Creating Crowdsourced Research Talks at Scale

Rajan Vaish*
Stanford University
rvaish@cs.stanford.edu

Shirish Goyal
Stanford University
shirish.goyal@stanford.edu

Amin Saberi
Stanford University
saberi@stanford.edu

Sharad Goel
Stanford University
scgoel@stanford.edu

ABSTRACT

There has been a marked shift towards learning and consuming information through video. Most academic research, however, is still distributed only in text form, as researchers often have limited time, resources, and incentives to create video versions of their work. To address this gap, we propose, deploy, and evaluate a scalable, end-to-end system for crowdsourcing the creation of short, 5-minute research videos based on academic papers. Doing so requires solving complex coordination and collaborative video production problems. To assist coordination, we designed a structured workflow that enables efficient delegation of tasks, while also motivating the crowd through a collaborative learning environment. To facilitate video production, we developed an online tool with which groups can make micro-audio recordings that are automatically stitched together to create a complete talk. We tested this approach with a group of volunteers recruited from 52 countries through an open call. This distributed crowd produced over 100 video talks in 12 languages based on papers from top-tier computer science conferences. The produced talks consistently received high ratings from a diverse group of non-experts and experts, including the authors of the original papers. These results indicate that our crowdsourcing approach is a promising method for producing high-quality research talks at scale, increasing the distribution and accessibility of scientific knowledge.

ACM Reference Format:
Rajan Vaish, Shirish Goyal, Amin Saberi, and Sharad Goel. 2018. Creating Crowdsourced Research Talks at Scale. In *Proceedings of The Web Conference 2018 (WWW 2018)*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3178876.3186031

1 INTRODUCTION

There is growing demand for learning and consuming scientific information through video [38]. This demand has in part been met by MOOCs [20], which typically focus on in-depth presentations of established areas, and by efforts such as "Two Minute Papers" and "Papers We Love", which distill scientific ideas for viewers with limited technical expertise. But the vast majority of contemporary research is still available only in the form of text, as traditional academic papers, in part because individual researchers often have limited time, resources, and incentives to produce video-based summaries of their work. This gap prompts a challenge: distilling the content of academic papers into short presentations suitable for students and researchers, and doing so at scale.

Here we introduce and evaluate a system for creating an open, multilingual repository of 5-minute lightening talks developed collaboratively by volunteers worldwide. These talks are catered to technically knowledgeable viewers, who after watching the video summary might read the original papers or attend a longer conference presentation. The initial videos are produced by distributed teams of individuals working in close collaboration; the videos can subsequently be edited and improved by any interested participant. Our project increases the accessibility of scientific knowledge by converting English-language research papers into video-based talks produced in multiple languages—and all without involving the authors of the paper or other domain experts. In the process of creating this content, volunteer contributors learn collaboratively, furthering educational opportunities and incentivizing participation.

Crowdsourcing such an open-ended expert task poses two key challenges. First it is not immediately clear how to collaboratively produce editable videos. Second, one must facilitate extended and complex coordination between large, distributed groups of individuals of varying expertise. To address the first task, we standardize each talk to consist of slides, a written script, and voice-overs; we then programmatically stitch these components together to produce a complete video presentation. To streamline this process, we created an online tool that lets people collaborate and seamlessly record audio on a slide-by-slide basis. Our modular approach supports efficient editing and reduces retake time, both during and after the initial videos are created. We address the second challenge by designing a structured scaffolding process to coordinate volunteers [28, 41]. Specifically, we divide the talk creation process into three discrete phases spanning a period of 21 days (three weeks): (1) on-boarding the crowd and forming teams; (2) generating a slide deck that includes both the talk slides and a slide-by-slide script of the talk; and (3) converting the script to slide-by-slide audio recordings, and reviewing the complete video presentation.

To test this system, we issued an open call for participation, attracting 840 people from 52 countries. This crowd of volunteers created 107 lightening talks in 12 languages based on 40 recent papers from top-tier computer science conferences and scientific journals. These talks were entirely created by the crowd, from designing the structure to producing the content. To evaluate the talks,

*The author's current affiliation is Snap Inc. This work was done while he was a postdoc at Stanford University.

WWW 2018, April 23–27, 2018, Lyons, France

ACM ISBN 978-1-4503-5639-8/18/04.
https://doi.org/10.1145/3178876.3186031

we solicited over 300 responses from outside reviewers, including 73 responses from the authors of the original papers. The talks were rated highly on both presentation quality and utility, receiving a median score of 4 out of 5 on both dimensions.

To explore the applicability of our approach for creating longer and more in-depth content, we experimented with developing technical tutorials on Python and machine learning. These tutorials ranged in length from 30 minutes to 5 hours, and were produced in multiple languages. As with the short research talks, these technical tutorials received high marks by outside evaluators.

Our primary contribution in this work is developing an end-to-end process for producing short, technical research talks from academic papers. To accomplish this, we introduce a new crowdsourcing workflow for achieving open-ended creative goals, built an online tool to facilitate collaborative video production, and analyzed a large-scale, long-term deployment of the method. Our work points to the potential for scalable creation and dissemination of video-based technical content, which we hope will help increase the exposure and accessibility of academic research.

2 RELATED WORK

Our work relates to and builds on several intertwined threads of research. We draw on work in collaborative communities to build the crowd [30], research in organizational behavior to coordinate the crowd [23], ideas in video production to engineer online collaboration tools [51], and studies in microproductivity to scale the system [49]. Below we briefly survey these areas.

Communication of scientific ideas

The presentation of scientific ideas—for both experts and non-experts—is a core function of the academic community. Gaps in communication may be partly responsible for widespread science illiteracy among non-experts [57], and may hinder scientific progress among experts. This communication challenge has been addressed in part by the rise of data journalism and blogging [13], and by tools for article discovery and summarization [50].

Among experts, the primary mode of science communication is text, in the form of papers. But in many domains, a rapid shift is underway towards consuming information in multimedia formats [10, 44]. In advertising [36], education [20, 25], and health [21], videos are becoming an important means of communication and learning [8]. In response to this shift, there are several initiatives to create videos that introduce research papers [59] and present their findings [1]. Those efforts, however, are generally driven by a small group of individuals, and lack the resources necessary to scale. Similarly, conference organizers like the ACM often have limited video production resources, and equipment is provided to conferences on a first-come, first-serve basis [7]. By utilizing crowdsourcing, our work helps to scale the process of video production and to disseminate research.

Crowd workflows and organizational behavior

Crowdsourcing techniques have helped researchers and industry professionals scale a variety of efforts, from microtasks like image labeling [43, 52] and translation [48] to creative expert tasks like producing animations [30] and designing software prototypes [41]. The successful crowdsourcing of expert tasks often relies on multi-stage crowd workflows and organizational structures to facilitate complex collaboration.

Research utilizing crowd workflows has inspired the design of our approach. Soylent [11] showed that splitting tasks into the find-fix-verify stages can improve the quality and accuracy of crowd workers' results. Multi-stage crowdsourcing workflows have also been designed to improve the tone of emails [53], to provide critiques to designers [33], and to improve the learning experience of existing how-to videos with step-by-step annotations [27]. These applications have demonstrated that crowd workflows can yield results comparable to those of experts. Our work contributes to this line of research by introducing and evaluating a new three-phased workflow tailored to the production of short research talks, an open-ended expert task.

Along with workflow design, research on organizational behavior offers insights for enabling effective team coordination. We draw on such work to develop our scaffolding process for crowd collaboration [15, 23, 31, 32, 41]. Past work has identified several obstacles to effective team coordination—such as technology-mediated communication, geographic dispersion, and dynamic team membership [23, 24, 39]—and has proposed solutions for mitigating attrition and motivating volunteers. Many of these findings are based on existing platforms with a critical mass of crowd workers, like Wikipedia, NewGround, and oDesk. For our real-world deployment, we built a community from scratch, and thus adapted these findings to our setting. For example, unlike asynchronous collaboration [14] that is popular when a large crowd is available, we rely on synchronous team-based collaboration [35] to ensure successful task completion. Noting the importance of roles and leadership in online collaborative environments [32], we rely on "directly responsible individuals" (DRIs) [41] to lead the dynamic teams.

Crowdsourced creative production

Crowdsourcing creative tasks in an online, distributed environment is inherently challenging. Past work has typically utilized a paid-crowdsourcing marketplace, a competition-based approach, or engaged the crowd only for specific, well-scoped subtasks. Projects like Ensemble [26] mobilized a crowd from Amazon Mechanical Turk to create stories, while Flash Teams [41] used a crowd of Upwork workers to design software prototypes and animation videos. On Tongal [6], requesters can crowdsource multiple ideas and detailed pitches for video production. However, unlike collaborative crowd-based idea generation systems like the IdeaHound [45], Tongal uses a paid, competition-based approach. Rather than collaborating, teams—often including professional media studios—compete against one another for the best idea, best pitch and a contract to produce the video. Winners for each phase are chosen by the Tongal staff and requesters, not the community. Initiatives like the Johnny Cash Project [3] mobilized volunteers to draw their own portrait of Johnny Cash to be integrated into a collective music video. In this case, though the creative content was crowdsourced, the task was quite specific and there was little collaboration between crowd members. In contrast to past efforts, our work mobilizes a volunteer crowd to collaborate with participants worldwide throughout the

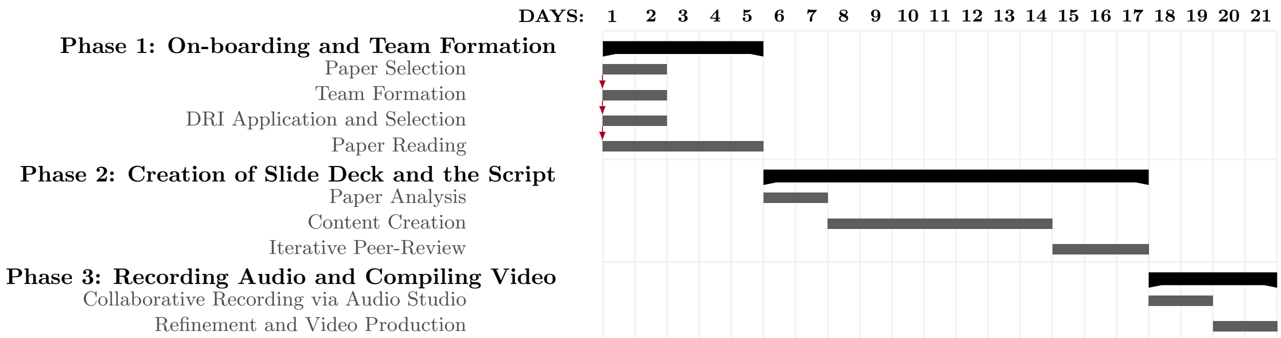

Figure 1: We use a three-phase, three-week long structured workflow to convert research papers to video talks.

entire creative process, from selecting research papers to final video production.

Collaborative video creation

Despite recent advances in audio and video production technology [42, 51], options for collaborative editing are still limited [2] and are often proprietary [18]. Most existing video production systems rely on a traditional timeline model, in which contributors make frame-by-frame edits. We introduce a new approach to collaborative video production. By drawing on and advancing research in microtasking [11, 49], we produce research talks by separately creating talk slides, written scripts, and audio recordings. These components are then programmatically stitched together to create a complete video presentation.

Online learning

Collaborative, online learning is an active area of research [16, 56]. Studies show that collaborative learning on a global scale enhances critical thinking [22] and engagement [9], and varies along cultural dimensions [54]. Its positive effects have been applied in academia and industry [4]. Additionally, research in education and "learnersourcing" [58] suggests that presenting learners with subgoals for procedural tasks improves learning. Margulieux et al. [34] showed that instructions including both specific steps and subgoals resulted in improved learning and transfer compared to those with the specific steps alone. Combining these findings, we set up a collaborative learning environment in Slack, and designed the workflow with subgoals such as paper reading and analysis. The crowd collaboratively worked toward these goals, learning about the latest research in computer science while producing educational content for others.

3 SYSTEM DESIGN

As outlined in Figure 1, our approach to crowdsourcing the creation of research videos proceeds in three phases, spanning 21 days in total. To achieve scale, multiple talks are produced in parallel following the same timeline. The first phase involves on-boarding the crowd, forming teams, and reading the papers to be converted. The second phase entails creating a slide deck and a written script for each paper. In the third phase, the crowd converts the scripts to slide-by-slide audio recordings, and the completed video talks are then reviewed for final improvements. Before the formal talk creation process begins, we recruit a crowd through an open call for participation, as discussed in Section 4.1. Our deployments typically involve creating 10 talks in parallel, with a total of 50–100 active participants. We describe this process in detail below.

3.1 Phase 1: On-boarding, paper selection, team formation, and paper reading

This first phase spans a period of five days. We start by providing participants documentation about the tools we use, instructions about the workflow, and best practices to follow. As people worldwide participate and contribute, the best practices continue to evolve and grow. The crowd next selects papers to convert into talks. This process proceeds in a free-form fashion, with participants proposing papers and soliciting votes of support from others. The most popular papers are selected for conversion, and participants then choose which paper team to join. Participants are free to join any team, irrespective of their location; there is no cap on team size. Each team works toward one talk, based on one paper. After joining a team, each member begins reading the paper. Communication between team members is primarily carried out on Slack, a popular tool for text-based messaging.

By letting participants select the papers themselves, it helps ensure crowd workers have adequate knowledge in the topic. It also seems like a reasonable policy since participants are volunteering their time. This flexibility, however, can result in favoring papers that are already well-known, exacerbating the so-called Matthew effect in science [37], in which the "rich get richer and the poor get poorer". Indeed, some of the papers selected by the crowd (e.g., "Mastering the Game of Go with Deep Neural Networks and Tree Search" [46]) had already garnered significant attention in the media. By and large, though, we found that the talks selected by the crowd span a diverse range.

After selecting the papers and joining teams, each team selects two to three *directly responsible individuals* (DRIs) to oversee the talk creation process [29, 32]. DRIs set the tone of the talk, help team members with any problems they might encounter, and ensure the team adheres to the timeline while meeting quality standards. DRIs are a sought-after role, as it confers decision-making power to execute one's vision. Any team member can apply to be a DRI by

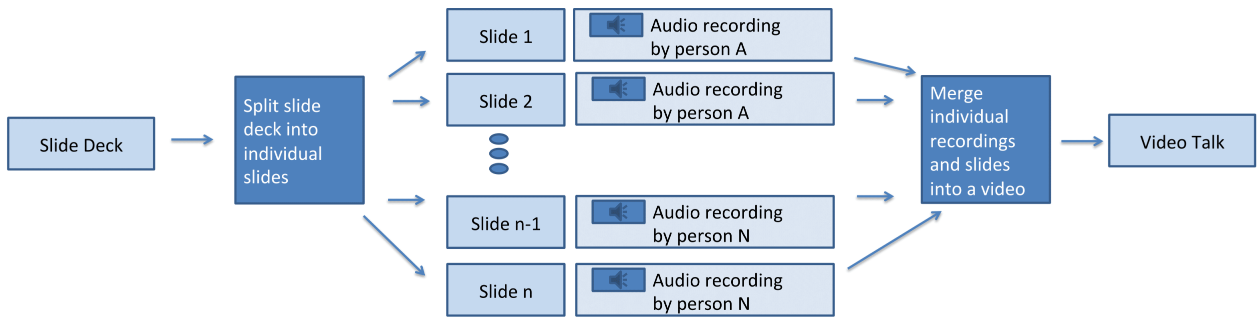

Figure 2: Workflow of Audio Studio, a tool to facilitate collaborate video production. Starting from a slide deck and script, Audio Studio first splits the deck into individual slides and accompanying scripts. Any crowd member can then record audio clips for each slide by reading the script. The slides and audio are then programmatically stitched together to create a complete video.

stating their interest, availability, and expertise in the topic. After a 24-hour window, the team votes and choses a group of DRIs. Having multiple DRIs helps ensure availability across all timezones. To encourage participation, any inactive DRIs are replaced by an active member of the team; this dynamic strategy helps keep DRIs active, while motivating other team members to work consistently to get a chance to become a DRI.

3.2 Phase 2: Creation of slide deck and script

In the second phase, participants begin actively collaborating with their teammates. Phase 2 is itself broken down into three steps. First, the team collectively completes a questionnaire to help participants make sense of the paper they selected and read in Phase 1. Second, they work together to put together the talk slides and slide-by-slide written scripts of what will ultimately be converted to audio. Finally, teams offer feedback on one another's work.

Step 1: Paper analysis. To help crowd members—who are typically non-experts—think critically about the paper they are tasked to present, we require them to collectively answer a series of questions. What is the contribution of the paper? Why is the problem hard? How did authors evaluate the experiment? In total we pose 17 questions, and allot two days to complete this task. Crowd members collaborate and are encouraged to build off of and edit each other's work, and the DRIs in particular help to synthesize answers into a coherent whole. The tone of the talk is largely determined by responses to these questions.

Step 2: Content creation. This second step, which spans seven days, involves creating the entire slide deck, along with written scripts to accompany each slide. Communication is carried out on Slack, and the slides themselves are created with Google Slides, which allows efficient real-time collaboration. The team is free to organize in any manner they see fit to accomplish the task. To promote high-quality content that is consistent across teams, we encourage crowd members to adhere to the following guidelines.

(1) Limit talks to approximately five minutes. Based on past literature [20] and informal pilot studies, we found that short, five-minutes videos were sufficient to convey the key ideas while still maintaining audience engagement and interest.
(2) Target an audience of viewers with technical expertise comparable to those who might read the paper or attend a conference presentation on the research. The papers often assumed a certain level of technical sophistication, and we mirrored this assumption for the talks, limiting discussion of background material. However, particularly for such short talks, this still meant focusing on the key contributions of the paper rather than on detailed technical descriptions.
(3) Write the slide-by-slide scripts exactly as they should be spoken. As described below, the audio recordings are created by individuals who may not have been directly involved in creating the slides. As such, it is critical that the written scripts indicate precisely what should be recorded for each slide. These scripts are added to the "notes" section of each slide of the slide deck.

In addition to providing these guidelines, we point participants to past talks produced by the crowd, as it is often easiest to learn by example. We further provide teams with a set of best practices compiled by previous teams. These include, for example, suggestions to use simple language and to explain the key aspects of figures on slides. Finally, we provide crowd members with a pair of articles on creating effective presentations [19, 55].

Team members were encouraged to improve each other's work by editing or commenting. If there are diverging opinions or conflicts, DRIs serve as arbiters and have final decision-making power. To help motivate crowd members through positive reinforcement, we introduced a simple mechanism for peer acknowledgement. Throughout the talk creation phase, anyone could give a "+1 thank you" to anyone else for their contribution. These acknowledgements were not intended to result in any specific tangible rewards, but rather were meant to show appreciation for a job well done.

Step 3: Peer review. In the final three days of Phase 2, crowd members offer feedback to other talk teams, and work on addressing the comments that they in turn receive. This review process gets

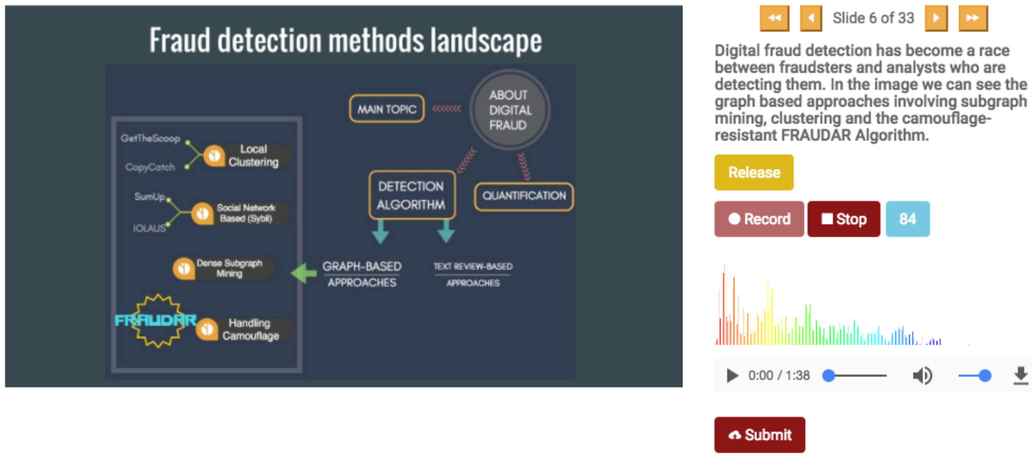

Figure 3: Audio Studio's recording interface, with slide and script next to one other. Contributors can review and re-record audio clips before submitting.

fresh eyes on each presentation, and garners valuable input from individuals who were not involved in creating the talks they are evaluating. Reviewing happens in a free-form iterative process, with talk creators rapidly incorporating feedback and then soliciting more reviews [17]. Teams that address feedback faster can iterate more times; there is no limit to the number of iterations a team can go through.

3.3 Phase 3: Recording audio and compiling the video presentation

At this point in the process, teams have completed their slides and scripts. As current text-to-speech systems [40] are not yet able to produce fully natural audio, crowd participants record the audio themselves. This recording process spans two days, and is facilitated by an online tool that we built, which we call Audio Studio. Audio Studio is a Django application, with front-end built on AngularJS. As shown in Figure 2, Audio Studio starts with the slide deck as input and proceeds in three steps. First, individual slides are extracted and paired with their accompanying scripts. Second, crowd members record audio clips for each slide. Finally, these audio clips are programmatically stitched together with the slides into a complete video presentation. We describe these three steps in more detail below.

(1) **Splitting the deck.** Audio Studio is designed to minimize editing and retake time by splitting the slide deck into individual slides for which audio can be recorded in isolation. Given the URL to a presentation in Google Slides, Audio Studio uses the Google Drive API to extract individual slides as PDFs together with the slide-by-slide scripts stored in the "notes" section of each slide.
(2) **Audio recording.** Any team member can record audio for any slide by reading off the narrative script displayed next to it (as shown in Figure 3). Contributors can replay and re-record audio until they are satisfied with the quality. We do not impose a strict cap on the number of people who can contribute audio to each presentation, but we encourage teams to limit recordings to two people. We further encourage contributors to record continuous slide sections. For example, one individual might record the first half and a second might record the latter half. These contributors are selected by vote using SimplePoll [5] on Slack after an informal audio audition, where the interested team members record a one-minute long introduction of themselves.
(3) **Video creation.** After audio recordings are submitted for every slide in the deck, DRIs review them for quality, and can request improvements if necessary. Our modular approach makes retakes relatively easy. Once the DRIs are satisfied with the recordings, the audio and slides are stitched together with the FFmpeg library to generate a complete video; this compilation step is done automatically within Audio Studio at the click of a button.

After the initial video presentation is produced, teams have two final days to review it and incorporate any additional edits. Our modular approach to the creation of slides, scripts, and audio facilitates rapid editing of all talk components. Once the DRIs approve the final talk, the video is published on YouTube and on the platform's website for the general public.

3.4 Localization

The initial creation of talk presentations is carried out in English, as described in the three-phase process above. Crowd members can then self-organize to develop localized versions of any previously created talk. The crowd forms teams dynamically based on their language of expertise and interest. They are free to utilize any existing assets—including the slides and scripts—and can edit these as they see fit. The localization process can be challenging for several reasons: certain scientific terms or phrases in English do not always have suitable translations; some languages (e.g., Arabic and Japanese) had few contributors in our community; and the crowd must translate both the written scripts as well as text on the slides. Localization was typically carried out by individuals working alone or in small teams, peer-reviewed for quality, and was completed in an unstructured fashion without specific deadlines or DRIs.

Figure 4: Sample slides from talks produced in Japanese, English, Oriya, Chinese, Spanish and Hindi (from left to right, top to bottom). Of the 107 research talks produced, 67 were in foreign languages.

4 REAL-WORLD DEPLOYMENT

To evaluate our system for creating crowdsourced research presentations, we launched an online platform called Stanford Scholar[1] that was open to individuals worldwide.

4.1 An open call for participation

Before the formal talk creation process began, we recruited a crowd through an open call for participation. This global call was made online via social media platforms (e.g., public Facebook and LinkedIn groups and Twitter), public mailing lists (e.g., UW Change and Berkeley TIER), online platforms (e.g., LetMeKnow.in, StudentCompetitions.com), and emails sent to universities worldwide. Such solicitations were repeated periodically. The call sought to motivate potential participants with a chance to help disseminate research ideas while learning about cutting-edge work in a collaborative environment. Contributors were not paid or otherwise compensated, and their participation was voluntary.

4.2 Participant demographics

Our open call for volunteers resulted in 840 sign ups from 52 countries on 6 continents. The majority of participants came from the United States (29%) and India (48%). Participants spanned the educational spectrum. The distribution of highest degree attained or in progress was: 11% high school diploma, 59% undergraduate degree, 24% masters, and 6% Ph.D. 25% of participants were women, and the median age of crowd members was 22 years-old. These participants primarily had a background in computer science or other engineering-related fields. Participants typically had moderate technical expertise in the topic of the papers they worked on. Based on self-reports, the mean expertise was 2.7 (median was 3) on a scale from 1 to 5, with 1 being "novice" and 5 being "expert". Two-thirds of participants had no prior experience preparing or giving a research talk before.

4.3 Outcomes

We carried out two 21-day talk creation rounds following the procedure outlined above. In each iteration, the crowd formed 10 teams to create 10 English-language talks. In preliminary work to develop the process, we ran two additional rounds that approximated our final approach, but which differed in some aspects (e.g., the earlier rounds had a less defined structure). In total, the four iterations of our deployment resulted in the creation of 40 English-language research presentations derived from 40 distinct papers. An additional 67 videos were created in 11 foreign languages, for a total of 107 presentations. To date, these videos have attracted over 50,000 views, 400 shares, and 800 subscribers on YouTube.

The presentations were based on 40 papers published at top-tier computer science conferences and scientific journals in the past two years. These papers spanned the spectrum of computer science, including human-computer interaction, data mining, machine learning, and security. Specifically, the papers were published in: WWW (7), CHI (4), UIST (3), KDD (2), AAAI (2), IJCAI (2) CSCW (2), Nature (1), ICML (3), NIPS (1), SIGMOD (1), CVPR (1), WSDM (1), EuroCrypt (1), OOPSLA (1), VLDB (1), ICWSM (1), SIGIR (1), ECCV (1), IROS (1), ICLR (1), NAACL-HLT (1), and UBICOMP (1). Papers were selected by crowd members, typically from among the best-paper award winners at the conference.

To make scientific knowledge accessible around the world, it is important to distribute research findings in multiple languages. Among the strengths of crowdsourcing is its scalability, and its ability to leverage diverse skills and expertise. Based on the 40 English-language presentations, crowd members created 67 versions in 11 foreign languages: Chinese, Hindi, Spanish, Catalan, Romanian, Oriya, Nepali, Malayalam, Japanese, Filipino and Tamil. As our initiative continues, crowd members have started localizing content into several more languages, including Asante Twi, Albanian, Assamese, Greek, French, Arabic, and Persian. Figure 4 illustrates some examples of localized content produced by the crowd.

[1] https://scholar.stanford.edu

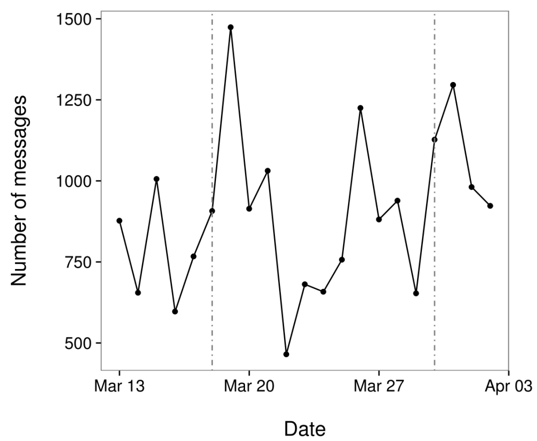

Figure 5: Slack activity during one 21-day video production round, in which 10 talks were simultaneously produced. The vertical lines separate the round into the three development phases. Activity is typically higher on weekends and after a new phase begins.

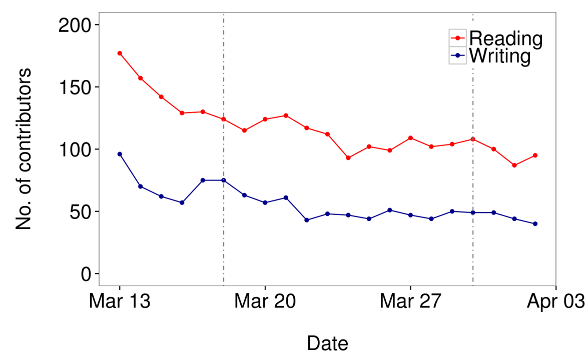

Figure 6: Slack activity during one video production round, in which 10 videos were simultaneously generated, broken down by number of people reading and writing messages each day. The vertical lines separate the round into the three development phases. Participation gradually declines but there is a critical mass of contributors throughout the development cycle.

4.4 Engagement

A campaign's success depends on the continued participation by its members in multiple capacities. In our initiative, Slack was the primary medium of communication and collaboration. In addition to a common channel, each talk had its own public channel. During the course of the project, over 100,000 messages were posted to Slack.

During each 21-day cycle of video production, the crowd worked on 10 talks in parallel and an average of approximately 1,000 messages were exchanged every day. As shown in Figure 5—which corresponds to one specific 21-day round—participation and discussion about the talk increased over the weekend and subdued during weekdays. This behavior likely stems from the fact that this is a voluntary activity, with participants busy at school and work during the week.

On average, talk teams consisted of about 25 crowd members, with about 10 who were regularly active and made substantial contributions. Across the batch of 10 talks created in parallel in a single round, approximately 50 people posted messages to Slack each day, and more than 100 people read the posted messages. Figure 6 shows a gradual decline in the number of people participating each day, which is typical of voluntary initiatives. Nonetheless, we maintained a large contingent of contributors throughout the talk development process.

5 EVALUATION

We conducted several surveys to assess the effectiveness and value of our crowdsourcing approach. These surveys measured: (1) presentation quality and educational value for talk viewers; and (2) process quality and learning experiences for talk creators. To do so, we surveyed three separate groups.

(1) **Paper authors**. We reached out to the authors of the 40 papers on which the presentations were based. As clear subject-matter experts, their feedback provided valuable insight and assessment of the quality of the crowd-generated talks. In total, 73 authors responded to our survey.
(2) **External evaluators**. Besides authors, we reached out to people not affiliated with the papers, from non-experts to domain experts. We solicited responses from various university groups, and also directed people who visited the project website to take our survey after watching a video. We requested input on overall presentation quality and communication of the research results. In total, we received 260 responses.
(3) **Participating crowd**. We requested feedback on the initiative from participants who created the talks, and received 95 responses in total. This group provided feedback on the talk creation process and their experiences, but did not evaluate the quality of the completed talks.

5.1 Presentation quality and educational value

Overall talk quality was evaluated on a Likert scale from 1 to 5, with 5 being "excellent" and 1 being "poor". Among the 73 paper authors we surveyed, the talks were typically rated highly (mean = 4.1, median = 4). Authors also generally thought the talks were useful for someone trying to get an overview of their papers. Talks received a mean of 4.2 and median of 4 on this dimension, with 5 being "extremely useful" and 1 being "not useful". The quotes below illustrate some of the reactions we received from paper authors.

> "The talk is extremely thorough despite its brevity. It's certainly better than the talk I gave at WWW."

> "It does such a great job at motivating the research problem and covers the gist of the paper very well, in language that is engaging to the broader audience."

Not all authors, however, thought the 5-minute format was sufficient to communicate a paper's contribution, as the following quote indicates.

> "The presentation is overall way too fast, so I am not sure how helpful it is for someone who is not already

quite deep into the topic. On the positive side, the presentation might indeed give some intuition for the key contributions of the paper."

External evaluators—who were not affiliated with the papers—echoed the feedback we received from paper authors. Based on 260 responses from this group, the quality of the talks was again rated highly (mean = 3.7, median = 4). Respondents also indicated that they successfully learned the key contributions of the paper, rating their understanding at a mean of 3.7 and median of 4, with 5 being "extremely well" and 1 being "not at all". Notably, a majority of respondents (65%) indicated that they would rather spend five minutes watching the video than skimming the paper. The quotes below show some of the feedback we received.

> "I found this paper to be quite abstract and hard to understand, so the video was a helpful summary. This video format is ideal for cases like this where the message of the paper can be hard to skim."

> "I was pleasantly surprised by how much substance was covered in five minutes. The video seems to have a nice balance between controlling the total duration and still going into a reasonable amount of depth."

> "Sometimes it's difficult to understand the methodology, flow of system or algorithm while reading paper, but in videos with visualization, using charts, flow diagrams or by showing an example it's easy to understand."

Figure 7 indicates that ratings by external evaluators were somewhat higher for those with basic technical expertise in the topic, likely because the talks were explicitly tailored to this group. Of the 260 responses from external evaluators, 73% came from people who rated their expertise as 3, 4 or 5. This subgroup of respondents with domain expertise rated the talk quality at a mean of 3.8 (median = 4) and its understandabilty at a mean of 3.8 (median = 4). In contrast, evaluators without such domain expertise rated talk quality at a mean of 3.4 (median = 3) and its understandability at a mean of 3.3 (median = 3). A t-test shows that these group differences are statistically significant at the 1% level ($p = 0.005$ for quality, and $p = 0.0002$ for understanding).

5.2 Crowd participation and learning experiences

Our approach was designed to attract and retain volunteers to create high-quality talks in a collaborative, educational environment. It is thus important to measure participants' perceptions of the initiative to gauge its sustainability. Overall, participants indicated the initiative was a high-quality experience (mean rating = 4.5, median = 5, with 5 being "great"). Crowd members also indicated support for various structural elements of the process. For example, participants found iterative talk review very useful (mean rating = 4.4, median = 5, with 5 being "extremely useful"). And participants rated the Audio Studio recording system as easy to use (mean rating = 4.0, median = 4, with 5 being "absolutely simple").

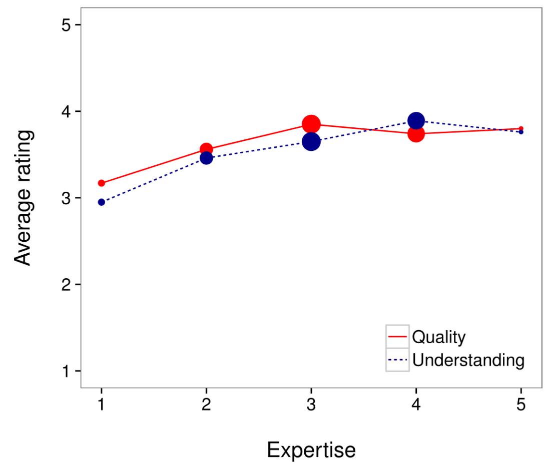

Figure 7: Ratings for video quality and understandability as a function of the evaluator's expertise in the topic.

Our survey indicates that crowd volunteers were motivated to participate in the initiative by an opportunity to learn about research while contributing to a public good. Based on a multiple-choice question answered by 75 crowd members, we found that the top three reported motivations were learning about different research topics (93%), collaborating with people worldwide (82%), and increasing research access (69%). These quantitative results are also reflected in the written feedback we received from participants.

> "I had always wanted to read through research papers on hot topics of computer science. But I could never get started. This program not only inspires me to read through a paper but requires me to understand it enough, so as to create a talk on it. And learning is always fun when more people are learning with us."

> "Before joining the program, I had wanted to contribute to the cause of education and separately learn about the latest trends and research in Computer Science and Technology. The initiative came as a single opportunity for both. 9 months after joining and having contributed to creation of 3 versions of talk creation, I have gained insights into Machine Learning, Deep Learning and HCI research. I am now in a position to contribute to research in a more informed way. I also learned collaboration and teamwork by working with incredible teams of smart and knowledgeable people working from locations all around the world. I also learned presentation skills as we had to interpret and present the real essence of complex research papers in 5 min talks. All these skills will surely be very valuable to me in the future."

As the quotes above suggest, reading papers and creating talks helped crowd participants learn about research topics. Following established self-assessment techniques in education [12, 47], we conducted a survey to better understand this learning experience. We found that before working on the talk, participants rated their expertise in the topic of the paper at a mean of 2.6 on a scale from

1 to 5 (median = 3, N = 29), where 1 is "novice" and 5 is "expert". After working on the talk and researching the topic for three weeks, these ratings rose to a mean of 3.4 (median = 4, N = 29). This difference is statistical significant at the 1% level ($p = 4e{-9}$) with a paired t-test. Further, among 75 surveyed crowd members, they generally indicated having a positive learning experience, rating their experience at a mean of 4.2 (median = 4), where 1 meant "did not learn anything" and 5 meant "learned a lot".

6 CREATING IN-DEPTH LECTURES

Our primary focus in this work was on developing short research presentations; however, our structured crowdsourcing approach also has promise for producing longer and more in-depth lectures. To investigate this potential, we had the crowd create four extended video tutorials in several languages, ranging from 30 minutes to nearly 5 hours. These tutorial were: (1) "A Short Introduction to Python" (37 minutes); (2) "An Extended Introduction to Python" (50 minutes); (3) "An Introduction to Algorithms" (195 minutes); and (4) "Practical Machine Learning with Python" (265 minutes). All videos were originally created in English, and the first lecture ("A Short Introduction to Python") was additionally translated into Arabic, Hindi, Spanish, and Catalan.

These extended lectures were created following the same basic process as that carried out for the 5-minute research talks, but with two key differences. First, the paper analysis step was replaced with syllabus creation. The goal of this step was two-fold: to determine the overall structure of the lecture, and to partition the lecture into short segments of approximately 5-10 minutes that could be carried out in parallel by different teams. Second, given the inherently interrelated nature of the video content, teams would more regularly interact with one another to ensure consistency and flow. The extended videos were rated by 44 external evaluators. In line with reviews for the research talks, these lectures were typically rated highly, receiving a mean rating of 3.9 and median of 4 out of 5, with 5 being "excellent".

7 DISCUSSION & CONCLUSION

There is pressing need to develop new ways to make scientific knowledge available to diverse global audiences. To address this challenge, we developed and evaluated a structured, three-phase approach to crowdsourcing the creation of 5-minute research talks. In the first phase, volunteer crowd members learned about the process, selected papers to present, and formed teams. In the second phase, team members collaborated to critically analyze their selected paper, and to create a slide presentation along with a written slide-by-slide script of the talk. In the third phase, participants used our Audio Studio application to convert the written scripts to audio clips. These audio clips and the slides were programmatically stitched together to create a complete video presentation. In total, volunteers from 52 countries created over 100 talks in 12 languages on papers from top-tier computer science conferences, including WWW, KDD, CHI, and CSCW. Ratings and comments from both the papers' authors and outside evaluators indicate the created videos were consistently high-quality. To date, these videos have attracted over 50,000 views, 400 shares, and 800 subscribers on YouTube.

Reflections and lessons from our study

We believe three key design choices contributed to the success of our method: modularity, structure, and community. By decomposing talks into slides, written scripts, and audio recordings, we could successfully overcome common challenges of collaborative video production. These pieces were further separated into micro-segments whose creation could be efficiently parallelized. Though not a universal solution for collaboratively making all types of videos, we demonstrated that such modularity works well for creating research talks and in-depth technical lectures. This general design pattern may prove useful in a variety of open-ended, presentation-based applications, from education to journalism.

A structured, fault-tolerant workflow—with a well-defined timeline and dynamic role allocation—helped maintain engagement and commitment from the volunteer crowd. This structure helped non-experts quickly learn difficult technical material and produce high-quality content in a relatively short period of time. We note that in informal pilot studies without this structure, it was difficult to achieve high quality—or even complete videos—as crowd members lacked direction and were often working alone. Our dynamic DRI system ensured that teams were continually led by active members, replacing inactive DRIs with active participants.

Finally, we prioritized creating a vibrant community of participants to facilitate sustainability. For example, we adopted a team-based approach to encourage meaningful interactions between participants. We also allowed the crowd to self-organize and provided freedom for self-expression and control over the final product. Many participants appreciated such design choices, stating that their primary motivation for participating was this opportunity to learn challenging scientific concepts in a supportive community. Understanding and responding to such motivations is critical in the case of voluntary crowd work.

Future work

By creating an open repository of multilingual research talks, we have sought to advance the dissemination of scientific knowledge while simultaneously providing a collaborative learning environment for individuals across the world. Moving forward, we plan to focus our efforts on three fronts. First, we aim to create more long-form videos, and to refine our process for collaboratively producing complex, interconnected material. Our preliminary efforts in this direction have produced encouraging results. Second, we seek to continue improving talk quality. Although our current approach consistently yields high-quality talks, there is always room for improvement. For example, one might experiment with animation or interaction to increase audience engagement. Finally, we plan to investigate the longer-term sustainability and scalability of our approach. The results of our study demonstrate it is possible to create a vibrant community of hundreds of volunteers who can work together productively with little external oversight; and to date, this community has thrived for over a year. The next challenge is building an even larger community, one that can produce thousands of presentations on diverse topics, both within and beyond computer science. Our approach is one step toward increasing access to and understanding of scientific research; we hope our work spurs further such efforts.

ACKNOWLEDGMENTS

We thank hundreds of participants of the Stanford Scholar community for their contributions. This work was supported by the Office of Naval Research awards N00014-16-1-2893 and N00014-15-1-2711.

REFERENCES

[1] 2016. Papers We Love. (2016). http://paperswelove.org/.

[2] 2016. WeVideo. (2016). https://www.wevideo.com/.

[3] 2017. The Johnny Cash Project. (2017). http://www.thejohnnycashproject.com/.

[4] 2017. NovoEd. (2017). http://novoed.com/.

[5] 2017. SimplePoll. (2017). https://simplepoll.rocks/.

[6] 2017. Tongal. (2017). https://tongal.com/how.

[7] ACM. 2017. SIGCHI Live and Recorded Video Policy. (2017). http://www.sigchi.org/conferences/Conferences/Policies/sigchi-live-and-recorded-video-policy.

[8] Catherine Adams, Yin Yin, Luis Francisco Vargas Madriz, and C Scott Mullen. 2014. A phenomenology of learning large: the tutorial sphere of xMOOC video lectures. *Distance Education* 35, 2 (2014), 202–216.

[9] Lai Hung Auyeung. 2004. Building a collaborative online learning community: A case study in Hong Kong. *Journal of Educational Computing Research* 31, 2 (2004), 119–136.

[10] Thomas Balslev, Willem S De Grave, Arno MM Muijtjens, and AJJA Scherpbier. 2005. Comparison of text and video cases in a postgraduate problem-based learning format. *Medical education* 39, 11 (2005), 1086–1092.

[11] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. 2015. Soylent: a word processor with a crowd inside. *Commun. ACM* 58, 8 (2015), 85–94.

[12] David Boud. 2013. *Enhancing learning through self-assessment.* Routledge.

[13] Geoff Brumfiel. 2009. Science journalism: Supplanting the old media? *Nature News* 458, 7236 (2009), 274–277.

[14] Axel Bruns. 2008. *Blogs, Wikipedia, Second Life, and beyond: From production to produsage.* Vol. 45. Peter Lang.

[15] Catherine Durnell Cramton. 2001. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science* 12, 3 (2001), 346–371.

[16] David D Curtis and Michael J Lawson. 2001. Exploring collaborative online learning. *Journal of Asynchronous learning networks* 5, 1 (2001), 21–34.

[17] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work.* ACM, 1013–1022.

[18] David A Dudas, James H Kaskade, and Kenneth W O'flaherty. 2007. System and methods for online collaborative video creation. (Jan. 5 2007). US Patent App. 12/159,736.

[19] DukeUniversity. 2016. How to Convert your Paper into a Presentation. (2016). http://twp.duke.edu/uploads/media_items/paper-to-talk.original.pdf.

[20] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14).* ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2556325.2566239

[21] R Hawkins and K Price. 1992. The effects of an education video on patients' requests for postoperative pain relief. *The Australian journal of advanced nursing: a quarterly publication of the Royal Australian Nursing Federation* 10, 4 (1992), 32–40.

[22] William J Haynie III et al. 1998. Collaborative Learning Enhances Critical Thinking. *Volume 7 Issue 1 (fall 1995)* (1998).

[23] Pamela Hinds, Lei Liu, and Joachim Lyon. 2011. Putting the global in global work: An intercultural lens on the practice of cross-national collaboration. *Academy of Management Annals* 5, 1 (2011), 135–188.

[24] Robert S Huckman, Bradley R Staats, and David M Upton. 2009. Team familiarity, role experience, and performance: Evidence from Indian software services. *Management science* 55, 1 (2009), 85–100.

[25] Robin Kay and Ilona Kletskin. 2012. Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers & Education* 59, 2 (2012), 619–627.

[26] Joy Kim, Justin Cheng, and Michael S Bernstein. 2014. Ensemble: exploring complementary strengths of leaders and crowds in creative collaboration. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.* ACM, 745–755.

[27] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J Guo, Robert C Miller, and Krzysztof Z Gajos. 2014. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems.* ACM, 4017–4026.

[28] Aniket Kittur. 2010. Crowdsourcing, Collaboration and Creativity. *XRDS* 17, 2 (Dec. 2010), 22–26. https://doi.org/10.1145/1869086.1869096

[29] Adam Lashinsky. 2012. *Inside Apple.* Hachette Book Group, New York.

[30] Kurt Luther and Amy Bruckman. 2010. Flash collabs: Collaborative innovation networks in online communities of animators. *Procedia-Social and Behavioral Sciences* 2, 4 (2010), 6571–6581.

[31] Kurt Luther, Kelly Caine, Kevin Ziegler, and Amy Bruckman. 2010. Why it works (when it works): Success factors in online creative collaboration. In *Proceedings of the 16th ACM international conference on Supporting group work.* ACM, 1–10.

[32] Kurt Luther, Casey Fiesler, and Amy Bruckman. 2013. Redistributing leadership in online creative collaboration. In *Proceedings of the 2013 conference on Computer supported cooperative work.* ACM, 1007–1022.

[33] Kurt Luther, Amy Pavel, Wei Wu, Jari-lee Tolentino, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2014. CrowdCrit: crowdsourcing and aggregating visual design critique. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing.* ACM, 21–24.

[34] Lauren E Margulieux, Mark Guzdial, and Richard Catrambone. 2012. Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In *Proceedings of the ninth annual international conference on International computing education research.* ACM, 71–78.

[35] Elaine McCreary and Madge Brochet. 1992. Collaboration in international online teams. In *Collaborative learning through computer conferencing.* Springer, 69–85.

[36] Tao Mei, Xian-Sheng Hua, Linjun Yang, and Shipeng Li. 2007. VideoSense: towards effective online video advertising. In *Proceedings of the 15th ACM international conference on Multimedia.* ACM, 1075–1084.

[37] Robert K Merton et al. 1968. The Matthew effect in science. *Science* 159, 3810 (1968), 56–63.

[38] Jeff Nevid and Alejandro Franco Jaramillo. 2011. Teaching the millennials. *APS Observer* 24 (2011), 53–56.

[39] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human-computer interaction* 15, 2 (2000), 139–178.

[40] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. *arXiv preprint arXiv:1609.03499* (2016).

[41] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S. Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S. Bernstein. 2014. Expert Crowdsourcing with Flash Teams. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14).* ACM, New York, NY, USA, 75–85. https://doi.org/10.1145/2642918.2647409

[42] Steve Rubin, Floraine Berthouzoz, Gautham J Mysore, Wilmot Li, and Maneesh Agrawala. 2013. Content-based tools for editing audio stories. In *Proceedings of the 26th annual ACM symposium on User interface software and technology.* ACM, 113–122.

[43] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.

[44] SearchEngineLand. 2015. YouTube Ranking Factors: Getting Ranked In The Second Largest Search Engine. (2015). https://goo.gl/FS8Ki4.

[45] Pao Siangliulue, Joel Chan, Steven P Dow, and Krzysztof Z Gajos. 2016. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-powered Real-time Semantic Modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology.* ACM, 609–624.

[46] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

[47] Dominique Sluijsmans, Filip Dochy, and George Moerkerke. 1998. Creating a learning environment by using self-, peer-and co-assessment. *Learning environments research* 1, 3 (1998), 293–319.

[48] Gwyneth Sutherlin. 2013. A voice in the crowd: Broader implications for crowdsourcing translation during crisis. *Journal of information science* (2013), 0165551512471593.

[49] Jaime Teevan, Shamsi T Iqbal, Carrie J Cai, Jeffrey P Bigham, Michael S Bernstein, and Elizabeth M Gerber. 2016. Productivity Decomposed: Getting Big Things Done with Little Microtasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 3500–3507.

[50] Simone Teufel and Marc Moens. 2002. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Comput. Linguist.* 28, 4 (Dec. 2002), 409–445. https://doi.org/10.1162/089120102762671936

[51] Anh Truong, Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. 2016. Quickcut: An interactive tool for editing narrated video. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology.* ACM, 497–507.

[52] Rajan Vaish, Sascha T Ishikawa, Sheng Lundquist, Reid Porter, and James Davis. 2013. Human Computation for Object Detection. *Tech Report UCSC-SOE-15-03* (2013).

[53] Rajan Vaish and Andrés Monroy-Hernández. 2017. CrowdTone: Crowd-powered tone feedback and improvement system for emails. *MSR-TR-2017-1* (2017).

[54] Ravi Vatrapu and Dan Suthers. 2007. Culture and computers: A review of the concept of culture and implications for intercultural collaborative online learning. In *Intercultural Collaboration.* Springer, 260–275.

[55] VirginiaTech. 2016. How To Make an Oral Presentation of Your Research. (2016). http://www.virginia.edu/cue/presentationtips.html.

[56] Mark Warschauer. 1997. Computer-mediated collaborative learning: Theory and practice. *The modern language journal* 81, 4 (1997), 470–481.

[57] Michael F Weigold. 2001. Communicating science: A review of the literature. *Science communication* 23, 2 (2001), 164–193.

[58] Sarah Weir, Juho Kim, Krzysztof Z Gajos, and Robert C Miller. 2015. Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 405–416.

[59] Karoly Zsolnai-Feher. 2016. Two Minute Papers. (2016). https://www.youtube.com/user/keeroyz/playlists.