DynamicDuo: Co-presenting with Virtual Agents

Ha Trinh, Lazlo Ring, Timothy Bickmore
College of Computer and Information Science
Northeastern University
Boston, MA, USA
{hatrinh, lring, bickmore}@ccs.neu.edu

ABSTRACT
The quality of most professional oral presentations is often poor, owing to a number of factors, including public speaking anxiety. We present DynamicDuo, a system that uses an automated, life-sized, animated agent to help inexperienced speakers deliver their presentations in front of an audience. The design of the system was informed by an analysis of TED talks given by two human presenters to identify the most common dual-presentation formats and transition behaviors used. In a within-subjects study (N=12) comparing co-presenting with DynamicDuo against solo-presenting with conventional presentation software, we demonstrated that our system led to significant improvements in public speaking anxiety and speaking confidence for non-native English speakers. Judges who viewed videotapes of these presentations rated those with DynamicDuo significantly higher on speech quality and overall presentation quality for all presenters.

Author Keywords
Co-presentation; Virtual Agent; Slideware; PowerPoint; Embodied Conversational Agent

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces.

INTRODUCTION
Despite the ubiquity and importance of oral presentations, many presentations have very low quality, resulting in their failing to engage, inform, and persuade the audience, and even damaging the credibility and social standing of the speaker. There are many reasons for these failures, including: deficiencies in language, speech, and presentation skills; lack of content mastery; time and resource constraints; lack of preparation and rehearsal; and public speaking anxiety. Prior research has demonstrated that these problems are very prevalent. An extensive survey of 2,501 professionals [9] revealed that 35% of respondents rarely or never rehearse for their presentations, and because of this and many other problems, respondents gave an average grade of "C-" (2.9 on a 1-to-5 scale) for presentations they had attended. Other research has reported that public speaking anxiety affects at least 35% of the population [4].

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04…$15.00
http://dx.doi.org/10.1145/2702123.2702584

Imagine if all speakers could give their presentations in the company of a polished, flawless, co-presenter – one who not only shared the cognitive burden of presentation preparation and delivery, but whose presence and delivery was engaging enough to keep the audience spellbound. The mere presence of a talented co-presenter should also be comforting to those who experience speech anxiety. Social impact theory predicts that the presence of a co-presenter reduces performance anxiety, and this effect has been demonstrated in empirical studies [11]. A co-presenter also enables dialogical presentation formats that are impossible in single speaker talks, in which the speakers perform a rehearsed conversation to dramatically illustrate a point. Finally, a polished co-presenter could function as a real-time role model, providing both a helpful demonstration of presentation skills and an increase in self-confidence that comes from being part of a talented team.

We have designed and developed an automated co-presenter that can perform many of these roles and functions, as an aide to less experienced presenters. The co-presenter is an embodied conversational agent that appears in the form of a life-sized human character that can seamlessly present part of a talk given with conventional presentation software (Figure 1). The co-presenter agent uses appropriate verbal and nonverbal behavior for content delivery, highlighting and emphasis, speaker hand-offs (turn-taking), and attentive listening when the human presenter is speaking. We also developed an authoring tool to allow human presenters to easily control the verbal and nonverbal behavior of the agent. We call the overall presentation technology system "DynamicDuo".

Our contributions include:

1. Identification of common presentation formats used by "gold standard" human co-presenters, and description of the verbal and nonverbal behavior used during introductions and turn transitions.
2. Development of the DynamicDuo authoring system, which allows non-technical users to easily define agent behavior as part of presentation authoring in a widely-used, commercially-available presentation support application.

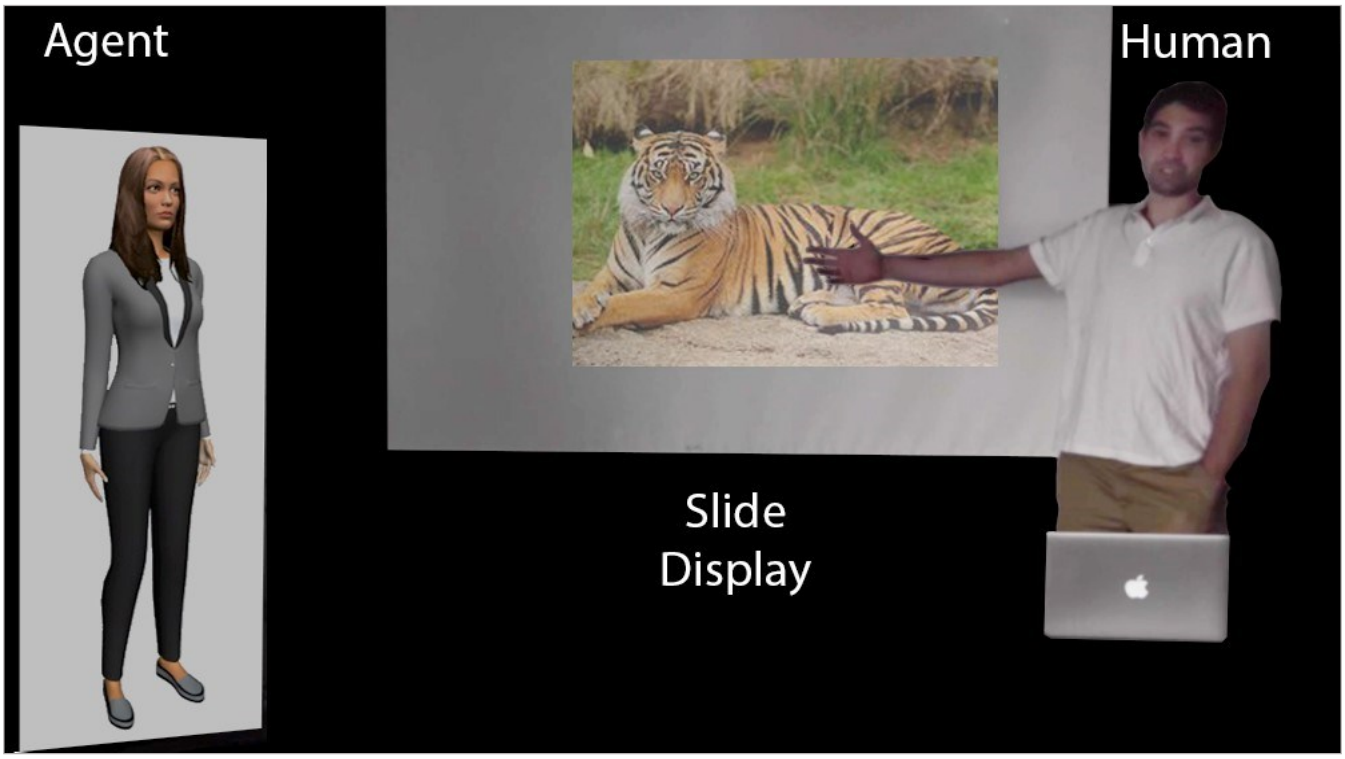

Figure 1. Collaborative presentation environment with slideshow display and Angela appearing on a standing display

3. Development of the DynamicDuo runtime system, to support co-presented oral presentations in front of an audience.
4. Validation of DynamicDuo in a comparative study that demonstrates its potential to improve the overall experience of both presenters and audiences.

RELATED WORK

In this section we explore the causes of public speaking anxiety in light of social impact theory, followed by a review of related presentation technologies and existing virtual presenter systems.

Public Speaking Anxiety

Public speaking is often regarded as one the most anxiety provoking situations. In a survey of 803 American college students, over 35% of respondents expressed the need for assistance with public speaking anxiety (PSA) [4]. Characterized by an increase in physiological arousal (e.g. increased heart rate) and an intense fear of what other people are thinking (e.g. I'm appearing incompetent) [6], PSA can lead to ineffective speech preparation [7], degraded performance quality [16] and increased speech disturbances [13]. Often categorized as either a trait or a state, PSA has been linked to environmental factors such as the size of the audience as well as the number of co-presenters present [11].

One of the seminal works into this phenomenon was Jackson and Latané's research on the impact of co-performing on stage fright [11]. In their study, 60 participants were recruited to imagine themselves in one of 72 different performance scenarios. In each of these scenarios, participants were shown images of the co-presenters and audience they should imagine themselves performing in front of. These images depicted different possible audiences and co-presenters varied in both size and social status. Their results showed that as the size of the audience increased, the participants' imagined tension grew. However, as the number of co-presenters increased, the imagined tension decreased logarithmically. To further explore this finding, a second experiment was conducted in which 48 student performers were asked to fill out a questionnaire prior to giving a live performance. Results of this experiment matched the previous findings, in which presenters performing with as few as one co-presenter experienced an exponential decrease in reported nervousness.

The main explanation behind this finding was social impact theory [14], which explains phenomena such as public speaking anxiety through the concept of social force fields. By imagining a force field projected from an individual, one's social impact can be measured based on the importance, immediacy, and number of people in presence. In addition to specifying the influence of the audience, the theory also accounts the dispersion of social impact felt when other people are co-performing, thus explaining why negative evaluation or embarrassment feels divided up if more than one person is performing at a time. This concept acts as one of the fundamental motivations behind our work,

since it suggests that a co-presenter agent may help diffuse the social impact felt by a human presenter.

Presentation Technologies

There have been several commercial products designed to improve oral presentations, with the most ubiquitous being Microsoft's PowerPoint [17] and Apple's Keynote [2]. A number of research projects have also proposed methods to support various presentation activities, from authoring [8] to rehearsal [24] and delivery [22]. A recent example of these is the PitchPerfect system [24], which provided an integrated rehearsal environment for structured presentation preparation. The system enables presenters to break down speaking notes into a series of note segments corresponding to specific visual elements on slides. It also includes a special note segment called transition note, which encourages presenter to prepare for verbal linkage between slides. Although not specifically designed for co-presenter note authoring, this structured note approach could potentially be adopted for distribution of speaking notes among co-presenters and preparing for transitions between presenters.

Virtual Agents and Presentations

Virtual Agents as Presenters

Over the last 20 years there have been a number of attempts to use virtual agents in presentation technology. One of the earliest attempts is the WebPersona system [1], which presented information automatically generated from the World Wide Web to a user via an animated cartoon character. In a user study with 30 participants, the virtual presenter was found to be easier to understand and more entertaining than the same information presented without the presenter. However, users were limited in the way they could interact with the character, since the only interaction option was to navigate to other generated scripts by clicking on embedded hyperlinks in the agent's response.

In addition to fully automated presentation systems, there has been a number of studies on the use of virtual agents to present manually authored speech text on behalf of a human presenter. One such example is Noma et al.'s 3D virtual presenter system [19]. The system enables users to annotate the presenter's speech text with various gesture commands, which could be performed by a 3D animated computer character capable of non-verbal behaviors and synthesized speech. Additional interaction options could also be programmed into the system using a menu-based scripting template. The interface to the system was, however, highly technical and was not evaluated in a user study.

Following from this research, Nijholt et al. explored the creation of an embodied virtual presenter agent for use in a virtual meeting room [18]. Similar to Noma's work, a virtual presenter system was developed to parse manually annotated presentations into animation scripts for a 3D character. Unlike the previous work however, this system was designed for real-time use in a virtual meeting environment. The system could capture audience motions via cameras placed in a meeting room, thereby allowing for a more realistic simulation of audience members and the person controlling the virtual presenter. However, similar to Noma's study, there was no reported user evaluation of the system.

Although all of these systems acted as virtual presenters, their main goal was to replace the human presenters instead of augmenting their performance through human-agent collaboration. Additionally, the interfaces of these systems were highly technical in nature, requiring users to be familiar with the scripting languages or programming toolkits specifically developed for the systems.

Virtual Agents as Audiences

Besides giving presentations for people or acting as their virtual avatar, virtual agents have also been used as virtual audiences to support public speaking training [12, 21]. Pertaub et al. conducted a study investigating the effects of virtual audiences on public speaking anxiety [21]. 40 students were recruited to give two talks in front of a virtual audience of eight male agents. Participants were randomly assigned to one of three groups in which they presented to either a neutral, positive or negative audience. Their results showed that the negative audience significantly decreased the users' confidence after their presentation, suggesting that the system could be used as a virtual environmental therapy tool for those experiencing fear of public speaking.

Virtual Agents as Coaches

Virtual presenters have also been used to coach people on other communication skills, such as job interviewing. A recent example of this is the MACH system [10], a highly realistic virtual interview coach that provides real-time visual feedback on various verbal and non-verbal behaviors of human interviewees, including speech, prosody and facial expressions. The design of the virtual coach was informed by an exploratory study that recorded interactions between a professional career counselor and students. The system was evaluated with 90 undergraduates, who were asked to practice their interviews by either interacting with MACH or watching an educational video. Students who interacted with MACH showed a significant improvement in interview performance compared to those in the control group, as rated by human experts.

UNDERSTANDING CO-PRESENTATION MODELS

To motivate and inform the design of our co-presenter agent, we analyzed ideal examples of talks given by two presenters. Our aim was to identify common collaboration models between co-presenters, as well as their verbal and non-verbal behaviors exhibited during the co-presentation process. We focused on TED talks (www.ted.com) since they are widely considered to be exemplary oral presentations, are freely available, and short in duration. Of the 1,732 talks available online (accessed 9/1/14), only 34 (1.9%) were given by two presenters. Of these, we selected 15 for detailed analysis.

From our analysis we identified the following interaction formats:

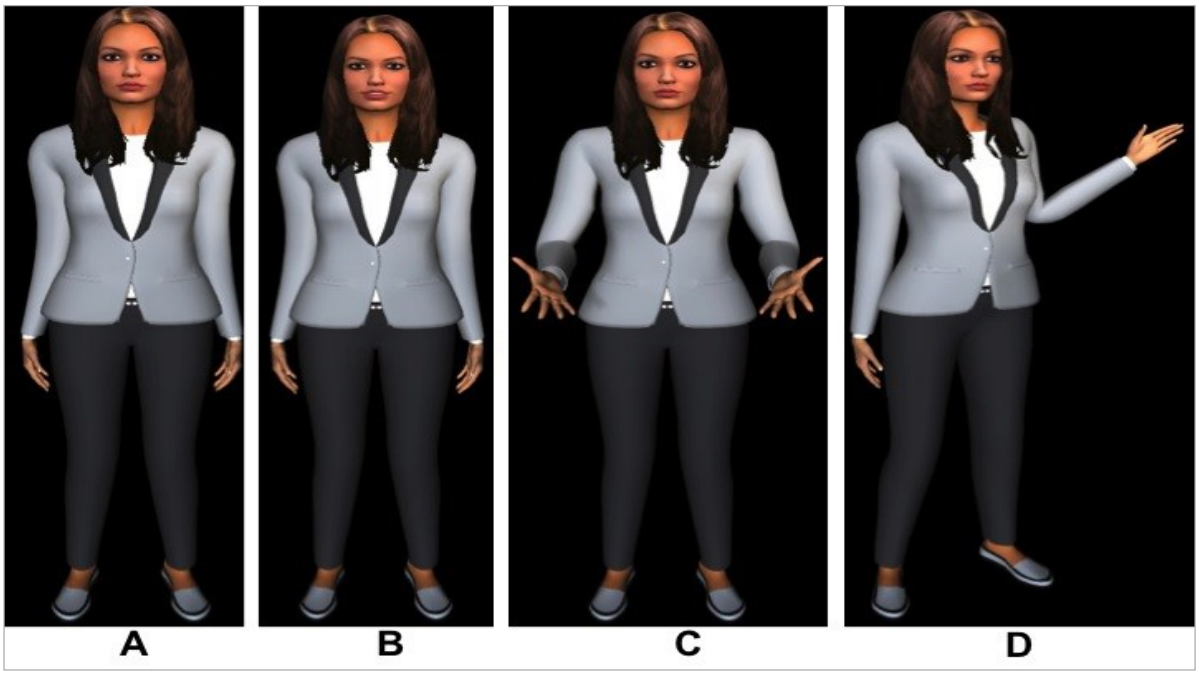

Figure 2. Virtual presenter: (A) relaxing posture; (B) smile; (C) contrast gesture; (D) pointing to slides

Iterative turn-taking. This was the most common format observed (47%), in which presenters take turns giving parts of the presentation. Transitions normally occur at presentation slide boundaries. Occasionally (in 57% of these talks), an embedded dialogue is used, for example, in which one speaker asks a question of the other. However, speaker transitions mostly occur by one speaker simply stopping and the other starting. The average speaking turn lasted 43.7 seconds (SD 27.8), although there was significant variability between dyads (18.9 seconds to 171.6 seconds).

Single turn. The second most common format observed (27%) was one in which each presenter spoke exactly once, sometimes after a brief introduction, with each sharing the floor approximately half of the time.

Dialogue. Although several talks in the iterative turn-taking format had brief embedded dialogues, one of the presentations we analyzed featured a staged dialogue for the entire talk. Turn length was considerably shorter than in iterative turn-taking, averaging 11.0 seconds (SD 8.5).

Interview. One of the talks followed an interview format, in which an interviewer asked each of the co-presenters a question as a prompt for the next part of their talk.

Debate. One of the presentations followed a debate format, in which a moderator introduced the co-presenters, gave them each a fixed time to make an argument, then opened the floor for interaction.

We also analyzed the verbal and non-verbal behavior used by co-presenters, particularly in the iterative turn-taking and single turn formats. Introductions of any kind at the start of talks were rare, occurring in only 20% of our dyads. The majority of talks began with one speaker simply taking their first turn. Explicit verbal turn transitions (e.g., "Sean's going to tell you…") were also rare, occurring in only 30% of talks, and only once or twice in each of these. Nonverbal behavior was obscured in 29% of the turn transitions (with the video showing a presentation slide). Of the visible transitions, the current presenter gazed at the next speaker when it was their turn to speak 36% of the time, and gestured at them 6% of the time. Gaze transitions appeared to most frequently represent listener behavior – in which the current speaker is passively attending to the next speaker once they start – and only rarely represent proactive signals to the next speaker that it was their turn.

DEVELOPMENT OF DYNAMICDUO

Informed by the findings of our presentation corpus analysis, we designed the DynamicDuo system, which supports dual presentations between human and virtual presenters. Implemented as an add-in in PowerPoint 2013 [17], our system consists of three primary components: (1) a *life-size co-presenter agent* that exhibits a range of verbal and non-verbal behaviors; (2) a *collaborative note authoring* tool that enables human-agent note scripting at slide level; (3) a *collaborative presentation environment* that is integrated into the PowerPoint's slideshow delivery mode. Unlike other virtual presenter systems, DynamicDuo enables presenters to incorporate different interaction formats into a dual presentation using a simple interface seamlessly integrated into PowerPoint, without the need for any markup scripting commands. This eliminates any cognitive overhead associated with learning complex scripting languages.

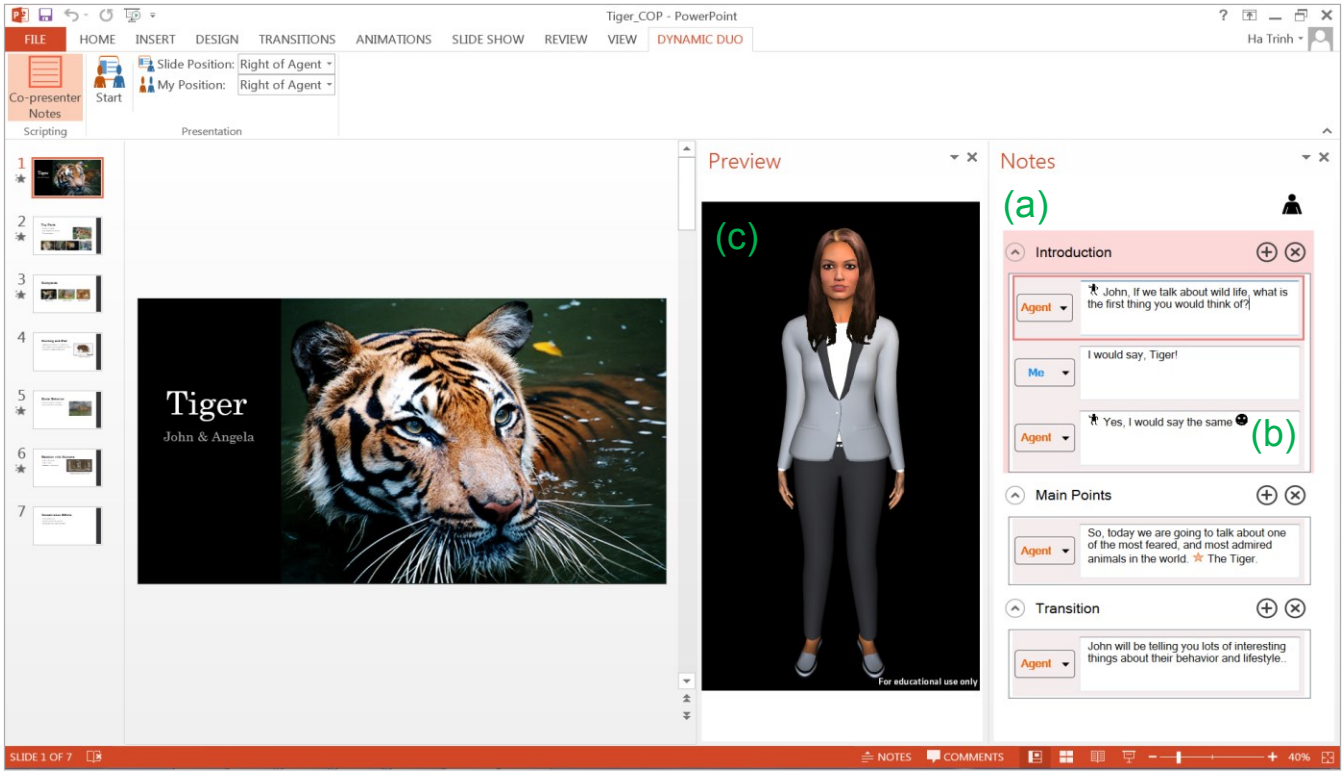

Figure 3. Collaborative note authoring interface with: (a) human-agent note segments; (b) icons representing manually added non-verbal behaviors; (c) agent speaking preview.

Co-presenter Agent: Angela

Our virtual presenter, Angela, is an animated human-like character developed using the Unity game engine (Figure 2). In terms of visual appearance, Angela was designed with a highly realistic look and appears to be multicultural in ethnicity. Synthesized speech is used to generate Angela's speech in real-time using CereProc's Speech Synthesis Engine [3]. Using viseme callbacks from the text to speech engine, Angela lip-syncs the synthesized speech using blend shape animations.

Angela is also capable of displaying a variety of non-verbal behaviors, including facial expressions of affect (smile, neutral, concern), eyebrow movement, directional gazes, head nods, posture shifts, and contrastive, beat (emphasis), and deictic gestures (e.g. pointing to a slide). A number of these behaviors are parameterized to allow for variable degrees of expression. Figure 2 illustrates four of Angela's non-verbal behaviors.

The majority of the agent's non-verbal behaviors are automatically generated using BEAT (Behavior Expression Animation Toolkit) [5]. Given the agent's speaking text, BEAT performs linguistic and contextual analysis of the text and automatically inserts five types of nonverbal behavior: *beats* (formless hand waves for emphasis, that account for approximately 50% of naturally occurring gestures), *eyebrow raises* (for emphasis), *gaze toward/away from* the conversational partner (for turn-taking), *contrastive hand gestures* (marking two items being contrasted, such as "Is it good or bad?"), and *posture shifts* to mark topic boundaries (based on analysis of discourse markers such as "anyways").

Collaborative Note Authoring

DynamicDuo enables the human presenter to enter human-agent speaking notes using the *Co-presenter Notes* side pane built in PowerPoint (Figure 3), which can be activated by clicking the corresponding control on the PowerPoint ribbon. As the presenter selects a slide, the system automatically creates placeholders for three note sections: *Introduction*, *Main Points*, and *Transition* (Figure 3a). These structured notes provide enhanced support for speech organization, while encouraging presenters to prepare for transitions, which are particularly important for smooth human-agent interaction.

Each note section is further divided into a series of note segments, each of which can be assigned to either the co-presenter agent or the human presenter (Figure 3a). The presenter can insert or remove selected note segments by clicking appropriate buttons on the side pane. Our note authoring interface provides a simple, yet flexible mechanism to incorporate different dyadic interaction formats identified in our presentation corpus analysis. For instance, Figure 3a shows an example slide note that contains an embedded dialogue in the introduction section, followed by main points and transition sections assigned to the agent.

In addition to BEAT's automatically generated non-verbal behaviors, our system also enables the presenter to manually

insert additional non-verbal behaviors into the agent's speaking notes. Right-clicking on one of Angela's note segments brings up a context menu of additional non-verbal behaviors. The menu consists of 10 presentation-specific behavior options, including: *gazing* towards the audience/ towards the human presenter, *pointing* to slides/ to the human presenter, *turning* towards the audience/towards the human presenter, *facial expressions*, and *playing selected animation* on the current slide. Each inserted behavior is represented by a visual icon in the note segment (Figure 3b).

To determine appropriate directions of gazes, posture shifts and deictic gestures, the presenter can specify their spatial position and the slide's position in reference to the agent's position for a given presentation room (i.e. either at the left or right to the agent), using appropriate controls on the PowerPoint ribbon (Figure 3).

While authoring agent's notes, the presenter can preview her verbal and non-verbal behavior from the *Preview* side pane (Figure 3c), which can be displayed by clicking the corresponding button on the *Co-presenter Notes* pane. Our system allows the presenter to preview the agent's speech at both note segment, section, and slide levels. Previewing a section will play all the agent's segments within that section sequentially, while previewing at slide level will play all the agent's segments in the current slide.

Collaborative Presentation Environment

To present with the co-presenter agent, the presenter enters our *collaborative presentation environment* by clicking the Start button on the PowerPoint ribbon (Figure 3). Once started, PowerPoint projects its standard slideshow view to the first external display and optionally displays the standard presenter view (with timing and speaking notes) onto to presenter's computer screen (Figure 1). The speaking notes are arranged into human-agent's note segments clearly labeled and separated by blank lines. The co-presenter agent is projected onto a second external standing display, allowing Angela to appear life-size, as shown in Figure 1.

In the presentation mode, the presenter advances the slides and cues the co-presenter agent's speech using an IR remote control from Turning Technologies [25] with four buttons: Next Slide, Previous Slide, Speak, and Stop. The Next and Previous Slide buttons advance the PowerPoint slides as usual, while the Speak button cues Angela to present the next available agent segments on the current slide. Once cued to speak, Angela presents her segments while automatically advancing the slides and playing pre-specified animations, and stopping when she reaches the next note segment of the human presenter. Angela performs a posture shift to turn and gaze toward the human presenter as an indicator of turn-giving. At any point during the presentation, the human presenter can pause the agent's speech using the Stop button. This allows the presenter to spontaneously elaborate on the agent's speech, or navigate to different slides in response to feedback and questions from the audience. The presenter can also resume the agent's speech using the Speak button.

While not actively presenting, Angela turns toward the human presenter based on the orientation selected within the collaborative authoring tool, and goes into an attentive listening mode. While in this mode, Angela randomly performs one of four nonverbal behaviors every 10 seconds. These behaviors include smiling, head nodding, turning toward the audience and gazing at the main slideshow display (Figure 1).

EVALUATION OF DYNAMICDUO

To examine the effectiveness of our co-presentation approach in improving the experience of both presenters and audiences, we conducted two user studies. Our *Presenter* study evaluated the presenter experience of preparing and co-presenting with Angela against solo-presenting with conventional presentation software (PowerPoint 2013). In our follow-up *Judge* study, we compared the quality of co-presentations with Angela against sole presentations with standard PowerPoint, as perceived by an audience.

Presenter Study: Procedure

We asked participants to rehearse and deliver two 7-minute presentations on comparable topics (Lions and Tigers) in English using prepared PowerPoint slide decks and notes. Each slide deck contained 6 slides and approximately 850-word supporting notes, covering 16 key points, in addition to introduction and transition notes. One of the presentations was co-presented with Angela, while the other was a sole presentation with the standard version of PowerPoint 2013. This study scenario is well-grounded since delivering pre-scripted and rehearsed presentations with minor or no variance is common, as evidenced in our analysis of TED talks and previous research on presentation rehearsal [24].

The study was a within-subject, counterbalanced design across two sessions. Each session lasted between 60-90 minutes, with 1 to 6 days between sessions. The ordering of the conditions (Human-Agent vs. Human-only) and the slide decks were randomly assigned and counterbalanced. Each presentation was videotaped for later evaluation.

Human-Agent Session: We asked participants to rehearse and deliver a presentation with Angela using the DynamicDuo system. We added the notes into the *Co-presenter Notes* side pane in advance. We distributed the slides and notes between the human presenter and Angela as follows: In slide 1, Angela and the human presenter engaged in a scripted dialogue in the introduction, followed by the human presenter's parts. At the end of slide 1, the human presenter transitioned to Angela, who presented slides 2 and 3 before transitioning back to the human presenter. Slides 4 and 5 were presented by the human presenter, and Angela took turn again to present slide 6.

At the beginning of the session, we introduced participants to the scenario of presenting a pre-made slide deck, and gave them brief instructions on presenting with Angela using DynamicDuo. We then gave the participants 30 minutes to rehearse their presentation, before delivering their

videotaped presentation. We conducted a semi-structured interview at the end of the session eliciting their attitudes towards presentation preparation and delivery with Angela.

Human-only Session: In this session, we asked participants to rehearse and deliver their presentation without Angela, using the standard version of PowerPoint. Notes were pre-loaded into the PowerPoint notes section, in which introduction, main points, and transitions were clearly separated using blank lines.

We gave participants brief instructions on the notes and Slide Show mode of PowerPoint, then gave them 30 minutes to prepare before delivering their videotaped presentation. We concluded the session with a semi-structured interview.

Presenter Study: Measures

Presenters were assessed using the following self-report measures:

Communication Competence: Assessed at intake using the Self-Perceived Communication Competence Scale [15].

State Anxiety: Assessed prior to each presentation using the State Anxiety questionnaire [23].

Speaker Confidence: Assessed at intake and after each presentation using the Personal Report of Confidence as a Speaker questionnaire [20].

Virtual Co-presenter Rating: Assessed after the human-agent presentation using a 6-question, 7-point scale measure.

Presenter Study: Participants

We recruited 13 students and professionals (4 male, 9 female, ages 23-62, mean 39) via an online advertisement. Of these 13 participants, 8 were native English speakers and 5 were non-native English speakers. The inclusion of both native and non-native English speakers allowed us to examine the effect of speech and language abilities on the presenter's presentation experience. Our participants had varying presentation experience and occupational backgrounds, ranging from law, medicine to administration. Participants were compensated for their participation.

Presenter Study: Quantitative Results

12 presenters successfully completed the study resulting in 24 total interactions, 12 with Angela and 12 without. Participant P1 (a female native English speaker) had to be dropped from the study due to technical issues.

Of the 12 remaining participants, 4 were categorized as high competence public speakers, 2 were categorized as low competence public speakers, and 6 had moderate competence according to the Self-Perceived Communication Competence Scale.

Our data was normal according to results of Shapiro-Wilk tests, thus we performed parametric tests (repeated-measure ANOVAs and t-tests) to examine the effects of our DynamicDuo system on the presenters' state anxiety and speaker confidence.

State Anxiety

There was a significant interaction effect of *Condition* (human-agent vs. human-only) and *Native Language* (native vs. non-native English speakers) on state anxiety, ($F_{1,10}$=8.909, p=.014, partial η^2=.471) (Figure 4). The virtual presenter significantly decreased state anxiety for non-native English speakers (t(4)=3.492, p=.025, d=1.562). There were no significant effects of condition on state anxiety for native English speakers.

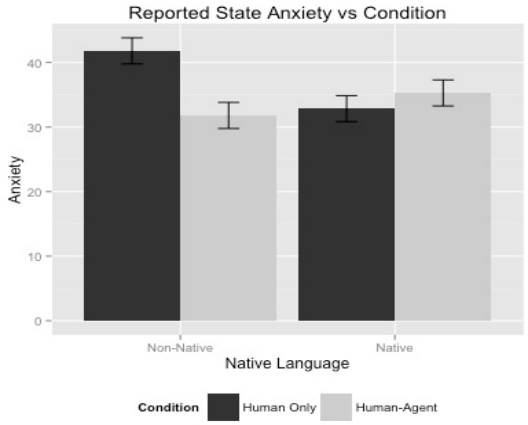

Figure 4. Self-reported state anxiety for human-only and human-agent conditions from non-native and native participants.

Speaker Confidence

There was a significant interaction effect of *Condition*Native Language* on speaker confidence ($F_{1,10}$=11.934, p=.006, partial η^2=.544) (Figure 5). The co-presenter agent significantly increased confidence for non-native speakers (t(4)=4.083, p=.015, d=1.826), while there were no significant differences for native speakers.

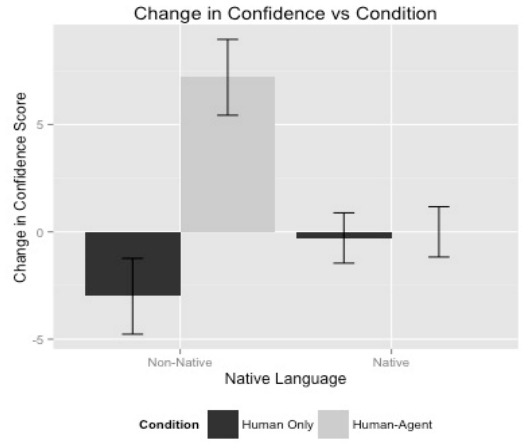

Figure 5. Self-reported speaker confidence for human-only and human-agent conditions from non-native and native participants.

Virtual Co-presenter Ratings

The ratings for the co-presenter agent were highly positive across all participants (Table 1). The virtual presenter was rated to be satisfying to work with (M = 6.5, SD = .67), likeable (M = 5.83, SD = 1.27), trustworthy (M = 6.58, SD = .9), helpful (M = 6.3, SD = 1.15) and participants expressed that they desired to use the system in their future presentations (M = 6.3, SD = 1). The one rating for which we received mixed results was ease of use (M = 4.25, SD = 2.56), since some participants found the system easy to use while others found it difficult.

Rating of Co-Presenter: ***(Scale Measures from 1-7)*** *1 – Not at All* *7 – Very Much*	**Mean (SD)**
How ***satisfied*** are you with…?	6.5 (.67)
How much do you ***like***…?	5.83 (1.27)
How much do you feel you ***trust***…?	6.58 (.9)
How ***helpful*** was…?	6.3 (1.15)
How much would you like to ***give future presentations*** with…?	6.3 (1)
How **easy** was it to use…?	4.25 (2.56)

Table 1. Average ratings of the virtual co-presenter.

Presenter Study: Qualitative Findings

We conducted semi-structured interviews at the end of each session, eliciting the presenter's experience of preparing and delivering a presentation with the agent and suggestions for improvement. We performed high-level coding on our transcribed interviews and derived three main themes related to workload, stage pressure and feelings of assurance.

Reducing Workload through Division of Labor

The most cited advantage of co-presenting with a virtual presenter was that it *"relieved you from a lot of workload"* [P13] during both presentation preparation and delivery. By sharing the cognitive *"burden of absorbing a lot of material"* [P3] under strict time pressure, our co-presenter agent reduced the human's memorization load, allowing them to focus and polish other aspects of their presentation, alleviating the nervousness of *"forgetting so many things"* [P9] and consequently helping them *"feel more confident"* [P6]. This reduced workload was especially appreciated by our non-native speakers, who often operate under a heavy cognitive load when presenting in their second language.

While reducing workload through shared content and responsibility, co-presenting with the agent imposed an extra coordination load as the presenters had to *"memorize the proper time to interact with the agent"* [P6]. During delivery, it required the presenters to *"multi-task, remember facts, deliver facts to the audience, try to interact with the audience and at the same time remember which button to push"* [P5]. For some participants, this added *"mental gymnastics"* [P12] caused additional stress during delivery. However, many participants expressed the confidence to *"master pretty quickly"* [P5] with practice. These findings suggest the need for further research to minimize the amount of coordination required to use the system, with possible integration of a natural user interface, such as speech or gesture recognition.

Reducing Stage Pressure through Shared Attention

In line with social impact theory and previous research, many participants stated that Angela helped with sharing *"the audience attention, and drew a little bit of the heat away"* [P5], which was particularly helpful for highly anxious presenters. It also allowed the human presenters more time to *"figure out what to say next"* [P8] while the agent was speaking, to *"observe the audience for their reaction"* [P6], enabling them to mentally review and rehearse their notes, leading to reduced note reliance and boosted confidence in their delivery.

Increasing Assurance with a Reliable Assistant

The dependability was also cited as being a key characteristic of our co-presenter agent. Comparing to a human co-presenter, the agent provided a high level of reliability, assuring the participants that *"there are going to be no curveballs"* [P5] and that *"she is not going to mess up the presentation"* [P2].

While being highly reliable and consistent, our agent still lacked the ability to provide spontaneous support, e.g. providing prompts when her co-presenters forget their lines. Thus, a future improvement could be to provide a monitoring and backup mechanism to track the presentation progress and automatically provide timely, appropriate cues for the human presenter.

To summarize, our qualitative findings demonstrated that our co-presenter agent helped improve the presentation experience for the majority of our participants, particularly for novice and non-native speakers. One of our participants emphasized: *"this was one of the greatest experience of my presenting"* [P9], while P2 stated: *"It helped me as much as someone could help me"*.

Judge Study: Procedure

In this study, we aimed to evaluate the relative quality of the 12 pairs of videotaped presentations (Human-Agent vs. Human-only), which were delivered by the 12 participants in our Presenter study.

We ask each judge participant to watch two pairs of videotaped presentations. After watching each pair, judges compared each presentation pair on six criteria adopted from [24], including: organization, content coverage, note reliance, speech, timing and pacing, and overall quality. Each criterion was judged on a 4-point ordinal scale of "no difference", "slight difference", "moderate difference", and "substantial difference", with an indication of the superior presentation, if any. The ordering of the presentations were randomly assigned and counterbalanced across the judge participants. In addition, after watching each presentation of

a pair, judges were also asked to evaluate the *human presenters* using a 7-item, 7-point scale questionnaire assessing competency, engagingness, nervousness, understandability, excitement, entertainingness and satisfaction. Each session lasted approximately 60 minutes.

Judge Study: Participants

12 judges were recruited for the study (4 male, 8 female, ages 24-58, mean 39, J1-J12) via an online advertisement. Our judges had varying presentation experience and occupational backgrounds, with four of them reported as expert presenters (researchers and lecturers in the field of communication). Our participants were compensated for their participation.

Judge Study: Quantitative Results

12 judges successfully completed the study resulting in 24 comparisons. We performed non-parametric tests (Wilcoxon signed-rank tests) to examine the effects of our co-presenter agent on the judges' ratings of presentation quality and human presenters.

Presentation Rating

There were significant differences on the judges' ratings of *note reliance* ($p<.05$), *speech* ($p<.01$) and *overall presentation quality* ($p<.05$), in favor of the Human-Agent condition. No significant differences were found for organization, timing and pacing, and content coverage.

The inter-rater reliability of the judges was moderate (Kappa = 0.439), which could be explained by the highly subjective nature of the rating task, the varying degrees of presentation experience across the judges, and the limited training given on the task.

Audience Perception of Human Presenters

There were no significant differences between the conditions on the judges' ratings of the human presenters for any of the seven measures. These results indicate that employing the support of the virtual co-presenter did not result in significant lower ratings of the human presenters, while significantly improving the overall presentation quality.

Judge Study: Qualitative Findings

We conducted a semi-structured interview at the end of each judging session, prompting for the judges' impressions of the overall presentation quality as well as their evaluation of the human presenters and the co-presenter agent. We performed high-level coding on our transcribed interviews and categorized the feedback into three themes.

Improving Content Understandability

Our judges consistently reported the role of the co-presenter agent in elevating the understandability of the presentations with her clear speech delivery that was timely synchronized with slide visuals: *"The character was able to deliver a lot of information without any uhm or ah, and clearly knew what was going on, and there was synchrony with the presentation aid"* [J7]. When presenting with inexperienced human speakers, *"who are not as strong, who are less confident with the material and less confident with their overall public speaking"* [J2], the professional and clear speech delivered by the agent provided a *"nice balance"* [J2], helping the audience *"take your mind off of the human's mistakes"* [J6]. Clearly planned transitions between human-agent sections also resulted in a *"more structured"* [J4] presentation with *"better previews"* [J4], further aiding the presentation understandability.

Injecting Variety into Presentations

As "*it can get monotonous listening to one voice the whole time"* [J8], having co-presenters on stage helped to *"break up the monotony of a PowerPoint lecture"* [J4], thereby maintaining the audience engagement. The variety of interaction formats, such as rehearsed dialogue and iterative turn-taking, also created a *"good interplay"* [J2] between the two presenters when the interactions were well-coordinated. Switching between the different presenters with different voices using different interaction models, however, could impose an extra cognitive load on the audience. Therefore, in the design of dual presentations, it is important to maintain a balance in the amount and the formats of interactions.

Additionally, many of the judges reported the lack of inflection and emotion in Angela's voice, a common problem of synthesized voices. Further research is warrant to add emotional variance to our agent's speech.

Increasing Audience Engagement

Most judges reported the novelty effects of a virtual co-presenter in attracting the audience's attention: *"People that may automatically yawn at a human may sit up and like 'oh what is it about' when they see an animated character. So it can spice it up a little"* [J1]. Judges were generally attracted to the agent's highly realistic appearance and human-like behaviors, as well as her charisma: *"She had more charisma, she seemed much more posh, much more prepared […]. I liked her mannerism, I liked the way she turned around and looked at the screen but made eye contact with the audience" [J4]*.

This novelty effect from the audience's initial exposure to the agent could, however, divert the audience attention from the presented content: *"it makes the listeners more engaged in the newness of her rather than what you are actually presenting about"* [J8]. Longitudinal studies are required to understand the effectiveness of the co-presenter agent on the audience engagement when they are exposed to the agent over a longer-term period.

Assessment of Classroom Acceptance

In addition to controlled studies, we have also further assessed our system in classroom settings. One of our co-authors gave a 90-minute lecture to 20 undergraduates with the virtual co-presenter. Students commented on the engaging nature of the co-presenter agent: *"It was a fresh approach to class"*, *"It forced me to pay attention a little more"*, and *"The change in presenter changed things up nicely and led to better attention"*. This positive feedback further confirmed the acceptability and practicality of our system.

LIMITATIONS AND FUTURE WORK

Results of our user studies provided a strong support for the acceptability and potential of a virtual co-presenter to improve the overall experience of both audiences and presenters, especially for novice and non-native speakers. However, with the small sample size and the presence of confounding factors such as novelty effects, it is difficult to generalize these findings without further studies. As future work, we aim to evaluate our system in longitudinal studies within more naturalistic settings, using larger and more diverse user groups who have varying speech, language and public speaking skills.

Our current co-presenter system was relatively simplistic and only provided limited support for extemporaneous presentations. We plan to improve the usability of our system through the incorporation of more natural human-agent interaction modalities, potentially using speech and gesture recognition. We also aim to employ sensing technologies for real-time monitoring of presentation progress, human presenter's performance and audience interaction. This would allow our system to offer timely support for the human presenter and dynamically adapt to changes in the presentation context. Finally, we wish to explore the applicability of virtual agents to various presentation formats, from presenting in front of an audience to remote presentations in virtual environments and online lectures.

ACKNOWLEGEMENTS

We thank Elizabeth Bates and Karlyle Pilones for their help on our user studies, and Zach Berwaldt for his support on the design of our virtual presenter.

REFERENCES

1. André, E., Rist, T., & Müller, J. (1998). WebPersona: a lifelike presentation agent for the World-Wide Web. Knowledge-Based Systems, 11(1), 25-36.
2. Apple Keynote. http://www.apple.com/iwork/keynote.
3. Aylett, M. P., & Pidcock, C. J. (2007). The CereVoice characterful speech synthesiser SDK. Lecture Notes in Computer Science, 4722, 413.
4. Bishop, J. B., Bauer, K.W., & Becker, E.T. (1998). A survey of counseling needs of male and female college students. Journal of College Student Development, 39, 205-210.
5. Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2004). BEAT: The behavior expression animation toolkit. Life-Like Characters, 163-185. Springer Berlin Heidelberg.
6. Daly, J. A., McCroskey, J. C., Ayres, J., Hopf, T., & Ayres, D. M. (1997). Avoiding communication: shyness, reticence, & communication apprehension (2nd ed.). Hampton Press.
7. Daly, J. A., Vangelisti, A. L., & Weber, D. J. (1995). Speech anxiety affects how people prepare speeches: A protocol analysis of the preparation processes of speakers. Communication Monographs, 62, 383 -397.
8. Edge, D., Savage, J., Yatani, K. (2013). HyperSlides: dynamic presentation prototyping. CHI'13, 671-680.
9. Goodman, A. (2006). Why bad presentations happen to good causes. Andy Goodman & Cause Communication.
10. Hoque, M. E., Courgeon, M., Martin, J. C., Mutlu, B., & Picard, R. W. (2013). MACH: my automated conversation coach. UBICOMP'13, 697-706.
11. Jackson, J. M., & Latané, B. (1981). All alone in front of all those people: Stage fright as a function of number and type of co-performers and audience. Journal of Personality and Social Psychology, 40(1), 73.
12. Kang, N., Brinkman, W. P., van Riemsdijk, M. B., & Neerincx, M. A. (2013). An expressive virtual audience with flexible behavioral styles. IEEE Transactions on Affective Computing, 4(4), 326-340.
13. Kasl, S. V., & Mahl, G. F. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. Journal of Applied Behavioral Analysis, 1, 425-433.
14. Latané, B., & Nida, S. (1980). Social impact theory and group influence: A social engineering perspective. Psychology of Group Influence.
15. McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. Communication Research Reports, 5, 108-113.
16. Menzel, K. E., & Carrell, L. J. (1994). The relationship between preparation and performance in public speaking. Communication Education, 43, 17 - 26.
17. Microsoft PowerPoint. http://office.microsoft.com/en-us/powerpoint.
18. Nijholt, A., van Welbergen, H., & Zwiers, J. (2005). Introducing an embodied virtual presenter agent in a virtual meeting room. Artificial Intelligence and Applications, 579-584.
19. Noma, T., Badler, N. I., & Zhao, L. (2000). Design of a virtual human presenter. Center for Human Modeling and Simulation, 75.
20. Paul G. L. (1966). Insight and desensitization in psychotherapy: An experiment in anxiety reduction. Stanford University Press.
21. Pertaub, D. P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. Presence: Teleoperators and Virtual Environments, 11(1), 68-78.
22. Saket, B., Yang, S., Tan, H., Yatani, K., & Edge, D. (2014). TalkZones: section-based time support for presentations. MobileHCI'14, 263-272.
23. Spielberger, C. D. (1989). State-trait anxiety inventory: bibliography (2nd ed.). Consulting Psychologists Press.
24. Trinh, H., Yatani, K., & Edge, D. (2014). PitchPerfect: integrated rehearsal environment for structured presentation preparation. CHI'14, 1571-1580.
25. Turning Technologies. http://www.turningtechnologies.com/response-solutions/responsecard-ir.