THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

PhD Qualifying Exam

A Survey on Intelligent User Interfaces for the Learning of Verbal Communication Skills

Presenter: Xingbo Wang
Supervisor: Prof. Huamin Qu

2020 June 9

Outline

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Introduction

Motivation

Challenges

Automatic Assessment

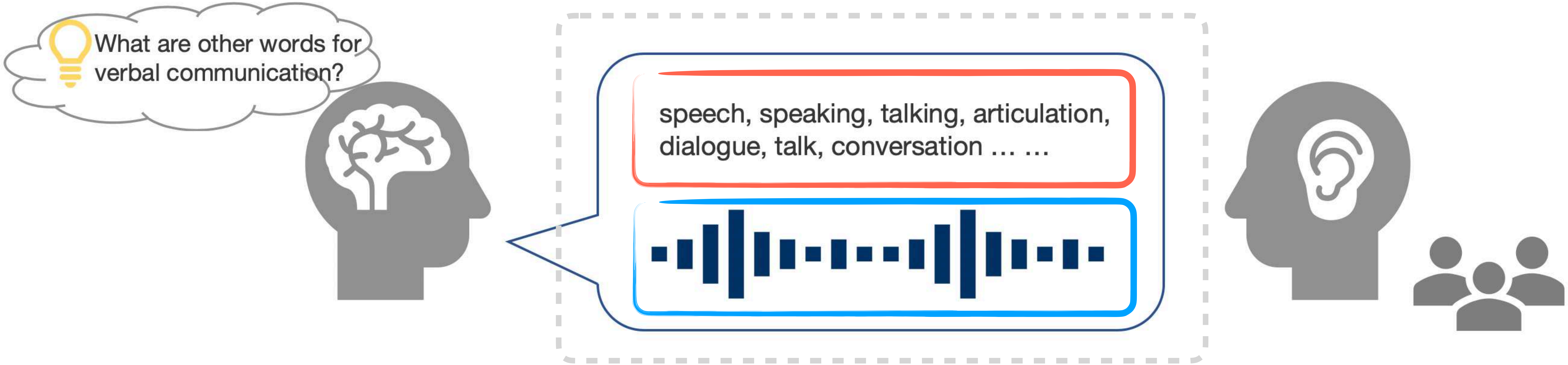
User Interfaces

Conclusion

Motivation

Background

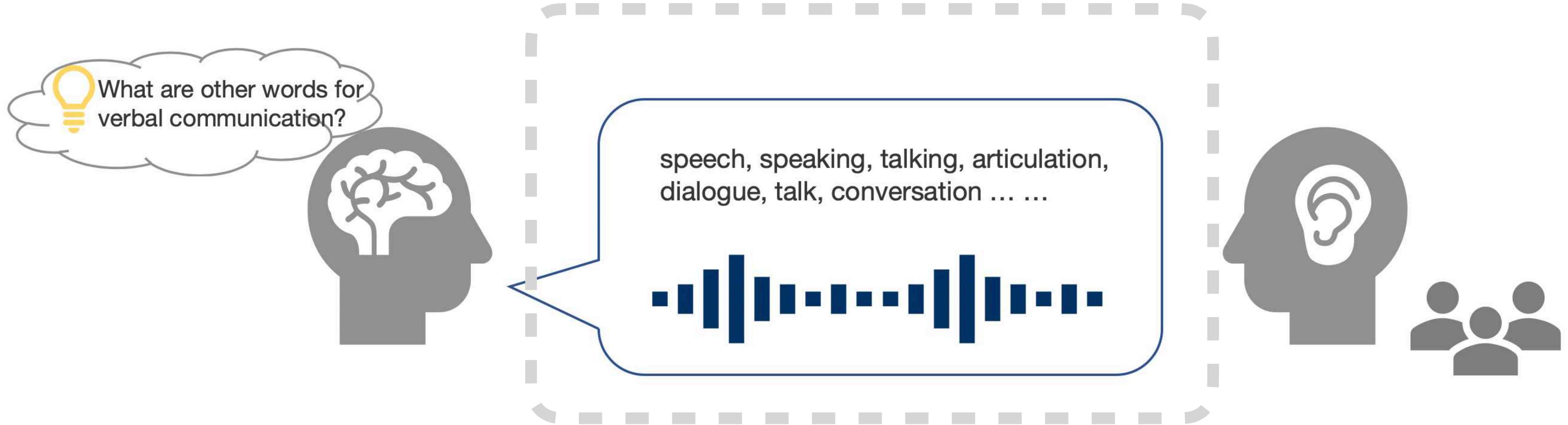
Verbal communication skills :Proper usage of words and sounds to deliver message



speech, speaking, talking, articulation, dialogue, talk, conversation … …

Motivation

Background



What are other words for verbal communication?

speech, speaking, talking, articulation, dialogue, talk, conversation … …

Verbal communication skills

Adaptive speech content

Engaging vocal delivery

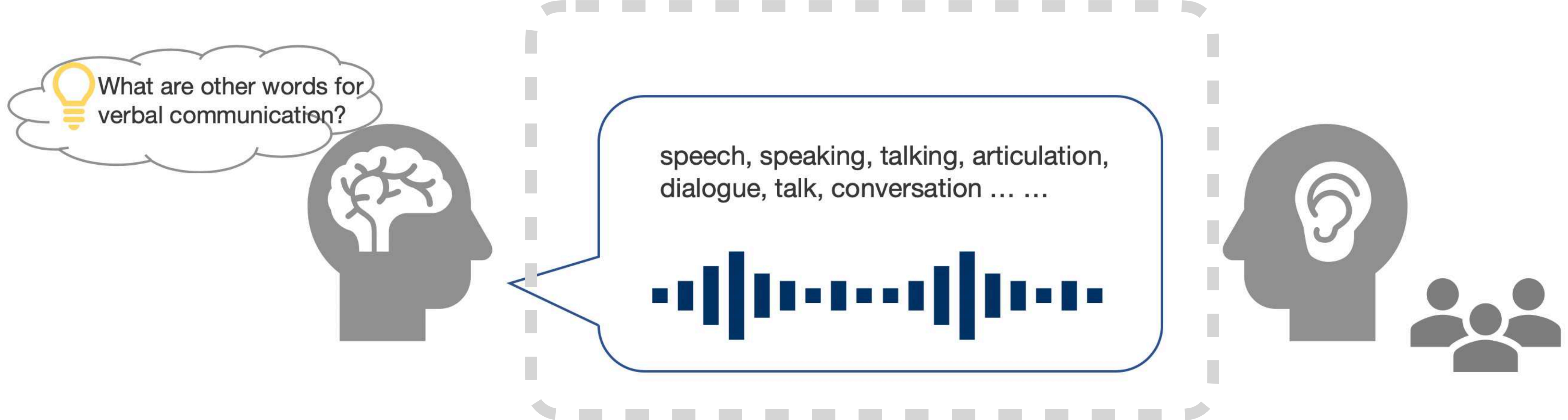
Public speaking

Job interview

Everyday conversation

Motivation

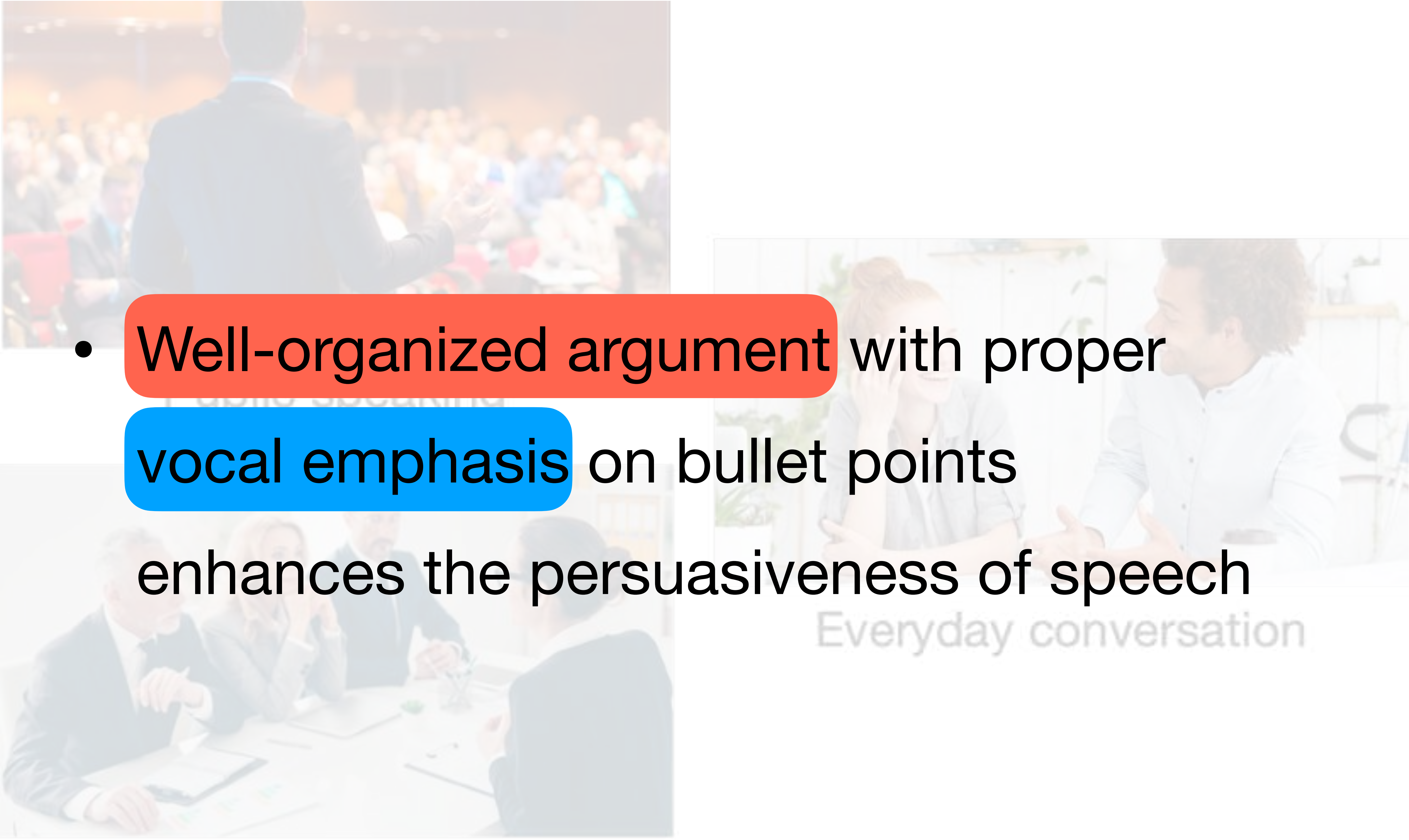
Background


What are other words for verbal communication?
speech, speaking, talking, articulation, dialogue, talk, conversation … …

Verbal communication skills

Adaptive speech content

Engaging vocal delivery



- Well-organized argument with proper vocal emphasis on bullet points enhances the persuasiveness of speech

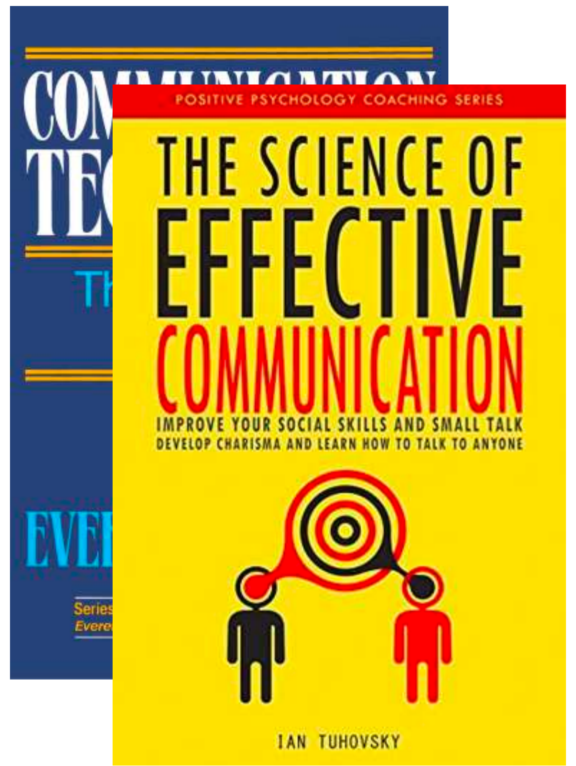
Job interview

Motivation

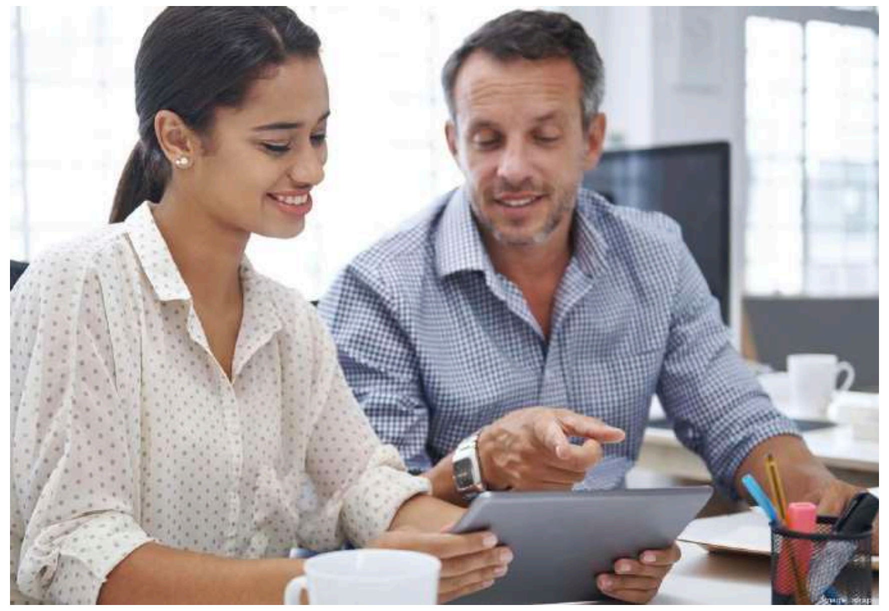
Verbal communication skills learning

- **Self-learning**: guidelines from books

  No feedback


POSITIVE PSYCHOLOGY COACHING SERIES
THE SCIENCE OF
EFFECTIVE
COMMUNICATION
IMPROVE YOUR SOCIAL SKILLS AND SMALL TALK
DEVELOP CHARISMA AND LEARN HOW TO TALK TO ANYONE
IAN TUHOVSKY

- **Professional training**: feedback from coaches
  - Qualitative
  - Inflexible

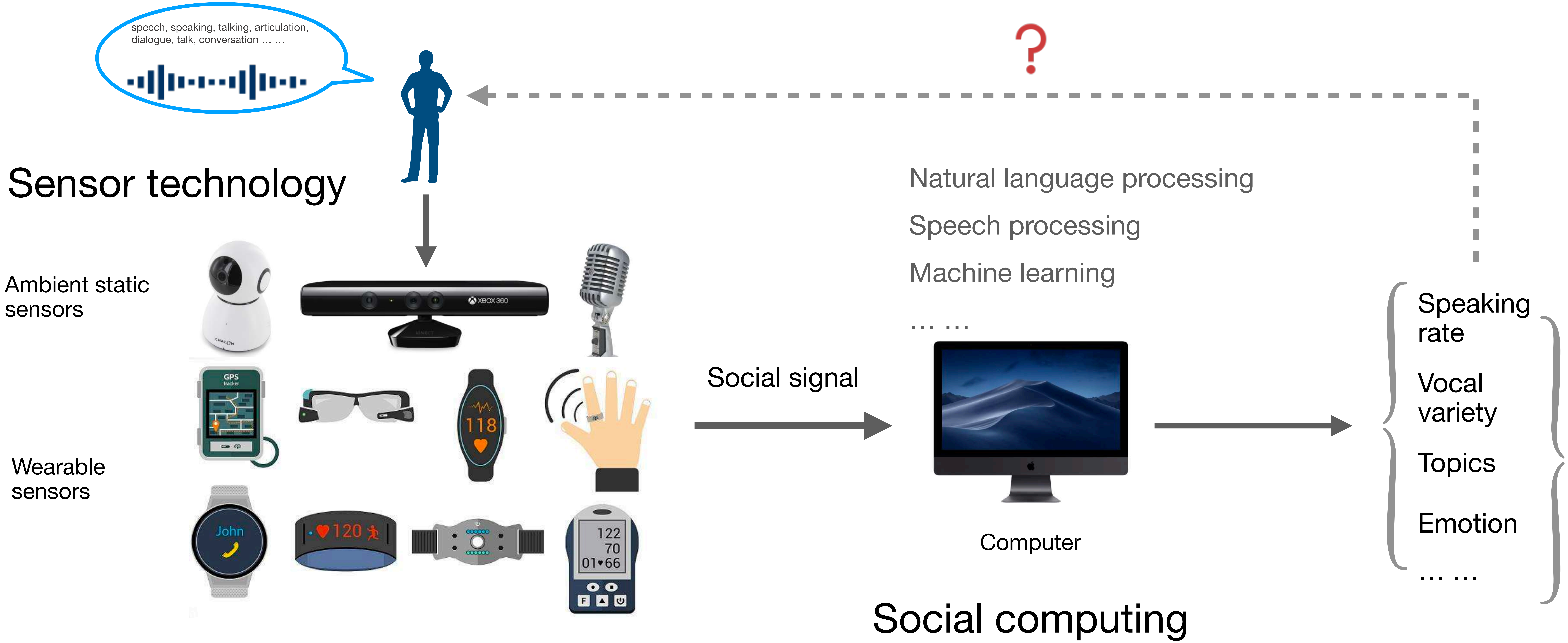


There is a lack of

- Quantitative & automated feedback
- Tool support for effective learning

Motivation

Quantitative automatic feedback



speech, speaking, talking, articulation, dialogue, talk, conversation … …

Sensor technology

Ambient static sensors

Wearable sensors

Social signal

Natural language processing

Speech processing

Machine learning

… …

Computer

Speaking rate

Vocal variety

Topics

Emotion

… …

Social computing

Motivation

User interfaces for learning

Knowledge learning platforms

Language learning softwares

Not for “soft skills”

- They have more clear criteria of what are “correct”
- Their feedback is mostly offline
- Their feedback is mostly in visual forms

Motivation

Challenges

Intelligence

Quantitative automatic feedback

- Derive quantitative descriptors about speech behavior
- Assess multimodal speech behavior

Learning

User interfaces for effective learning

- Offer valuable guidance on users’ behavior
  - When ? (e.g., realtime or post hoc feedback)
  - How ? (e.g., through visuals or other forms)
  - …

Outline

Introduction

Automatic Assessment

Competence rubrics

Computational features

Performance assessment

User Interfaces

Conclusion

Automatic Assessment

Competence rubrics

Researchers have conducted a number of studies to identify core communication competency and its rubrics for the practice of communication skills

(Quianthy, 1990; Lucas, 2007; Morreale et al., 2007; Rhodes, 2010; Thomson & Rucker, 2002)

	Core aspects of communication proficiency
Adaptive speech content	Topic selection, support material usage, idea organization, word choices
Engaging vocal delivery	Vocal variety, articulation, non-verbal behavior

Automatic Assessment

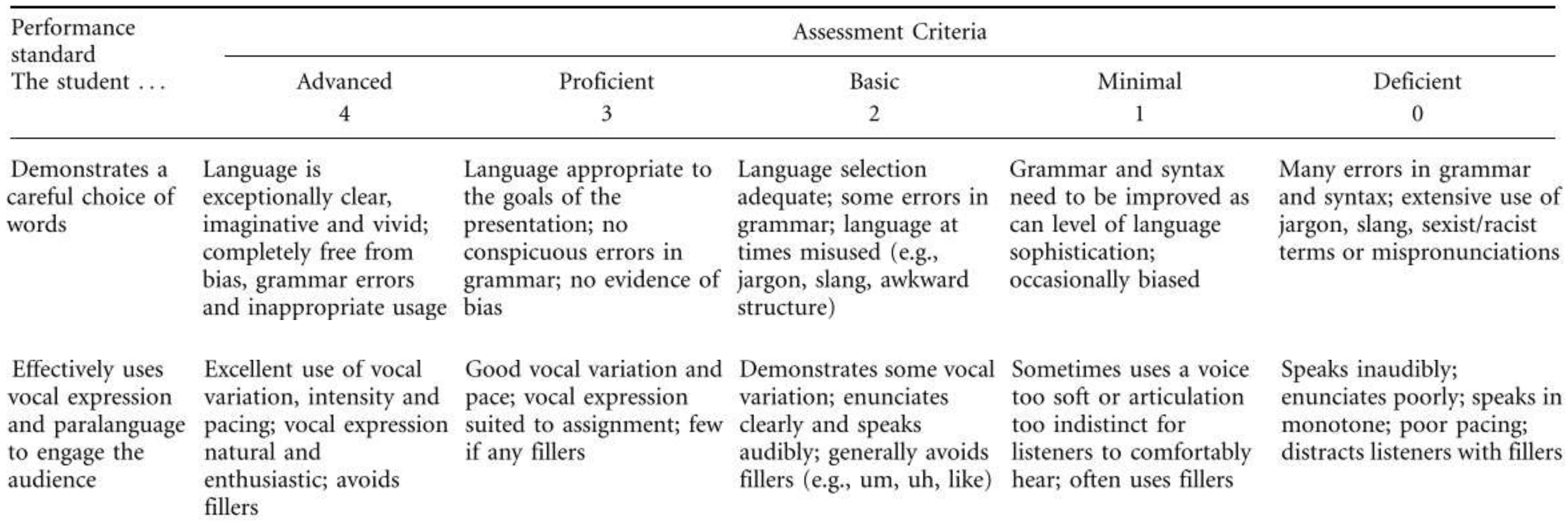
Competence rubrics

Researchers have conducted a number of studies to identify core communication competency and its rubrics for the practice of communication skills

(Quianthy, 1990; Lucas, 2007; Morreale et al., 2007; Rhodes, 2010; Thomson & Rucker, 2002)

Adaptive speech content

Engaging vocal delivery



Performance standard The student . . .	Assessment Criteria				
	Advanced 4	Proficient 3	Basic 2	Minimal 1	Deficient 0
Demonstrates a careful choice of words	Language is exceptionally clear, imaginative and vivid; completely free from bias, grammar errors and inappropriate usage	Language appropriate to the goals of the presentation; no conspicuous errors in grammar; no evidence of bias	Language selection adequate; some errors in grammar; language at times misused (e.g., jargon, slang, awkward structure)	Grammar and syntax need to be improved as can level of language sophistication; occasionally biased	Many errors in grammar and syntax; extensive use of jargon, slang, sexist/racist terms or mispronunciations
Effectively uses vocal expression and paralanguage to engage the audience	Excellent use of vocal variation, intensity and pacing; vocal expression natural and enthusiastic; avoids fillers	Good vocal variation and pace; vocal expression suited to assignment; few if any fillers	Demonstrates some vocal variation; enunciates clearly and speaks audibly; generally avoids fillers (e.g., um, uh, like)	Sometimes uses a voice too soft or articulation too indistinct for listeners to comfortably hear; often uses fillers	Speaks inaudibly; enunciates poorly; speaks in monotone; poor pacing; distracts listeners with fillers

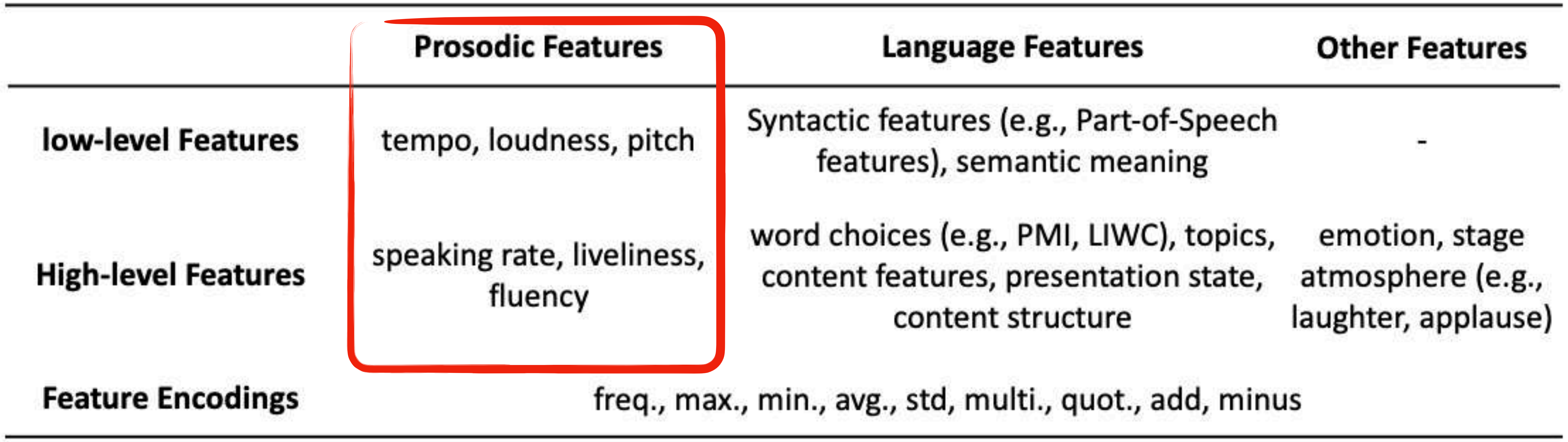
Public Speaking Competency Rubric (PSCR) (Schreiber et al., 2012)

Automatic Assessment

Computational features

Vocal delivery

Speech content



	Prosodic Features	Language Features	Other Features
low-level Features	tempo, loudness, pitch	Syntactic features (e.g., Part-of-Speech features), semantic meaning	-
High-level Features	speaking rate, liveliness, fluency	word choices (e.g., PMI, LIWC), topics, content features, presentation state, content structure	emotion, stage atmosphere (e.g., laughter, applause)
Feature Encodings	freq., max., min., avg., std, multi., quot., add, minus		

Computational features summarized from previous work

Tempo, loudness, pitch -> auditory perception of a speech

Articulation

- *Speaking rate: syllables/ words/sentences per minute*
- *fluency: smoothness of speech -> filled pauses, filler words (e.g., "em","hmm")*

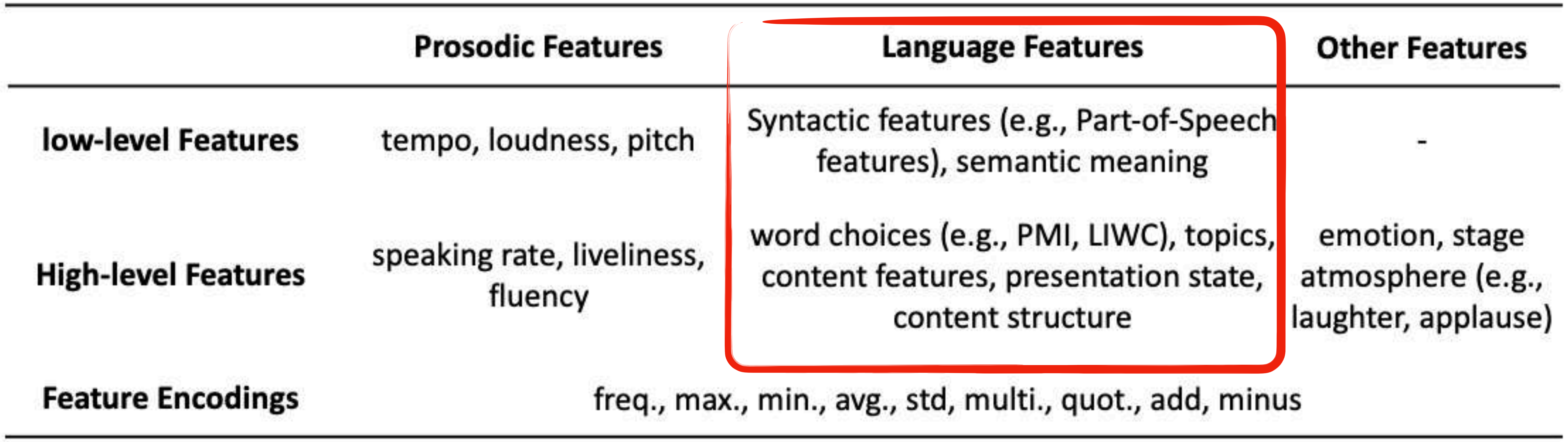
Vocal variety

- *liveliness: expressiveness of voice -> intonation -> variation of pitch and volume*

Automatic Assessment

Computational features

Vocal delivery   Speech content



	Prosodic Features	Language Features	Other Features
low-level Features	tempo, loudness, pitch	Syntactic features (e.g., Part-of-Speech features), semantic meaning	-
High-level Features	speaking rate, liveliness, fluency	word choices (e.g., PMI, LIWC), topics, content features, presentation state, content structure	emotion, stage atmosphere (e.g., laughter, applause)
Feature Encodings	freq., max., min., avg., std, multi., quot., add, minus		

Computational features summarized from previous work

Word usage

=> commonness: *PMI*

=> psychology: *LIWC, sentiment/subjectivity lexicons*

Topics => LDA

Content organization => BoW, Word2Vec

Adaptation

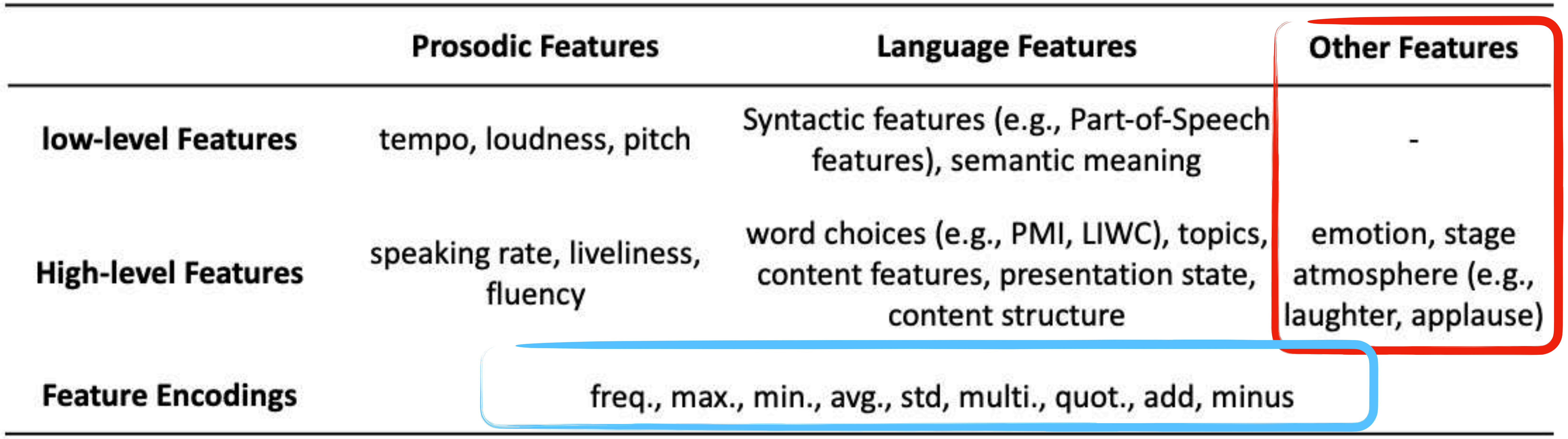
=> *Presentation state: presentation & QA*

Automatic Assessment

Computational features

Vocal delivery

Speech content



	Prosodic Features	Language Features	Other Features
low-level Features	tempo, loudness, pitch	Syntactic features (e.g., Part-of-Speech features), semantic meaning	-
High-level Features	speaking rate, liveliness, fluency	word choices (e.g., PMI, LIWC), topics, content features, presentation state, content structure	emotion, stage atmosphere (e.g., laughter, applause)
Feature Encodings	freq., max., min., avg., std, multi., quot., add, minus		

Computational features summarized from previous work

Speaker engagement

=> emotion

Audience engagement

=> stage atmosphere (e.g., laughter, applause, booing)

Automatic Assessment

Performance assessment

Rule-based methods

Based on the statistical properties of features to set the range of “good”/“bad” performance

(e.g., std, freq., mean)

Vocal delivery

Pitch variety, speech speed => mean+std of words/sentences

Speech content

Content coverage => spotted keywords in speech and their weights (tf.idf) / text in slides

It is simple and useful for basic features. However,

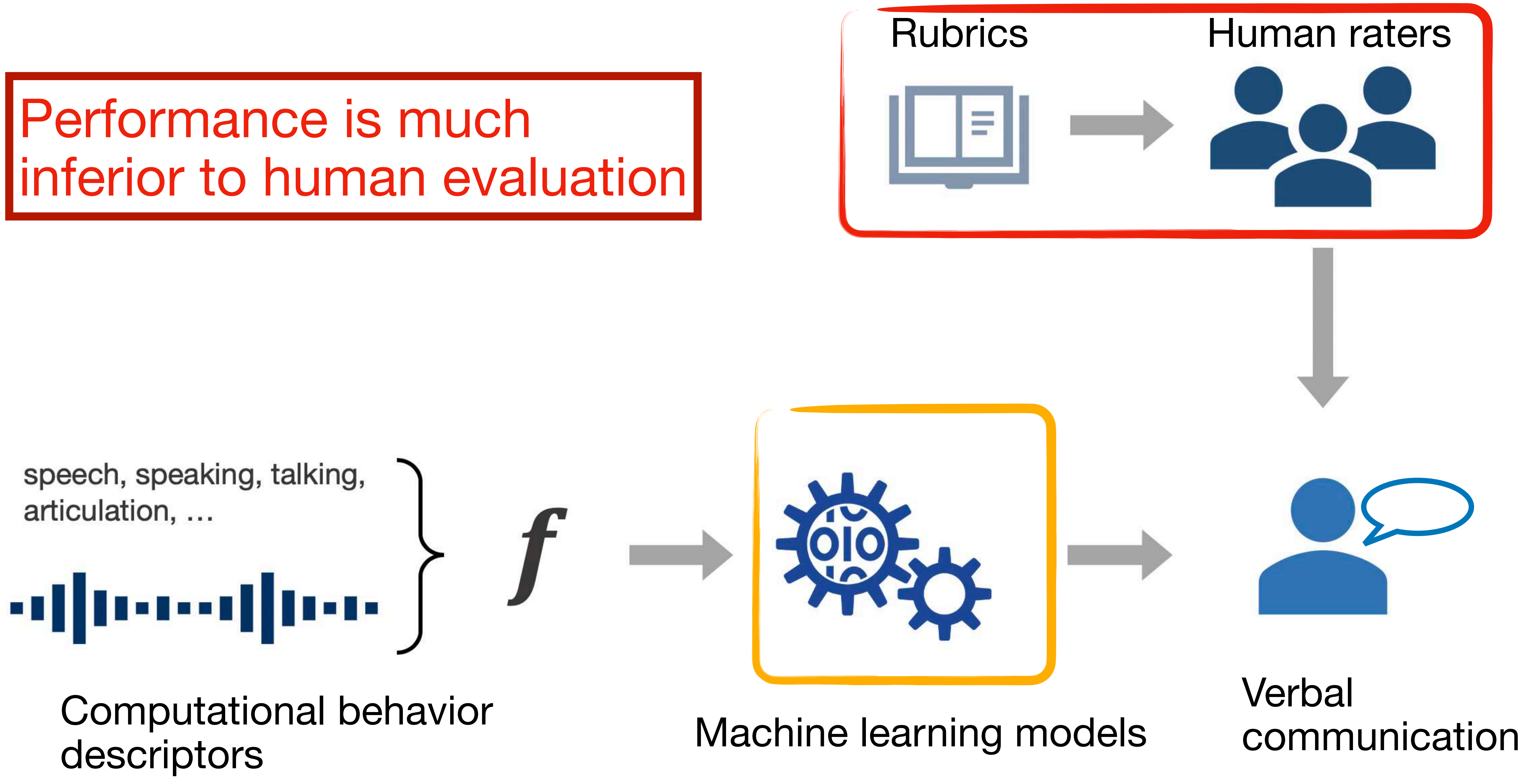
- it is intricate to decide thresholds for complex features (e.g., emotion)
- it fails to adapt to different speakers and different speaking scenarios

Automatic Assessment

Performance assessment

Machine learning (supervised learning)

Performance is much inferior to human evaluation


Rubrics
Human raters
speech, speaking, talking, articulation, …
f
Computational behavior descriptors
Machine learning models
Verbal communication

□ Human evaluation examples

Quality control: inter-agreement

□ Supervised learning models

Q1: Models <=?=> judgements

- SVM/SVR, L1/L2 Regularized Logistic Regression, Lasso, tree-based models (e.g., RF)
- BN, HMM

Q2: Features <=?=> judgements

- Correlation Coefficients

Automatic Assessment

Performance assessment

Deep learning

CNN (Krizhevsky et al., 2012), LSTM (Hochreiter and Schmidhuber, 1997), Transformer (Vaswani et al., 2017) and their variations achieve impressing results on complex analytical tasks of human communication understanding

- CNN (Hershey et al., 2017) => Audio event detection and classification
- BERT (Devlin et al., 2018) => Various NLP tasks
- MFN (Zadeh et al. 2018) => Multimodal feature fusion

It is difficult for human to understand and interpret the model results

Outline

Introduction

Automatic Assessment

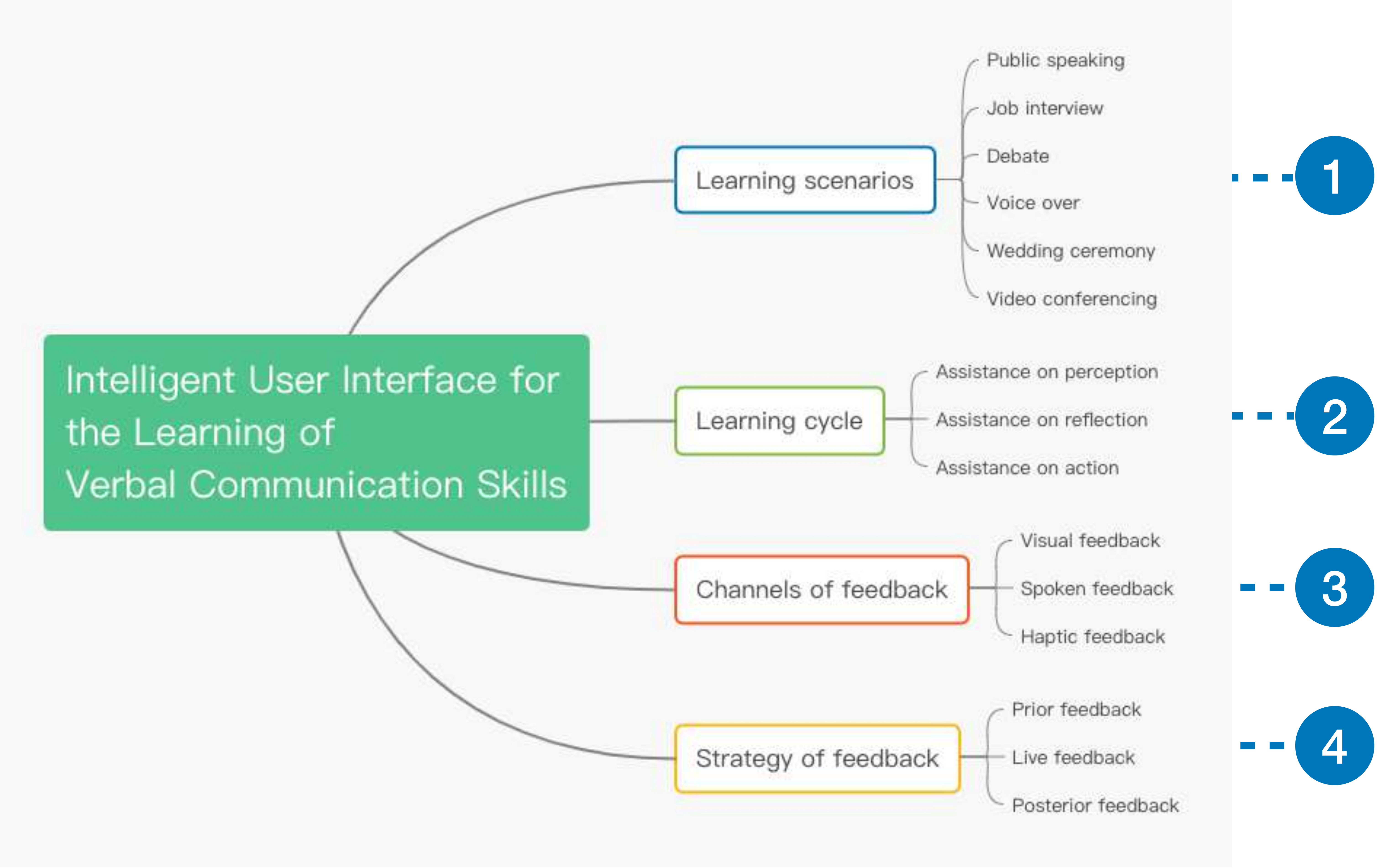
Intelligent User Interface

- Taxonomy
- Prior feedback
- Live feedback
- Posterior feedback

Conclusion

Intelligent User Interface

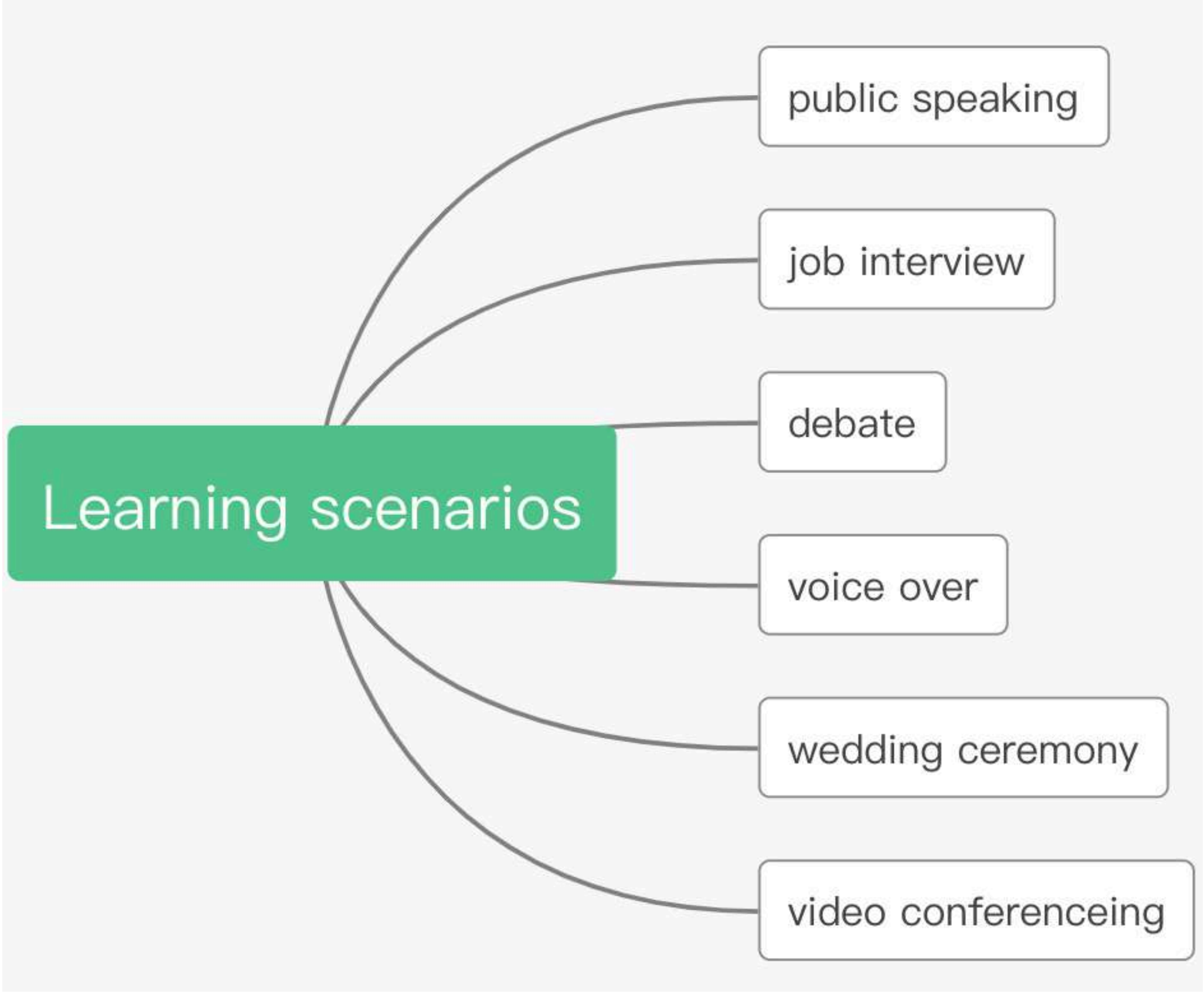
Taxonomy



Intelligent User Interface

Taxonomy

1 Considering learning scenarios,


Learning scenarios

public speaking

job interview

debate

voice over

wedding ceremony

video conferenceing

Speaking anxiety

Content organization

Vocal delivery

Stage management

Group interactions & dynamics

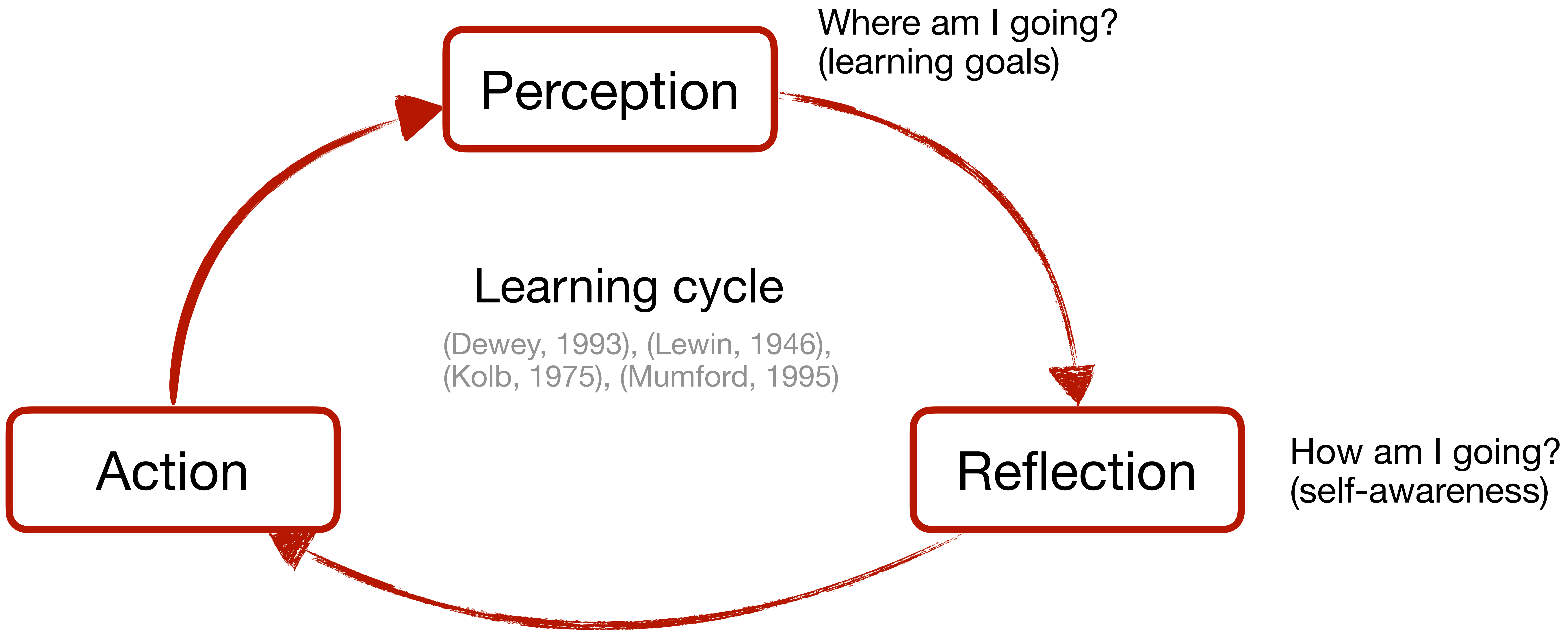
Design focus

Intelligent User Interface

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Taxonomy

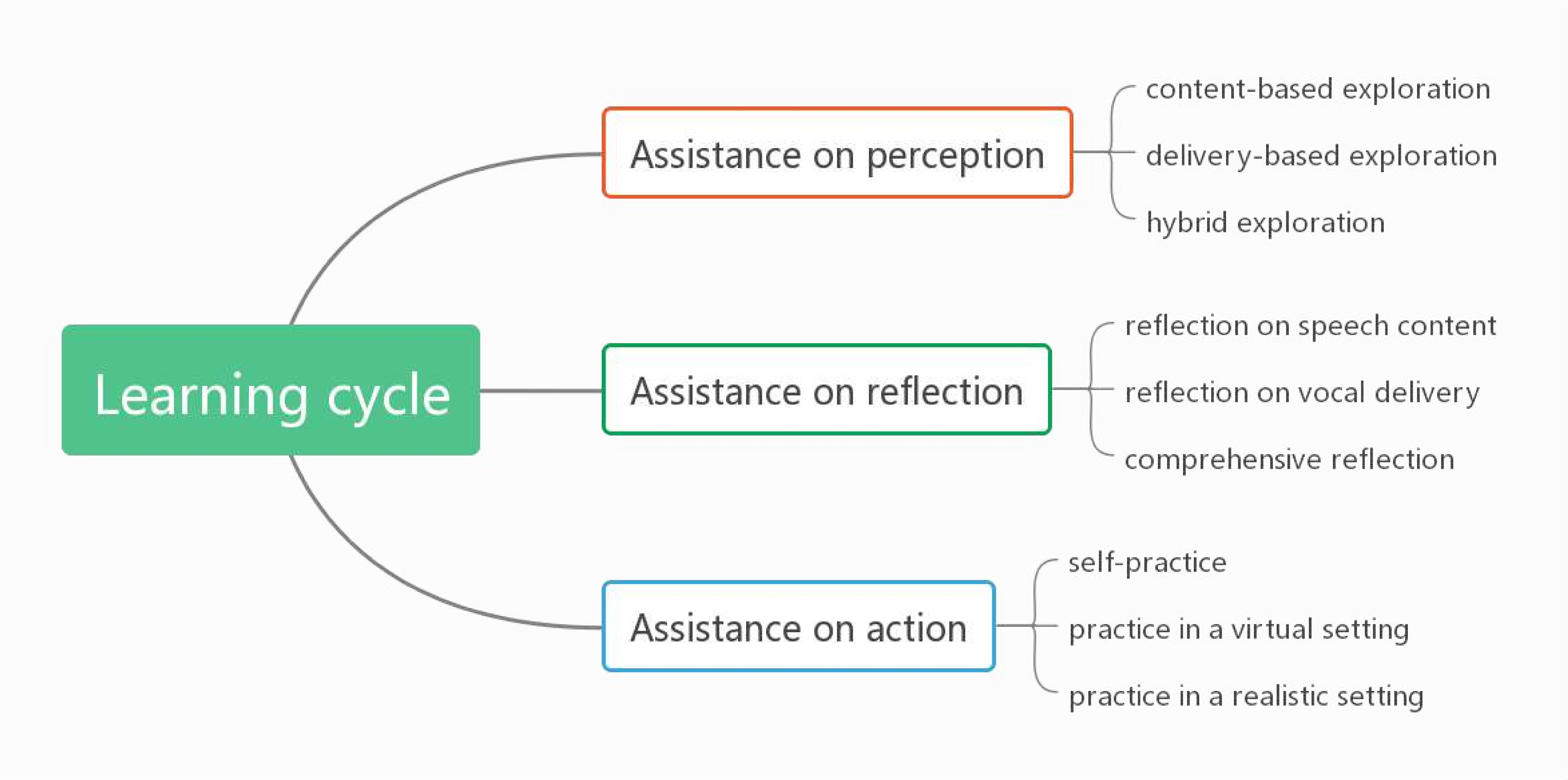
2 Considering the learning process,



Intelligent User Interface

Taxonomy

2 Considering the learning process,



Learning cycle
Assistance on perception
content-based exploration
delivery-based exploration
hybrid exploration
Assistance on reflection
reflection on speech content
reflection on vocal delivery
comprehensive reflection
Assistance on action
self-practice
practice in a virtual setting
practice in a realistic setting

Exploration of knowledge base

Reflection of vocal/verbal behavior

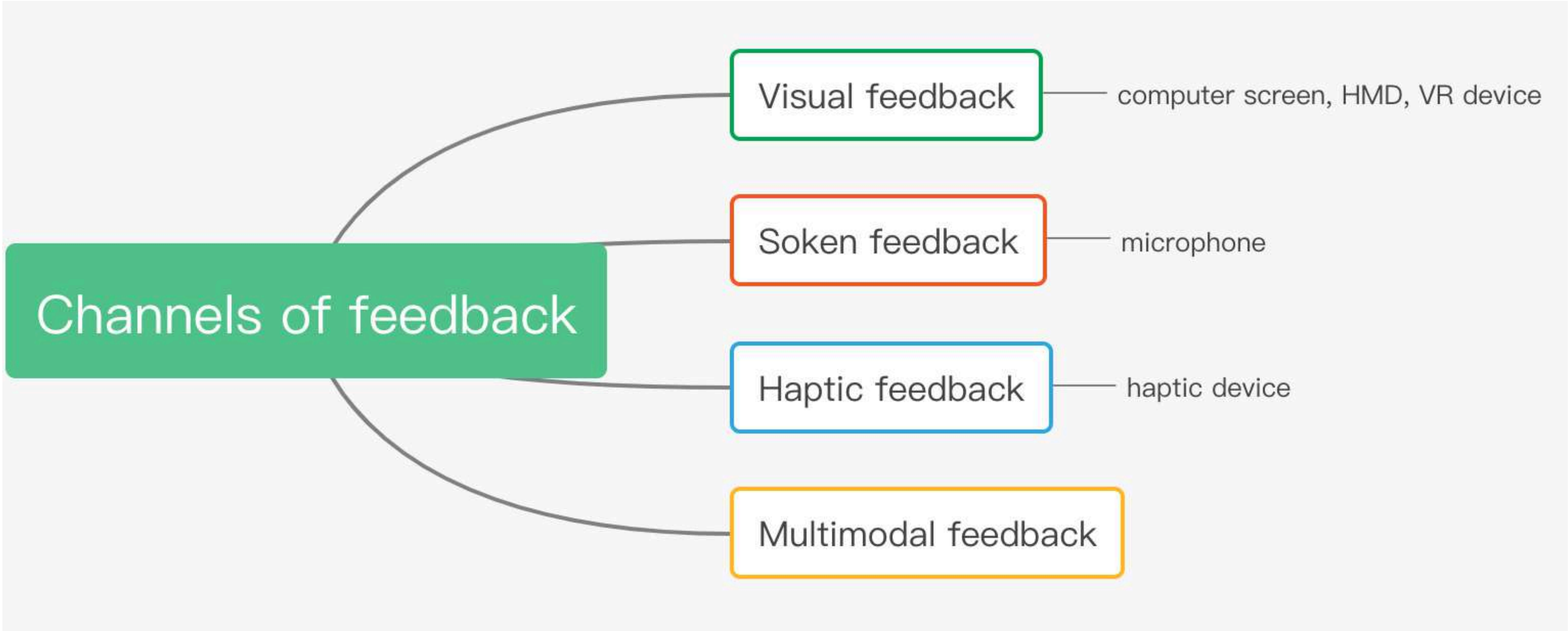
Putting knowledge into action

Intelligent User Interface

Taxonomy

Feedback has been considered as an effective intervention in skills learning and a key consideration of learning interfaces

3 Considering the channels of feedback,



Channels of feedback

Visual feedback — computer screen, HMD, VR device

Soken feedback — microphone

Haptic feedback — haptic device

Multimodal feedback

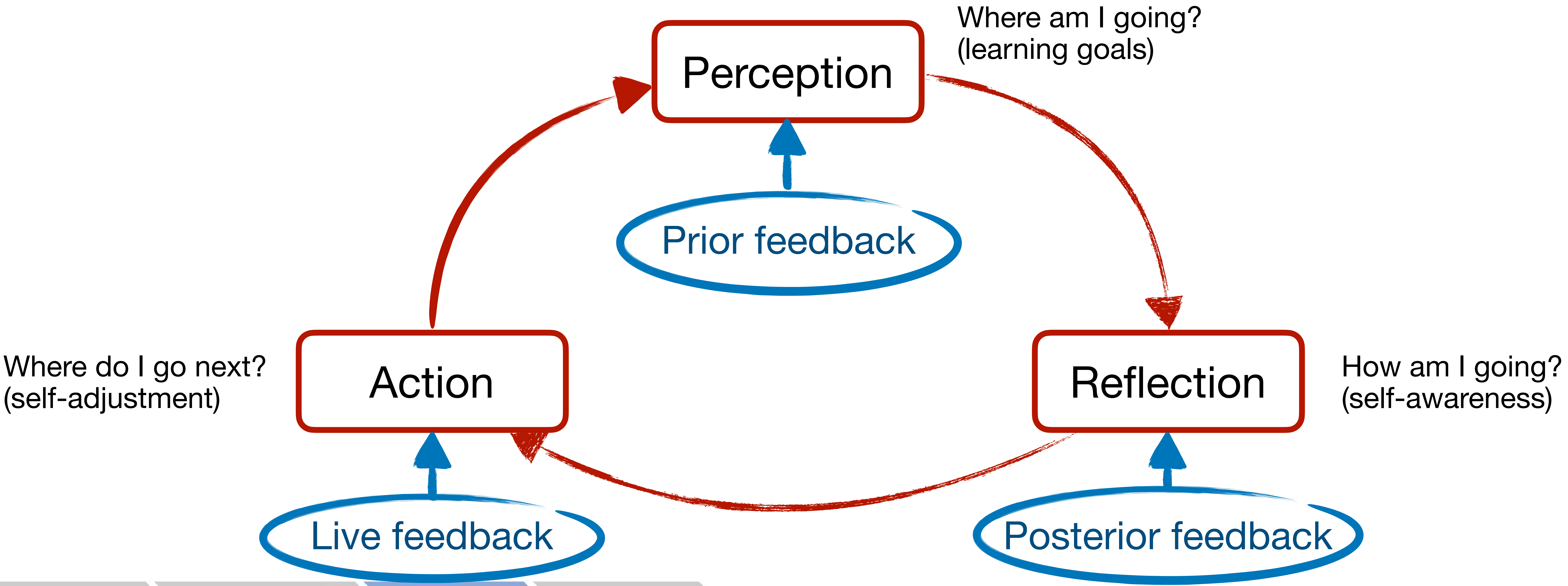
Most widely used for feedback on speech content and vocal delivery

Reduce cognitive load

Intelligent User Interface

Taxonomy

4 Considering the feedback strategy (in the learning cycle),



Where am I going?
(learning goals)

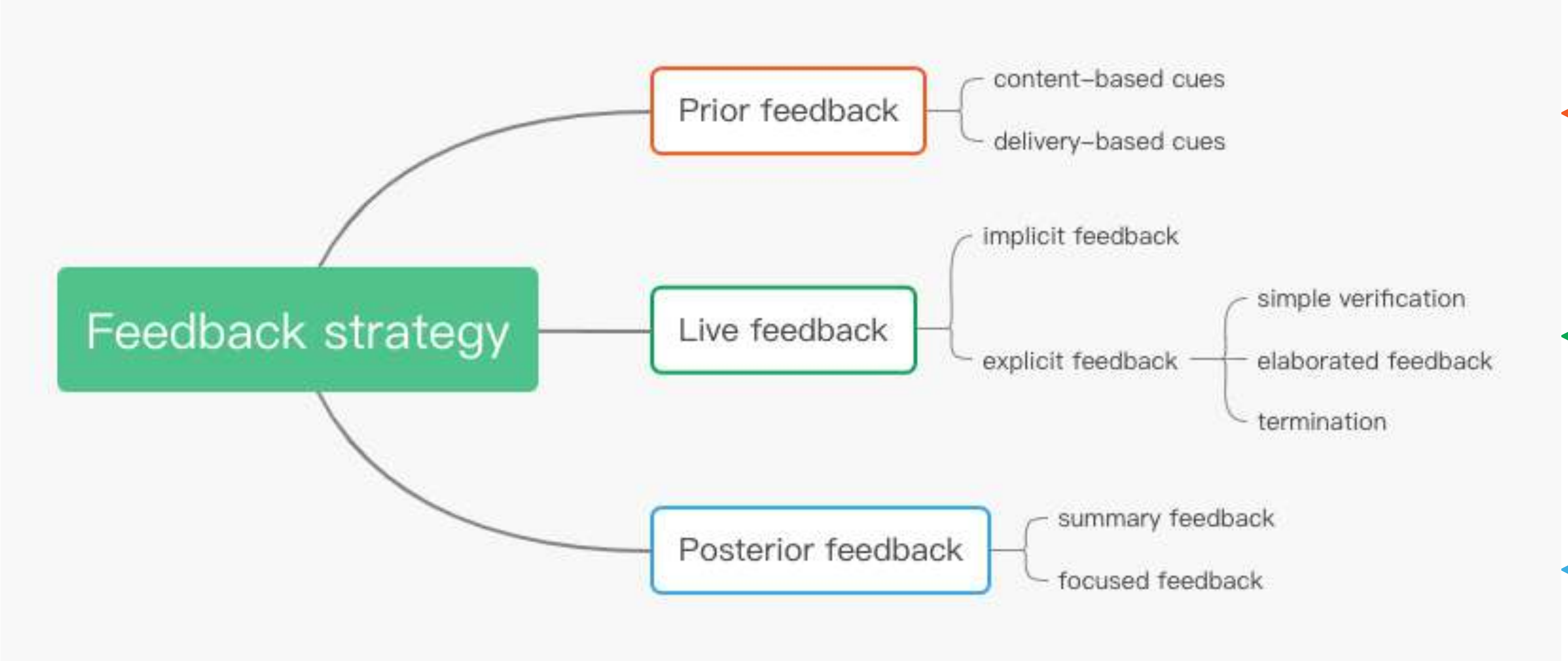
Where do I go next?
(self-adjustment)

How am I going?
(self-awareness)

Intelligent User Interface

Taxonomy

4 Considering the feedback strategy (in the learning cycle),



Evidences & consequences

Excellent learning examples

↓

Good speech delivery needs varied intonation

Feedback complexity

Avoid information overload

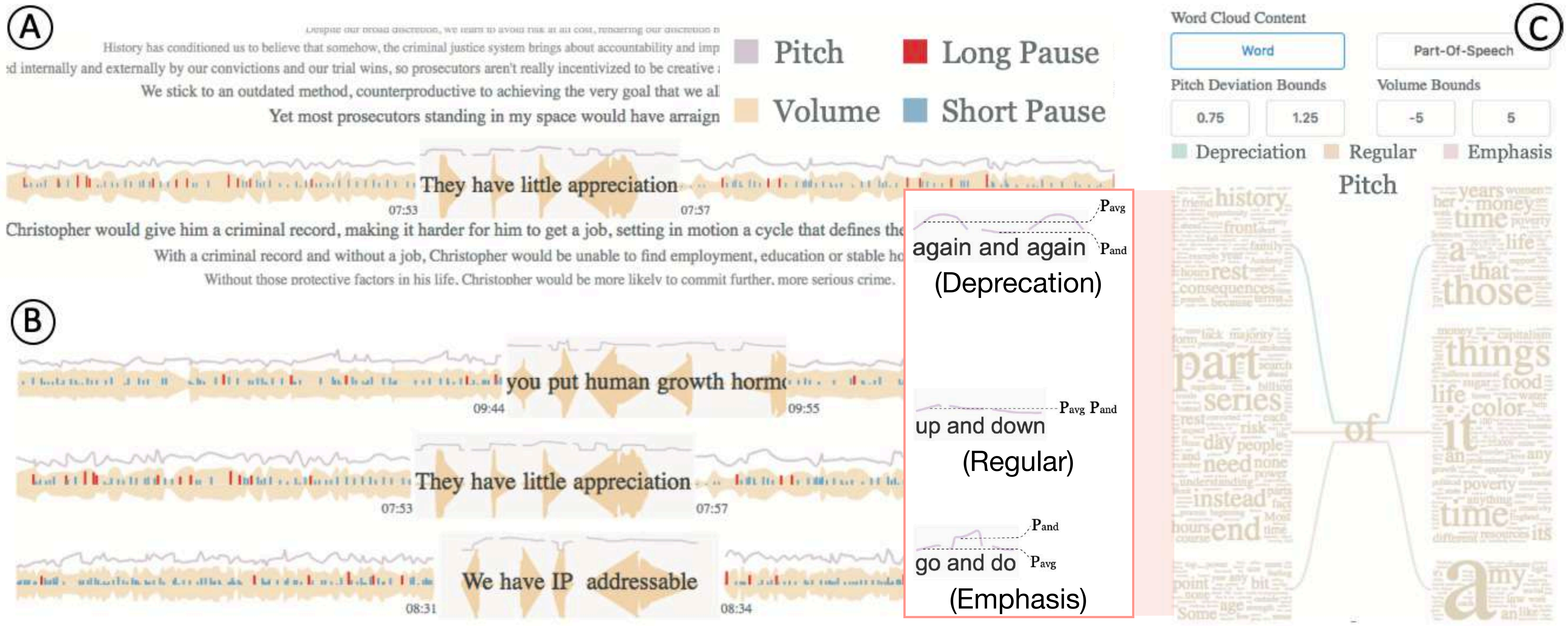
Focus of self-reflection

Exploration from coarse to fine

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Exploring narration strategies (pitch, pause, volume)



Despite our broad discretion, we learn to avoid risk at all cost, rendering our discretion n

History has conditioned us to believe that somehow, the criminal justice system brings about accountability and imp

d internally and externally by our convictions and our trial wins, so prosecutors aren't really incentivized to be creative

We stick to an outdated method, counterproductive to achieving the very goal that we al

Yet most prosecutors standing in my space would have arraign

Christopher would give him a criminal record, making it harder for him to get a job, setting in motion a cycle that defines the

With a criminal record and without a job, Christopher would be unable to find employment, education or stable ho

Without those protective factors in his life, Christopher would be more likely to commit further, more serious crime.

Sentence-level

A: context+focus design for visualization of prosodic features

B: Structural query

Word-level

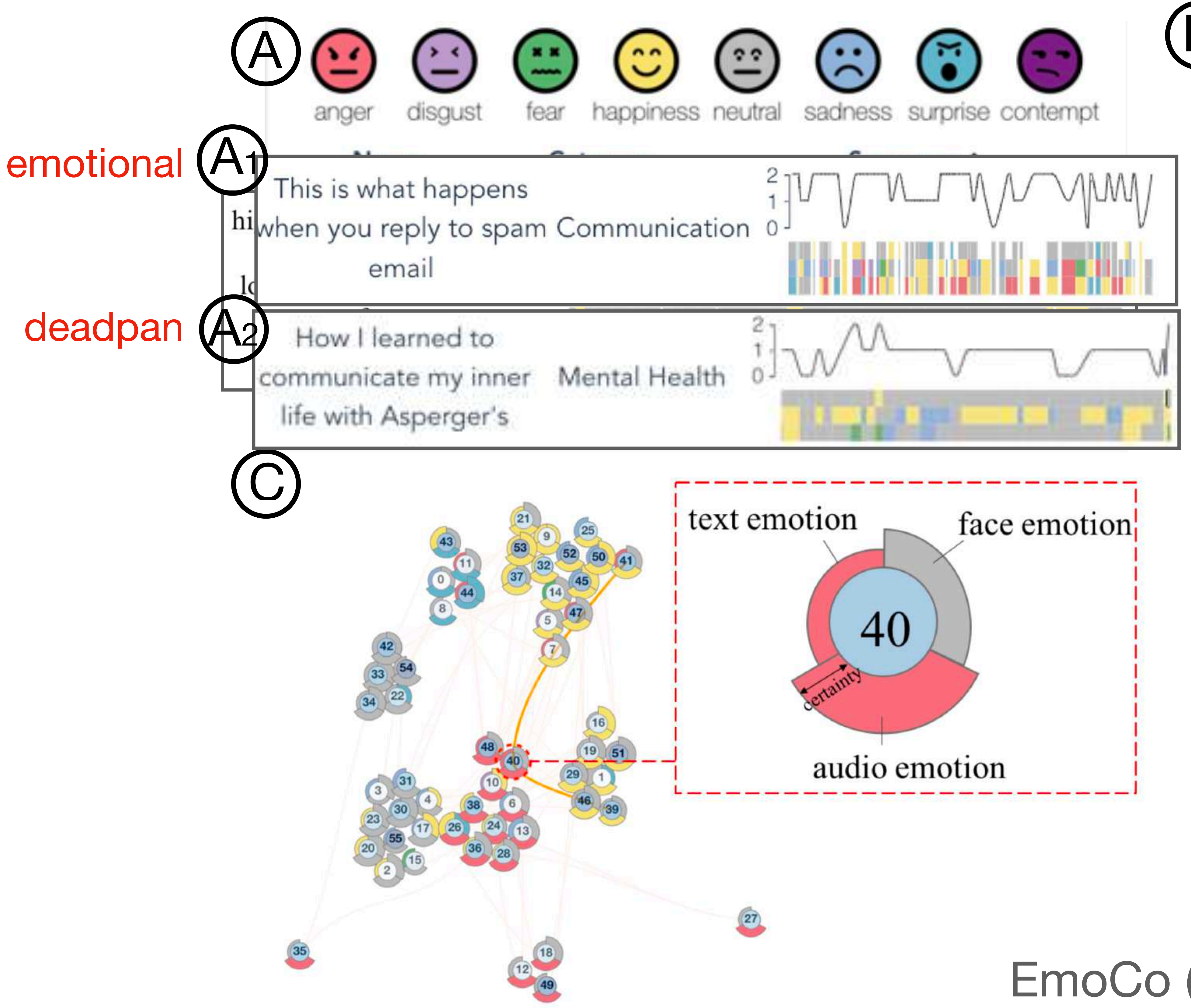
C: Word clouds for phrase intonation

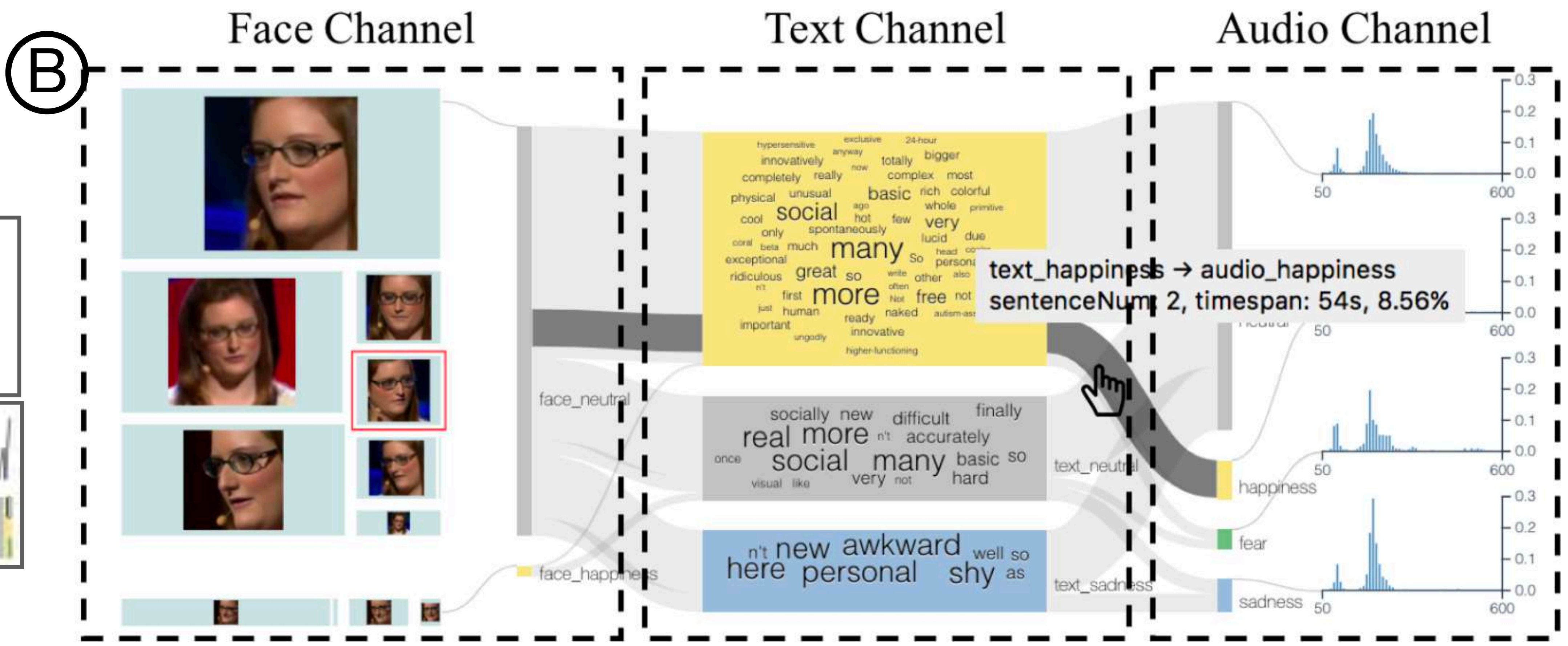
SpeechLens (Yuan et al., 2019)

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Exploring emotion coherence in presentation





A: video barcode charts & coherence curves → overview

B: an augmented Sankey diagram → channel connection

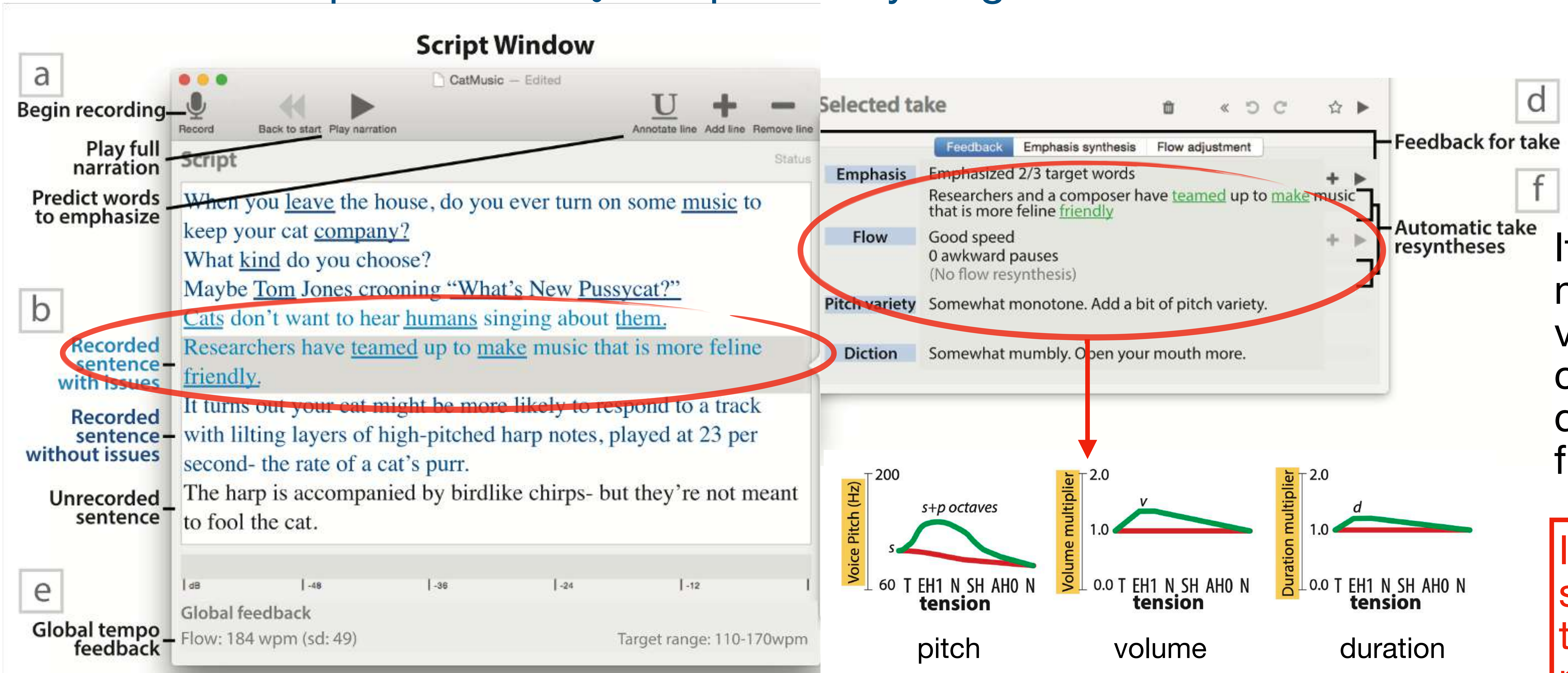
C: sentence clustering view → emotion combinations & its distribution

EmoCo (Zeng et al., 2019)

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

From feature exploration → Speech styles generation for voice-over



When you leave the house, do you ever turn on some music to keep your cat company?
What kind do you choose?
Maybe Tom Jones crooning "What's New Pussycat?"
Cats don't want to hear humans singing about them.
Researchers have teamed up to make music that is more feline friendly.
It turns out your cat might be more likely to respond to a track with lilting layers of high-pitched harp notes, played at 23 per second- the rate of a cat's purr.
The harp is accompanied by birdlike chirps- but they're not meant to fool the cat.

Global feedback
Flow: 184 wpm (sd: 49) Target range: 110-170wpm

Emphasis: Emphasized 2/3 target words
Researchers and a composer have teamed up to make music that is more feline friendly

Flow: Good speed
0 awkward pauses
(No flow resynthesis)

Pitch variety: Somewhat monotone. Add a bit of pitch variety.

Diction: Somewhat mumbly. Open your mouth more.

NarrationCoach (Rubin et al., 2015)

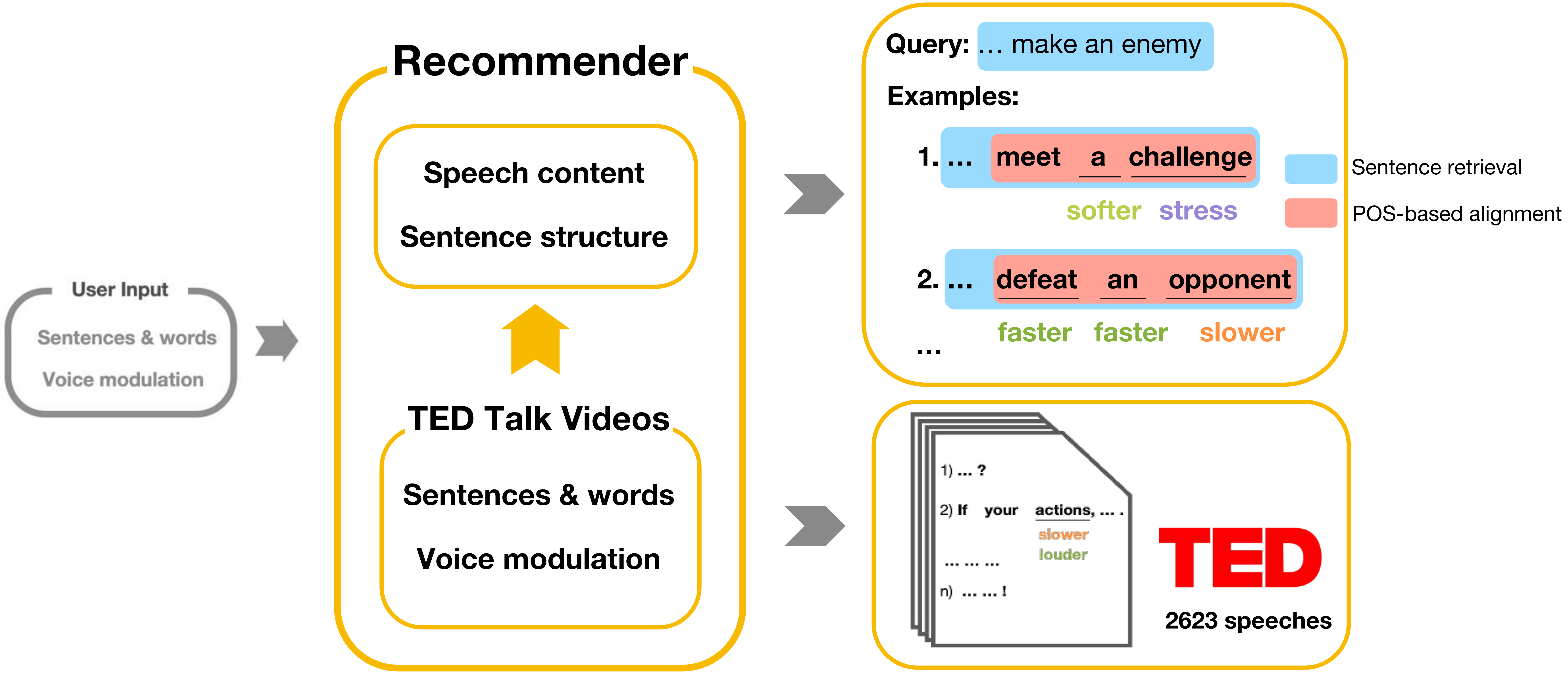
It can automatically modify the pitch, volume and duration curves to generate desired emphasis and flow.

It is difficult for novice speakers to identify the words for resynthesis

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Data-driven recommendation of voice modulation techniques



User Input

Sentences & words

Voice modulation

Recommender

Speech content

Sentence structure

TED Talk Videos

Sentences & words

Voice modulation

Query: … make an enemy

Examples:

1. … meet a challenge

softer stress

2. … defeat an opponent

faster faster slower

…

🟦 Sentence retrieval

🟥 POS-based alignment

1) … ?

2) If your actions, … .

slower

louder

… … …

n) … … !

TED

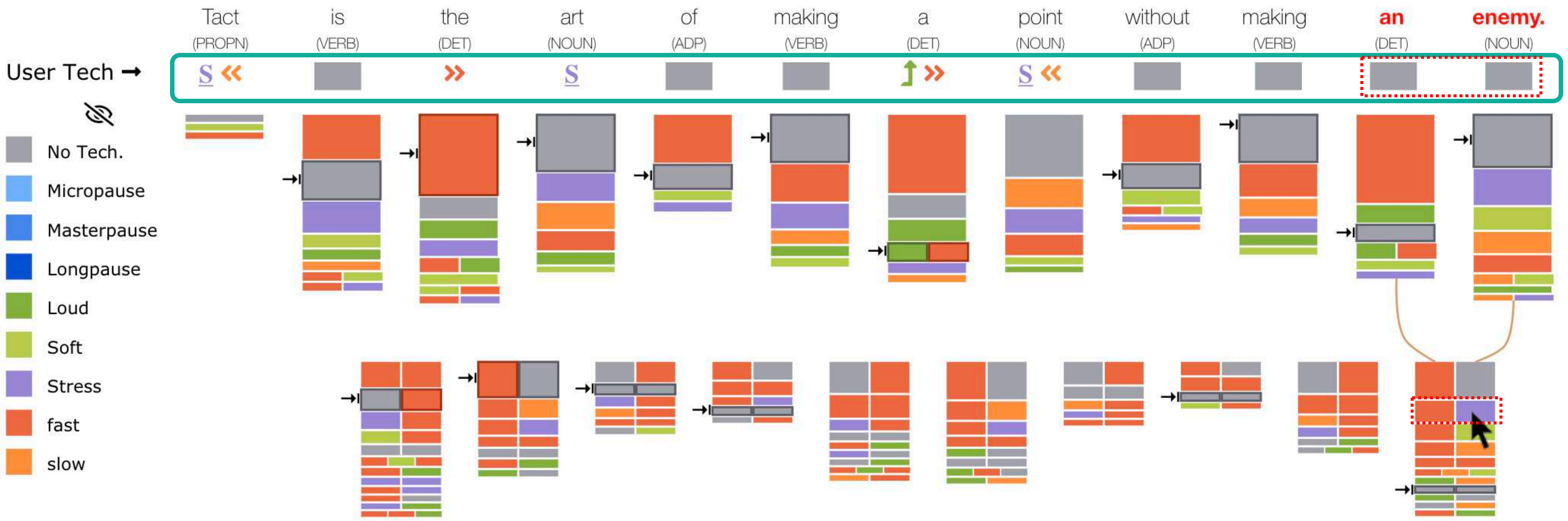
2623 speeches

Overview of VoiceCoach (Wang et al., 2020)

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Data-driven recommendation of voice modulation techniques



N-gram based hierarchical summary

VoiceCoach (Wang et al., 2020)

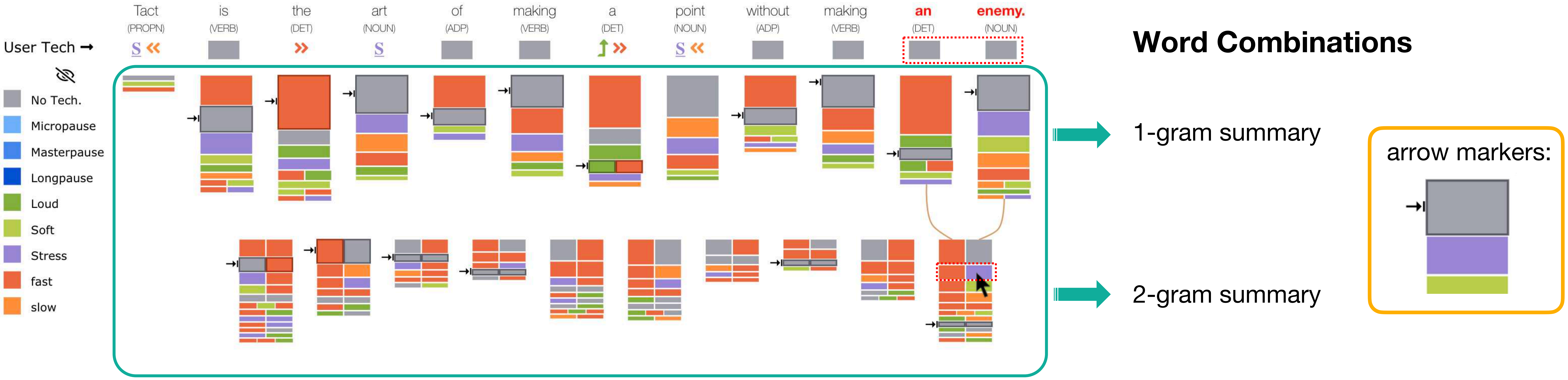
Users' performance

Speed	Faster	»
	Slower	«
Volume	Louder	↑
	Slower	↓
Pause	Brief Pause	🟦
	Master Pause	🟦 🟦
	Long Pause	🟦 🟦 🟦
Pitch	Stress	S
	Normal	▨

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Data-driven recommendation of voice modulation techniques



N-gram based hierarchical summary

VoiceCoach (Wang et al., 2020)

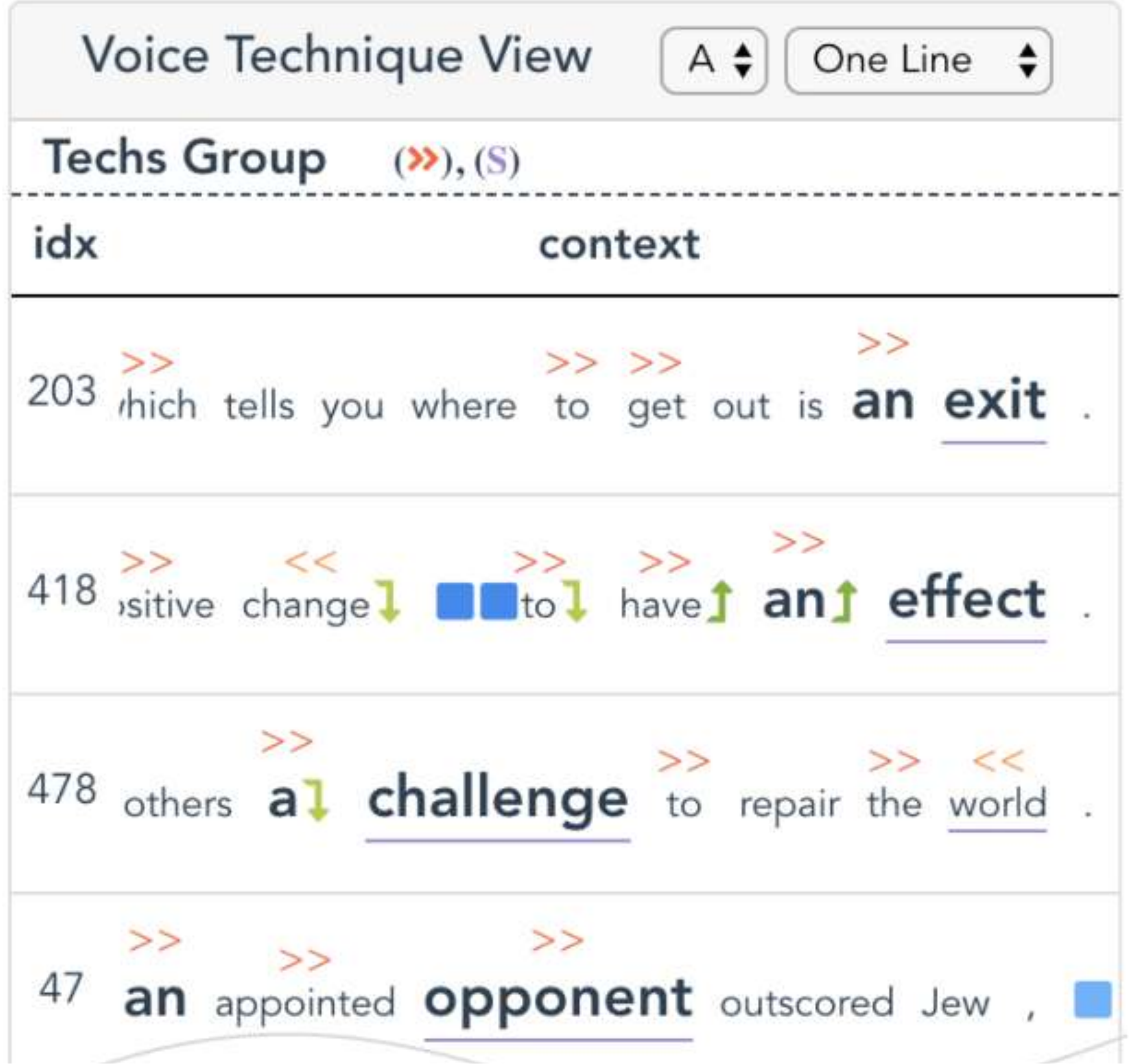
Intelligent User Interface

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Prior feedback: Delivery-based cues ≫ Content-based cues

Data-driven recommendation of voice modulation techniques

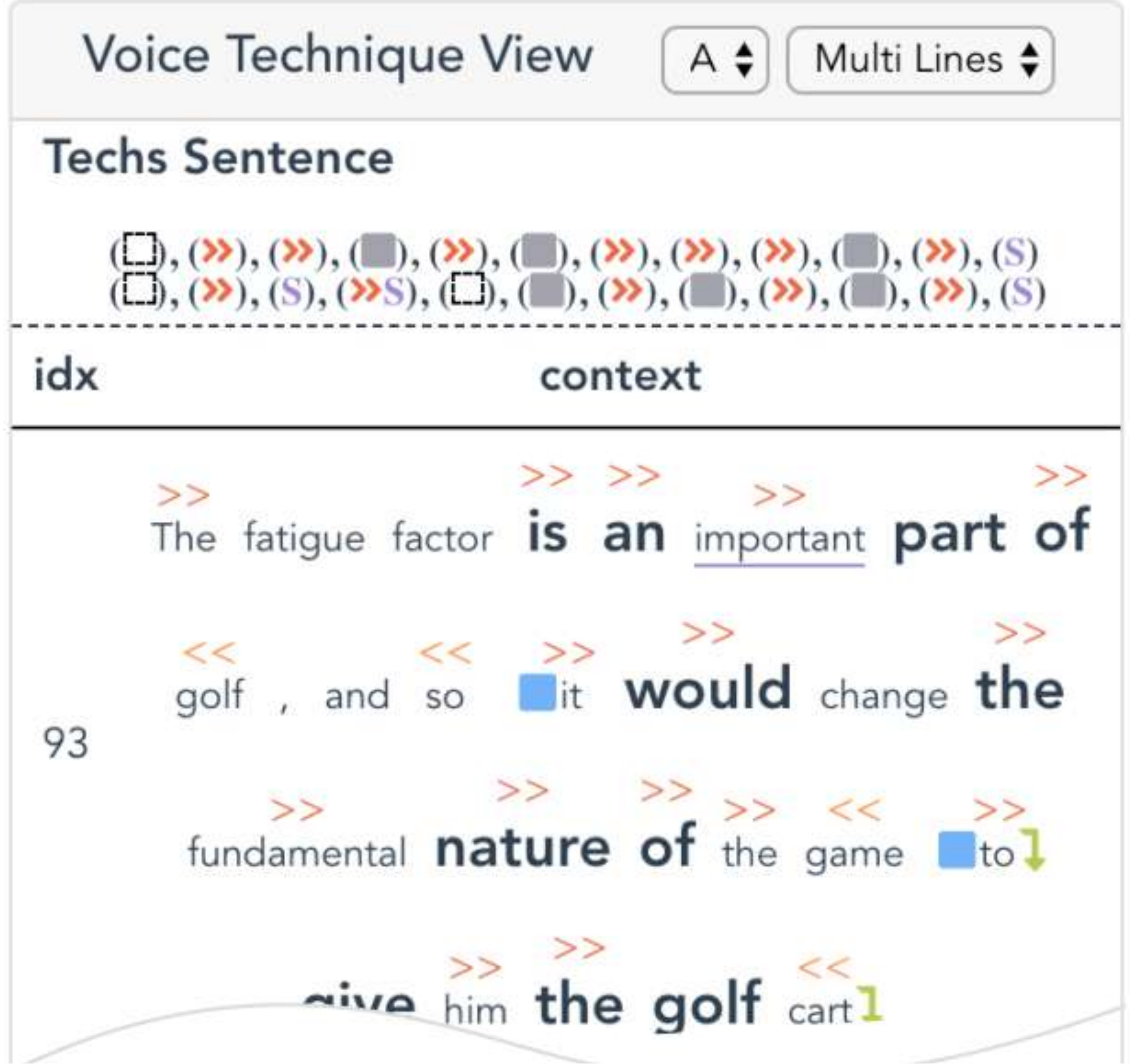
Query: Tact is … making **an enemy**



Voice Technique View | A | One Line

Techs Group (»), (S)

idx	context
203	hich tells you where to get out is an exit .
418	sitive change to have an effect .
478	others a challenge to repair the world .
47	an appointed opponent outscored Jew ,



Voice Technique View | A | Multi Lines

Techs Sentence

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

One-line mode

Multi-line mode

VoiceCoach (Wang et al., 2020)

Speech examples

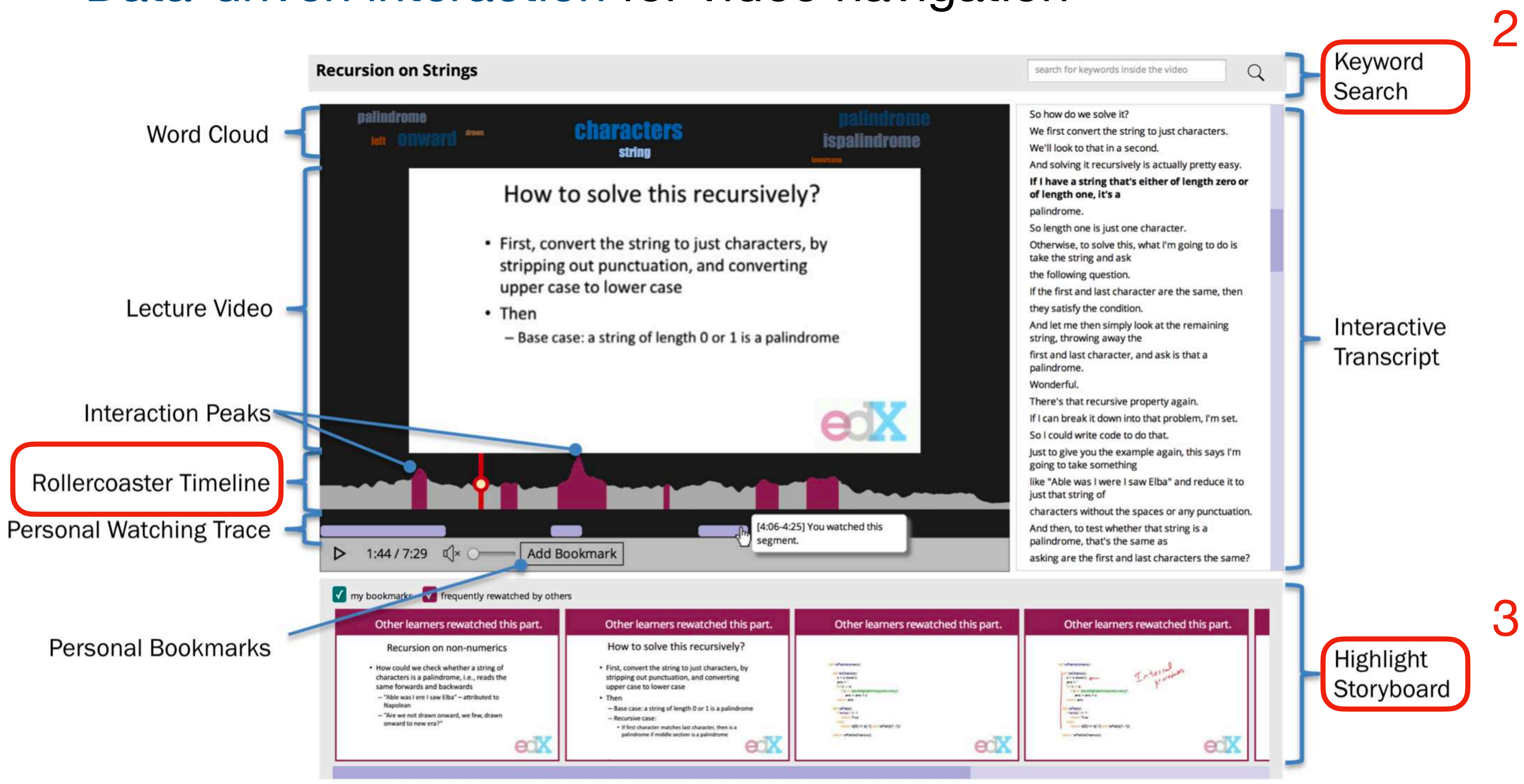
- modulation of interest is highlighted
- context for phrase of interest
- listening to original audio clips

Intelligent User Interface

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Prior feedback: Delivery-based cues ≫ Content-based cues

Data-driven interaction for video navigation



Recursion on Strings

How to solve this recursively?

- First, convert the string to just characters, by stripping out punctuation, and converting upper case to lower case
- Then
  - Base case: a string of length 0 or 1 is a palindrome

[4:06-4:25] You watched this segment.

So how do we solve it?
We first convert the string to just characters.
We'll look to that in a second.
And solving it recursively is actually pretty easy.
If I have a string that's either of length zero or of length one, it's a
palindrome.
So length one is just one character.
Otherwise, to solve this, what I'm going to do is take the string and ask
the following question.
If the first and last character are the same, then
they satisfy the condition.
And let me then simply look at the remaining string, throwing away the
first and last character, and ask is that a palindrome.
Wonderful.
There's that recursive property again.
If I can break it down into that problem, I'm set.
So I could write code to do that.
Just to give you the example again, this says I'm going to take something
like "Able was I were I saw Elba" and reduce it to just that string of
characters without the spaces or any punctuation.
And then, to test whether that string is a palindrome, that's the same as
asking are the first and last characters the same?

☑ my bookmarks ☑ frequently rewatched by others

Other learners rewatched this part.

Recursion on non-numerics

- How could we check whether a string of characters is a palindrome, i.e., reads the same forwards and backwards
  - "Able was I ere I saw Elba" – attributed to Napolean
  - "Are we not drawn onward, we few, drawn onward to new era?"

Other learners rewatched this part.

How to solve this recursively?

- First, convert the string to just characters, by stripping out punctuation, and converting upper case to lower case
- Then
  - Base case: a string of length 0 or 1 is a palindrome
  - Recursive case:
    - If first character matches last character, then is a palindrome if middle section is a palindrome

Other learners rewatched this part.

Other learners rewatched this part.

Internal procedures

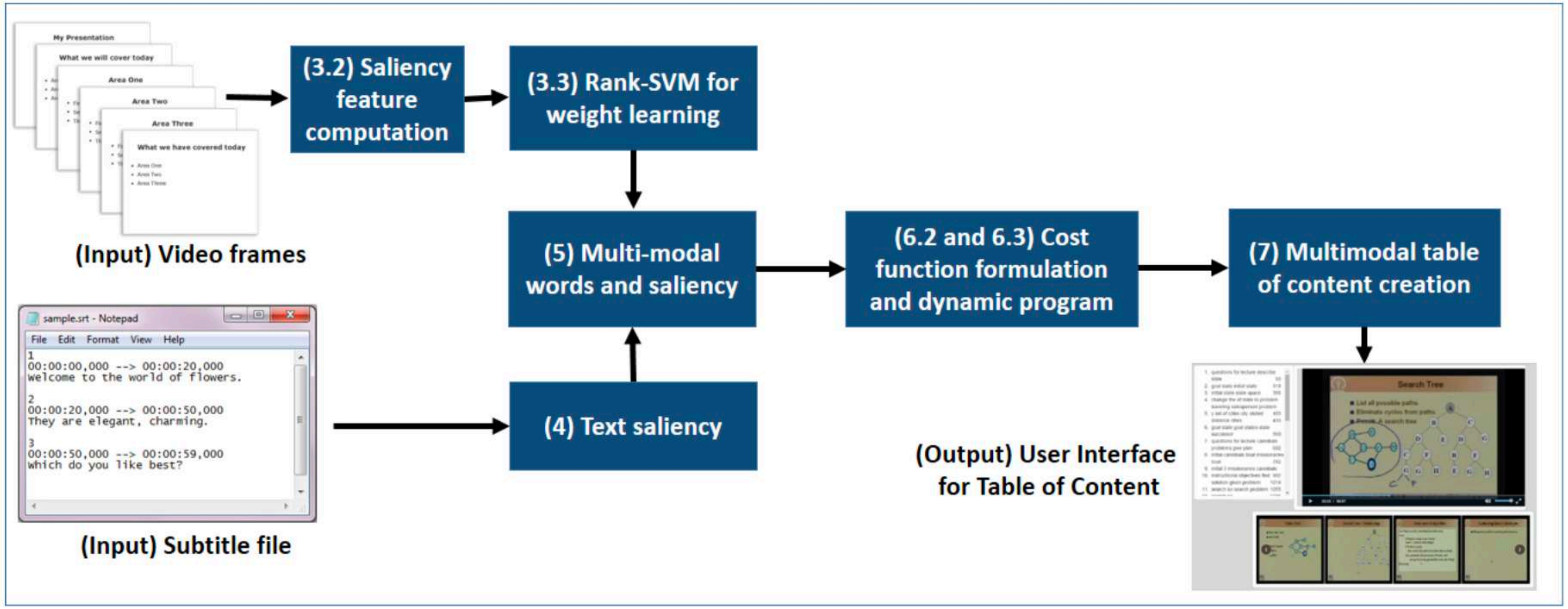
LectureScape (Kim et al., 2014)

- Timeline (1)
- Search (2)
- Summarization (3)

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Topic-based content summarization


My Presentation
What we will cover today
Area One
Area Two
Area Three
What we have covered today
(Input) Video frames
(3.2) Saliency feature computation
(3.3) Rank-SVM for weight learning
(5) Multi-modal words and saliency
(6.2 and 6.3) Cost function formulation and dynamic program
(7) Multimodal table of content creation
(4) Text saliency
sample.srt - Notepad
File Edit Format View Help
(Input) Subtitle file
(Output) User Interface for Table of Content
Search Tree

```
1
00:00:00,000 --> 00:00:20,000
Welcome to the world of flowers.

2
00:00:20,000 --> 00:00:50,000
They are elegant, charming.

3
00:00:50,000 --> 00:00:59,000
Which do you like best?
```

MMToC (Biswas et al., 2015)

Visual & spoken word fusion

- Extend visual salient words with a group of spoken salient words based on the semantic similarity

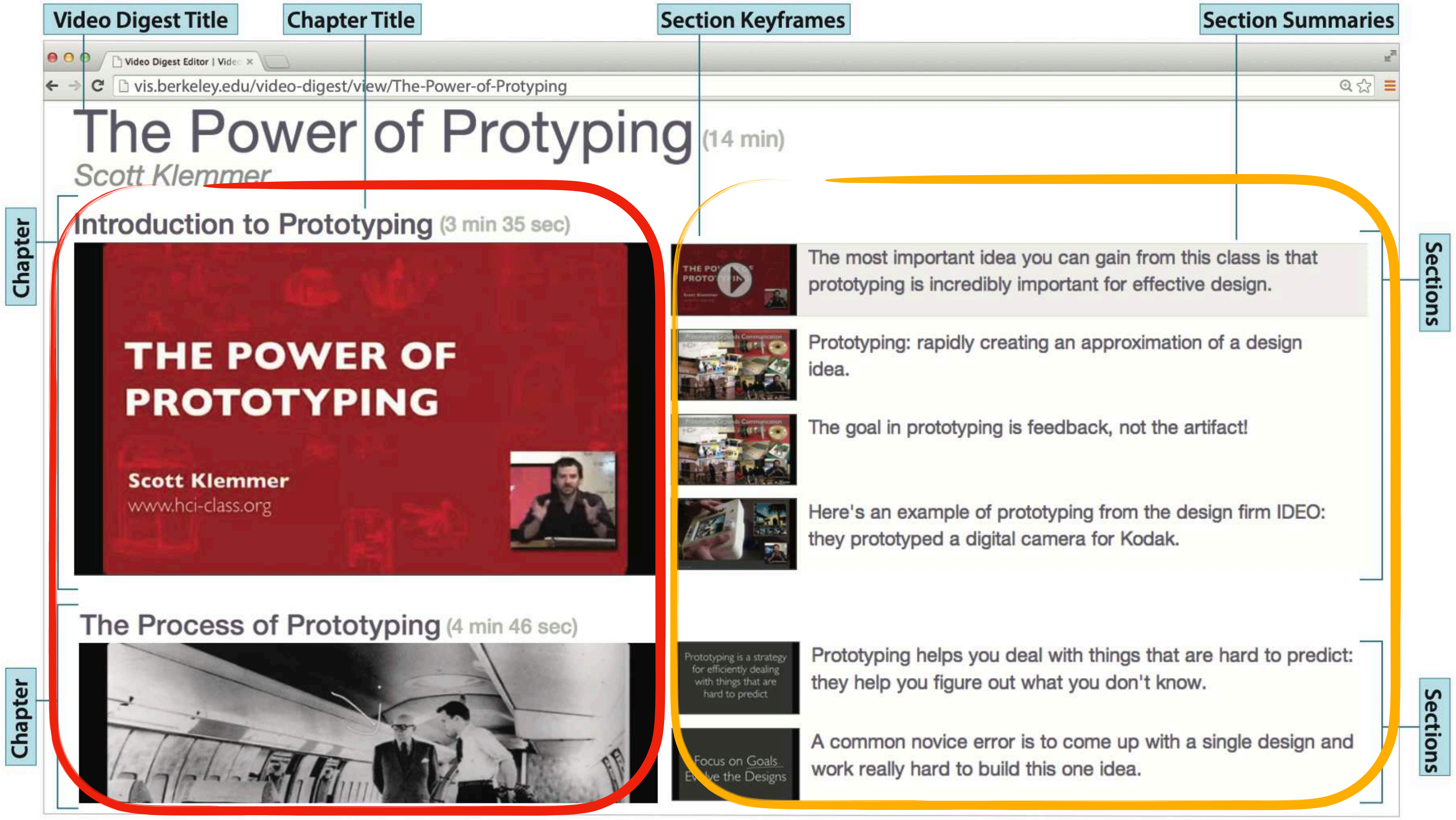
Content segmentation

- Minimize the inner-group difference
- Maximize the inter-group difference

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

Textbook-inspired chapter/section content organization



The Power of Protyping (14 min)

Scott Klemmer

Introduction to Prototyping (3 min 35 sec)

The most important idea you can gain from this class is that prototyping is incredibly important for effective design.

Prototyping: rapidly creating an approximation of a design idea.

The goal in prototyping is feedback, not the artifact!

Here's an example of prototyping from the design firm IDEO: they prototyped a digital camera for Kodak.

The Process of Prototyping (4 min 46 sec)

Prototyping helps you deal with things that are hard to predict: they help you figure out what you don't know.

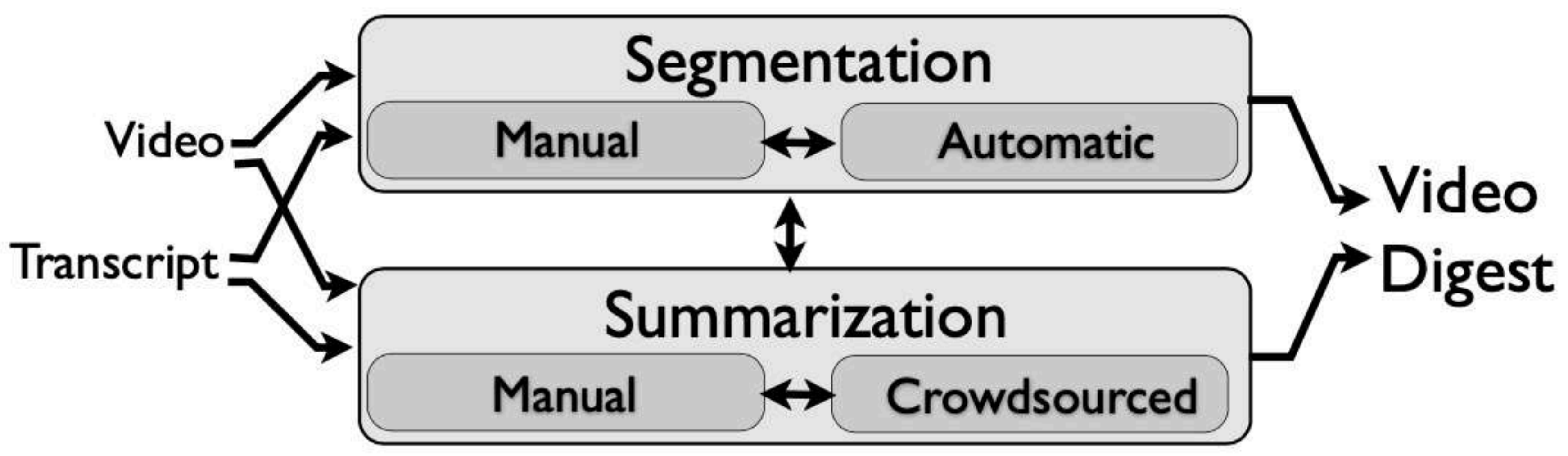
A common novice error is to come up with a single design and work really hard to build this one idea.

Video Digest (Pavel et al., 2014)

🟥 Chapter — Topically coherent sections

🟧 Section — A set of varying topics

Bayesian topic segmentation



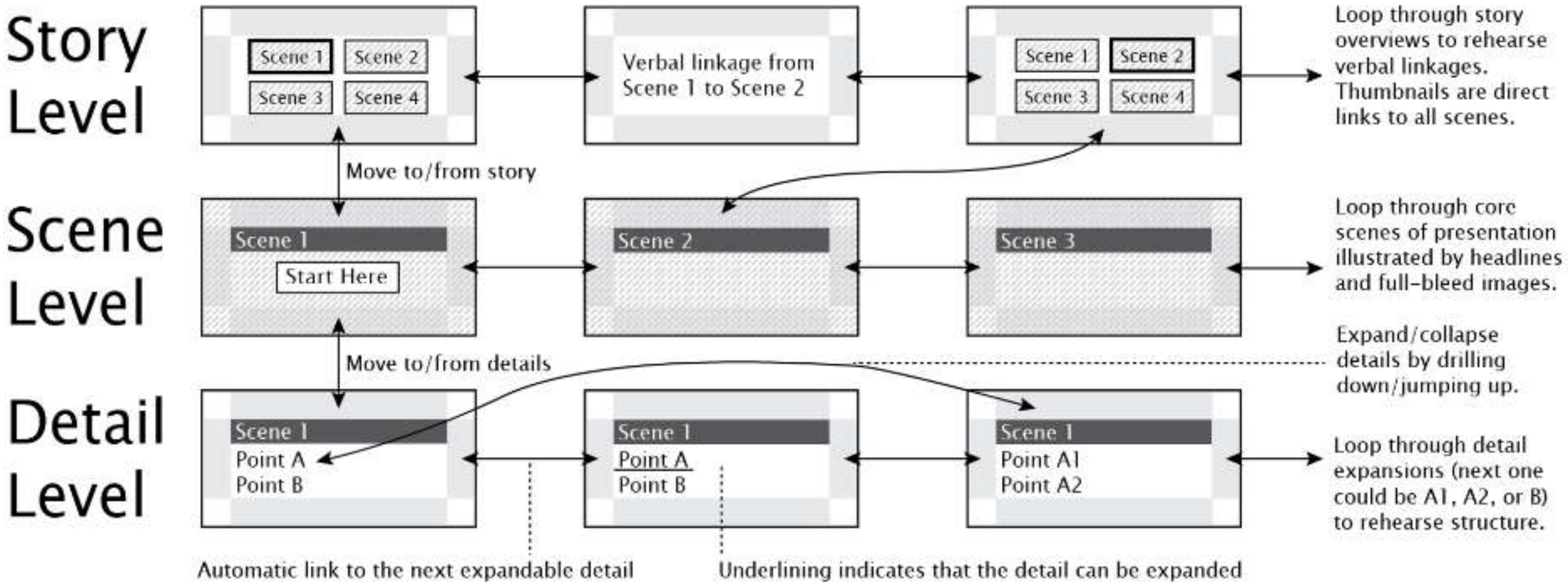
Crowdsourcing summary and ranking

Intelligent User Interface

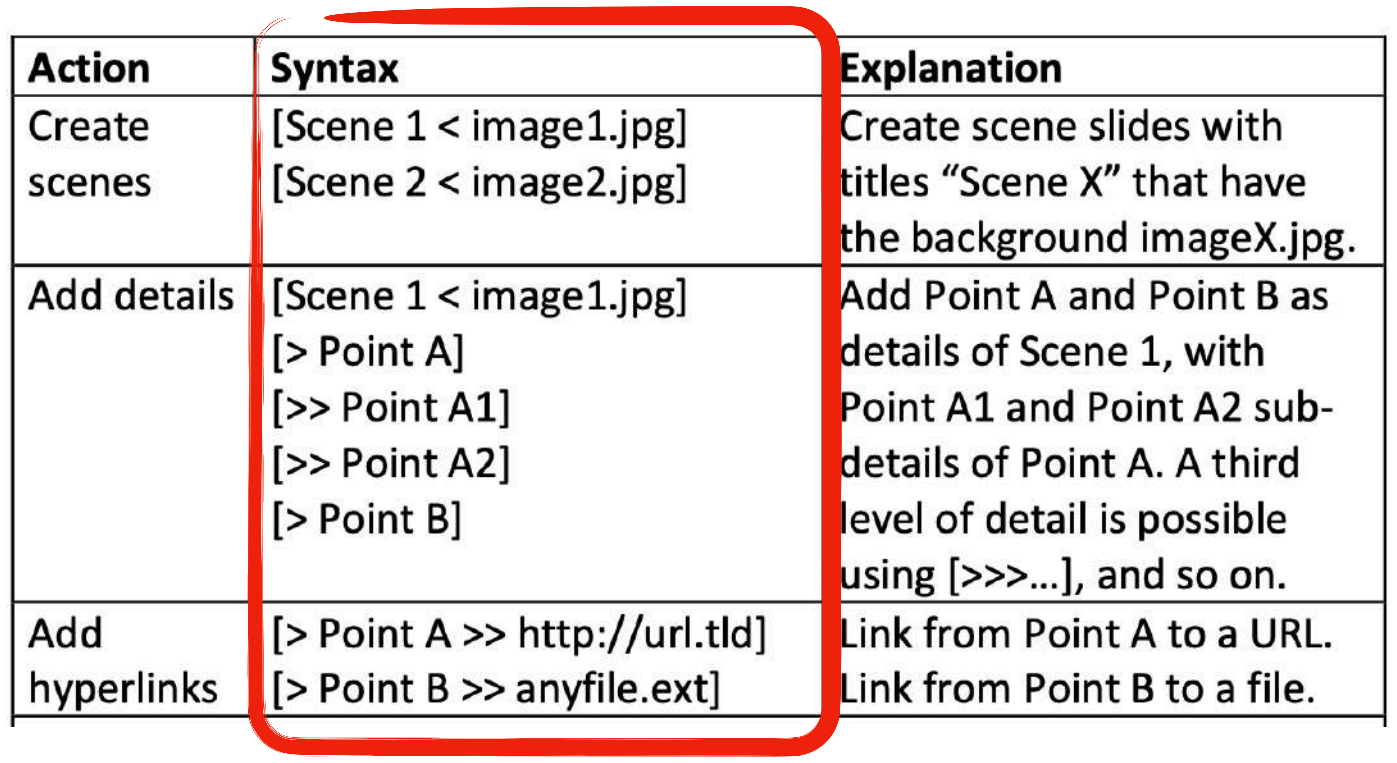
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Prior feedback: Delivery-based cues ≫ Content-based cues

From content exploration → Hierarchical content structure planning


Story Level
Scene Level
Detail Level
Scene 1
Scene 2
Scene 3
Scene 4
Verbal linkage from Scene 1 to Scene 2
Scene 1
Scene 2
Scene 3
Scene 4
Loop through story overviews to rehearse verbal linkages. Thumbnails are direct links to all scenes.
Move to/from story
Scene 1
Start Here
Scene 2
Scene 3
Loop through core scenes of presentation illustrated by headlines and full-bleed images.
Move to/from details
Expand/collapse details by drilling down/jumping up.
Scene 1
Point A
Point B
Scene 1
Point A
Point B
Scene 1
Point A1
Point A2
Loop through detail expansions (next one could be A1, A2, or B) to rehearse structure.
Automatic link to the next expandable detail
Underlining indicates that the detail can be expanded

HyperSlides (Edge et al., 2014)



Action	Syntax	Explanation
Create scenes	[Scene 1 < image1.jpg] [Scene 2 < image2.jpg]	Create scene slides with titles "Scene X" that have the background imageX.jpg.
Add details	[Scene 1 < image1.jpg] [> Point A] [>> Point A1] [>> Point A2] [> Point B]	Add Point A and Point B as details of Scene 1, with Point A1 and Point A2 sub-details of Point A. A third level of detail is possible using [>>>...], and so on.
Add hyperlinks	[> Point A >> http://url.tld] [> Point B >> anyfile.ext]	Link from Point A to a URL. Link from Point B to a file.

Mark-up language to create hierarchically structured scenes

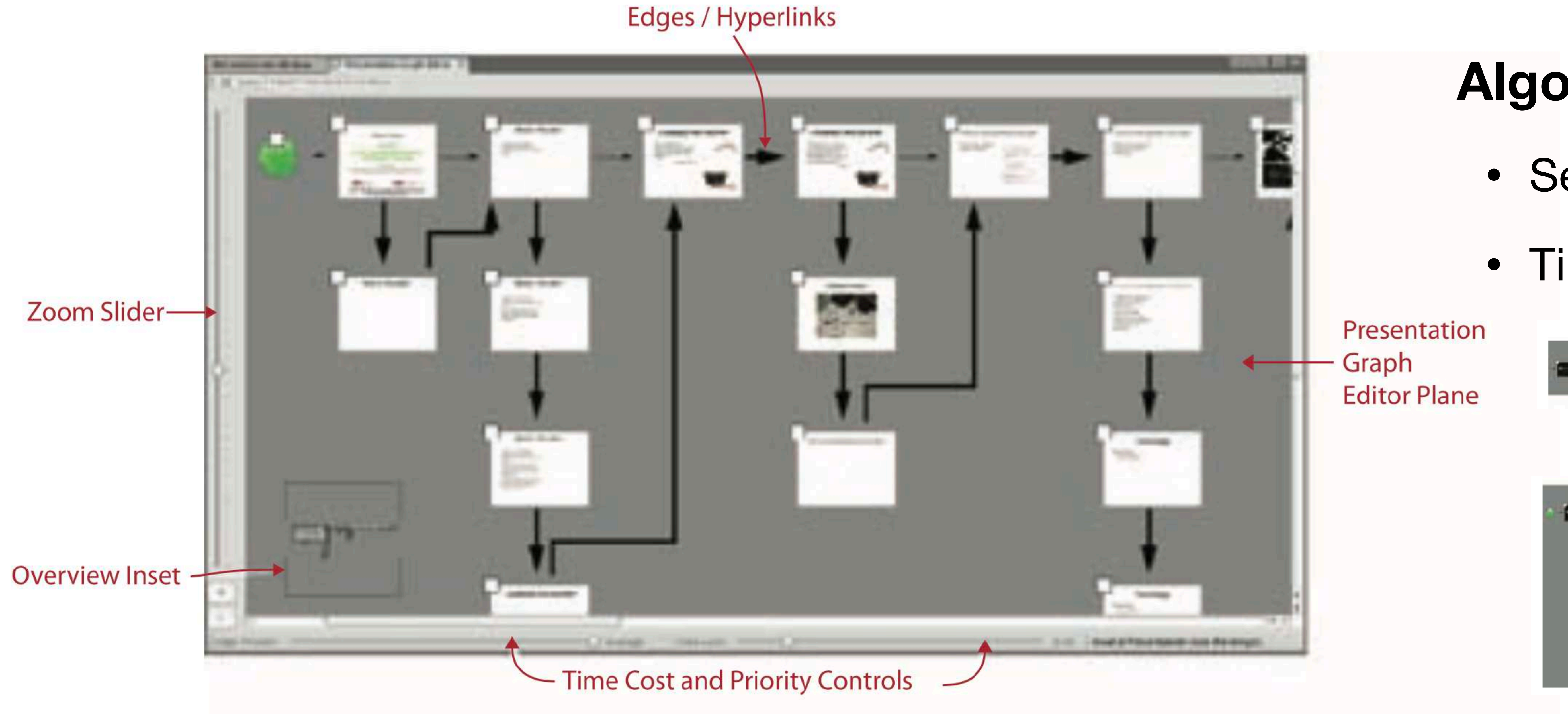
↓

hyperlinked slides of a consistent and minimalist style

Intelligent User Interface

Prior feedback: Delivery-based cues ≫ Content-based cues

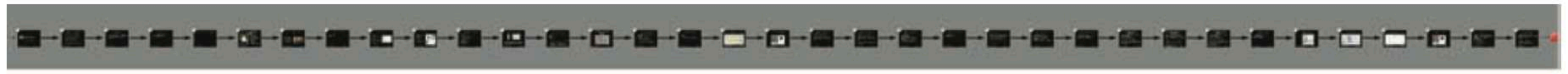
From manual planning → Automatic structure generation & path suggestions


Edges / Hyperlinks
Zoom Slider
Presentation Graph Editor Plane
Overview Inset
Time Cost and Priority Controls

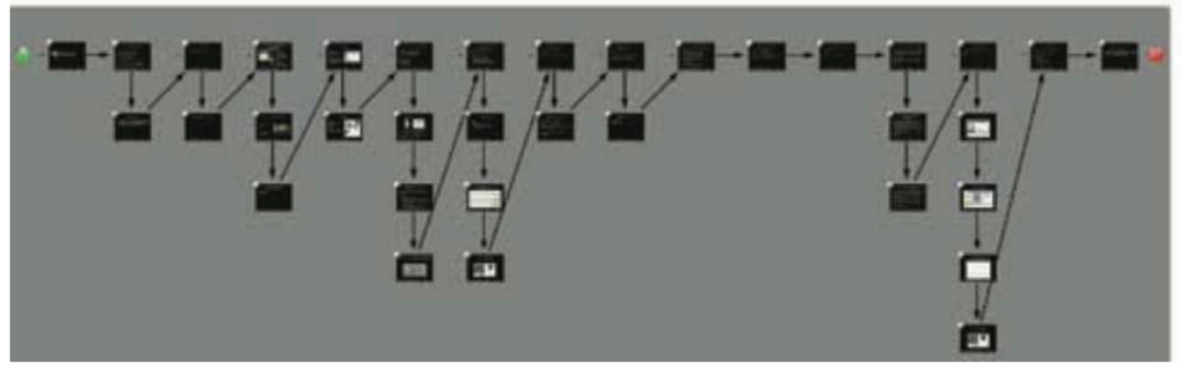
NextSlidePlease (Spicer et al., 2012)

Algorithm support

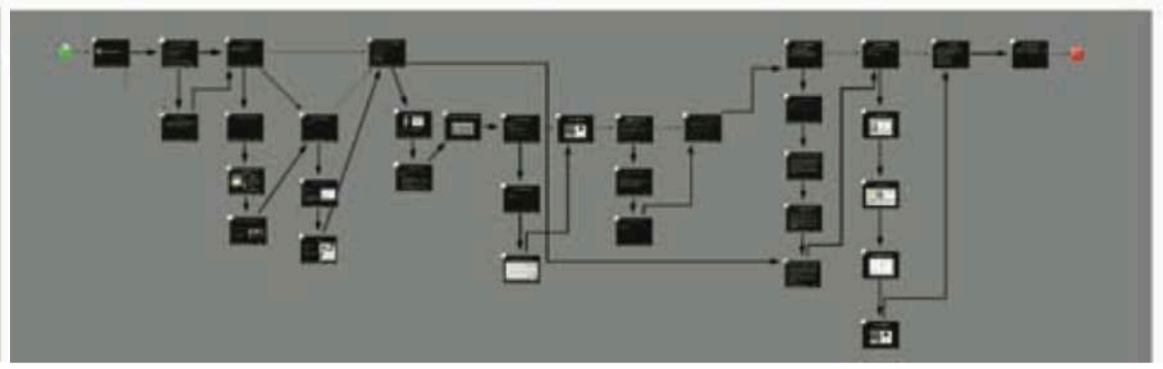
- Semantic similarity between adjacent slides => presentation graph
- Time constraints, priority => path suggestions



Linear layout



Structure generation



Post editing

Intelligent User Interface

Prior feedback:

Limitations

- Do not consider learning from BAD examples
- Do not consider learning from multimodal speech styles

NEXT: Live feedback (implicit feedback)

Intelligent User Interface

Live feedback: Implicit feedback ≫ Explicit feedback

Simulate nonverbal behavior of virtual audience

- Posture (e.g., straight, relaxed, forward)
- Head orientation
- Gaze

Virtual audiences in Cicero (Batrinca et al., 2013)

MACH (Hoque et al., 2013)

Acknowledgement

- nodding, changing posture
- spoken acknowledgement "That's very interesting"

Intelligent User Interface

Live feedback: Implicit feedback ≫ Explicit feedback

Investigate the impact of nonverbal behavior of virtual audience

Two characters of virtual coaches (Gebhard et al., 2013)

	Gestures	*Facial expression*	*Pause*	*Gaze*	*Comments*
Understanding	Narrow	Positive	Shorter	Friendly	Many polite phrases
Demanding	Space-taking	Negative	Longer	Dominant	Few polite phrases

Within-subject study

- Participants perceive the differences and they reported that **demanding** character induced higher level of stress
- Demanding condition: more breathing **pauses**, higher **movement** energy

Intelligent User Interface

Live feedback: Implicit feedback ≫ Explicit feedback

Limitations

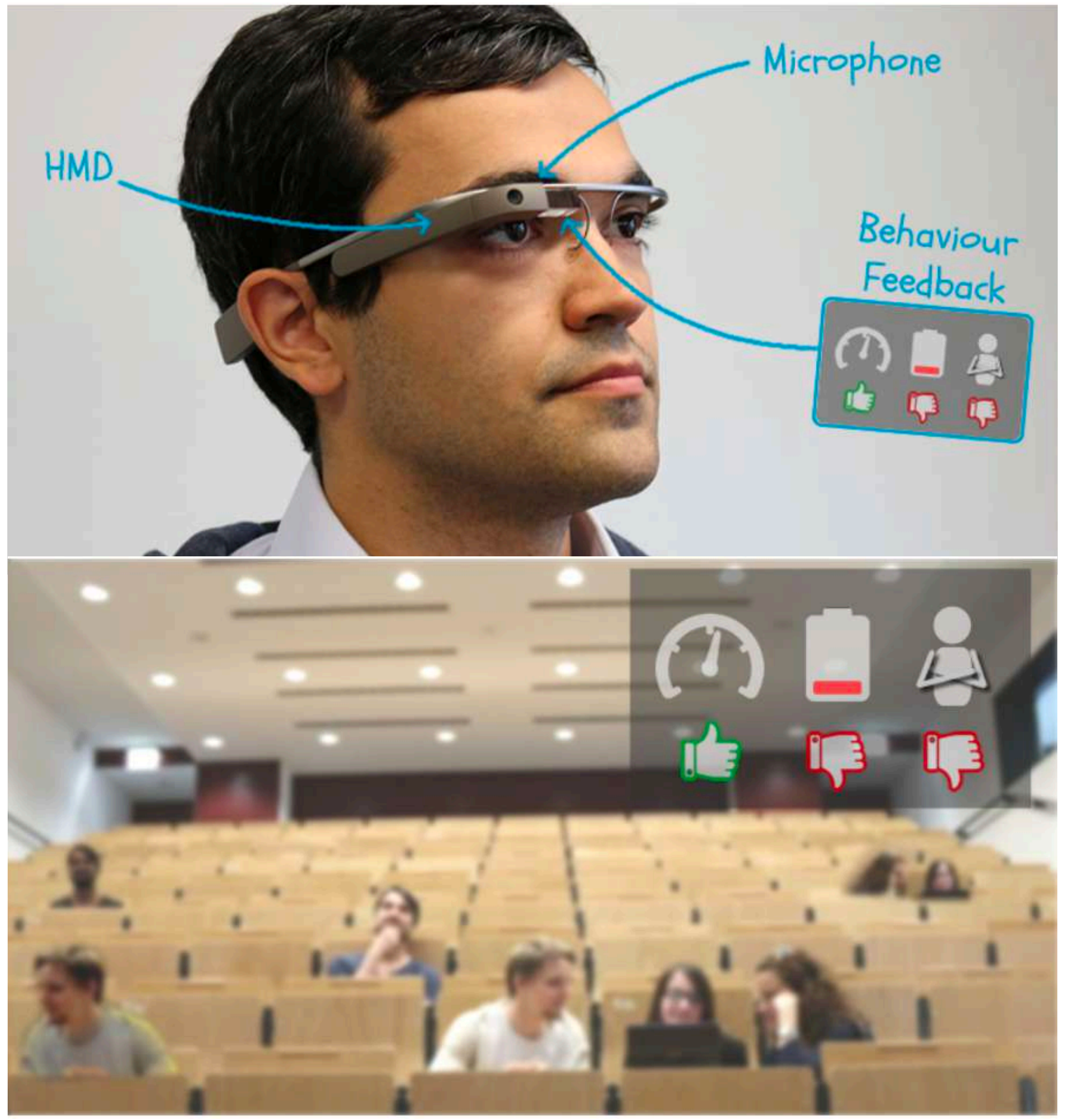
- Most listening behaviors of virtual audience are controlled by finite state machines. There is a lack of more intelligent models to simulate affective states of listeners

NEXT: Explicit feedback (simple verification)

Intelligent User Interface

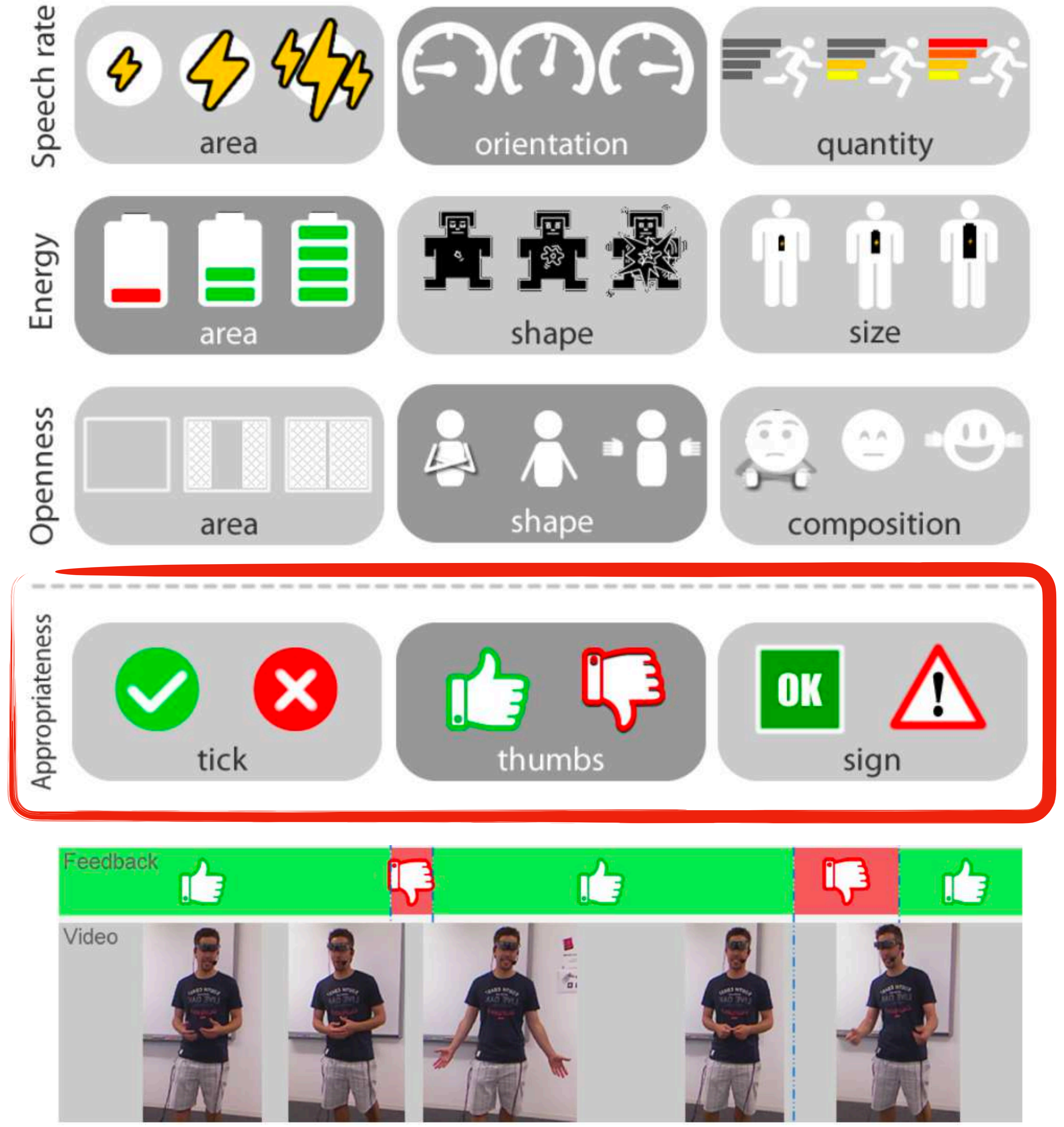
Live feedback: Implicit feedback ≫ Explicit feedback - simple verification

Realtime behavioural checking on speech delivery with Google Glass



Logue (Hoque et al., 2013)

Feedback icon alternatives

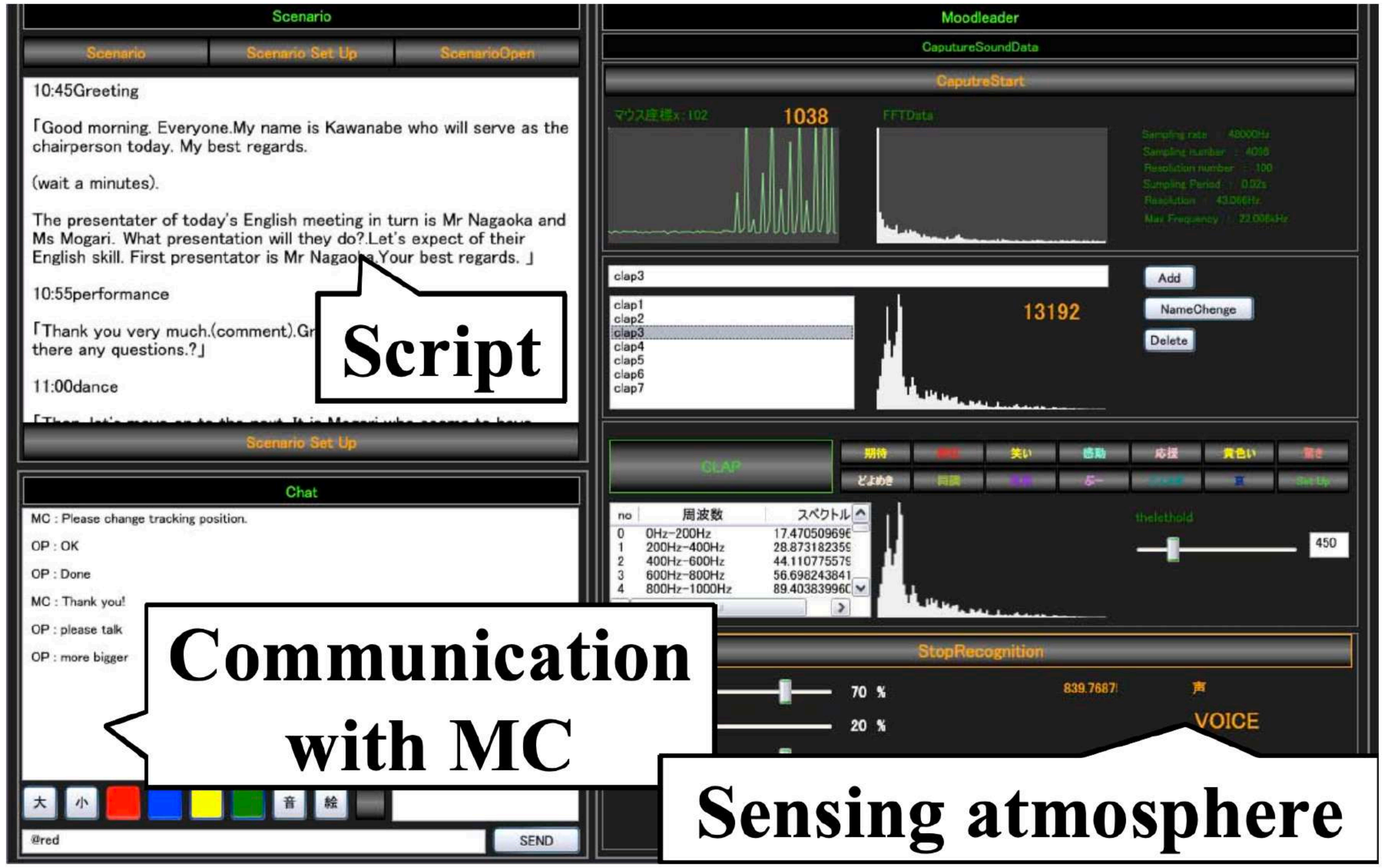


▨ Final design

Intelligent User Interface

Live feedback: Implicit feedback ≫ Explicit feedback - simple verification

Realtime behavioural checking on speech content with HMD



10:45Greeting

「Good morning. Everyone.My name is Kawanabe who will serve as the chairperson today. My best regards.

(wait a minutes).

The presentater of today's English meeting in turn is Mr Nagaoka and Ms Mogari. What presentation will they do?.Let's expect of their English skill. First presentator is Mr Nagaoka.Your best regards. 」

10:55performance

「Thank you very much.(comment).Gr
there any questions.?」

11:00dance

MC : Please change tracking position.

OP : OK

OP : Done

MC : Thank you!

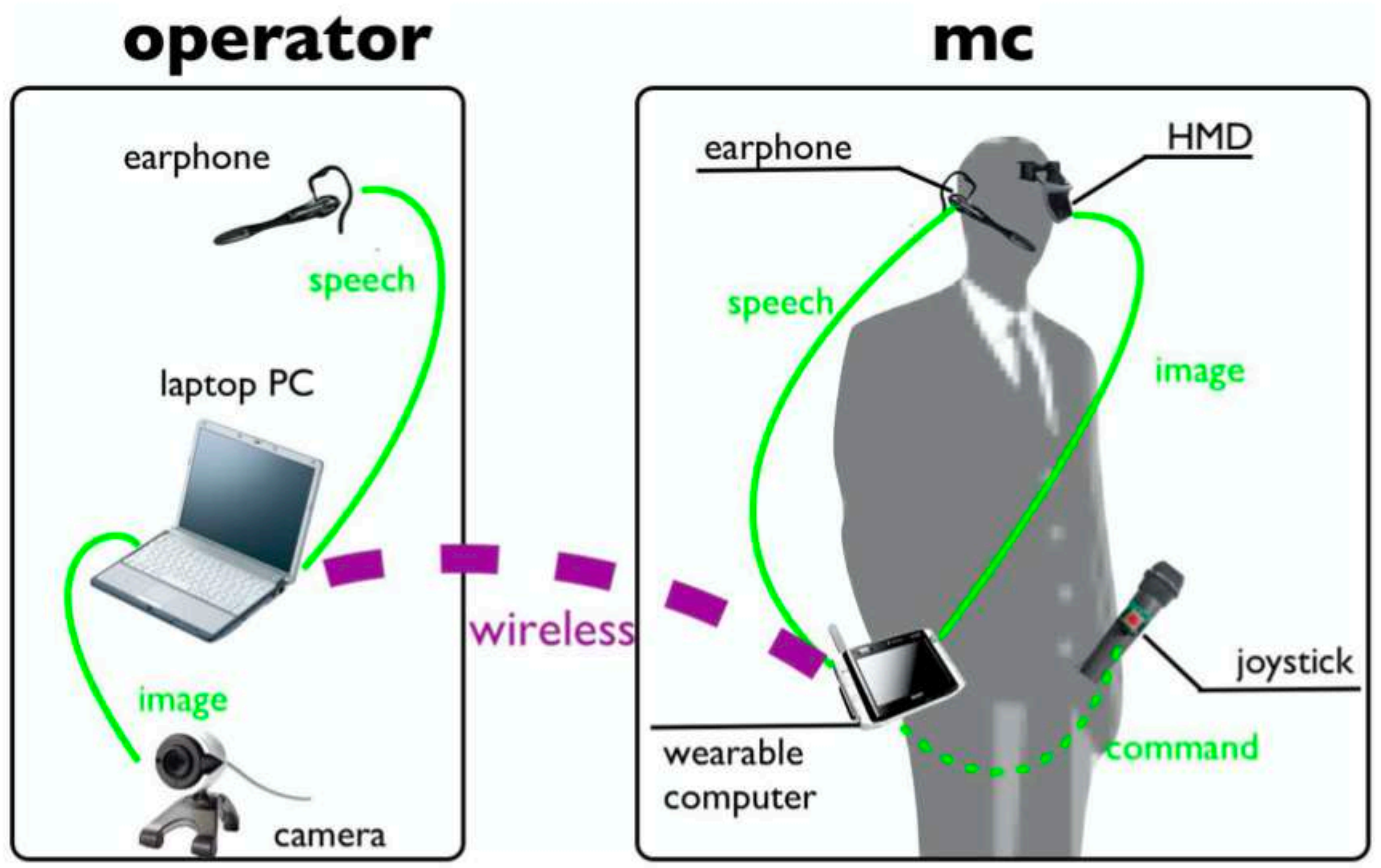
OP : please talk

OP : more bigger

Sampling rate : 48000Hz
Sampling number : 4096
Resolution number : 100
Sampling Period : 0.02s
Resolution : 43.066Hz
Max Frequency : 22.006kHz

no	周波数	スペクトル
0	0Hz-200Hz	17.47050969€
1	200Hz-400Hz	28.87318235€
2	400Hz-600Hz	44.11077557€
3	600Hz-800Hz	56.698243841
4	800Hz-1000Hz	89.40383996C

A wearable MC system (Okada et al., 2011)



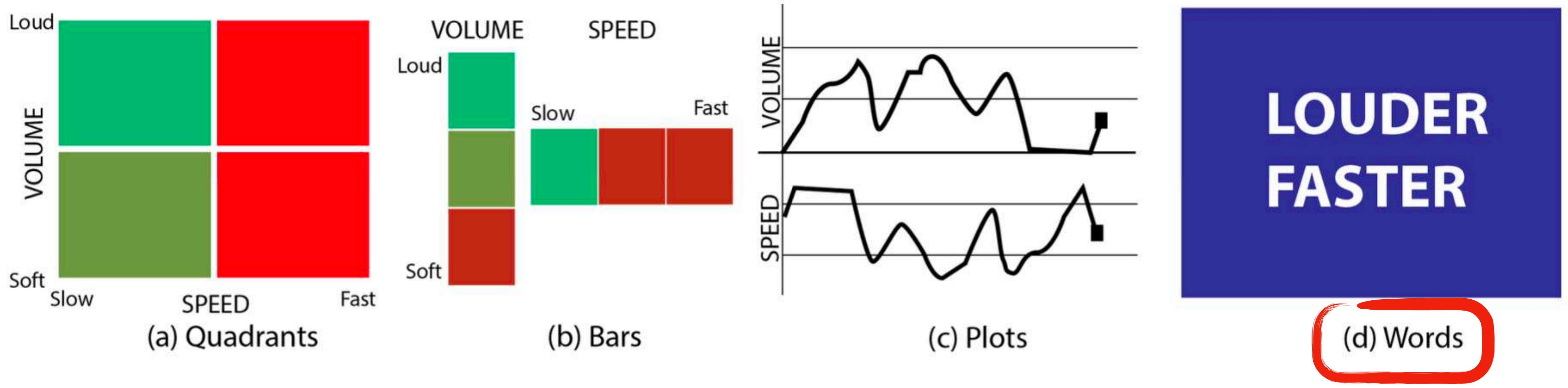
Manage stage

- Realtime tracking of MC's speech
- Communication with operators
- Sensing atmosphere (e.g., buzzing, laugh)

Intelligent User Interface

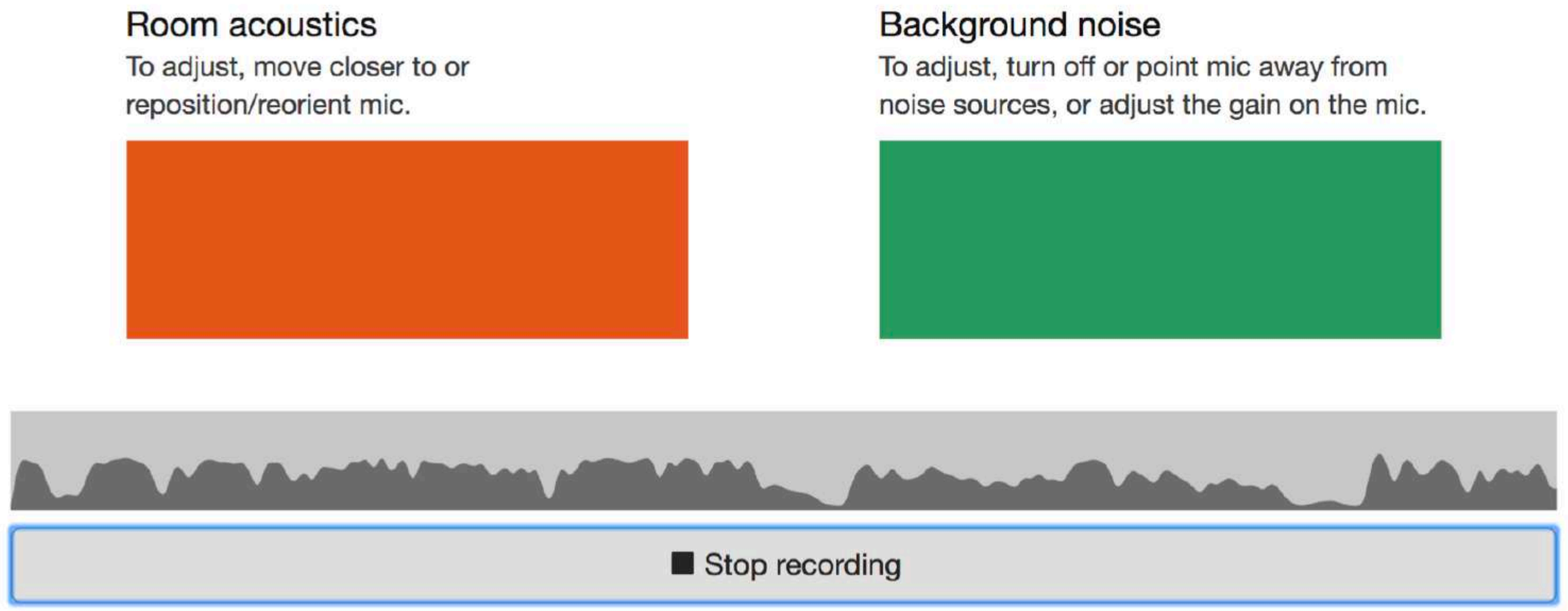
Live feedback: Implicit feedback ≫ Explicit feedback - elaborated feedback

Realtime instructions on speech delivery with visual feedback



(a) Quadrants (b) Bars (c) Plots (d) Words

Rhema (Okada et al., 2011)



Room acoustics
To adjust, move closer to or reposition/reorient mic.

Background noise
To adjust, turn off or point mic away from noise sources, or adjust the gain on the mic.

VoiceAssist (Okada et al., 2011)

User study

- Verbal feedback is most favored
- Participants prefer sparse feedback to continuous feedback

Room acoustics: speech transmission index

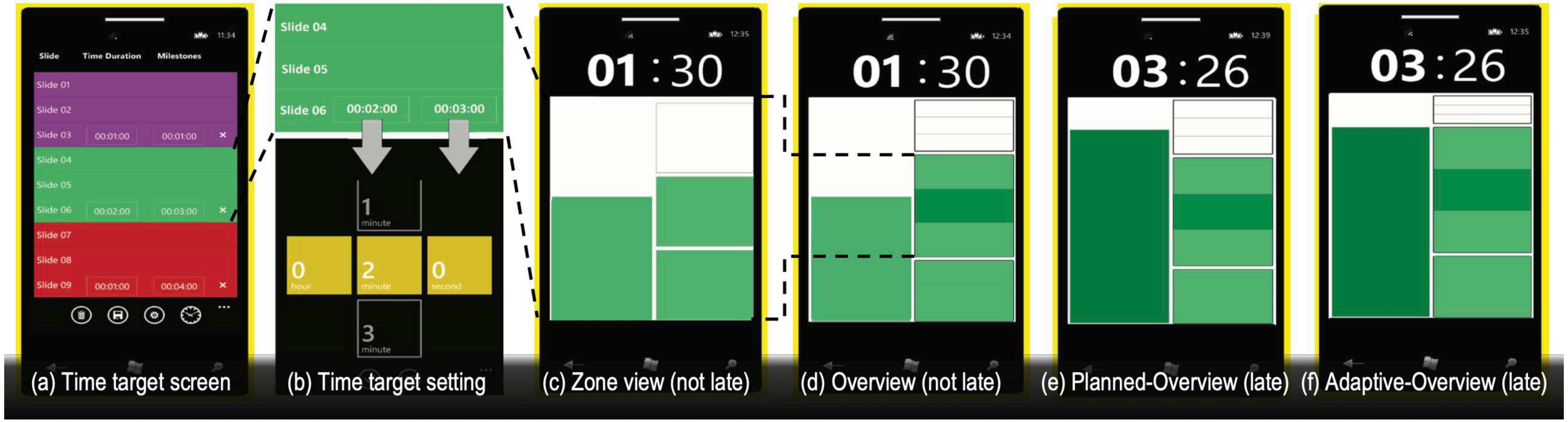
Background noise: signal to noise ratio

Box: red/green indicates low/high audio quality

Intelligent User Interface

Live feedback: Implicit feedback ≫ Explicit feedback - elaborated feedback

From behavior awareness → Dynamic time control during presentation



(a) Time target screen (b) Time target setting (c) Zone view (not late) (d) Overview (not late) (e) Planned-Overview (late) (f) Adaptive-Overview (late)

TalkZones (Saket et al., 2014)

Two types of timing support

Less flexible

- Planned rehearsal (e)

More flexible

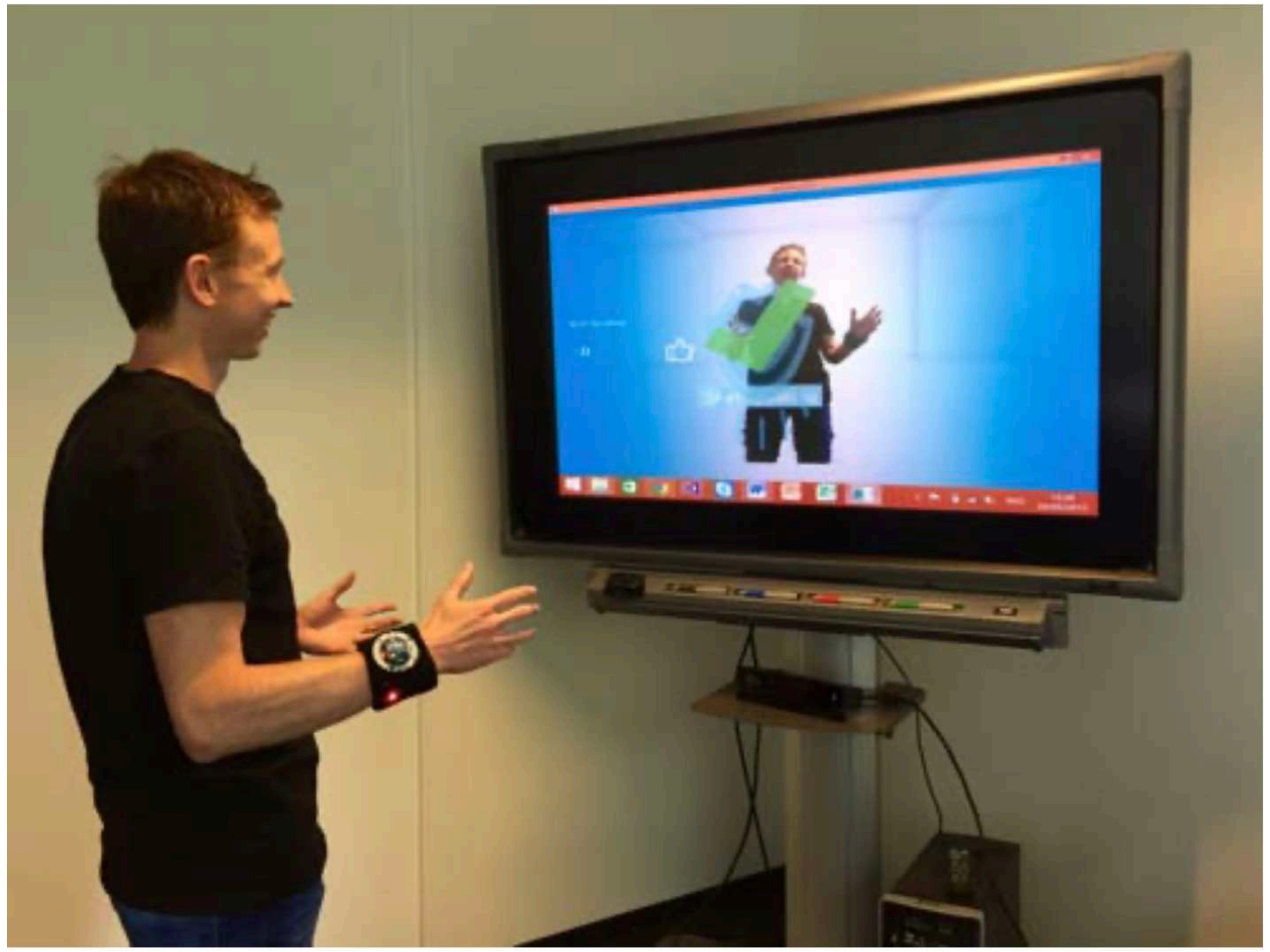
- Adaptive guidance (f)

Haptic feedback is enabled for redundant representation and reminder of lateness

Intelligent User Interface

Live feedback: Implicit feedback » Explicit feedback - termination

Interruption for improving specific skills



Presentation Trainer (Schneider et al., 2014)

Action list (e.g., voice modulation)

- Volume: loud, soft, normal
- Pause: long narration without pauses
- Filler sounds: “ehm”, “hmm”

→ **Corrective feedback** (realtime visuals)

Severe mistakes

(vibration, pause sound, stops the program)

- Repetition of same mistakes
- Mistakes without being corrected for too long
- Predefined severe mistakes

→ **Interruptive feedback**

Intelligent User Interface

Live feedback

Limitations

Implicit feedback

Explicit feedback

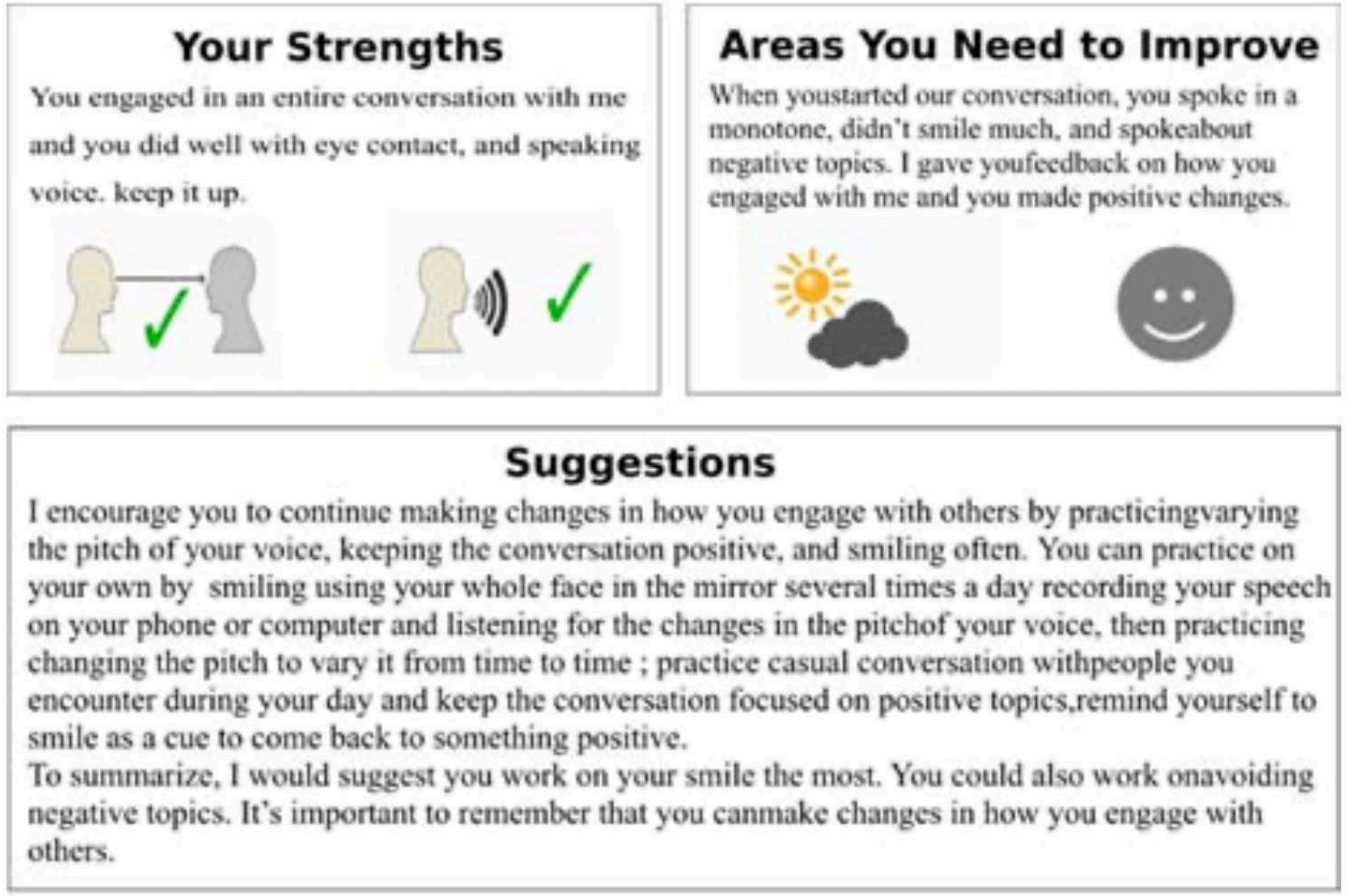
- Simple verification
- Elaborated feedback
- Termination

- Most systems focus on providing timely suggestions about users' performance. They do not consider how to help them effectively and efficiently correct their mistakes

Intelligent User Interface

Posterior feedback: Summary feedback ≫ Focus feedback

Summary of strengths & weaknesses



Your Strengths

You engaged in an entire conversation with me and you did well with eye contact, and speaking voice. keep it up.

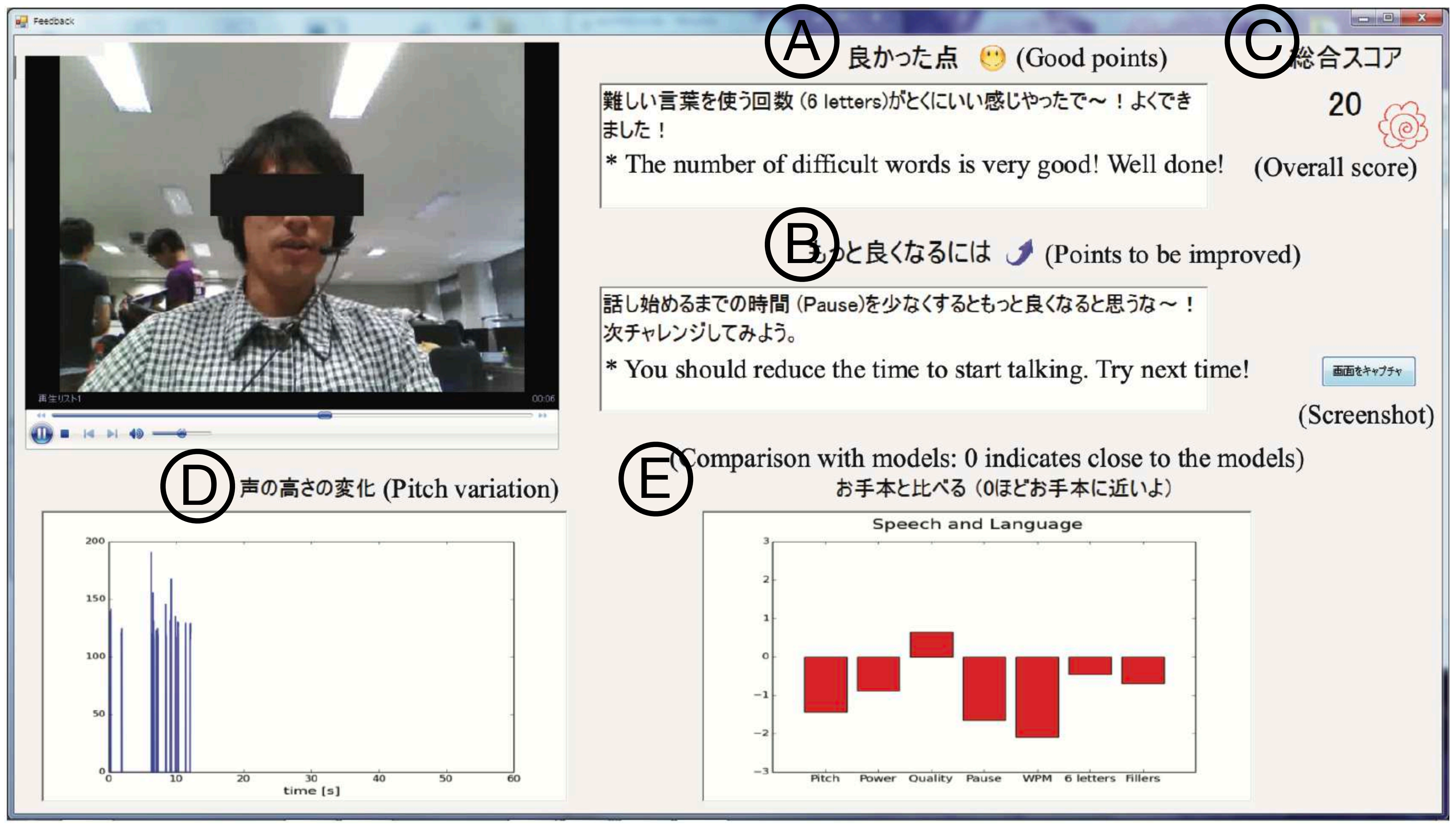
Areas You Need to Improve

When youstarted our conversation, you spoke in a monotone, didn't smile much, and spokeabout negative topics. I gave youfeedback on how you engaged with me and you made positive changes.

Suggestions

I encourage you to continue making changes in how you engage with others by practicingvarying the pitch of your voice, keeping the conversation positive, and smiling often. You can practice on your own by smiling using your whole face in the mirror several times a day recording your speech on your phone or computer and listening for the changes in the pitchof your voice, then practicing changing the pitch to vary it from time to time ; practice casual conversation withpeople you encounter during your day and keep the conversation focused on positive topics,remind yourself to smile as a cue to come back to something positive.
To summarize, I would suggest you work on your smile the most. You could also work onavoiding negative topics. It's important to remember that you canmake changes in how you engage with others.

Aging and Engaging (Ali et al., 2018)



[redacted]

良かった点 (Good points)

難しい言葉を使う回数 (6 letters)がとくにいい感じやったで～！よくできました！

* The number of difficult words is very good! Well done!

総合スコア

20

(Overall score)

もっと良くなるには (Points to be improved)

話し始めるまでの時間 (Pause)を少なくするともっと良くなると思うな～！
次チャレンジしてみよう。

* You should reduce the time to start talking. Try next time!

(Screenshot)

声の高さの変化 (Pitch variation)

(Comparison with models: 0 indicates close to the models)
お手本と比べる (0ほどお手本に近いよ)

Automated Social Skills Trainer (Tanaka et al., 2015)

A: good points, B: bad points

C: overall score

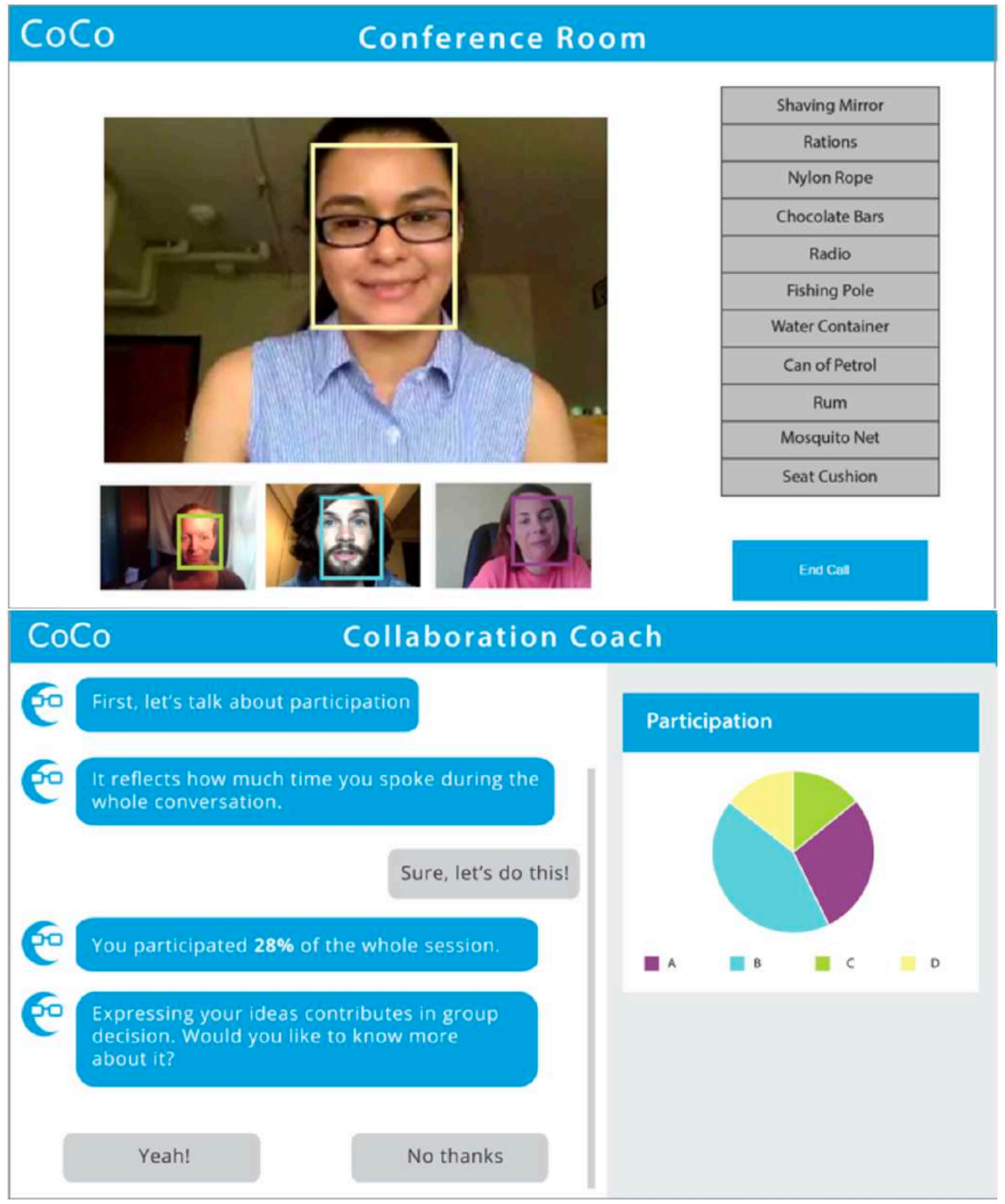
D: pitch variation

E: comparison with model persons (pitch, power, energy, pause, WPM, 6 letters, fillers)

Intelligent User Interface

Posterior feedback: Summary feedback ≫ Focused feedback

Explaining affective behavioral performance on demand



CoCo

Conference Room

CoCo

Collaboration Coach

First, let's talk about participation

It reflects how much time you spoke during the whole conversation.

Sure, let's do this!

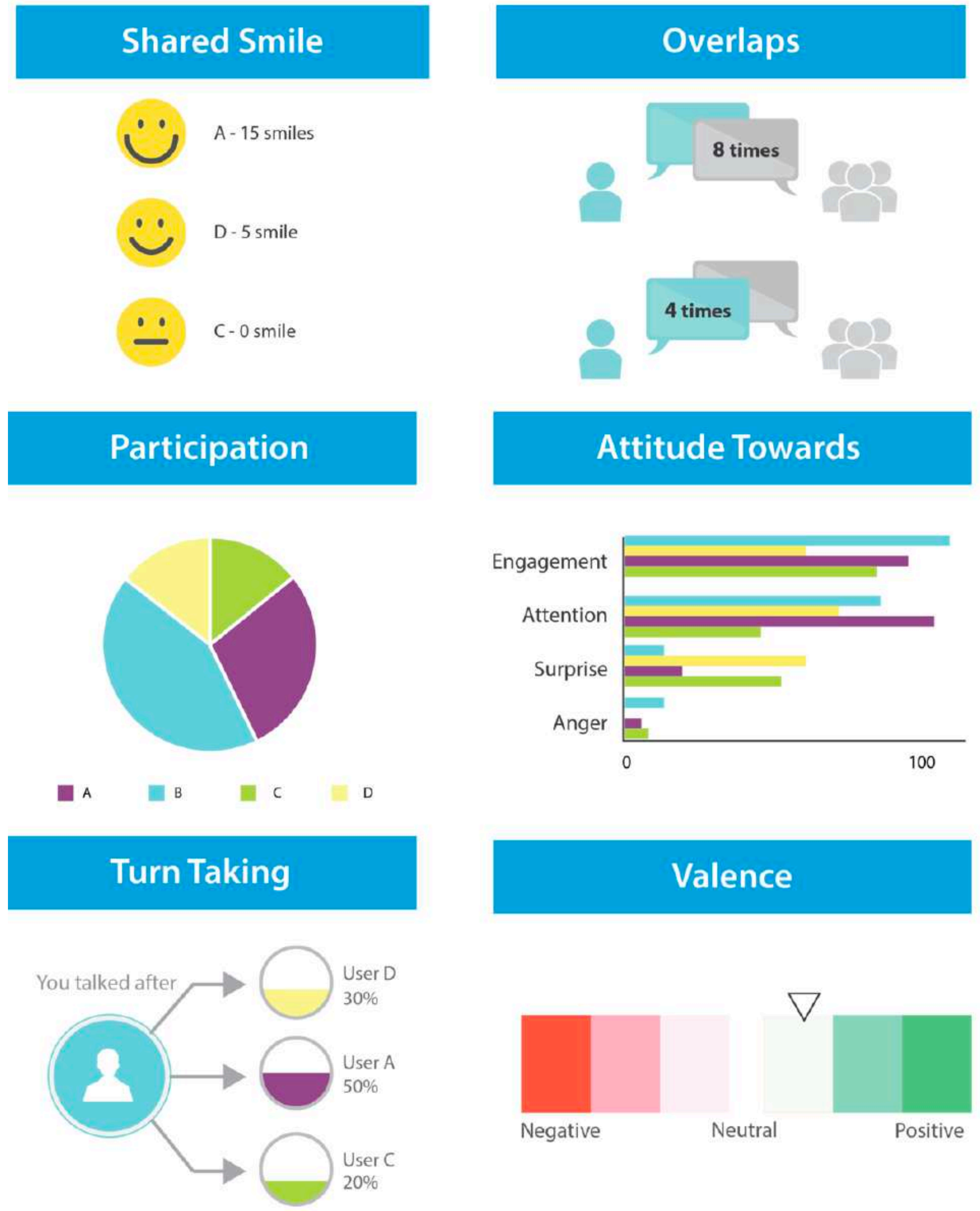
You participated **28%** of the whole session.

Expressing your ideas contributes in group decision. Would you like to know more about it?

Yeah!

No thanks

CoCo (Samrose et al., 2017)



Performance graph

Chat-based, post-conversation feedback

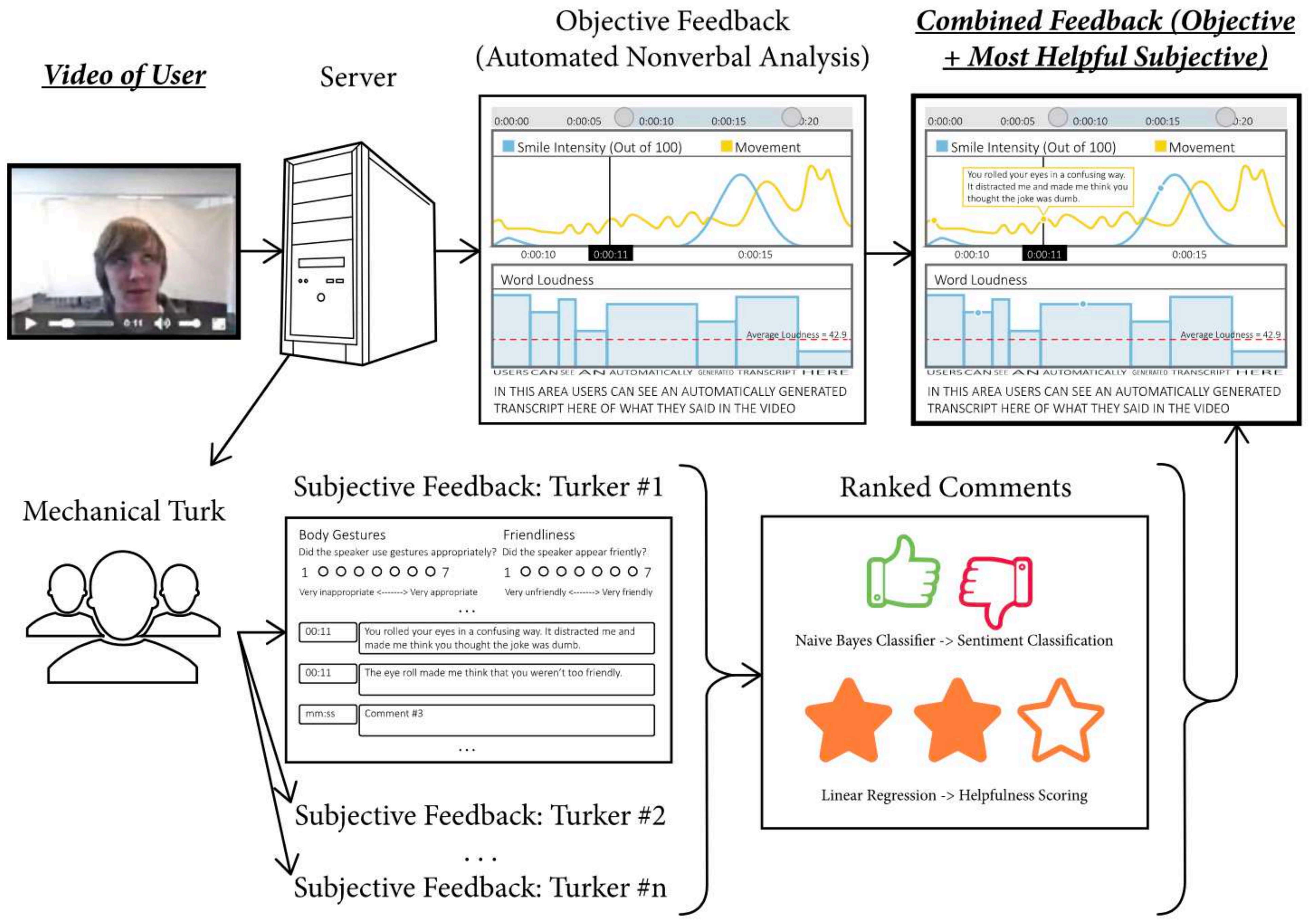
User study (with feedback)

- Balanced participation
- Skills awareness
  - How often they let others talk
  - Teammates's communicative skills

Intelligent User Interface

Posterior feedback

Comprehensive feedback from the machine and crowdsourced workers


Video of User
Server
Objective Feedback
(Automated Nonverbal Analysis)
Combined Feedback (Objective
+ Most Helpful Subjective)
0:00:00
0:00:05
0:00:10
0:00:15
0:20
Smile Intensity (Out of 100)
Movement
You rolled your eyes in a confusing way. It distracted me and made me think you thought the joke was dumb.
0:00:10
0:00:11
0:00:15
Word Loudness
Average Loudness = 42.9
USERS CAN SEE AN AUTOMATICALLY GENERATED TRANSCRIPT HERE
IN THIS AREA USERS CAN SEE AN AUTOMATICALLY GENERATED TRANSCRIPT HERE OF WHAT THEY SAID IN THE VIDEO
Mechanical Turk
Subjective Feedback: Turker #1
Body Gestures
Did the speaker use gestures appropriately?
1 ○ ○ ○ ○ ○ ○ ○ 7
Very inappropriate <-------> Very appropriate
Friendliness
Did the speaker appear frientl y?
1 ○ ○ ○ ○ ○ ○ ○ 7
Very unfriendly <-------> Very friendly
...
00:11
You rolled your eyes in a confusing way. It distracted me and made me think you thought the joke was dumb.
00:11
The eye roll made me think that you weren't too friendly.
mm:ss
Comment #3
...
Subjective Feedback: Turker #2
...
Subjective Feedback: Turker #n
Ranked Comments
Naive Bayes Classifier -> Sentiment Classification
Linear Regression -> Helpfulness Scoring

Overview of ROC Speak (Fung et al., 2015)

Motivation

Machine Consistently & objectively sense subtle human behavior

Human Interpreting contextual behavior

Gather human feedback

- score overall performance, voice modulation, friendliness, body gestures from 1 to 7

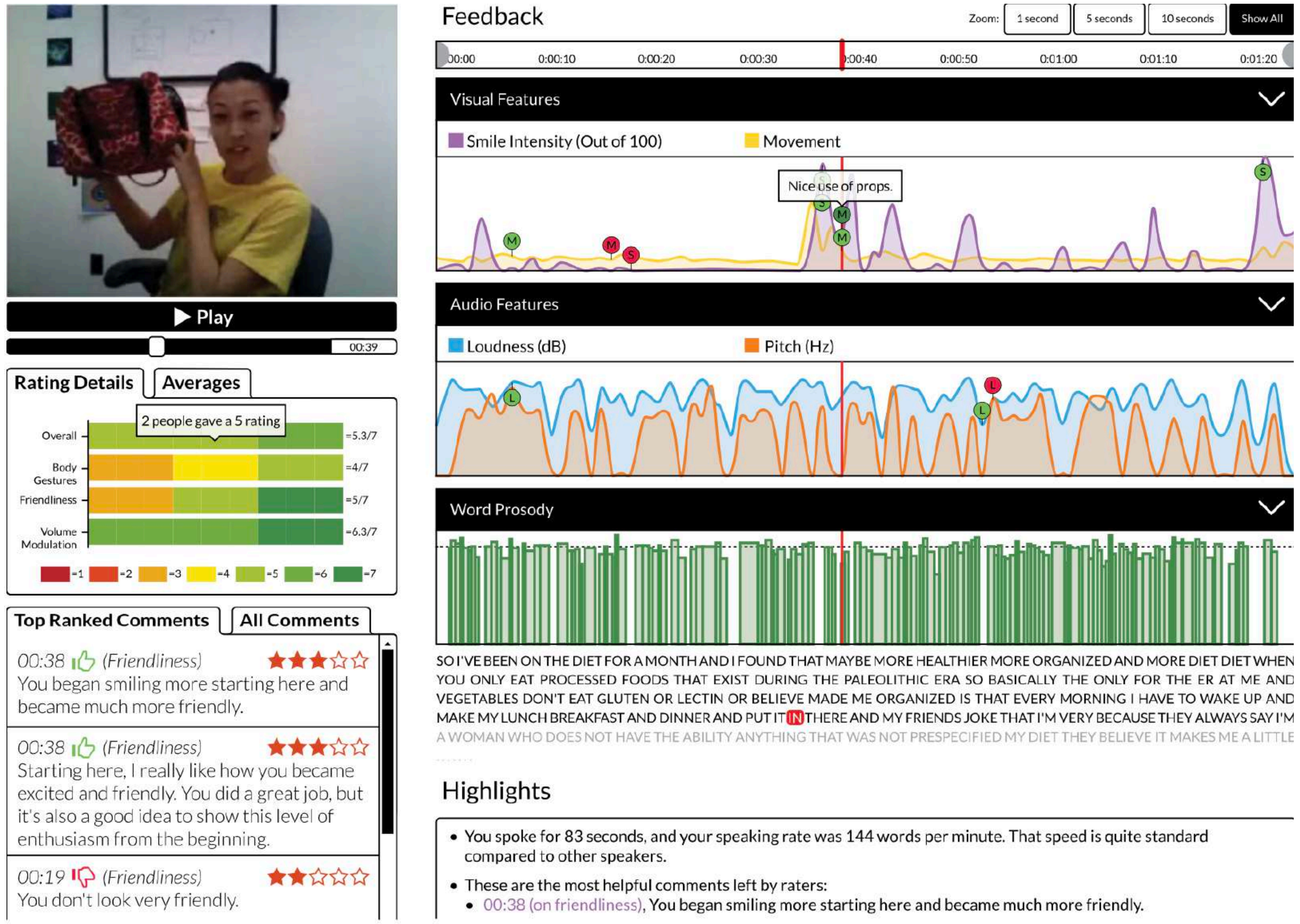
Automated ranking

- Label *helpfulness* & *sentiment*
- Train classifiers for prediction

Intelligent User Interface

Posterior feedback

Comprehensive feedback from the machine and crowdsourced workers



00:38 (Friendliness) ★★★☆☆
You began smiling more starting here and became much more friendly.

00:38 (Friendliness) ★★★☆☆
Starting here, I really like how you became excited and friendly. You did a great job, but it's also a good idea to show this level of enthusiasm from the beginning.

00:19 (Friendliness) ★★☆☆☆
You don't look very friendly.

SO I'VE BEEN ON THE DIET FOR A MONTH AND I FOUND THAT MAYBE MORE HEALTHIER MORE ORGANIZED AND MORE DIET DIET WHEN YOU ONLY EAT PROCESSED FOODS THAT EXIST DURING THE PALEOLITHIC ERA SO BASICALLY THE ONLY FOR THE ER AT ME AND VEGETABLES DON'T EAT GLUTEN OR LECTIN OR BELIEVE MADE ME ORGANIZED IS THAT EVERY MORNING I HAVE TO WAKE UP AND MAKE MY LUNCH BREAKFAST AND DINNER AND PUT IT IN THERE AND MY FRIENDS JOKE THAT I'M VERY BECAUSE THEY ALWAYS SAY I'M A WOMAN WHO DOES NOT HAVE THE ABILITY ANYTHING THAT WAS NOT PRESPECIFIED MY DIET THEY BELIEVE IT MAKES ME A LITTLE

Highlights

- You spoke for 83 seconds, and your speaking rate was 144 words per minute. That speed is quite standard compared to other speakers.
- These are the most helpful comments left by raters:
  - 00:38 (on friendliness), You began smiling more starting here and became much more friendly.

ROC Speak (Fung et al., 2015)

Overview of human feedback

Ranked comments

Quantitative visual graphs

+

Embedded human feedback

red: negative

green: positive

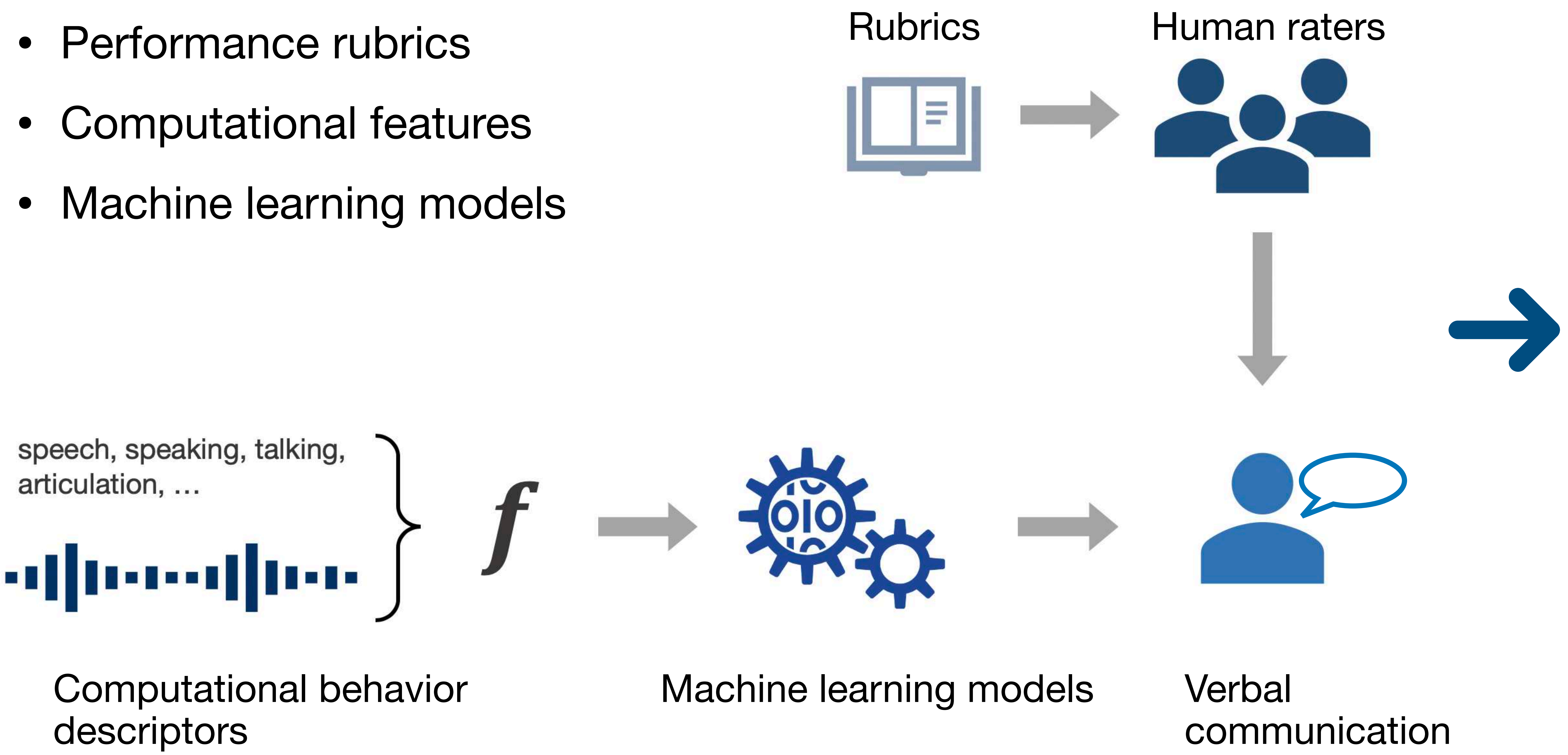
Most helpful comments for each category

Conclusion

Summary & Future work

Machine Intelligence ≫ Learning Interface

- Performance rubrics
- Computational features
- Machine learning models



Rubrics

Human raters

speech, speaking, talking, articulation, …

f

Computational behavior descriptors

Machine learning models

Verbal communication

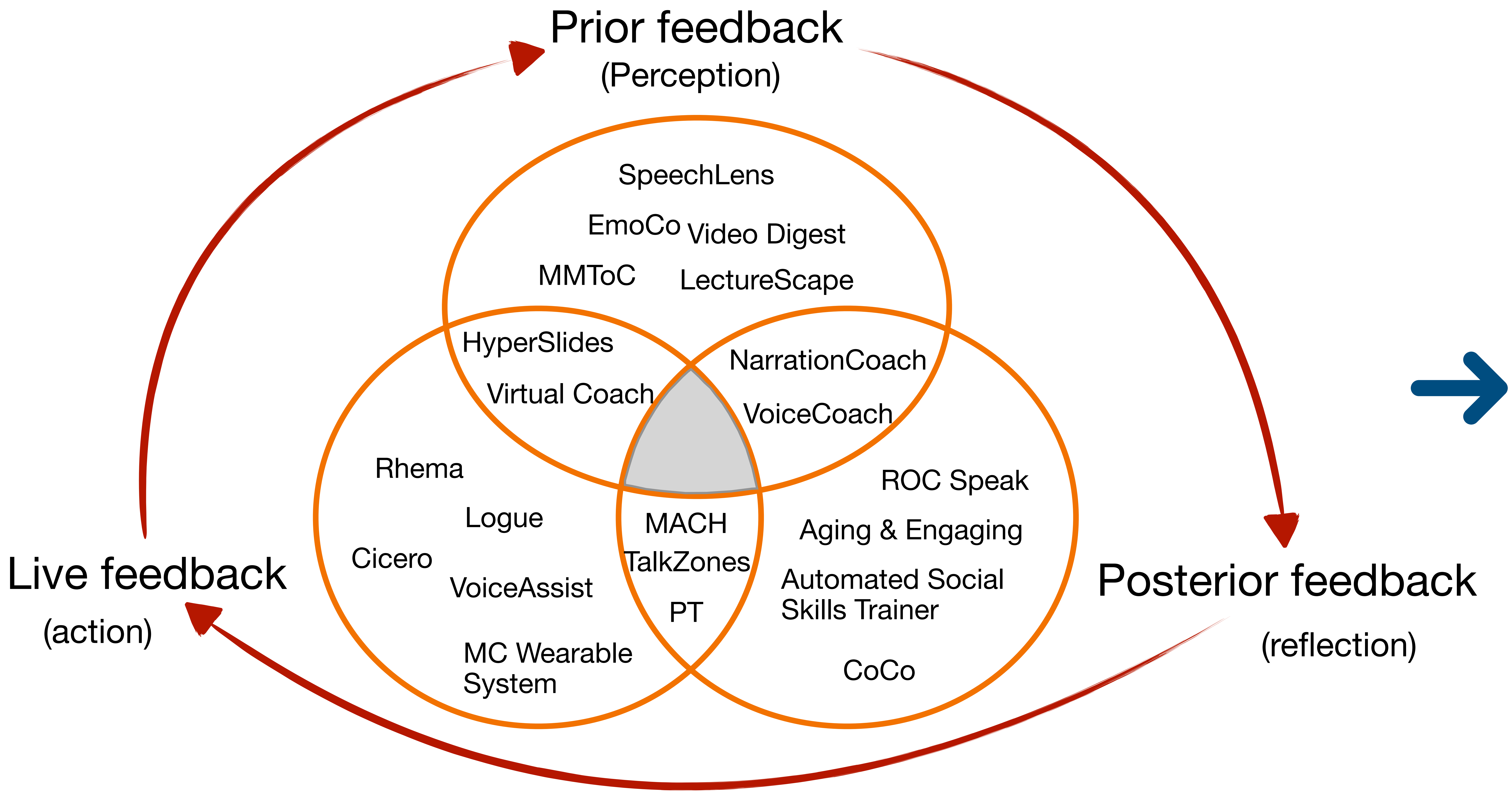
- Developing more advanced and interpretable models for verbal communication assessment

- Investigating interactions among different modalities

Conclusion

Summary & Future work

Machine Intelligence ≫ **Learning Interface**



→

- Providing comprehensive feedback at all stages of learning cycle
- Engaging users in an iterative learning process