Creating Crowdsourced Research Talks at Scale

Rajan Vaish
Stanford University
rvaish@cs.stanford.edu

Sharad Goel
Stanford University
scgoel@stanford.edu

Amin Saberi
Stanford University
saberi@stanford.edu

ABSTRACT
There has been a shift towards learning and consuming information through video, but most academic research is still distributed only in text form. Limited time and resources prevent researchers from creating video versions of their work. Crowdsourcing this task is a promising alternative, but it requires solving complex coordination and collaborative video production problems. In this paper, we propose a scalable, end-to-end solution for crowdsourcing the creation of short research videos based on academic papers. 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 system through 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. We evaluated the produced talks in part by soliciting feedback from the authors of the original papers: the authors typically rated the talks “very good”, giving a mean score of 4.1 out of 5. These results suggest that our crowdsourcing approach is a viable method for producing high-quality research talks at scale, increasing the distribution and accessibility of scientific knowledge.

INTRODUCTION
There is growing demand for learning and consuming scientific information through videos [25], as evidenced by the popularity of MOOCs [12] and efforts such as “Two Minute Papers” and “Papers We Love”. Most scholarly material, however, is available only in the form of text, as books and papers. Despite this disconnect, individual academic researchers often have little incentive to invest the time and resources to produce video-based versions of their work.

In this paper, we present and evaluate a self-sustaining system for creating an open, multilingual repository of research talks developed collaboratively by volunteers worldwide. The initial research videos are produced by distributed teams of individuals working in close collaboration; these 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 their educational mobility and incentivizing their continued 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. We further created an online system 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, 18]. 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 5-minute 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, we solicited 72 responses from authors of the original papers, and an additional 154 responses from outside reviewers. The talks were rated “very good” and “very useful”, 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 also 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. In line with results from our research talks, 44 outside evaluators rated these tutorials “very good”, giving them a median rating of 4 out of 5.

Our primary contribution in this paper is developing an end-to-end process for producing short research talks from academic papers. To accomplish this, we introduce a new crowdsourcing technique 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 research content. Though still in its early stages, we hope that our initiative will help bring academic work to broader audiences.

RELATED WORK
In the past decade, there has been a shift towards consuming multimedia information rather than text [4, 31]. In advertising [24], education [12, 16], and health [13], videos are becoming the primary means of communication. Currently, over 65% of smartphone users regularly watch video [11]. Realizing this shift, there are a number of initiatives to create videos that introduce research papers [37] 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 ACM often have limited video production resources, and equipment is provided to conferences on a first-come, first-serve basis [3].

Meanwhile, crowdsourcing techniques have helped researchers and industry professionals scale a variety of efforts, from microtasks like image labeling [35, 30] and translation [32], creative expert tasks like writing stories [17], producing animations [20], and designing software prototypes [28].

The successful crowdsourcing of expert tasks often relies on structure to facilitate complex collaboration. To develop our scaffolding process for crowd coordination, we draw inspiration from such projects [28, 21, 22, 14, 7]. Research on collaboration and organizational behavior has identified several obstacles to effective team coordination—such as technology-mediated communication, geographic dispersion, and dynamic team membership [26, 14, 15]—and has proposed solutions to 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 [6] that is popular when a large crowd is available, we rely on team-based collaboration [23] to ensure successful task completion.

Despite recent advancements in video and audio production technologies [34, 29], options for collaborative video editing tools are still limited [2] and often proprietary [9]. Most existing systems rely on a traditional timeline model, in which contributors make frame-by-frame edits. One of our primary contributions is reformulating the process of collaborative video production. By utilizing and advancing research in microtasking [33, 5], we produce presentations by separately creating talk slides, written scripts, and audio recordings. These
components are then programmatically stitched together to create a complete video.

Overall, our work draws on research in collaborative communities to build the crowd [20], organizational behavior to coordinate the crowd [14], video production to build our online collaboration tools [34], and microproductivity to scale the system [33].

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 reviewed for final improvements. We describe this process in detail below.

We note that before the formal talk creation process begins, we recruit a crowd through an open call for participation, which we describe in later sections. For context, our real-world deployments typically involve creating 10 talks in parallel, with a total of 50–100 active participants.

Phase 1: On-boarding, 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 work flow, and best practices to follow.

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 or expertise; 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.

Each team selects two to three directly responsible individuals (DRIs) to oversee the talk creation process [19, 22]. DRIs set the tone of the talk, 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 stating their interest, availability, and expertise in the topic. After a 24 hour window, the team votes and choses a selected group of DRIs. To maximize availability of DRIs across all timezones, two to three are chosen for each team. To encourage participation, any inactive DRIs are replaced by an active member of the team; this strategy helps keep DRIs active, while motivating other team members to work consistently to get a chance to become a DRI.

Phase 2: Creation of slide deck and script
In this 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 them make sense of the paper they have just read. 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 (the complete list is included in the Appendix), and allot two days to complete this task. Crowd members collaborate and are encouraged to build off of and edit each others 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 three guidelines.

1. Limit talks to approximately five minutes. Based on the past literature [12] and informal pilot studies, we found that shorter videos of five minutes was sufficient to convey the key ideas in most research papers, while still maintaining audience engagement and interest.

2. In the slide scripts, exactly specify the voice-overs. 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.

3. Focus talks on the paper itself, not on background material. The papers often assumed a considerable level of technical expertise, and we mirrored this assumption for the talks. In other words, we established the intended audience for the talks to be the same as those who might read the paper.

In addition to these guidelines, we point participants to past talks produced by the crowd, as it’s 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. (We include the full list in the Appendix.) Finally, we provide crowd members with a pair of articles on creating effective presentations [10, 36].
Figure 2. Workflow of Audio Studio, a tool to facilitate collaborative 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.

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

As always, all 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 acknowledgments were not intended to result in any specific tangible outcomes, but rather were meant to show appreciation for a job well done.

Step 3: Peer-review

In the final three days of this phase, crowd members offer feedback to other talk teams, and work on addressing the feedback that they in turn receive. This review process gets fresh eyes on each presentation, and provides valuable feedback 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 [8]. Teams that address feedback faster can iterate more times; there is no limit to the number of iterations a team can go through.

Phase 3: Recording audio and compiling video

At this point in the process, teams have completed their slides and scripts. As current text-to-speech systems [27] 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 to create a complete video.
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 Google slide presentation, 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 2). 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 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. Due to editing capabilities with Audio Studio, it is easy to implement improvements and reproducing the video again. Our modular approach to the creation of slides, scripts, and audio facilitate edits to any of those components. Once the DRIs approve the final talk, the video is published on YouTube and on the platforms website for the general public.

**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. As the localization process is generally straightforward, this process is unstructured, with no specific timeline, and does not depend on a DRI.

**REAL-WORLD DEPLOYMENT**

To evaluate our system for creating crowdsourced research presentations, we launched an online platform that was open to individuals worldwide. Contributors were not paid for taking part in this effort; their participation was completely voluntary.

**Demographics**

We attracted participants via an open call for volunteers, which resulted in 840 sign ups from 52 countries in 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 and engineering related fields.

Participants 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 Likert scale of 1 to 5, with 1 being a “novice” and 5 an “expert”. Two-thirds of participants had no prior experience preparing or giving a research talk before.

**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. For example, 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 papers were created in 11 foreign languages, for a total of 107 presentations. To date, these videos have attracted over 10,000 views on YouTube.

**Paper Selection**
The crowd produced talks for 40 papers published at top-tier computer science conferences and scientific journals in the past two years. These papers spanned a wide-range 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.

**Localization**
To make scientific knowledge accessible around the world, it is important to distribute it 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, Catalan, Romanian, Oriya, Nepali, Malayalam, Japanese, Filipino, Japanese, and Tamil. Figure 4 illustrates some examples of localized content produced by the crowd.

**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 subsided during the 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.

**EVALUATION**
We surveyed three groups to assess the effectiveness of our approach to producing high-quality crowdsourced talks.

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. **Community members.** A large number of people signed up to participate in the initiative, but were never involved in any talk creation team. For example, they may not have been
available to participate at the start date. We solicited feedback on talk quality from these individuals, and received a total of 133 responses.

3. **Talk creators.** We requested general feedback on the initiative from all participants, and received 95 responses in total. This group provided feedback on the talk creation process and their experiences, but did not evaluate the completed talks.

**Presentation quality**

Talks were 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 “very good” (mean = 4.1, median = 4). Authors indicated appreciation for the initiative, and found the talks to be of high quality, as the following quotes demonstrate.

“The talk is extremely thorough despite its brevity. It certainly better than the talk I gave at WWW.”

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

However, not all authors thought that the 5-minute format was ideal to understand a paper’s contribution.

“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.”

Authors also generally thought the talks were “very useful” for someone trying to get an overview of their papers. Specifically, 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 133 community members we surveyed likewise gave the talks positive quality ratings, with a mean of 3.7 and median of 4, corresponding to “very good”. The slightly lower ratings from community members relative to paper authors may be attributable to their lack of technical expertise. Figure 7 shows that average ratings increase with the expertise of the evaluator. Expertise is on a five-point scale from “novice” to “expert”. Evaluators knowledgeable in the general subject area rated the talks as highly as the paper authors.

**Participant experiences**

Our approach was designed to attract and retain volunteers to create high-quality talks in a collaborative, educational environment. It was thus important to measure participants’ perceptions of the initiative to gauge its sustainability.

Overall, participants indicated the initiative was a “great experience” (average 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 “extremely useful” (average rating = 4.4, median = 5, with 5 being “extremely useful”). Among the 59 participants who rated the Audio Studio recording system, most found it “very simple” to use (average rating = 4, median = 4, with 5 being “absolutely simple”).

The 133 community members we surveyed likewise gave the talks positive quality ratings, with a mean of 4.2 and median of 4 on this dimension, with 5 being “extremely useful” and 1 being “not useful”.

The crowd participants were motivated to contribute, as the effort provided them an opportunity to learn about research while contributing to a public good. Based on a multiple-choice survey conducted with 75 crowd members, we found that their top motivation was learning about different research topics (93%), followed by working with people worldwide towards something interesting (82%), then research access impact (69%), followed by building resume and profile (51%), and lastly, certificate for career advancement (47%). Their reflection contributed to their positive experience, as the quotes below indicate.

“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.”

“First of all, I liked the idea of increasing accessibility to research topic. Secondly, I’m a fan of volunteering. Last but not least, to work with others on some interesting topics.”

“To learn. Wanted to start a way to develop interest in research and the process associated in delivering research talks. Finally finding inspiration to work on my own ideas for research paper in the future.”

The crowd utilized the paper reading and talk creation process to learn about the research topics. Before working on the talk, participants rated their expertise in the topic of the paper at a median of 3 on a scale of 1 to 5 (mean = 2.6, N = 29), where 1 is a “novice” and 5 is an “expert”. After working on the talk and researching the topic for three weeks, these ratings rose to a median of 4 (mean = 3.4, N = 29). Among 75 surveyed crowd members, they generally indicated having a positive learning experience, and rated it at a median of 4 (mean = 4.2), where 1 meant “no learning experience” and 5 meant “learned a lot.”

Upon request we provided crowd members statements that documented their participation in the effort, and which outlined their specific contributions. These contributions were
compiled and verified by their team members. In total, we provided these statements of contribution to 14 participants. We latter surveyed these participants and received responses from 11 of them. Of these 11, 8 received an offer of acceptance from the institution or company to which they applied. It is difficult to determine the extent to which our initiative contributed to this success—and it may well be that we had little direct impact—but feedback from participants indicates that the initiative may have helped. For example, one high school student from Singapore who applied and was admitted to Stanford University shared the following comment.

“Thank you so much for the reference and opportunity to join this initiative. I’ve been taking courses for so long but being able to create my own and help out with education was very special to me... I think seizing the opportunity definitely helped me get admission at Stanford.”

We also note that those who were not granted admission to the programs of their choice acknowledged the educational benefits of the program.

“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.”

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 evaluated by 44 community members who were not involved in their creation. In line with reviews for the research talks, these lectures were typically rated “very good”, receiving a mean rating of 3.9 and median of 4 out of 5.

DISCUSSION AND CONCLUSION

There is a shift in preferences for learning through video. Individual researchers, however, often lack the resources and incentives to produce such content. To address this gap, we developed and evaluated a structured, three-phase approach to creating research talks from papers. In the first phase, volunteer crowd members learned about the process, selected papers, and formed teams. In the second phase, teams worked collaborative 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. Ratings and comments from both the papers’ authors and outside evaluators suggest the created videos were consistently high-quality.

The core scientific contribution of our work is designing a scalable and sustainable approach to creating research talks accessible to global audiences. Two key design choices contributed to the success of our method: modularity and structure. 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 videos, our approach works well for creating research talks. To ensure a critical mass of volunteers were present and active to create each talk, we adopted a structured workflow with a well-defined beginning, middle, and end. 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.

By creating an open repository of multilingual research talks, we hope to advance the dissemination of scientific knowledge. This effort also provides a collaborative learning environment for individuals across the world to support one another. Indeed, the primary motivation for many participants was this opportunity to learn challenging scientific concepts in a supportive community.

Moving forward, we plan to focus our efforts on three fronts. First, we plan to increase the reach of our initiative by mobilizing volunteers to produce thousands of research talks on diverse topics in dozens of languages. Second, 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. Finally, 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. There is a pressing need to develop new ways to make scientific knowledge available to diverse global audiences. Our crowdsourced approach to this problem is one step in that direction, and we hope it spurs further such efforts.

ACKNOWLEDGMENTS
We thank the hundreds of crowd participants who collaboratively created research talks and tutorials. We also thank to the outside evaluators and authors who helped in the assessment of the talks.

REFERENCES
APPENDIX
Below we include the set of questions we used to spur the crowd to think critically about papers, and best practices for creating talks. We also include a crowd-generated sample slide deck (Figures 8–11) for a paper published at CHI; the crowd-generated video of this talk is included in the online supplementary material.

Questionnaire
The 17 questions below were posed to each team to help them understand and analyze the paper they were tasked to present. Teams were instructed to answer each question in 50–100 words. The crowd was further encouraged to rely primarily on paraphrasing, not on direct quotes from the paper.

1. What does the paper’s title mean?
2. What is the current state of the world or problem that this paper is trying to solve?
3. Why is this problem important?
4. Why is this problem hard to solve?
5. What are the primary research questions that this paper is trying to answer?
6. Why are these research questions interesting?
7. What is this paper about? Rewrite the abstract of the paper in your own words.
8. Identify the purpose and the background of this research.
9. Who else did similar or related work, and how is the work presented in this paper different?
10. What are the main contributions of this paper?
11. What are some real-world applications of this contribution?
12. What did the authors do? Describe the paper’s contribution. If it is a systems paper, describe the system; if an experiment is carried out, describe the experimentation setup or methods (design, participants, materials, procedure); if it’s a data science paper, describe the data analysis done or the algorithms used.
13. How did the authors do it? Identify and describe the methodology or approach used and the reason why you think the authors chose this methodology.
14. How is the paper evaluated? Systems, algorithms, and experiments all need to be evaluated to verify the claims made. What methods or techniques did authors utilize to evaluate their claims?
15. What does this paper report? What are the results?
16. How does the design of the study address the research questions? What does this study contribute toward answering the original questions?
17. Share your interpretation of the results. Please discuss and summarize.

Best Practices
These 10 best practices were curated by the crowd over the course of our deployment.

1. Write short scripts per slide.
2. Use multiple slides for different concepts.
3. Always support an image with some explanatory text.
4. If the talk exceeds the 5-minute limit, consider dropping “related work” and “summary” slides.
5. Scripts should not simply repeat the words on the slides.
6. Read the script in a natural, conversational tone.
7. Don’t speak fast to meet the 5-minute limit. If the talk is long, adjust the script and the content.
8. A good talk is more important than a comprehensive talk; no need to cover the entire paper.
9. Create expository talks and use simple language.
10. Cite all the images used in the slides.
Figure 8. Slide deck for a crowd-generated presentation in human-computer interaction (slides 1–8).
Aim
- To allow visually impaired audio editors to conveniently produce and edit audio content

The community of visually impaired users is increasing

Challenges
- The visually impaired are limited to one mode of interaction (i.e., audio) as of current scenario while interpreting sound waves
- Cross-modal mapping of audio by substituting visual aspects of digital audio editing to another sensory modality is required
- Hard to make general solutions that fit everyone’s unique needs
- It takes time to adapt to a new technical solution

Primary Research Questions
If visual waveform displays allow sighted users to see the sound, could we build an alternative interface for visually impaired users to feel the sound?

The paper: Introduction

Idea
- Design, development, and deployment of the Haptic Wave
- Haptic Wave renders audio into kinesthetic (force feedback) information
- Cross-modal mapping of audio into the haptic domain

Figure 9. Slide deck for a crowd-generated presentation in human-computer interaction (slides 9–16, continued from above).
Figure 10. Slide deck for a crowd-generated presentation in human-computer interaction (slides 17-24, continued from above).
Findings
- We can think of the users as partners in research and treat them as experts in non-visual interfaces for all users.
- This comes from a survey of 53 sighted musicians, many of whom described the screen as a distraction.
- This opens up the possibility of using Haptic Wave even for sighted users

Audio Data as Kinesthetic Information
- Visual modality can be translated to other sensory modalities
- An accessible interface for audio editing and audio visualization.
- Users can literally “Feel the sound”

Use of the Haptic Wave
- As a fully functional device alongside advanced equipment in a recording studio
- Potential to be developed as a commercial product
- Eyes-free interface for sighted users. Eg: DJs
- Can be used for multimodal tactile feedback to enhance
  o navigation, and
  o translation of graphic user interfaces.
- Can be used for better use of input devices like
  o Mouse pointers
  o Touch screen interfaces

Findings
- Increase in the speed and accuracy of editing tasks
- Users able to form a mental image of the task they carried out

Real World Applications

Re-visitng Research Questions
What does this study contribute toward answering the original question?

To Summarize . . .
- Audio editing for visually impaired is difficult
- Authors aimed at cross modal mapping of audio to haptic domain
- User centered design involved 3 main stages - Design, Development and Deployment
- Series of 3 workshops and studio trials were conducted
- Feedback from each workshop helped prototype, build and improve the product
- Positive results obtained from lab workshop and studio trials
- Device sped up tasks and improved accuracy
- Enthusiasm to adapt device as it integrated into daily workflows

Figure 11. Slide deck for a crowd-generated presentation in human-computer interaction (slides 25–32, continued from above).