Dongwook Yoon · Research Statement

My research vision is to ***bring the richness of human-to-human interaction into online collaboration tools***. Face-to-face interaction, as compared to meeting in a virtual way, offers unmatched expressivity for conveying complex ideas and nuanced emotions (e.g., emotions embedded in voice inflection or the unspoken meaning of a pointed finger). To enhance expressivity of virtual collaboration tools, several studies have leveraged multi-modal user inputs, i.e., interactions through multiple communication channels (e.g., speech, writing, and gesture), but the key challenge is that people using the systems have difficulty creating, managing, and sharing the resulting multi-media content (e.g., editing a recorded voice comment is not as easy as editing text). To address such problems, my approach establishes ***a research framework for creating rich collaboration systems*** by:

- translating natural human interactions into combinations of intuitive multi-modal user input (e.g., voice + inking + gesture in RichReview [4], and hand grasp + tablet motion in Grasp + Motion [6]).
- designing fluid interfaces for lightweight and high performance interactions by integrating the interface into the flow of users' physical and cognitive operations [1,3].
- building high-fidelity systems for deployment in real world contexts, especially classrooms, where we can compare expected outcomes with what actually happens on the ground [5].

Through my projects, I demonstrate the potential of rich collaboration systems for ***enriching the way people learn, communicate, and collaborate online.*** Successive rounds of field deployments targeting both students and instructors have evaluated the efficacy and applicability of my rich collaboration systems [5]. The premier example is a large-scale deployment of the RichReview system in one of Cornell's massive open online courses (MOOC) where more than 150 students used the tool for peer discussion assignments. We are currently working with Cornell Academic Technology to launch RichReview as an official discussion tool for university classes.

Multi-Modal Commenting for Enriching Document-Centered Collaboration.

At present, digital text comments are the default option for collaborating on shared online documents. However, people prefer to meet in person, often at great cost in time and effort, because discussion is most effective when people can talk and gesture over the document. To enhance expressivity of online collaboration over documents, my research translates the communication channels of face-to-face meetings into multi-modal components of digital comments.

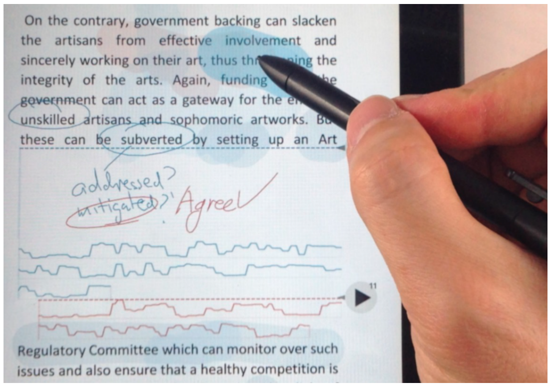

On the contrary, government backing can slacken
the artisans from effective involvement and
sincerely working on their art, thus th ing the
integrity of the arts. Again, funding he
government can act as a gateway for the e
unskilled artisans and sophomoric artworks. Bu
these can be subverted by setting up an Art
addressed?
mitigated?
Agree
11
Regulatory Committee which can monitor over such
issues and also ensure that a healthy competition is

Figure 1. RichReview integrates natural communication modalities of face-to-face meetings, such as voice (waveform), pen-writing, and pointing gestures (blobs in blue), into digital document annotation.

RichReview[1], a multi-modal commenting system, ***simultaneously records multiple aspects of users' communicative expressions and replays them in a synchronized fashion at remote locations*** [4]. For instance, as shown in Figure 1, a commenter can record speech to provide verbal descriptions while creating digital ink markups and hovering the stylus over the screen to point and refer to different parts of the page. On the recipient side, these multi-media components are replayed in synchronicity, thereby delivering vivid and engaging sensations as if the commenter is narrating over, writing on, and pointing at the page.

The key to the success of rich collaboration systems is to minimize the effort required to produce and consume high-density multi-modal content. The following studies illustrate my approach to facilitating easy production and enabling lightweight consumption.

[1] Video of RichReview: https://youtu.be/twSTqxghHNQ

Enhancing Production for Easy Speech Editing.

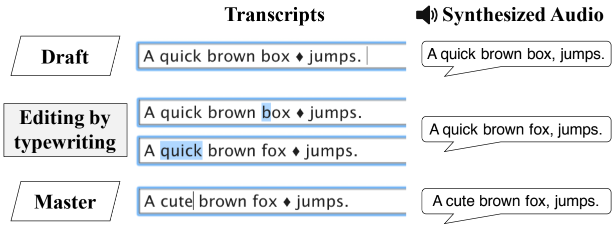

Transcripts
Synthesized Audio
Draft
A quick brown box ♦ jumps.
A quick brown box, jumps.
Editing by typewriting
A quick brown box ♦ jumps.
A quick brown fox ♦ jumps.
A quick brown fox, jumps.
Master
A cute brown fox ♦ jumps.
A cute brown fox, jumps.

Figure 2. TypeTalker synthesizes voice from auto-transcribed text to make speech editing as effortless as text editing.

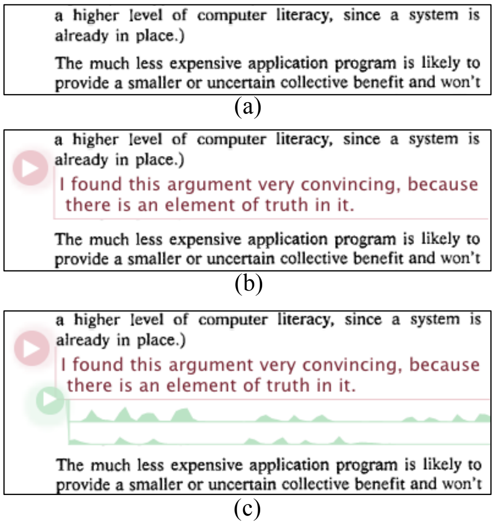
To examine how people produce speech comments, we deployed RichReview to a class at Cornell for a peer discussion activity where all students were asked to create comments [5]. We found that the difficulty of speech editing was the major barrier for students in creating multi-modal comments. Existing speech editing tools are based on a waveform interface, which is too laborious for student users, or on timestamped auto-captions as a surrogate for editing [2], which doesn't support text-like insertion of content or word-level changes (e.g., editing articles or adding the plural suffix 's').

To reduce the effort of speech editing, our new interface, TypeTalker, narrates the auto-transcribed content of the original speech using the synthesized voice of a text-to-speech engine in lieu of the user's own voice [1]. Any edits made on the text captions are reflected back to the audio through text-to-speech technology (see Figure 2). At the same time, TypeTalker goes beyond mere text-to-speech approaches by retaining the core richness of the original speech, such as subtle pauses and pointing gestures. With this approach, ***editing voice becomes as fluid and effortless as editing text.*** In our follow-up study, we evaluated TypeTalker against the previous interface [2] for voice commenting tasks in an educational setting. The results showed that our voice synthesis-based approach ***lowers participants' effort level*** for creating voice comments, and furthermore ***reduces their anxiety*** about making speech disfluencies while recording, because now they can correct mistakes post hoc.

Enhancing Consumption for Lightweight Speech Browsing.

Listening to a recorded voice comment using a traditional scroll-bar interface is tedious and time-consuming. The bar tells little about its content, so the user has to listen through the audio from beginning to end to find the target content. To address this speech consumption problem, RichReview offers ***fluid interactions for semantic audio indexing interfaces such as time-stamped auto-captions*** (or waveforms when automatic transcription is not available, see Figure 3). For instance, the user can visually spot a keyword and simply tap on it to jump to the corresponding time frame and replay the speech with all other modalities synchronized.

The system also employs the TextTearing interaction [3] that ***interlaces the voice comments in-between lines of text by dynamically adjusting the page layout*** (see Figure 3). This fluid layout approach has several benefits. First, it facilitates speech browsing by making all available information visible by default instead of hiding it behind pop-up buttons. Second, it locates comments near their anchor location without occluding the surrounding text. Third, it enables a voice comment to be commented on as if it was a text comment, forming a conversation thread.


a higher level of computer literacy, since a system is already in place.)

The much less expensive application program is likely to provide a smaller or uncertain collective benefit and won't

(a)

a higher level of computer literacy, since a system is already in place.)

I found this argument very convincing, because there is an element of truth in it.

The much less expensive application program is likely to provide a smaller or uncertain collective benefit and won't

(b)

a higher level of computer literacy, since a system is already in place.)

I found this argument very convincing, because there is an element of truth in it.

The much less expensive application program is likely to provide a smaller or uncertain collective benefit and won't

(c)

Figure 3. The TextTearing technique interlaces speech comments in the fluid document layout [3]. (a) Text before commenting. (b) Creating a speech comment opens a space, and inserts it in-between lines of text. (c) Another user (green, waveform) can comment upon the existing voice comment.

Field deployment. A field deployment to a small seminar class at Cornell indicated several ways in which the system facilitates students' understanding of the instructor's feedback on their writing [5]. The students found that the rich comments successfully conveyed the nuances and emotions of the instructor. More importantly, the accessible consumption interface allowed them to re-listen to parts of a comment until they reached full understanding of the instructor's intention.

Sensing Tablet Grasp + Motion for Collaborative Reading.

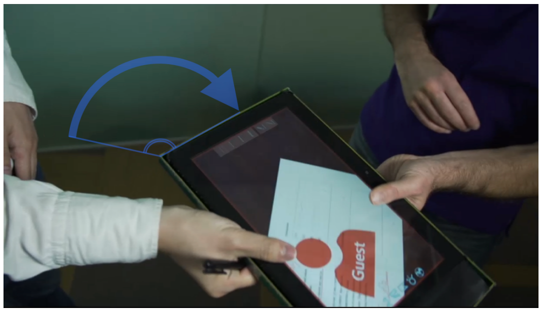


Figure 4. The Grasp + Motion system senses a sequence of tablet rotation and hand grips to capture hand-off interactions in face-to-face collaboration.

Beyond verbal comments, people in collaborative contexts often communicate over a shared document through physical interactions such as grasp and motion. For example, a person can slant a document toward others to draw their attention to a given page; and as collaborators take turns, the document they are talking about may move from the grasp of one person to that of another. My tablet Grasp + Motion system[2] demonstrates ***how capacitive around-the-body touch sensing and motion sensing can be combined to capture and address collaborative contexts*** (see Figure 4) [6]. The design of these grasp interactions is directly informed by people's physical interactions in natural collaborative contexts. The resulting fluid interactions aim to keep users in the flow of their primary tasks while affording embodied ways to interact with the tablet and the collaborators.

Future Research on Innovative Classroom Applications.

Innovating work processes in the education sector.

My research is uniquely targeted to innovate the core work process of instructors and students in the education sector, such as feedback or discussion activities. This approach will have a strong impact on a broad user base because it responds to what people need by innovating existing processes that they are already doing, rather than asking them to do something new. In the future, I will continue to seek opportunities to apply my rich collaboration approach to enhance many different aspects of educational activities, including design feedback, classroom presentation, assignment grading, and course evaluation.

Technical intervention to emotions of students and instructors.

In an educational setting, the emotional response from an instructor to a student is a double-sided sword that can either boost or discourage learning significantly. My studies have shown that carefully designed speech interfaces not only better convey an instructor's emotion but also reduce the student's anxiety and self-consciousness. While my past studies explored the systems with richer media characteristics for enhanced emotional supports, I want to move beyond the adjustment of media characteristics and into active intervention techniques. In the future, I will work on building future generations of classroom communication systems that can intelligently capture and interpret on-the-fly emotional disturbances or outburst of instructors and students to give personalized feedback for their own reflection in educational activities such as feedback or discussion.

Future Research on Rich Collaboration Systems.

Expanding communication vocabularies.

One unique aspect of my research program is that it views new input modalities as added communication channels that connect people by conveying subtle nuances and emotions, rather than as a way to operate a system. In the near future, I plan to leverage advanced sensing techniques to add additional forms of user input to the vocabulary of remote communication. Examples include augmenting the RichReview system with indicators of facial expression and extending the Grasp + Motion system with full body orientation.

Mixed reality-based rich collaboration.

The devices and software for mixed reality (MR, i.e., virtual or augmented reality) are inherently designed to capture and reproduce bodily, personal, and immersive interactions. Previous studies on virtual collaborative

[2] Video of the Grasp + Motion system: https://youtu.be/mTSfoh-M88w

environments focused primarily on connecting distant workers across spatial barriers. Here I see opportunities of applying my multi-modal commenting approach to build ***MR-based asynchronous collaboration systems*** that convey the presence of co-workers across time constraints. As in 2D settings, the key to success will be to balance the workloads for consuming and producing rich contributions.

A new challenge in ***consuming*** 3D multi-modal comments is the design of ***navigational cues***. In document annotation, the navigational cues (e.g., waveforms, captions, and gesture traces) can be flattened on the 2D page. Visualizing navigational cues in 3D, however, demands a hyper-space (4D) representation, which is difficult to comprehend. My solution, inspired by multiple exposure photography, is to reduce the dimensions of the navigational cues by selectively visualizing snapshots of the 3D comments (e.g., the moment when the speaker offers a deictic pronoun, or when a significant change of the scene is detected).

For effective ***production***, easy editing of recorded comments is as critical as ever. Yet, switching to the desktop setting for text-like editing takes extra effort in MR settings (e.g., taking off the headset). In future work, I plan to support fluid editing interfaces for MR content by (1) repurposing snapshots of 3D comments from the consumption interface as delimiters for editing, (2) designing spatial interactions for speech editing (e.g., manipulating auto-captions as if they are physical objects), and (3) supporting a fluid switch between the MR interface and the traditional interface (e.g., stylus + interactive surfaces as a part of the MR environment).

The extension of my rich collaboration systems approach has the potential to offer the experience of working together in the same physical space at the same time. and, thereby, it will change how online collaborators work together in a wide variety of domains in design, business, military, and education.

References

1. Ian Arawjo, **Dongwook Yoon**, and François Guimbretière. TypeTalker: A Speech Synthesis-Based Multi-Modal Commenting System. In *Proc. CSCW 2017: ACM Conference on Computer-Supported Cooperative Work & Social Computing*.
2. Venkatesh Sivaraman, **Dongwook Yoon**, and Piotr Mitros. Simplified Audio Production in Asynchronous Voice-Based Discussions. In *Proc. of CHI '16: ACM Conference on Human Factors in Computing Systems*.
3. **Dongwook Yoon**, Nicholas Chen, and François Guimbretière. TextTearing: Opening White Space for Digital Ink Annotation. In *Proc. of UIST '13: ACM Symposium on User Interface Software and Technology*.
4. **Dongwook Yoon**, Nicholas Chen, François Guimbretière, and Abigail Sellen. RichReview: Blending Ink, Speech, and Gesture to Support Collaborative Document Review. In *Proc. of UIST '14: ACM Symposium on User Interface Software and Technology*.
5. **Dongwook Yoon**, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E Löckenhoff, Steven J Jackson, Abigail Sellen, and François Guimbretière. RichReview++: Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. In *Proc. of CSCW '16: ACM Conference on Computer-Supported Cooperative Work & Social Computing*.
6. **Dongwook Yoon**, Ken Hinckley, Hrvoje Benko, François Guimbretière, Pourang Irani, Michel Pahud, and Marcel Gavriliu. Sensing tablet grasp + micro-mobility for active reading. In *Proc. of UIST 2015: ACM Symposium on User Interface Software and Technology*.