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

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

Verbal communication skills

Adaptive speech content

Engaging vocal delivery

Public speaking

Everyday conversation

Job interview
Motivation

Background

Verbal communication skills

- Adaptive speech content
- Engaging vocal delivery

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

Verbal communication skills learning

- **Self-learning:** guidelines from books
  - No feedback

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

There is a lack of

- Quantitative & automated feedback
- Tool support for effective learning
Motivation

Quantitative automatic feedback

Sensor technology

Ambient static sensors

Wearable sensors

Natural language processing
Speech processing
Machine learning

Social computing

- Speaking rate
- Vocal variety
- Topics
- Emotion

… …
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
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
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)

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

#### Computational features

<table>
<thead>
<tr>
<th>Vocal delivery</th>
<th>Speech content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prosodic Features</strong></td>
<td><strong>Language Features</strong></td>
</tr>
<tr>
<td>tempo, loudness, pitch</td>
<td>Syntactic features (e.g., Part-of-Speech features), semantic meaning</td>
</tr>
<tr>
<td>speaking rate, liveliness, fluency</td>
<td>word choices (e.g., PMI, LIWC), topics, content features, presentation state, content structure</td>
</tr>
<tr>
<td><strong>Feature Encodings</strong></td>
<td>freq., max., min., avg., std, multi., quot., add, minus</td>
</tr>
</tbody>
</table>

**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

Computational features summarized from previous work
### Automatic Assessment

#### Computational features

<table>
<thead>
<tr>
<th>Vocal delivery</th>
<th>Speech content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosodic Features</td>
<td>Language Features</td>
</tr>
<tr>
<td>low-level Features</td>
<td>tempo, loudness, pitch</td>
</tr>
<tr>
<td>High-level Features</td>
<td>speaking rate, liveliness, fluency</td>
</tr>
<tr>
<td>Feature Encodings</td>
<td>freq., max., min., avg., std, multi., quot., add, minus</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Prosodic Features</th>
<th>Language Features</th>
<th>Other Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>low-level Features</td>
<td>tempo, loudness, pitch</td>
<td>Syntactic features (e.g., Part-of-Speech features), semantic meaning</td>
</tr>
<tr>
<td>High-level Features</td>
<td>speaking rate, liveliness, fluency</td>
<td>word choices (e.g., PMI, LIWC), topics, content features, presentation state, content structure</td>
</tr>
<tr>
<td>Feature Encodings</td>
<td>freq., max., min., avg., std, multi., quot., add, minus</td>
<td></td>
</tr>
</tbody>
</table>

- **Vocal delivery**
  - Computational features summarized from previous work

- **Speech content**
  - Speaker engagement => emotion
  - Audience engagement => stage atmosphere (e.g., laughter, applause, booing)
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

Computational behavior descriptors

\[ f \]

speech, speaking, talking, articulation, ...

Rubrics → Human raters → Machine learning models → Verbal communication

Human evaluation examples

Quality control: inter-agreement

Supervised learning models

Q1: Models \(\iff\) judgements
- SVM/SVR, L1/L2 Regularized Logistic Regression, Lasso, tree-based models (e.g., RF)
- BN, HMM

Q2: Features \(\iff\) judgements
- Correlation Coefficients

Performance is much inferior to human evaluation
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

1. Learning scenarios
   - Public speaking
   - Job interview
   - Debate
   - Voice over
   - Wedding ceremony
   - Video conferencing

2. Learning cycle
   - Assistance on perception
   - Assistance on reflection
   - Assistance on action

3. Channels of feedback
   - Visual feedback
   - Spoken feedback
   - Haptic feedback

4. Strategy of feedback
   - Prior feedback
   - Live feedback
   - Posterior feedback

Intelligent User Interface for the Learning of Verbal Communication Skills

Introduction  Automatic Assessment  User Interfaces  Conclusion
Intelligent User Interface

Taxonomy

1 Considering learning scenarios,

- Speaking anxiety
- Content organization
- Vocal delivery
- Stage management
- Group interactions & dynamics
Considering the learning process,

Perception

Learning cycle
(Dewey, 1993), (Lewin, 1946),
(Kolb, 1975), (Mumford, 1995)

Where am I going?
(learning goals)

Action

Where do I go next?
(self-adjustment)

Reflection

How am I going?
(self-awareness)
Intelligent User Interface

Taxonomy

2 Considering the learning process,

- Exploration of knowledge base
- Reflection of vocal/verbal behavior
- Putting knowledge into action
Feedback has been considered as an effective intervention in skills learning and a key consideration of learning interfaces.

Considering the channels of feedback,

- **Visual feedback**: Most widely used for feedback on speech content and vocal delivery.
- **Soken feedback**: Reduce cognitive load.
- **Haptic feedback**
- **Multimodal feedback**
Intelligent User Interface

Taxonomy

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)

Perception

Prior feedback

Posterior feedback

Live feedback

Action

Reflection
Intelligent User Interface

Taxonomy

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

- Prior feedback
  - content-based cues
  - delivery-based cues
- Live feedback
  - implicit feedback
  - explicit feedback
  - simple verification
  - elaborated feedback
  - termination
- Posterior feedback
  - summary feedback
  - focused feedback

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)

A: **context+focus** design for visualization of prosodic features

B: **Structural query**

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

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

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

---

**Recommender**

- Speech content
- Sentence structure
- TED Talk Videos
- Sentences & words
- Voice modulation

**Query:** … make an enemy

**Examples:**

1. … meet a challenge
   - softer stress
   - faster faster slower

2. … defeat an opponent

---

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

Users’ performance

- **Speed**
  - Faster ➞
  - Slower ➡

- **Volume**
  - Louder ➩
  - Slower ➩

- **Pause**
  - Brief Pause
  - Master Pause
  - Long Pause

- **Pitch**
  - Stress S
  - Normal

N-gram based hierarchical summary

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)
Intelligent User Interface

**Prior feedback:** Delivery-based cues ➢ Content-based cues

Data-driven recommendation of voice modulation techniques

Query: Tact is ... making **an enemy**

---

**Speech examples**

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

---

VoiceCoach (Wang et al., 2020)
Intelligent User Interface

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

Data-driven interaction for video navigation

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

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

MMToC (Biswas et al., 2015)
Intelligent User Interface

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

Textbook-inspired chapter/section content organization

Video Digest (Pavel et al., 2014)

Chapter: Topically coherent sections
Section: A set of varying topics

Bayesian topic segmentation

Segmentation
- Manual
- Automatic

Summarization
- Manual
- Crowdsourced

Video Digest

Crowdsourcing summary and ranking
Intelligent User Interface

Prior feedback: **Delivery-based cues** ➞ **Content-based cues**

From content exploration ➔ Hierarchical content structure planning

**Story Level**

**Scene Level**

**Detail Level**

HyperSlides (Edge et al., 2014)

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

Algorithm support

- Semantic similarity between adjacent slides ➞ presentation graph
- Time constraints, priority ➞ path suggestions
Intelligent User Interface

Prior feedback:

Limitations

• Do not consider learning from BAD examples

• Do not consider learning from multimodal speech styles
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)

**Acknowledgement**

- Nodding, changing posture
- Spoken acknowledgement: “That’s very interesting”

MACH (Hoque et al., 2013)
Intelligent User Interface

Live feedback: **Implicit feedback** ➞ **Explicit feedback**

Investigate the impact of nonverbal behavior of virtual audience

<table>
<thead>
<tr>
<th></th>
<th>Gestures</th>
<th>Facial expression</th>
<th>Pause</th>
<th>Gaze</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding</td>
<td>Narrow</td>
<td>Positive</td>
<td>Shorter</td>
<td>Friendly</td>
<td>Many polite phrases</td>
</tr>
<tr>
<td>Demanding</td>
<td>Space-taking</td>
<td>Negative</td>
<td>Longer</td>
<td>Dominant</td>
<td>Few polite phrases</td>
</tr>
</tbody>
</table>

**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

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

Live feedback: Implicit feedback ➔ Explicit feedback - simple verification

Realtime behavioural checking on speech delivery with Google Glass

Feedback icon alternatives

Logue (Hoque et al., 2013)
Intelligent User Interface

Live feedback: Implicit feedback ➞ Explicit feedback - simple verification

Realtime behavioural checking on speech content with HMD

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

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

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](image)

**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

**Aging and Engaging (Ali et al., 2018)**

- **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 you started our conversation, you spoke in a monotone, didn’t smile much, and spoke about negative topics. I gave you feedback 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 practicing varying 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 pitch of your voice, then practicing changing the pitch to vary it from time to time; practice casual conversation with people you encounter during your day and keep the conversation focused on positive topics, reminding 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 on avoiding negative topics. It’s important to remember that you can make changes in how you engage with others.

**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

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

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

**Overview of human feedback**

**Ranked comments**

**Quantitative visual graphs**

**Embedded human feedback**

red: negative

green: positive

Most helpful comments for each category

ROC Speak (Fung et al., 2015)
Conclusion

Summary & Future work

Machine Intelligence  ➔  Learning Interface

- Performance rubrics
- Computational features
- Machine learning models

Rubrics  ➔  Human raters

- Developing more advanced and interpretable models for verbal communication assessment
- Investigating interactions among different modalities

Computational behavior descriptors  ➔  Machine learning models  ➔  Verbal communication

speech, speaking, talking, articulation, …
Conclusion

Summary & Future work

Machine Intelligence ➔ Learning Interface

Prior feedback
(Perception)

• Providing comprehensive feedback at all stages of learning cycle

Live feedback
(action)

• Engaging users in an iterative learning process

Posterior feedback
(reflection)

- SpeechLens
- EmoCo
- Video Digest
- MMToc
- LectureScape
- HyperSlides
- Virtual Coach
- NarrationCoach
- VoiceCoach
- ROC Speak
- Aging & Engaging
- Automated Social Skills Trainer
- CoCo
- MACH
- TalkZones
- PT
- Rhema
- Logue
- Cicero
- VoiceAssist
- MC Wearable System