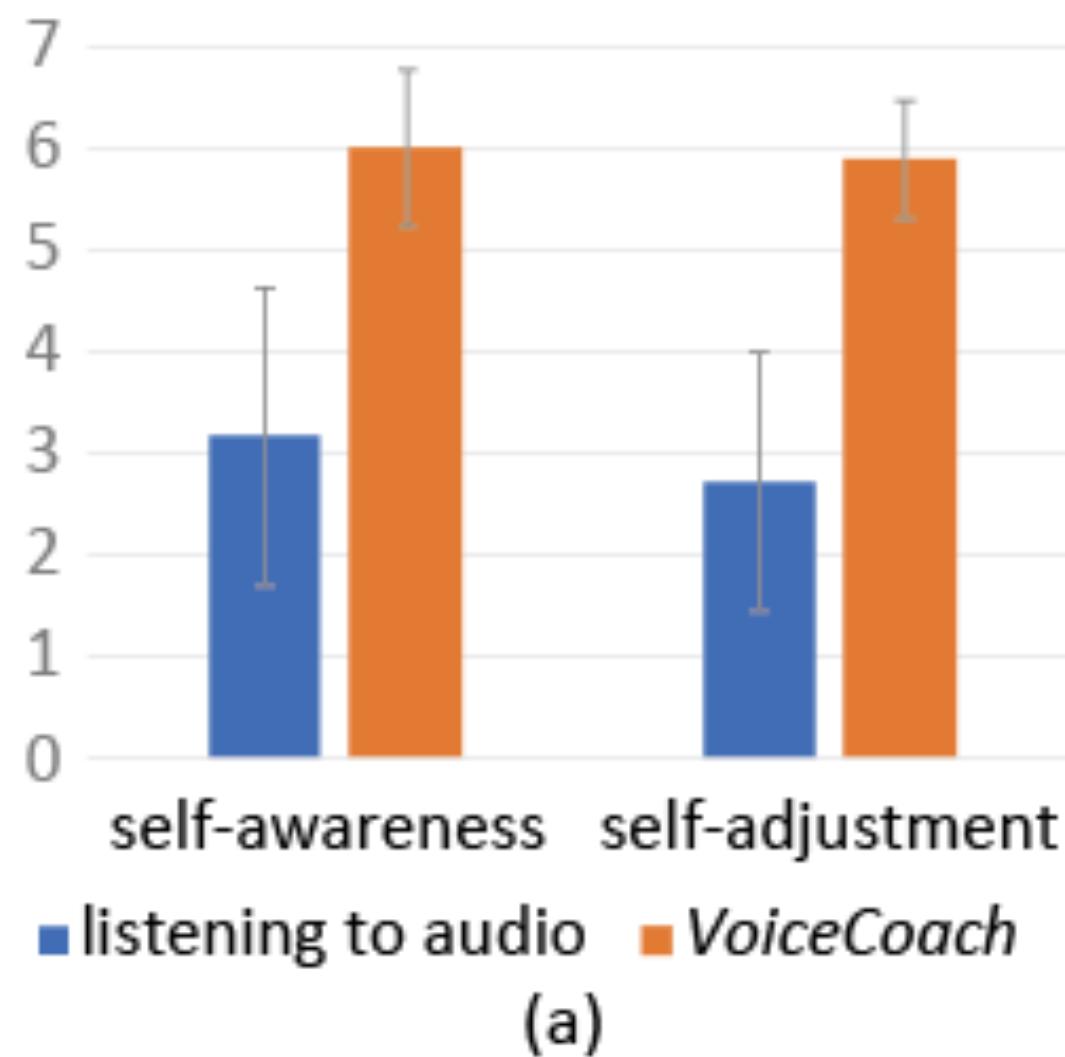


Fig. Results of questionnaire about user experience of practice:

VoiceCoach V.S. Listening to audio

Wilcoxon signed-rank test

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

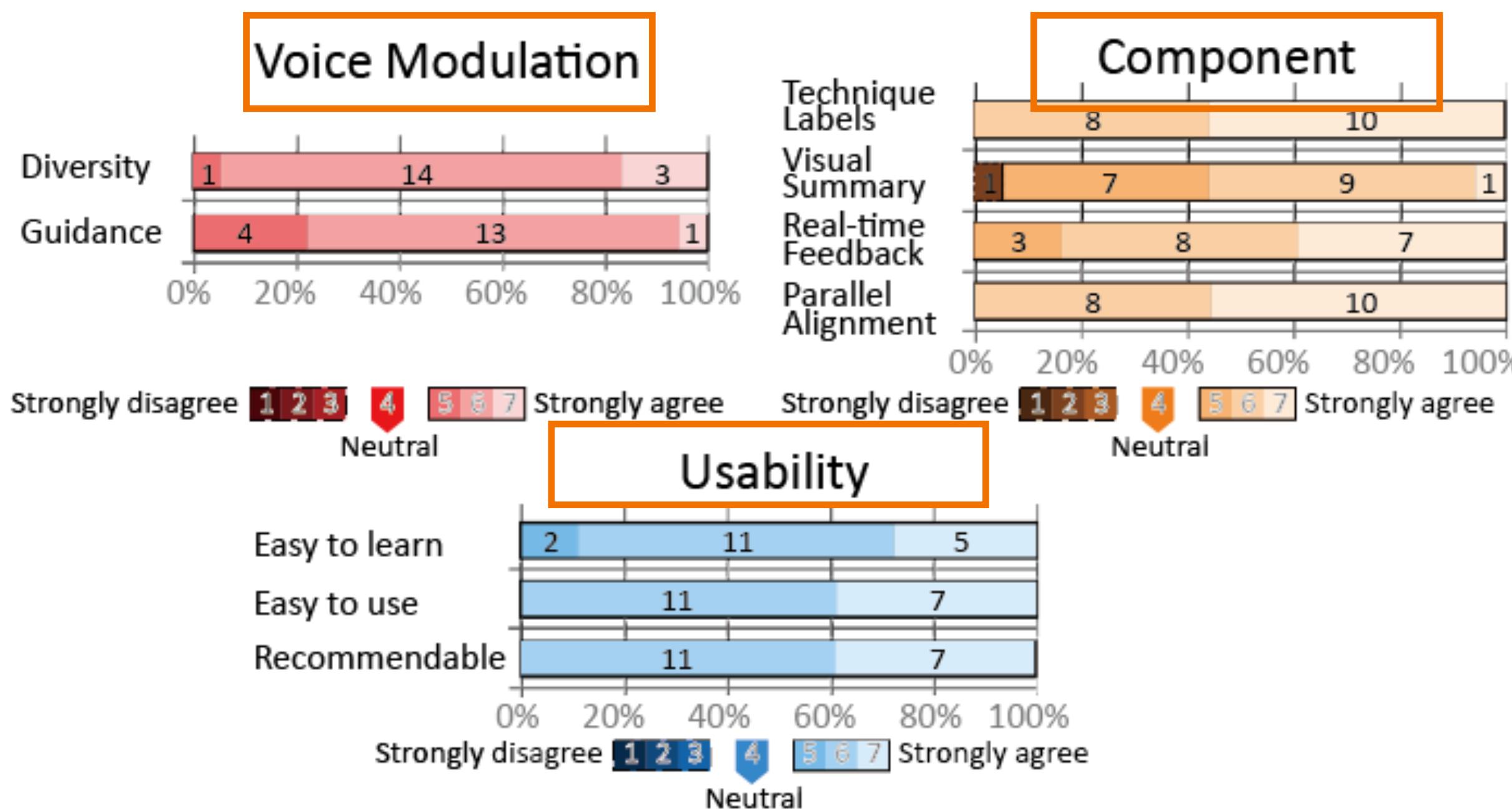
Overall feedback

- All participants prefer VoiceCoach

User Study

Overall Evaluation

1. Participants' feedback on VoiceCoach



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)

Fig. Results of questionnaire about system.

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

The screenshot displays the VoiceCoach application interface, which includes four main panels:

- User Panel:** Features a microphone icon for "Voice Input" and a blue play button for "Stop". Below are fields for "Choose File" (No file chosen), "Input Area", and a text box with the placeholder "If your actions inspire others to dream more," with a "Submit" button.
- Recommendation View:** Shows a grid of words with their parts of speech (e.g., NOUN, VERB) and associated voice techniques. A legend defines terms like "Tact", "Micropause", "Masterpause", "Longpause", "Loud", "Soft", "Stress", "fast", and "slow".
- Voice Technique View:** Displays a list of sentences with their contexts and corresponding voice techniques. Examples include sentence 203: "which tells you where to get out is **an exit**.", sentence 418: "positive change **to have** **an effect**.", sentence 478: "others **a challenge** to repair the world.", and sentence 47: "an appointed **opponent** outscored Jew, **to**".
- Practice View:** Allows users to practice sentences. It includes a "Start Practice" button, a playback slider, and a text input field containing the sentence "Tact is the art of making a point without making an enemy.". Below the text is a spectrogram and controls for "PauseTech", "SpeedTech", "StressTech", "VolTech", and "Focus".

Thank you!

Xingbo Wang

Ph.D. Student @VISLab, HKUST



Email: xwangege@cse.ust.hk

Home Page: <http://www.cse.ust.hk/~xwangege/>

