

HKUST SPD - INSTITUTIONAL REPOSITORY

Title	Designing an Artificial Agent for Cognitive Apprenticeship Learning of Elevator Pitch in Virtual Reality
Authors	Zhao, Zhenjie; Ma, Xiaojuan
Source	IEEE Transactions on Cognitive and Developmental Systems, 20 April 2021, article number 9409128
Version	Accepted Version
DOI	10.1109/tcds.2021.3073814
Publisher	IEEE
Copyright	© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This version is available at HKUST SPD - Institutional Repository (<https://repository.ust.hk>)

If it is the author's pre-published version, changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published version.

Designing an Artificial Agent for Cognitive Apprenticeship Learning of Elevator Pitch in Virtual Reality

Zhenjie Zhao and Xiaojuan Ma

Abstract—Developing social skills like elevator pitch requires being situated within authentic activities and contexts, which is difficult to achieve on a daily basis. In this work, we explore whether an artificial agent with embodied feedback in virtual reality (VR) can foster a situated learning experience. Previous works on computer-mediated feedback have shown that VR can foster oral presentation competence for pre-university and junior undergraduates through delivering feedback. However, it is unclear how well the learning experiences are and how well students perceive an artificial coach in VR, especially for senior undergraduates and postgraduates seeking job and research opportunities. To inform the design of such a system, we conducted interviews with experts and observed real elevator pitches. We then designed a proof-of-concept VR coaching system with three embodied feedback strategies: immediate, after-action, and the combination of both. Through a between-subject experiment with 40 participants, we studied learners' perceptions under the embodied feedback. We found that receiving embodied feedback can create a stronger sense of cognitive apprenticeship, *i.e.*, coaching and helping from experts, and help improve the perception of the artificial agent and the effect of learning. We further investigated the pros and cons of different strategies and discussed room for improvement.

Index Terms—Oral presentation skill development, cognitive apprenticeship, artificial agent, situated learning, embodied feedback, perception, user study, virtual reality

I. INTRODUCTION

Elevator pitch or elevator speech, *i.e.*, delivering a quick persuasive pitch to arouse interest in an idea or a product, is an important social skill that can potentially bring opportunities for jobs, collaborations, and investments [1]. Developing and mastering such a skill, however, is considered very challenging, as elevator pitch requires one to deliver or convey an idea within a short amount of time (usually 30 to 120 seconds) and often in an unfamiliar environment [2]. Many online tutorials have discussed how to perform a successful elevator pitch, and the last step is always “practice makes perfect”. However, practicing elevator pitch on one’s own is often unproductive

and tedious [3], because of the lack of constructive feedback and a convincing environment [4].

Situated learning is an instructional approach featuring learning in a real or simulated environment [5, 6, 7]. The focus on authentic contexts and convincing learning environment makes it an ideal approach for practicing elevator pitch. Cognitive apprenticeship (CA) is one type of situated learning strategies [6]. Different from other methods, CA expects the learning activity to be conducted with a professional coach or an expert [6]. Its core process – making expert thought “visible” to learners through *feedback* – is shown as an effective approach to improve one’s social skills [8], for example, elevator pitch skills. As a formative assessment strategy, feedback is considered as one of the main design principles for improving oral presentation competence [9]. However, in practice, it is difficult, if not impossible, for one to find a professional elevator pitch coach to guide repeated practices on demand.

The advances in virtual reality (VR) makes it possible to realize CA without the need of a real professional. Although previous research has incorporated situated learning in various domains such as medical training [10], literature education [11], public speech [12], and job interview training [13], they either did not adopt the concept of CA or did not use VR for an immersive experience. Virtual characters in those systems are often regarded as part of the situated environment that users interact with (not an advisor or expert coach) [14] and feedback in these training systems is limited to symbolical form [15] (*e.g.*, information tables or graphical visualizations). There was no embodied feedback – verbal or non-verbal feedback directly given by virtual characters to the users, which usually will be given if a real human coach is situated.

Although there have been general design guidelines of how to apply CA in learning activities in reality [16], little is known regarding how it can be applied to a VR-based social skill learning system, and how learners would perceive and perform with a virtual coach’s embodied feedback. In this paper, we propose to apply CA in VR, where an artificial agent can play the role of an expert coach while the virtual environment can simulate a real-life situation [17]. As the first step towards a ready-to-use system, our focus of this paper is to investigate how to design and offer embodied feedback via an artificial agent, *i.e.*, a virtual coach in VR, to improve a learner’s elevator pitch skills.

We first conducted semi-structured interviews with domain experts from the local language and career centre to derive key

Manuscript received October 5, 2020; revised December 2, 2020; accepted April 13, 2021. This work is supported by the Hong Kong General Research Fund (GRF) with grant No. 16204819. Zhenjie Zhao is supported by the Startup Foundation for Introducing Talent of NUIST. (*Corresponding author: Zhenjie Zhao.*)

Z. Zhao is with the Department of Computer Science and Technology, Nanjing University of Information Science and Technology, Nanjing, Jiangsu Province, 210044, China (e-mail: zzhaao@nuist.edu.cn).

X. Ma is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China (email: mxj@cse.ust.hk).

elements of a good elevator pitch, and the general strategies to give learner feedback. We then observed real elevator pitches at various occasions to supplement the interviews regarding to how to design the embodied behaviors of a virtual coach. These findings guided the design of an immersive VR coaching experimenting system. In terms of how to provide users feedback, we consider the dimensions of intensity and timing of feedback, which is also correlated with the design principles given by Van Ginkel *et al.* [9]. Our focus in this paper is to investigate how people perceive embodied feedback in VR. For intensity, we used an artificial agent in VR only with listening behaviors as a baseline, which can be seen as equipping an agent with weak feedback. We then compared the baseline with agents with explicit embodied feedback. For timing, based on our interviews, we first divided all mistakes that users may make during presentation into two groups: intermittent mistakes that occur at irregular intervals (*e.g.*, eye contact) and continuous mistakes that occur from the beginning to the end (*e.g.*, the use of time). We then designed three potential embodied feedback strategies: immediate feedback, after-action feedback, and the combination of both. The immediate strategy gives feedback whenever users make a mistake during the practice. The after-action strategy only makes suggestions after users complete their presentation. The combination strategy prompts immediate feedback upon intermittent mistakes on the fly, and summarizes continuous ones after the practice. The three strategies only differ on the time of giving feedback and the amount of feedback is the same.

We conducted an empirical, between-subject design study with 40 participants on this proof-of-concept system to assess how virtual character's embodied feedback can help improve a learner's elevator pitch skills. Results show that embodied feedback on learning activities creates a stronger sense of CA – participants agree that they are being coached and helped by the virtual character. It also improves users' perception of the virtual character – participants feel the virtual character is trustworthy and have empathy for the user. Moreover, the embodied feedback also plays an important role on the effect of learning, including increasing users' self-confidence, satisfaction of their elevator pitch and awareness of performance, decreasing their anxiety, and improving users' performance. We further investigated the underlying reasons, compare the pros and cons of each embodied feedback strategy, and identified room to improve through in-depth post-study interviews.

The contributions of this paper are: 1) with a design process, we have designed and implemented a VR coaching prototype with an artificial agent for developing elevator pitch skills; 2) we carefully conducted a user study with 40 participants to investigate the pros and cons of different embodied feedback strategies of an artificial agent quantitatively, and discussed design insights qualitatively for future system development in this field; 3) through the system design and user study, we show the potential of simulating an artificial agent in VR for soft skill development. An abstract and non-archival version of this paper is published as a poster at ACM Symposium on User Interface Software and Technology (UIST) 2020. We extended the abstract with more reviews on related works, details of the system design, and more about the user study and in-depth

data analysis.

II. BACKGROUND AND RELATED WORK

A. Elevator Pitch

An elevator pitch, also known as elevator speech or elevator statement, is a short description (30-120 seconds) of an idea, plan or concept that a listener can understand in a short period of time [1]. A good elevator pitch may help jobseekers convince potential employers that they are qualified for a particular job. Likewise, students may impress professors who are looking for interns or postdocs in the same manner at the coffee break of a conference. That is to say, an elevator pitch can serve as a “foot in the door” with key stakeholders and can lead to further opportunities [3]. There are many online guidelines summarizing how to conduct a good elevator pitch [18, 19, 20, 21], for example, “communicate your unique selling proposition”, “engage with a question”, or “be persuasive”. However, learning and developing elevator pitch skills is not all about memorizing those key concepts. It is a social skill that people will have to eventually exercise in front of others in real environments [1, 17]. It is thus recommended to practice elevator pitch with a partner, so that learners can get used to the scenario and feel more comfortable whenever they need to conduct it.

B. Situated Learning and Cognitive Apprenticeship

Situated learning that emphasizes social co-participation during the learning activity, which allows learners to experience a real scenario is one promising approach to learn elevator pitch skills [22]. Among existing situated learning strategies, cognitive apprenticeship that stresses learning through collaborative social interaction and the social construction of knowledge by a professional coach seems to be a particularly important approach to mastering elevator pitch [6, 8]. Cognitive apprenticeship usually contains six phases: modeling, coaching, scaffolding, articulation, reflection, and exploration [23]. Modeling refers to a conceptual model of what to learn. Coaching refers to the procedure of helping learners acquire knowledge and develop of a specific model, and is the basic phase of cognitive apprenticeship. Scaffolding refers to the development and stabilization of cognitive skills. Articulation refers to articulating thought processes verbally, *i.e.*, “thinking aloud”. Reflection refers to evaluating thought processes oneself. Exploration refers to exploring and solving new tasks. To acquire social skills like elevator pitch through the cognitive apprenticeship model of situated learning, coaching is the most essential part of the entire experience [3]. Effective coaching can correct learners' mistakes in a timely manner, accustoming them to how to behave in that particular situation [1, 17]. Therefore, as a first step of exploring how to design a VR system for practicing elevator pitch, we are mainly interested in coaching in this paper.

C. Feedback and Social Signal

Coaching emphasizes reaching learners' goals of knowledge acquisition [23]. One of the most important features of a

coaching system is feedback [24]. There are several displaying ways to give feedback, such as graphical visualization and information tables [15, 14], sound volume and rate [25]. However, previous works usually consider symbolically visual feedback that is far from a real human coaching experience. It is therefore hard to know how to design embodied feedback, *i.e.*, verbal or non-verbal feedback, directly given by a coach to the users, and how people will perceive it. Meanwhile, prior research works show that virtual signals that are expressed in certain social norms can bring the same feedback effects as real social signals [26, 27, 28], which demonstrates the potential of using embodied virtual signals as a feedback form. In this paper, we consider embodied feedback that a virtual coach uses to communicate his/her intentions.

D. Social VR and Its Applications

Social VR technology, namely, simulating a 3-dimensional (3D) interactive virtual character in VR is a promising way to implement cognitive apprenticeship in VR, where learners can perceive the social presence of the virtual character and have a real elevator pitch scenario with the character [29]. Several previous systems have tried to use VR to enable an immersive training environment. For example, in [29], several virtual patients are simulated to help doctors practice their negotiation abilities. In [30], a virtual character is simulated to help people overcome social interaction phobias. There are also works on public presentation training [31, 32]. But their systems are not in VR contexts that may be more affordable for general users and the presence feeling brought by VR systems can also better transfer skills learned in a virtual world to the real world [33, 29, 30]. Moreover, elevator pitch is usually conducted with only one audience (mostly an expert) and requires the presenter to pay close attention on the reaction of the audience.

III. RESEARCH QUESTIONS

Informed by CA, we explore how to design embodied feedback of an artificial agent or a virtual coach to support learning an elevator pitch in an immersive VR environment. We first bring up a straightforward research question regarding to embodied feedback of a virtual coach for cognitive apprenticeship learning of elevator pitches in VR.

- **RQ.1:** Will people perceive embodied feedback of a virtual coach in VR as a CA experience?

Presenting with a virtual coach is different from a real person [34, 35, 36], it is therefore critical to further investigate how people perceive the virtual character, such as trustworthy and empathetic. In particular, we have the following additional research question.

- **RQ.2:** With embodied feedback in an interactive VR coaching system, how will users perceive the virtual character?

Finally, we investigate how people feel the training effects in a VR context. Our research goal here is not on how to evaluate participants' performance. Instead, we focus more on users' perception of embodied feedback in VR and do not involve experts' judgement.

- **RQ.3:** How effective will it be for users to learn in our VR environment?

Previous works have shown that that feedback in VR can improve oral presentation competence for pre-university [37] and junior undergraduate students [38] through delivering feedback. Considering the influence of student perceptions of learning on their learning performance [39, 40, 41], we also investigated learning performance to verify the validation of the perception study.

IV. DESIGN PROCESS

We follow a design research process [42], which involves performing initial exploratory qualitative data collection by conducting semi-structure interviews with experts and observing real elevator pitches. Informed by these insights, we investigate how to develop and test the VR system.

A. Preliminary Qualitative Data Collection

To inform how to conduct a good elevator pitch and how to design an interactive VR system for practicing elevator pitches, we first interviewed four domain experts for deriving useful codes. Based on the codes, we further observed and videotaped real elevator pitches to infer how to quantify the performance of an elevator pitch and to design appropriate behaviors of a virtual coach.

1) *Interviewing Experts:* We conducted semi-structure interviews with four experts (E.1-E.4) from the career centre and language centre in a local university. The experts had experience in teaching spoken English and public presentations for at least five years. E.1 is experienced in evaluation of teaching and learning, literary studies and discourse analysis, and development of materials for teaching English as a foreign language. E.2 is experienced in coaching English skills, presentation skills and communication skills to senior management in industry, and is also a native English speaker. E.3 is experienced in second language acquisition, language curriculum and pedagogy, cross-cultural pragmatics, testing and assessment, and business communication. E.4 is experienced in English language learning strategies and English language teaching methodology. Each interview was conducted on-site and individually, and lasted for about half to one hour. For each interview, we asked:

- What is a good elevator pitch?
- How to practice an elevator pitch?
- If they had suggestions for designing an interactive virtual character in VR to coach elevator pitch skills.

Finding experts to interview is relatively challenging. Although the number of interviewed experts is small, we are able to identify a set of key points shared among the experts.

2) *Observing Real Elevator Pitches:* To further verify and supplement how to design the VR coaching system, especially behaviors of a virtual character, we attended workshops and seminars to observe real elevator pitches. In total, we attended four public talks, two big exhibitions, and one book club activity. Speakers in these activities include novelists, engineers, experts in machine learning and computer vision. Attendees are mainly students. We had two paper authors

to observe and videotape any potential elevator pitch. We obtained consent from all the involved participants in the videos. The observers also took pictures and notes to record the main speech contents. In total, we collected 154 videos (lasting from 30 seconds to 3 minutes) and 235 pictures, from which we extracted about 30 videos of typical elevator pitches with transcripts. The context of each video is different, as we aim to capture general principles of elevator pitches instead of a specific context.

B. Preliminary Qualitative Findings

We analyzed interview transcripts and real elevator pitch materials using inductive coding methods [43]. From this initial data collection, we identified the basic components of an elevator pitch, and the specific requirements for designing a coaching system to help people improve such skill.

1) *Quantify the Performance of an Elevator Pitch*: Through expert interviews, we summarized five principles of how to conduct an elevator pitch (EP). Each point is mentioned by at least two experts.

- **EP.1**: Tell a story about yourself;
- **EP.2**: Make a connection with the listeners by demonstrating why you and your story are important to them;
- **EP.3**: Speak clearly with proper rhythm and volume;
- **EP.4**: Use nonverbal cues properly such as hand gestures and eye contact;
- **EP.5**: Pay attention to the time and finish the speech in about 120 seconds.

Van Ginkel *et al.* summarized excellent rubrics aimed to foster oral presentation skills in terms of content of the presentation, structure of the presentation, interaction with the audience, and presentation delivery [44]. Our summarized rubrics mainly focus on the aspect of interaction with the audience and presentation delivery in Van Ginkel *et al.*'s taxonomy. In particular, **EP.1** and **EP.2** concern about the speech content and structure, which can be carefully prepared and polished before conducting the speech. Meanwhile, they are also not suitable for embodied feedback, otherwise it will cost too much cognitive workload for learners, *i.e.*, learners may not be able to understand and learn the instructional feedback in a short period of time. It is more critical, as stated by **EP.3**, **EP.4** and **EP.5**, to emphasize issues of improvisation ability during an elevator pitch, where embodied feedback in VR can play an role on users' learning experience. As a first step of exploring how to design a VR coaching system, in this paper, we mainly focus on **EP.3**, **EP.4**, and **EP.5**.

We quantify the performance of an elevator pitch in a way that it can be measured by existing sensors such as Kinect¹. Although there may be some kinds of mistakes in an elevator pitch that cannot be identified by our system, the types of issues covered by our system are generally enough for the purpose of exploring users' perception on virtual characters' embodied feedback in VR. More specifically, we first discuss as many as possible existing off-the-shelf sensors that can measure **EP.3-5**, and then derive specific parameters by analyzing the recorded videos and transcripts. We also verify proposals

from existing literatures [45, 13, 46, 47]. After several regular group meetings and discussions, we summarized five main metrics, as listed in Table I. The detailed implementation is presented in the system design section.

Name	Description	Sensor
Eye contact	Users should look at the character's face frequently, with no more than 2 seconds without eye contact	Inertial measurement unit (IMU)
Hand gesture	Users should use hand gestures frequently, with no more than 2 seconds without a hand gesture	Kinect
Rhythm	Users should pay attention to their speaking speed, and maintain about 120~180 words per minute	Speech recognizer
Volume	Users should pay attention to the volume of their voice, keep it at about 50 dB, neither too loud nor too low	Microphone
Timing	Users should finish an elevator pitch in about 120 seconds	Time clock

TABLE I
A COMPILED RULE LIST FOR QUANTIFYING AN ELEVATOR PITCH.

2) *Behaviors as a Recipient*: As E.2 stated, it is important for a coach to investigate what behaviors and feedback would be more proper to enhance speakers' self-awareness of their speaking performance. This matches our motivation by investigating the perception of embodied feedback for elevator pitch skills training in VR. Although the behaviors as an elevator pitch recipient in the recorded videos and pictures vary from individual to individual, there are general talking and listening behavior patterns of a recipient. We derived such general behaviors by analyzing existing literatures [48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60]. We then customized those behaviors for elevator pitches through verifying their usage empirically in the recorded videos and pictures. After several regular group meetings and discussions, we summarized verbal and non-verbal behaviors of a recipient during talking and listening, as shown in Table II. Talking behaviors can be used to design our embodied feedback, and listening behaviors are used universally to the listeners during an elevator pitch.

	Talking	Listening
Verbal	Communicate information	Umm, OK, Yes, Cool, ...
Nonverbal	Hand gesture, Smile, Eye contact, Body movement	Smile, Eye contact, Nod

TABLE II
TYPICAL TALKING AND LISTENING BEHAVIORS DERIVED FROM OUR RECORDED VIDEOS AND PICTURES.

3) *Situated Embodied Feedback from a Coach*: Our interviews show that all experts confirm the importance of feedback for improving elevator pitches but have different opinions on

¹<https://developer.microsoft.com/en-us/windows/kinect>

the exact time to give it. To be specific, experts want a virtual coaching system embedded with the expert experiences for users to more easily correct their mistakes. However, in terms of the exact time to give feedback, options vary according to the experts. E.2 thought that mistakes should be summarized and presented at the end for speakers to have time to understand, while E.3 and E.4 stated feedback should be immediate to correct mistakes in a timely manner. We consolidated these opinions into three feedback strategies based on the timing when designing a virtual coach: **immediate**, **after-action**, and the **combination** of both. The detailed behaviors are presented in section IV-C3.

C. System Design

Based on the initial exploratory study of elevator pitches, we designed a VR coaching prototype. The system consists of four components: sensing, ranking, feedback, and VR display. With the sensing module, we first detect the learner's mistakes during elevator pitch practice, and then push these mistakes into a queue asynchronously. Afterwards, the ranking module selects a highest priority mistake and sends to the feedback strategy module. With a pre-set feedback strategy and the mistake, we generate corresponding embodied behaviors, and display them in VR. The system architecture is shown in Figure 1. We implement it using Unity game engine ².

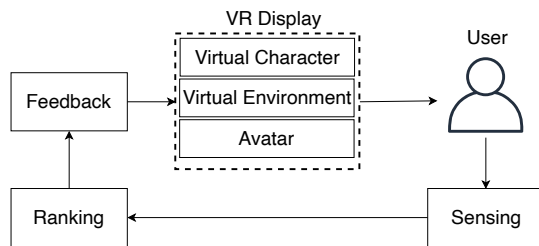


Fig. 1. System architecture of our VR coaching prototype.

1) *Sensing Elevator Pitch Mistakes*: Referencing to Table I, we detect five common mistakes using corresponding sensors.

Eye contact. We use the inertial measurement unit (IMU) in an Oculus Rift VR headset ³ to approximate eye contact detection. In particular, we use the `Physics.Raycast` API to detect whether a ray casted from the user's head camera hits the bounding box of the virtual coach's head. If the ray does not hit the bounding box, it is likely that the user does not look at the virtual coach and we then send an eye contact mistake.

Hand gesture. We use a Kinect for tracking the use of hand gestures with Unity SDK of Kinect ⁴. We detect the variation of hand gestures as a binary signal. If the displacement of the wrist is bigger than 10 centimeters, or a change of angle of the elbow is bigger than 50 degrees, we count it as being a

hand gesture. If there is no hand gestures for each time step, we send a hand gesture mistake.

Rhythm. For the rhythm of audio signals, we use the Google speech recognizer service ⁵ to transform audio signal to text input, and then calculate words per second (WPS) to indicate the rhythm of speaking. If WPS is above 5, we send a "too fast" mistake, and if WPS is below 1, we send "too slow".

Volume. At each time step, we find the peak value in a 128 size window of microphone inputs to measure the volume of audio signals. If the value is above 1, we send a "too loud" mistake, and if it is below 0.001, we send a "too quiet" mistake.

Timing. We count the time elapse when the user starts to talk. If it is above 120 seconds, we send a "too long" mistake. If users explicitly say that they finish the talk and the time elapse is below 70 seconds, we send a "too short" mistake.

2) *Ranking Mistakes to Select One for Generating Feedback*: Due to all sensing modules working asynchronously, within a short time window, it is better to rank the detected mistakes and select one that has the highest priority to trigger the current embodied feedback. For simplicity, we rank the mistakes by considering: 1). the accumulated number of each mistake category $number_{mistake}$, 2). the time elapse of each mistake category $elapse_{mistake}$, and 3) its pre-set priority $priority_{mistake}$ (eye contact, rhythm, and "too quiet" have high priorities, and the remaining ones have low priorities). For each time step, if one mistake m is detected the most frequently, lasts the longest time, and has higher priority, we select it. We then decide whether to trigger corresponding feedback behaviors of m empirically by:

if $number_m > 3$ and $elapse_m > 2$ then
trigger feedback behaviors of m

It is possible to further investigate better ranking and decision algorithms to present more intelligent behaviors. However, for simplicity and the study goal of this paper, we instead design a wizard-of-oz interface to bypass this challenge, as shown in Section **Wizard-of-oz User Interface**.

3) *Embodied Feedback and Feedback Strategy for Cognitive Apprenticeship Learning*: After regular group meetings and discussions, we summarized a codebook that records a set of verbal and non-verbal feedback behaviors for each elevator speech mistake. For example, a verbal signal "hand gesture" corresponds to the hand gesture mistake. We then implemented the three feedback strategies according to the codebook. In particular, based on expert interviews, we first divided all mistakes that users may make during presentation into two groups: 1) intermittent ones that occur at irregular intervals, including eye contact, rhythm, and no speech (volume); 2) continuous ones that always occur, including all the remaining mistakes. We set the priorities of intermittent mistakes as high because they are more appropriate for immediate feedback, and set the priorities of the continuous ones as low because they are more appropriate for delayed feedback [9]. Instead of evaluating and comparing these three strategies in details, we use them only for ensuring that appropriate embodied feedbacks are used in our study because we cannot conclude which

²www.unity.com

³https://www.oculus.com/rift/#oui-csl-rift-games=robo-recall

⁴www://assetstore.unity.com/packages/3d/characters/kinect-v2-examples-with-ms-sdk-and-nitrack-sdk-18708

⁵www://cloud.google.com/speech-to-text

one is the best from our qualitative study. For the immediate strategy, during the speech, it generates embodied feedback whenever users make a mistake, including nonverbal behaviors like weaving hands to attract users and verbal feedback like “look at me”. After the speech, the virtual character will say goodbye directly without any further comments. For the after-action strategy, during the speech, it simulates the normal reactions of a person, and only gives embodied suggestions (verbal and nonverbal) after users finish their presentation. The combination strategy hybridizes the immediate and after-action strategies. Specifically, during the speech, it only gives users immediate feedback for the intermittent mistakes, and summarizes continuous ones after the speech. We summarized the embodied behavior templates of different strategies in Table III. The virtual character will always have listening behaviors. An artificial agent with only listening behaviors is used as a baseline, which can be seen as equipping an agent with weak feedback. Our considerations on timing and intensity are also correlated with the design principles given by Van Ginkel *et al.* [9].







	Immediate	Combination	After-action
During the speech	 <p>“Look at me!” “Hand gesture!” “Come on!” “Faster!” “Slow down!”</p>	 <p>“Look at me!” “Come on!” ...</p>	 <p>“OK.” “Yes.” ...</p>
After the speech	 <p>“Remember what I said, and practice more.” “See you next time.” ...</p>	 <p>“OK That was pretty good but there are a few things we can work on. First, your <mistake> is <description>. You should <action>. Second, ...” ...</p>	 <p>“OK That was pretty good but there are a few things we can work on. First, your <mistake> is <description>. You should <action>. Second, ...” ...</p>

TABLE III

EMBODIED BEHAVIOR TEMPLATES OF DIFFERENT FEEDBACK STRATEGIES, WHERE *< mistake >* DENOTES A SPECIFIC MISTAKE IN TABLE I, *< description >* DESCRIBES THE WRONG BEHAVIOR OF THIS MISTAKE, AND *< action >* DENOTES HOW TO CORRECT THE MISTAKE.

4) *Displaying VR Content to Users:* We use Autodesk Character Generator⁶ to design virtual characters, and modify a free Unity asset⁷ to build the virtual environment. We use the SALSA package⁸ to transform text to speech.

a) *Virtual Coach:* An elevator pitch usually occurs in a more causal and relaxed atmosphere, and the virtual character should not give people too much pressure like job interviewers. We follow this principle and designed a female and male character called *Jane* and *David* separately to eliminate gender bias during the user evaluation [13]. In other words, male participants will interact with David and female participants will interact with Jane. We keep the characters smiling slightly to make them look friendly. The appearances are shown in Figure 2. We had regular meetings and designed the behaviors

of the characters iteratively. Referring to the collected videos and pictures, we designed more casual and friendly body language for the characters. One of the authors of this paper acted out those behaviors, and we used a Kinect sensor to record skeleton motions and then fine-tuned key-frames manually to make the behaviors look more natural.



Fig. 2. The appearances of the virtual coaches *Jane* (left) and *David* (right).

b) *Avatar:* To increase physical and social presence, we also needed to provide users virtual avatars of themselves in the virtual world [17]. Since the users cannot see their own faces in the virtual world, we did not worry too much about it and designed one boy and one girl separately to represent themselves. We used a Kinect to track body movements of users, and in the virtual world, they will see their own two hands.

c) *Environment:* We adopt a common hall environment where an elevator pitch is most likely to happen. The hall looks like a campus building. To bring the real elevator pitch experience, at the beginning, the virtual coach will go from outside to inside and meet the user in the front. The virtual character will then take on the coach role to let the user start an elevator pitch and provide feedback during the process.

D. Wizard-of-oz User Interface

The simple ranking and decision algorithms are not sophisticated enough to allow the virtual character to behave intelligently. We also designed an user interface that allows a tele-operator to control the behaviors of the virtual character by triggering buttons.

The user interface is shown on the right-hand side of Figure 3. In particular, the detected mistake with detailed information will be visualized in the left-hand side panel (b). In the right-hand side, users can prompt different feedback behavior panels (d), and listening behavior panels (e). When pressing a particular button of panels (d), the actual feedback depends on the current feedback mode. Listening behaviors are independent on the feedback mode. Meanwhile, in the left-hand side of Figure 3, users can select different feedback strategies, characters, and stages. The actual behaviors depend on those meta-parameters.

⁶[www://charactergenerator.autodesk.com](http://www.charactergenerator.autodesk.com)

⁷[www://assetstore.unity.com/packages/3d/environments/snapshots-prototype-office-137490](http://www.assetstore.unity.com/packages/3d/environments/snapshots-prototype-office-137490)

⁸<https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442>



Fig. 3. The user interface of our system. (a): the panel for selecting feedback strategy, character, and stage; (b): displaying mistakes, the operator can deduct a mistake by pressing the minus button; (c): start and end buttons; (d): immediate feedback buttons; (e): listening behavior buttons.

V. HYPOTHESES AND MEASUREMENTS

For RQ.1, we investigate embodied feedback of an interactive virtual character in VR. In particular, we consider the influence of embodied feedback on CA, as well as the perception of the virtual character and the effect of learning in VR. We have the following hypotheses:

- **H1:** A virtual character with embodied feedback will convey to users a stronger sense of cognitive apprenticeship than the one without.

We follow [16] to measure the *coaching* stage of CA, and ask participants to rate their experience on a 7-point Likert scale: 1) Feedback while observing: “Jane/David provided feedback as he/she observed me independently perform my elevator pitch”; 2) Help: “When I made an error conducting my elevator pitch, Jane/David provided hints/reminders that helped me complete the task”; 3) Coach: “When I had difficulties during conducting my elevator pitch, Jane/David was able to coach me through completing the task”; 4) Feedback for improvement: “Jane/David provided feedback for how I could improve my elevator pitch”; 5) Adaptation: “The more I increased my ability to conduct an elevator pitch, the less feedback I received from Jane/David”. Although the difference between question 1,2,3,4 are subtle, they have different focuses. In particular, question 1) emphasizes the availability of feedback, question 2) emphasizes the usefulness of the artificial agent, question 3) emphasizes the perceived capability of the artificial agent, and question 4) emphasizes the approach of the feedback provided by the artificial agent. Question 5) is an important element to verify the adaptability of the artificial agent.

Apart from CA, understanding how people perceive the virtual character also plays a vital role in deriving potential design implications [61]. For RQ.2, we hypothesize that:

- **H2:** Users find a virtual character with embodied feedback more trustworthy and empathetic, like it more, and treat it more like a coach, compared to the one without such feedback.

We measure the perception of the virtual character by asking [29]: 1). *Trust*: “I consider Jane/David as people who can

be trusted”; 2). *Like*: “Do you like Jane/David”; 3). *Role*: “To what extent can you accept Jane/David as a coach”; 4). *Empathy*: “How well do you think Jane/David shows empathy”, on a 7-point Likert scale.

Finally, for RQ.3, because of the immersive experience brought by VR, we hypothesize that our system will help users learn better with increasing use.

- **H3:** Virtual characters (with or without embodied feedback) in VR will gradually enhance users’ self-feeling with each round of practice. Moreover, virtual characters with embodied feedback will improve users’ performance of elevator pitches.

We measure the effect of learning in each round of practice by asking [4]: 1). *Confidence*: “Rate your confidence on your elevator pitch just now”; 2). *Satisfaction*: “Rate your satisfaction on your elevator pitch just now”; 3). *Anxiety*: “Rate your anxiety of your elevator pitch just now”; 4). *Self-awareness of performance*: “Rate your elevator pitch performance just now”, on a 7-point Likert scale. We also analyzed the logged performance quantitatively.

Apart from the quantitative study, we conducted in-depth interviews to understand the perception of different embodied feedback strategies and compare their advantages and disadvantages qualitatively.

VI. USER EVALUATION

We conducted a between-subject user study⁹ on our VR coaching system to explore the three hypotheses raised in Section **Hypotheses and Measurements**. In this study, we investigated the perceived situated learning experience with an interactive VR character quantitatively and compare different embodied feedback strategies qualitatively. We invited each participant to use our VR system, conducting a three-round practice of elevator pitches in front of a virtual character in one of the four feedback conditions: immediate, after-action, combination, and the baseline version with only listening

⁹The proposal ITS/011/19 (HPR #356) has been reviewed and approved by HPR Panel.

behaviors which can be seen as an artificial agent with weak feedback. While maintaining the content of the speech in each round, participants need to learn to adjust their presentation manners based on the interaction experience with the virtual character.

A. Participants

We recruited 40 volunteers (17 females, average age 22.8, SD: 2.2) from a local university through fliers and word-of-mouth. The participants are mainly senior undergraduate or postgraduates students with engineering or business backgrounds. They all reported to have the need to give elevator pitches on different occasions, such as at academic conferences, in business plan competitions, or during internships. On a 7-point Likert scale, participants indicated that they had limited understanding of elevator pitches (mean: 2.5, SD: 1.8) and were not very confident of public speaking (mean: 3.5, SD: 1.4). Participants also mentioned that they were not familiar with VR (mean: 3.4, SD: 1.3). We randomly assigned each participant to one of the feedback conditions (10 people for each condition) to minimize the learning effects. We labelled them in the immediate group as IM.1-10, the combination group as CB.1-10, the after-action group as AC.1-10, and the none group as NA.1-10.

B. Apparatus

We conducted the experiment in a quiet room, as shown in Figure 4. We also projected the scene to a 2D screen to monitor what was happening in VR. We ran the application on a laptop with an Intel CPU i7 2.8GHz, 16GB RAM, and Nvidia GeForce 1070 GPU. We set up a Kinect RGBD-camera to obtain users' body poses, and employed the microphone and the IMU of Oculus Rift to capture their speech and head pose. Two experimenters who are familiar with the feedback codebook monitored the whole experiment. Whenever the participant makes a mistake but the coaching system does not prompt a response or the participant does not make a mistake but the system is going to give a wrong feedback, the experimenters will correct them with the wizard-of-oz interface.



Fig. 4. Apparatus: a user wears a VR headset to practice elevator pitch skills with a virtual coach in VR.

C. Task

We asked participants to deliver an elevator pitch on their future research interest, or current research topic, or past internship experience to a virtual character. We instructed them to treat the virtual character as a coach who acts as a professor or an industrial leader (business or technology if applied) in their field of study. To let participants express their research or experience fully, we told them that the elevator pitch could last up to 120 seconds. Each participant repeated the practice three times.

D. Procedure

After getting consent from the participants, we first introduced the concept of elevator pitches, verbal and nonverbal behaviors involved, and the learning objective in the study (*i.e.*, able to deliver an elevator pitch that meets the standard derived in our qualitative study) based on materials obtained from our preliminary expert interviews. We then described the task in detail. Participants were given three minutes to prepare the content of speech. Once ready, we asked them to put on the VR headset. Participants then had 5 to 10 minutes to get familiar with the VR environment. Once they proceeded to the main task, we started the program and the virtual character *Jane* or *David* would enter the hall and greet the speaker. To alleviate gender bias [13], male participants would meet the male character *David* and female participants would interact with the female character *Jane*. Participants could then start their elevator pitches. After each round of practice, participants needed to fill in a questionnaire to report the immediate effect of their confidence, satisfaction, self-awareness of performance, and anxiety. To alleviate the negative effect of boredom as discovered in our pilot study, the characters would wear different clothes and use a slightly different wording in their opening, feedback, and closing speech every round. Upon the completion of all three rounds, participants filled in another questionnaire to report their feeling about CA, as well as the perception of the character, followed by an in-depth exit interview. For the baseline version, we asked participants to treat listening behaviors as feedback with weak intensity. Our system logged the number of issues users had during delivery of their elevator pitches.

E. Results and Analysis

We first study the participants' sense of CA, as well as their perception of the virtual character and the effect of learning quantitatively, and then analyzed interview scripts to further understand reasons behind the results and compare the different embodied feedback strategies.

1) *Cognitive apprenticeship*: Following the CA measurements, *i.e.*, *feedback during observing*, *help*, *coach*, *feedback for improvement*, and *adaptation*, we analyzed the corresponding questionnaire items statistically using one-way ANOVA test. As shown in Figure 5, all measurements we considered have significant differences. In particular, we have the following statistical findings:

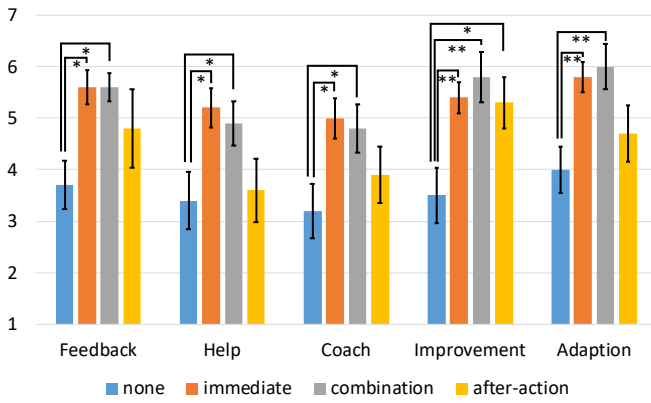


Fig. 5. The means and standard errors of CA measurements, including feelings of feedback during observing (feedback), help received from the virtual character (help), coaching by the virtual character (coach), feedback for improvement (improvement), and feedback adaptation based on the presentation (adaptation) (* : $p < .05$, ** : $p < .01$).

a) *Feedback during observing*: One-way ANOVA analysis shows that there is a significant effect of feedback while observing ($F(3, 36) = 3.292, p < 0.05$). LSD post-hoc test shows that feedback while observing of the immediate and combination strategies are significantly better than the none strategy ($p < 0.05$).

b) *Help*: One-way ANOVA analysis shows that there is a significant effect of help ($F(3, 36) = 3.174, p < 0.05$). Further LSD post-hoc test shows that the help of the immediate and combination strategies are significantly better than the none strategy ($p < 0.05$).

c) *Coach*: One-way ANOVA analysis shows that there is a significant effect of coaching ($F(3, 36) = 2.909, p < 0.05$). LSD post-hoc test reveals that coaching of the immediate and combination strategies is significantly better than the none strategy ($p < 0.05$).

d) *Feedback for improvement*: One-way ANOVA analysis shows that there is a significant effect of feedback for improvement ($F(3, 36) = 4.794, p < 0.01$). LSD post-hoc test results suggest that feedback for improvement of the immediate and combination strategies are significantly better than the none strategy ($p < 0.01$), and the after strategy is significantly better than the none strategy ($0.05 < p < 0.1$).

e) *Adaptation*: One-way ANOVA analysis shows that there is a significant effect of adaptation ($F(3, 36) = 4.593, p < 0.01$). LSD post-hoc test results indicate that adaptation of the immediate and combination strategies are significantly better than the none strategy ($p < 0.01$).

2) *Perception of Virtual Characters*: We also analyzed the perception of virtual characters quantitatively. As shown in Figure 6, following the measurements of *trust*, *like*, *role*, and *empathy*, in Section V, people find that the virtual character with embodied feedback more trustworthy, less favorable, and more like a role than the one without. The empathy level is dependant on feedback strategies, and the virtual character with the immediate feedback strategy is considered more empathetic than the one with other strategies. Unfortunately, no statistically significant difference is found. We further

investigated it in the interview section.

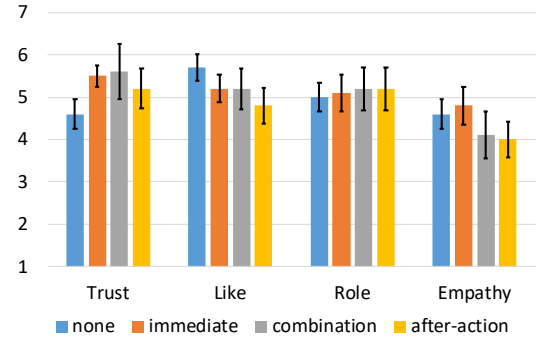


Fig. 6. The means and standard errors of measured perception of the virtual character with different embodied feedback strategies, including trust, like, role, and empathy.

3) *The Effect of Learning in VR*: We mainly consider participants' self-peception of their learning. A two-way ANOVA was conducted to compare the main effects of stage (round 1~3), strategy (none, immediate, combination, after-action), and the interaction effect between them on participants' self-awareness of learning in VR, including their confidence, satisfaction and anxiety on the presentation, and self-awareness of their performance. There is no significant interaction effect between stage and strategy on participants' learning experience. However, there are significant differences on all measurements ($p < 0.01$) in terms of stage. LSD post-hoc test further shows that participants become more confident, satisfactory and less anxious of their presentation, and feel their performance become better with the practice progresses. This result is also consistent with the previous finding that the situated experience brought on by VR can increase people's self-awareness of learning [4].

To further verify the effect of our system on users' performance (although this is not the main goal), we compared the performance of different embodied feedback strategies and the baseline by analyzing the recorded logs. Owing to some runtime errors, we lost logs of four participants (CB.1, CB.5, IM.3, AC.6), and only used logs of the remaining 36 participants and conducted a two-way ANOVA study to examine the effect of stage and strategy on learners' performance. There is generally no statistically significant interaction effect between stage level and strategy on learners' performance. However, as shown in Table IV (in reality, we found that users rarely made mistakes of speaking too loud and too fast), we can see a general decreasing trend in the four types of mistakes (eye contact, hand gesture, rhythm slow/"too slow", and volume small/"too quiet"), suggesting potential performance improvement over repeated practices. Moreover, LSD post-hoc test shows that the immediate strategy is significantly better than the baseline ($p < 0.05$) in term of the eye contact mistake, and marginally significantly better than the baseline ($0.05 < p < 0.1$) in term of the small volume mistake.

F. Interview Feedback and Discussions

To further investigate and compare pros and cons of different feedback strategies, we analyzed interview scripts using

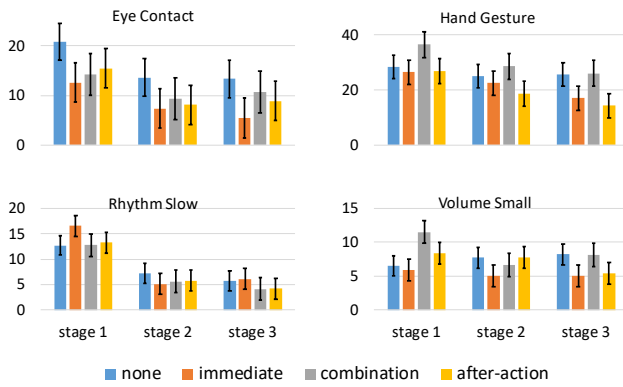


TABLE IV
THE MEANS AND STANDARD ERRORS OF USERS' PERFORMANCE, WHERE
y AXIS DENOTES THE NUMBER OF MISTAKES MADE BY USERS
(* : $p < .05$, + : $.05 < p < .1$).

inductive coding methods [43]. We leverage the qualitative findings to further interpret our quantitative results and identify other issues beyond the initial research questions. Overall, participants find our prototype system “interesting” and “well developed”, and has the potential to help people practice elevator pitches and other similar social skills on a daily basis. We use Don Norman’s seven design guidelines (DG) of applying CA [62] as a reference to discuss our system design.

1. provide a high intensity of interaction and feedback;
2. have specific goals and established procedures;
3. motivate users;
4. provide a continual feeling of challenge, neither too hopeless nor too boring;
5. provide a sense of direct engagement;
6. provide appropriate tools that fit the user and the task;
7. avoid distractions and interruptions that intervene and destroy the subjective experience.

1) *Understanding of Cognitive Apprenticeship*: Generally speaking, embodied feedback seems to evoke a stronger sense of CA, which is important for the design guidelines DG.1-3, DG.5 of cognitive apprenticeship learning [6]. However, our quantitative analysis shows that the after-action type of feedback is no better than the none condition in this regard from a statistical point of view (Figure 5). One potential reason we learned in the interviews is that the after-action feedback strategy does not fully match participants’ mental model of what coaching should be like, compared to the immediate and the combination methods. Four out of ten participants feel that after-action feedback is not given in a timely manner and as a result they lost track of their mistakes. “The (after-action) feedback is given too late. Sometimes I do not know what she is talking about.”-AC.5 (Female). “I do not know what mistakes I made. For example, when I break my speech, you (the virtual character) should point out immediately so that I can know it.”-AC.10 (Female).

Losing the context makes it harder to interpret retrospective feedback. A few participants thus suggested having a play-back function to point out where users make mistakes to make after-action review more informative. How to seamlessly link

video playback and embodied reflection in VR could be a direction for future research.

We have tried to minimize possible interference during the speech by shortening the virtual character’s verbal command, such as “faster” instead of “you should speak faster”, which is consistent with DG.7. Still, users may get distracted as they need bandwidth to process such information while speaking. “I couldn’t understand David’s intention until I got used to it.”-IM.2 (Male). How to balance the desire for intense feedback (immediate) and the need to alleviate interruption (after-action) is thus a critical design issue. The combination strategy seems like a good alternative as it takes the best of both worlds.

Sometimes, proficient speakers feel bored because the coach tends to repeat the same mistakes. However, for less fluent speakers, they often encounter new issues in each round and feel constantly challenged. Therefore, following DG.4, it is necessary to adapt feedback behaviors according to users’ professional levels.

In addition, we found that users generally expect instructions from the virtual character regarding language issues, such as organization, vocabulary, etc., of their speech, as most of our participants are not confident about the content they have drafted. “The structure of the speech is more important for me. I expect Jane gives me more tips on it, but she didn’t”-E.62 (Female). This is related to DG.2 of CA – having specific goals. However, we choose not to look into language-related issues in elevator pitches in our current prototype system, because these problems often need more elaborate explanations and thus are not suitable for having constant embodied feedback while practicing. Otherwise it would become a violation of DG.7 – avoiding distractions and interruptions.

2) *Understanding of Perception of Virtual Characters*: Previous research works have shown that students’ perceptions of learning influence their learning performance, including authenticity and alignment of study approaches [39], academic environment [40], and so on. Therefore, it is vital to investigate how well people perceive an artificial agent in VR. The analysis on CA reveals that the immediate feedback brings a better coaching experience. This is not surprising as most people report that they feel the virtual character with embodied feedback is “helping” them, and thus treat it as being more trustworthy and think it shows more empathy (Figure 6, although the effect is not significant). However, for feedback with after actions (after-action and combination), the trust and empathy levels drop a little bit. One possible reason is that sometimes people do not remember what mistakes they made. When the virtual character points out a mistake that they do not acknowledge, users may feel the character is artificial. A play-back function might thus be necessary in after-action reviews. Another potential cause is the varying accuracy of the detection algorithm. In situations where the system gives suggestions that do not match users’ own judgment especially in the immediate condition, they may also consider it insincere.

Even though a virtual character with embodied feedback appears more like a coach of social skills (Figure 6-Role), six out of 20 participants are not comfortable with such a style of teaching, especially when they have their first encounter

with the character. From interviews we found that a prominent reason is that most participants do not enjoy being interrupted by the immediate feedback during their presentation. *“I are not accustomed to the feedback David provided during my speech, especially at the beginning. I feel I am being interrupted. But latter when I am used to it, I feel better.”-IM.1 (Male).* Designing proper timing for feedback during practice may be a way to alleviate such discomfort. As reported by some participants, they feel it is more artificial if they found that the behavior timing is incorrect. For example, the virtual character should give acknowledgement feedback when the user finishes saying something [63]. If the timing is too advanced or too delayed, they will feel it is “unreal” and “interrupting”, which will further influence their perception of the learning experience. This is related to **DG.1** and **DG.7** of CA. We need to guarantee the behaviors from the virtual character are timed right so that the feedback does not distract users and people can perceive it as genuine and intelligent.

3) *Understanding of the Learning Effect in a VR Environment* : Our quantitative data show that users experienced increased self-awareness confidence, satisfaction, performance, and decreased anxiety when practicing in an authentic environment, even without coach’s feedback in the none condition. *“In fact, sometimes I can realize the problems myself, even if David did not point it out when I made a mistake.”-CB.2 (Male).* Interview results reveal that people usually self-monitor their behaviors, and the situated environment provided by VR solidifies such subconscious activities. With proper feedback, users may be able to learn social skills even more effectively in a VR environment [4].

Overall, immediate feedback can strengthen the sense of CA and make users feel the virtual character is more trustworthy. However, it may not give people enough time to appreciate the presentation of the virtual character. The after-action feedback gives enough time for summarizing users’ performance, but without the play-back function, it is more likely users would feel the virtual character is artificial. Therefore, the point is how to select what to present immediately and what to save for after-action review, as well as how to determine whether an issue should be brought up repeatedly, following **DG.2,4,7**. For example, context-sensitive issues such as gaze is more suitable for immediate feedback, while language related issues that require elaborations can be communicated after each round of practice. For a repeating mistake, the coach can pick it out once at its first occurrence and echo the point again afterwards.

VII. CONCLUSION

In this paper, through a design process, we investigated how to design embodied feedback of an artificial agent to improve people’s situated learning experience for developing elevator pitch skills in VR. In particular, we have designed and developed a proof-of-concept VR coaching system for practicing elevator pitches. We then conducted a between-subject user study to explore potential design considerations both quantitatively and qualitatively. Results show that embodied feedback helps create a stronger sense of cognitive

apprenticeship, as well as improving users’ perception of the virtual character and the effect of learning. In the future, we plan to experiment with a fully functional system built upon insights from this work.

REFERENCES

- [1] R. House, J. Livingston, S. Summers, and A. Watt, “Elevator pitches, crowdfunding, and the rhetorical politics of entrepreneurship,” in *2016 IEEE International Professional Communication Conference (IPCC)*, Oct 2016, pp. 1–4.
- [2] K. D. Pagana, “Ride to the top with a good elevator speech,” *American Nurse Today*, vol. 8, no. 3, pp. 14–16, 2013.
- [3] L. Blume, R. Baecker, C. Collins, and A. Donohue, “A communication skills for computer scientists course,” *SIGCSE Bull.*, vol. 41, no. 3, pp. 65–69, Jul. 2009.
- [4] M. Muratore, C. Tuena, E. Pedroli, P. Cipresso, and G. Riva, “Virtual reality as a possible tool for the assessment of self-awareness,” *Frontiers in Behavioral Neuroscience*, vol. 13, no. 62, pp. 1–7, 2019.
- [5] L. M. Lunce, “Simulations: Bringing the benefits of situated learning to the traditional classroom,” *Journal of Applied Educational Technology*, vol. 3, pp. 37–45, 2006.
- [6] H. McLellan, *Situated learning perspectives*. Educational Technology, 1996.
- [7] S. Sinnett, D. Smilek, and A. Kingstone, *Cognition*. Oxford University Press, 2016.
- [8] J. Miller and B. Strand, “The role of youth sport coaches in developing life skills,” *Ohio AAHPERD Future Focus*, vol. 36, no. 1, pp. 20–25, 2015.
- [9] S. Van Ginkel, J. Gulikers, H. Biemans, and M. Mulder, “Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education,” *Educational Research Review*, vol. 14, pp. 62 – 80, 2015.
- [10] H. H. Mei and L. S. Sheng, “Applying situated learning in a virtual reality system to enhance learning motivation,” *International Journal of Information and Education Technology*, vol. 1, no. 4, pp. 298–302, 2011.
- [11] R. C. M. Yusoff, H. B. Zaman, and A. Ahmad, “Design a situated learning environment using mixed reality technology-a case study,” *World Academy of Science, Engineering and Technology*, vol. 47, pp. 887–892, 2010.
- [12] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht, “Presentation trainer, your public speaking multimodal coach,” in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI ’15. New York, NY, USA: ACM, 2015, pp. 539–546.
- [13] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, “Mach: My automated conversation coach,” in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’13. New York, NY, USA: ACM, 2013, pp. 697–706.
- [14] M. Chollet, P. Ghate, C. Neubauer, and S. Scherer, “Influence of individual differences when training pub-

- lic speaking with virtual audiences,” in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, ser. IVA '18. New York, NY, USA: ACM, 2018, pp. 1–7.
- [15] I. Damian, C. S. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André, “Augmenting social interactions: Real-time behavioural feedback using social signal processing techniques,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 565–574.
- [16] A. D. Leimer, “Measuring a cognitive apprenticeship model of instruction in statistics education,” *The University of Southern Mississippi*, 2015.
- [17] H. J. Smith and M. Neff, “Communication behavior in embodied virtual reality,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 289:1–289:12.
- [18] Online Document from Small Farm Program of University of California, “The 30 second elevator speech,” Retrieved September 18, 2019. [Online]. Available: <http://sfp.ucdavis.edu/files/163926.pdf>
- [19] A. Doyle, “How to create an elevator pitch with examples,” Retrieved September 18, 2019. [Online]. Available: <https://www.thebalancecareers.com/elevator-speech-examples-and-writing-tips-2061976>
- [20] Indeed Career Guide, “How to give an elevator pitch (with examples),” Retrieved September 18, 2019. [Online]. Available: <https://www.indeed.com/career-advice/interviewing/how-to-give-an-elevator-pitch-examples>
- [21] Mind Tools Content Team, “Crafting an elevator pitch,” Retrieved September 18, 2019. [Online]. Available: <https://www.mindtools.com/pages/article/elevator-pitch.htm>
- [22] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.
- [23] N. M. Seel, S. Al-Diban, and P. Blumschein, *Mental Models & Instructional Planning*. Dordrecht: Springer Netherlands, 2000, pp. 129–158.
- [24] A. J. Amorose and M. R. Weiss, “Coaching feedback as a source of information about perceptions of ability: A developmental examination,” *Journal of sport and exercise psychology*, vol. 20, no. 4, pp. 395–420, 1998.
- [25] M. I. Tanveer, E. Lin, and M. E. Hoque, “Rhema: A real-time in-situ intelligent interface to help people with public speaking,” in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ser. IUI '15. New York, NY, USA: ACM, 2015, pp. 286–295.
- [26] A. S. Pentland, “Socially aware computation and communication,” *Computer*, vol. 38, no. 3, pp. 33–40, Mar. 2005.
- [27] S. Erhel and E. Jamet, “Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness,” *Computers & Education*, vol. 67, pp. 156 – 167, 2013.
- [28] N. Yee, J. N. Bailenson, M. Urbanek, F. Chang, and D. Merget, “The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments,” *CyberPsychology & Behavior*, vol. 10, no. 1, pp. 115–121, 2007.
- [29] X. Pan, M. Slater, A. Beacco, X. Navarro, A. I. Bellido Rivas, D. Swapp, J. Hale, P. A. G. Forbes, C. Denvir, A. F. de C. Hamilton, and S. Delacroix, “The responses of medical general practitioners to unreasonable patient demand for antibiotics - a study of medical ethics using immersive virtual reality,” *PLOS ONE*, vol. 11, no. 2, pp. 1–15, 02 2016.
- [30] X. Pan, M. Gillies, C. Barker, D. M. Clark, and M. Slater, “Socially anxious and confident men interact with a forward virtual woman: An experimental study,” *PLOS ONE*, vol. 7, no. 4, pp. 1–13, 04 2012.
- [31] H. Trinh, R. Asadi, D. Edge, and T. Bickmore, “Robocop: A robotic coach for oral presentations,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 27, pp. 1–24, 2017.
- [32] L. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer, “Cicero-towards a multimodal virtual audience platform for public speaking training,” in *International workshop on intelligent virtual agents*. Springer, 2013, pp. 116–128.
- [33] J. Bissonnette, F. Dubé, M. D. Provencher, and M. T. M. Sala, “Evolution of music performance anxiety and quality of performance during virtual reality exposure training,” *Virtual Reality*, vol. 20, no. 1, pp. 71–81, 2016.
- [34] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik, “Technologies for social augmentations in user-embodied virtual reality,” in *25th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST 19. New York, NY, USA: Association for Computing Machinery, 2019.
- [35] J. Lugin, S. Oberdorfer, M. E. Latoschik, A. Wittmann, C. Seufert, and S. Grafe, “VR-assisted vs video-assisted teacher training,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 625–626.
- [36] J. Lugin, M. Landeck, and M. E. Latoschik, “Avatar embodiment realism and virtual fitness training,” in *2015 IEEE Virtual Reality (VR)*, 2015, pp. 225–226.
- [37] S. Van Ginkel, D. Ruiz, A. Mononen, C. Karaman, A. Keijzer, and J. Siithiworachart, “The impact of computer-mediated immediate feedback on developing oral presentation skills: An exploratory study in virtual reality,” *J. Comput. Assist. Learn.*, vol. 36, pp. 412–422, 2020.
- [38] S. Van Ginkel, J. Gulikers, H. Biemans, O. Noroozi, M. Roozen, T. Bos, R. van Tilborg, M. van Halteren, and M. Mulder, “Fostering oral presentation competence through a virtual reality-based task for delivering feedback,” *Computers & Education*, vol. 134, pp. 78 – 97, 2019.
- [39] J. T. Gulikers, T. J. Bastiaens, P. A. Kirschner, and L. Kester, “Relations between student perceptions of assessment authenticity, study approaches and learning outcome,” *Studies in Educational Evaluation*, vol. 32, no. 4, pp. 381 – 400, 2006.
- [40] A. Lizzio, K. Wilson, and R. Simons, “University stu-

- dents' perceptions of the learning environment and academic outcomes: Implications for theory and practice," *Studies in Higher Education*, vol. 27, no. 1, pp. 27–52, 2002.
- [41] K. J. Suda, J. M. Sterling, A. B. Guirguis, and S. K. Mathur, "Student perception and academic performance after implementation of a blended learning approach to a drug information and literature evaluation course," *Currents in Pharmacy Teaching and Learning*, vol. 6, no. 3, pp. 367 – 372, 2014.
- [42] M. W. Easterday, D. G. R. Lewis, and E. M. Gerber, "The logic of design research," *Learning: Research and Practice*, vol. 4, no. 2, pp. 131–160, 2018.
- [43] M. Miles, A. Huberman, M. Huberman, and P. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE Publications, 1994.
- [44] S. Van Ginkel, R. Laurentzen, M. Mulder, A. Mononen, J. Kytä, and M. Kortelainen, "Assessing oral presentation performance : Designing a rubric and testing its validity with an expert group," *Journal of Applied Research in Higher Education*, vol. 9, pp. 474–486, 2017.
- [45] WikiHow Online Document, "How to measure decibels," Retrieved September 19, 2019. [Online]. Available: <https://www.wikihow.com/Measure-Decibels>
- [46] S. M. Smith and D. R. Shaffer, "Speed of speech and persuasion: Evidence for multiple effects," *Personality and Social Psychology Bulletin*, vol. 21, no. 10, pp. 1051–1060, 1995.
- [47] E. Kuhnke, *Body language for dummies*. John Wiley & Sons, 2012.
- [48] J. Allwood, J. Nivre, and E. Ahlsen, "On the Semantics and Pragmatics of Linguistic Feedback," *Journal of Semantics*, vol. 9, no. 1, pp. 1–26, 01 1992.
- [49] M. Cook, "Gaze and mutual gaze in social encounters: How long and when we look others "in the eye" is one of the main signals in nonverbal communication," *American Scientist*, vol. 65, no. 3, pp. 328–333, 1977.
- [50] J. K. Burgoon, D. B. Buller, J. L. Hale, and M. A. de Turck, "Relational Messages Associated with Nonverbal Behaviors," *Human Communication Research*, vol. 10, no. 3, pp. 351–378, 03 2006.
- [51] J. Cassell, O. E. Torres, and S. Prevost, "Turn taking vs. discourse structure: How best to model multimodal conversation," in *Machine Conversations*. Kluwer, 1998, pp. 143–154.
- [52] J. R. Curhan and A. Pentland, "Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes," *Journal of Applied Psychology*, vol. 92, no. 3, p. 802, 2007.
- [53] S. Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of personality and social psychology*, vol. 23, no. 2, p. 283, 1972.
- [54] S. Duncan Jr, "On the structure of speaker-auditor interaction during speaking turns," *Language in society*, pp. 161–180, 1974.
- [55] F. Kaplan and V. V. Hafner, "The challenges of joint attention," *Interaction Studies*, vol. 7, no. 2, pp. 135–169, 2006.
- [56] L. Meltzer, W. N. Morris, and D. P. Hayes, "Interruption outcomes and vocal amplitude: Explorations in social psychophysics," *Journal of Personality and Social Psychology*, vol. 18, no. 3, p. 392, 1971.
- [57] D. G. Novick, B. Hansen, and K. Ward, "Coordinating turn-taking with gaze," in *Proceeding of Fourth International Conference on Spoken Language Processing, ICSLP'96*, vol. 3. IEEE, 1996, pp. 1888–1891.
- [58] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, and I. Poggi, "A model of attention and interest using gaze behavior," in *International Workshop on Intelligent Virtual Agents*. Springer, 2005, pp. 229–240.
- [59] M. R. Key, *The relationship of verbal and nonverbal communication*. Walter de Gruyter, 1980.
- [60] H. H. Clark, "Language use and language users," *Handbook of social psychology (3rd ed.)*, pp. 179–231, 1985.
- [61] A. W. de Borst and B. de Gelder, "Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective," *Frontiers in psychology*, vol. 6, p. 576, 2015.
- [62] D. Norman, *Things that make us smart: Defending human attributes in the age of the machine*. Diversion Books, 2014.
- [63] G. Skantze, A. Hjalmarsson, and C. Oertel, "Turn-taking, feedback and joint attention in situated human-robot interaction," *Speech Communication*, vol. 65, pp. 50 – 66, 2014.



Zhenjie Zhao received the B.Eng. and M.Eng. degrees from Nankai University, in 2012 and 2015, respectively, and the Ph.D. degree in Computer Science and Engineering from Hong Kong University of Science and Technology (HKUST), in 2020. He is currently an assistant professor at the Department of Computer Science and Technology, Nanjing University of Information Science and Technology (NUIST). His research interests include natural language processing, developmental robotics, brain-like computing, and human-computer interaction.



Xiaojuan Ma is an assistant professor of Human-Computer Interaction (HCI) at the Department of Computer Science and Engineering (CSE), Hong Kong University of Science and Technology (HKUST). She received the Ph.D. degree in Computer Science at Princeton University. She was a post-doctoral researcher at the Human-Computer Interaction Institute (HCII) of Carnegie Mellon University (CMU), and before that a research fellow in the National University of Singapore (NUS) in the Information Systems department. Before joining HKUST, she was a researcher of Human-Computer Interaction at Noah's Ark Lab, Huawei Tech. Investment Co., Ltd. in Hong Kong. Her background is in Human-Computer Interaction. She is particularly interested in data-driven human-engaged AI and Human-Engaged Computing (HEC) in the domain of education, health, and design.