On Pause: How Online Instructional Videos are Used to Achieve Practical Tasks

Sylvaine Tuncer
DSV, Stockholm University
Stockholm, Sweden
sylvaine@dsv.su.se

Barry Brown
DSV, Stockholm University
Stockholm, Sweden
barry@dsv.su.se

Oskar Lindwall
Applied Information Technology
Göteborg University, Sweden
oskar.lindwall@gu.se

ABSTRACT
Instructional videos have become an important site of everyday learning. This paper explores how these videos are used to complete practical tasks, analyzing video-recorded interactions between pairs of users. Users need to repeatedly pause their videos to be able to follow the instructions, and we document how pausing is used to coordinate and interweave watching and doing. We describe four purposes and types of pausing: finding task objects, turning to action, keeping up, and fixing problems. Building on these results, we discuss how video players could better support following instructions, and the role of basic user interface functions in complex tasks involving different forms of engagement with the physical world and with screen-based activity.

Author Keywords
Video interface; Instructional videos; Pause button; Video players; Ethnomethodology

CCS Concept
• Human-centered computing ~ Human computer interaction (HCI) ~ HCI design and evaluation methods ~ User studies
• Human-centered computing ~ Human computer interaction (HCI) ~ Empirical studies in HCI

INTRODUCTION
Available in the millions on YouTube and other online sources, 'how-to' videos are one of the most popular uses of online video [23]. These videos can instruct us for many different purposes, and they have become a prevalent site of everyday pedagogy. This paper presents an in-depth study of how users use online instructional videos to achieve practical tasks, documenting the job of interlacing video and task. To balance the video and their activities, to manipulate artefacts that are part of the task, and to do the task itself, users need to repeatedly pause and resume videos. In this paper we address the central role of *pausing* in following video instructions. With ethnomethodology [11,37] as an approach, we uncover unremarkable aspects of ordinary action, and describe shared methods through which members shape and recognise actions within the unique features of a situation. While a simple device, the pause button is used for many interesting different functions, as participants quickly move between video and task at specific points in time. Closer study of the ordinary use of this familiar interface component, the pause button, reveals some interesting complexities. With empirical data, we describe how pausing allows users to locate important tools and artefacts, to turn to action before the video moves on to the next step, to catch up before it gets too far ahead of action, and to recover when things go wrong. We also describe the *alternate* and the *simultaneous* organizations, two distinctive ways of articulating video and task with one type of pausing for each.

We explore two directions based on these results. The first is to discuss how videos could better support instructional activities, including the organization of the videos themselves and the design of video players and tools. Second, having shown that this activity consists in embedding several loci of actions as part of one and the same course of action, we reflect on how our data gives us a view on this characteristic situation with a close interdependence between onscreen and physical (inter)action in the domain of everyday video watching and use.

BACKGROUND
There are now millions of videos available online providing step-by-step instructions on various practical skills such as how to apply makeup, change a bicycle tire, set up a network router, and repair a hole in drywall. In many cases, the making of these videos feature careful production, organization and editing – with techniques such as voice overs, slow motion, repeats, cuts, and so on, aiming to make instructions easier to follow. These videos are thus not just simply recordings of 'doings' – they are artful 'showing' of what to do and how to do it, combining verbal descriptions, manual demonstrations, textual annotations, commentaries about what is or should be done, and so on. These videos are watched through online video sharing services (notably YouTube), and through devices such as smartphones, providing a massive corpus of video training and support. They are a hugely popular media in their own right – not every watching of these video leads to practical activity, and often they are watched in their own right. Google has reported that 100 million hours of 'how-to' videos are watched every year [30]. These instructional videos are, thus, a site of 'everyday pedagogy' – a massive area of education

CHI '20, April 25–30, 2020, Honolulu, HI, USA

ACM 978-1-4503-6708-0/20/04...$15.00
DOI: https://doi.org/10.1145/3313831.3376759

taking place outside formal academic contexts. Our goal here was to empirically examine how users achieve practical tasks using online videos, how interact with the video interface to attend and make sense of the instructions.

General public interest in these videos has prompted, over the years, a number of attempts in HCI to develop tools to assist following instructions on videos. Attention has been paid to skimming over video and generating automatic annotations from video streams [5,6,22,28,35], as well as designs for playback controls to support summarizing video content and allowing users to search and move around a playing video stream [5,6,8,9,22,28,33,35]. Recently, Chang *et al* [5] studied how instructions are followed from video and designed a voice-controlled video pausing tool which allowed users to manage video while their hands were busy with following a task.

Three elements of instructional videos in use interested us. First, instructional videos are intriguing sites regarding video controls and navigation. Surprisingly, video controls have remained broadly the same over the years, and the pause button is a classic, pervasive, and humble user interface. We were particularly interested in how this simple interactional element comes to be used in complex tasks. Second, to follow an instructional video, users need to juggle between the digital online video and the requirements of whatever material activity is being instructed and achieved. Lastly, instructions have their own complexity: they are essentially incomplete, decontextualized renderings of courses of actions [11,26]. With this study, we also aim to show how users repeatedly and continuously establish correspondence between the instructions and physical action in the here-and-now, with pausing as a resource. Instructional videos are therefore also an interesting site to understand the engagements that take place across physical and digital worlds, and involve a fine multitasking between different worlds with different temporal properties.

On pause

The key problem that users of instructional videos face is the need to combine watching a video with carrying out the task itself. They need to coordinate the temporality of the video instructions with that of their own progress at achieving the task. Ethnomethodology takes temporality as a central feature of how members shape and understand actions, so that they fit into and progress a timely unfolding activity, reflexive of what happened before and what can be projected to happen next [27,40]. In turn, sequentiality is a core concept of Conversation Analysis [39] referring to the irremediably stepwise character of human interactions. Most tasks, considered as a goal-oriented series of actions, have some sort of sequence with some things needing to be done before others. As for watching, videos play forward in time, and thus follow the sequence and order composed by the video creator. But through pausing, users can coordinate their physical task in the 'here-and-now', with video instructions, bringing together two different 'temporalities', and two different material worlds, with tasks requiring particular objects that need to be manipulated in different ways. The pause button then has a simple role – it stops everything 'happening at once', as Wheeler quipped about time itself [40], allowing users to halt instruction while they deal with the task itself.

While the stop function loses the place in the video, pause allows users to temporarily suspend and resume the video at the same place. This supports interweaving onscreen and physical concerns – in the case of video, watching media and doing other activities. We are likely all familiar with using pause with video or audio media. More broadly, it is a feature of nearly all time-based interfaces – music, video editing, music production, time series data, system animations to name a few – which, one can assume, users effectively use, for different and equally crucial purposes as with instructional videos. While most interface elements are concerned with effecting a change on digital materials, pause is interesting in that it is about moving away from the screen in some way (or alternatively focusing closely on a single frame). Pausing, playing, and navigating the video are central resources to do this. Besides coordinating time, therefore, pausing is also what gives users time to engage in the work of embodied correspondence, to transform instructions into actions.

Existing HCI studies on pausing in various sorts of activities inform us on what kind of achievements pausing can support, mainly to enhance the effective use of videos. In language learning, for example, pausing a movie at specific moments allows eliciting students' imagination about what could happen next, and thus stimulate intercultural understanding [34]. Inserting artificial pauses in videos to add audio-descriptions for which time would have been lacking otherwise provides better access to video content to visually-impaired people [10]. Or, the possibility for audibly-impaired students to pause real-time captions gives them time to turn to other visual material, avoid falling behind, and get a better understanding of the lesson [24]. In other words, across a variety of video uses, pausing enables to make time to turn to other resources, to process video content, to accomplish related practical actions, and to understand content. It is no surprise therefore that pausing is prevalent in the use of instructional videos [5].

METHODS

Suchman's ground-breaking work on photocopier use [40] initiated a number of important contributions to HCI and CSCW, in particular the work of Paul Luff and colleagues [16]. Alongside its theoretical contributions, and broader discussions of situated action with technology, Suchman's work was also one of the first using in-depth video analysis to study interaction around technology. In Suchman's case, it was pairs of photocopy users attempting to make sense of, and use, a photocopier to copy various documents. Two users collaborating on a practical problem interact with each other and with the machine, and these interactions exhibit the

"observable-reportable" accountability of practical reasoning and practical action [13]. Suchman's work has inspired a whole generation of research making use of real time recordings of technology 'in the wild' in various ways [4].

We adopted the same method for this study, mainly for practical reasons. First, as mentioned above, having pairs of users collaborating on a task, instead of single users on their own, makes actions much more amenable to analysis because users, accountable to each other, verbalise their actions. Collaboration elicits a sort of natural accountability, somehow similar to the classic "think aloud" protocol. Second, as a series of exploratory interviews suggested, the use of instructional videos is often responsive to an emerging situation – there is a need to fix or do something there and then. This makes the activity difficult to capture as naturally-occurring data, so that we chose to elicit situations we would video-record, of participants using video instructions, something they would commonly engage in anyway.

We recruited ten pairs of participants among our personal contacts and through a student volunteer website, in exchange for small material rewards such as cinema tickets or money. We proposed them a specific task, and if they were willing to (attempt to) achieve it, we made sure that they had little or no particular know-how of it. We equipped a room with three video cameras, an external microphone, a computer with screen recording, and the tools and objects needed for the tasks. We then gave participants one hour to achieve the task using online instructional videos, and video-recorded them doing so. The data include a variety of practical tasks: replacing bicycle brakes, replacing a bicycle chain, picking a lock with paper clips, applying make-up (two sessions), practising yoga (two sessions), doing origami (two sessions), and cooking a dish. Participants could look for and choose any video tutorials they wished. Some of the participants are native English speakers, some are not, but all the data are in English. For all the cases here, the interface to the video itself was the same – our participants made use of the YouTube website to view the videos on a laptop and interact with them.

In group data analysis sessions, we pursued an 'unmotivated inquiry' approach to sensitize ourselves to seen but un-noticed patterns of action, to the resources that are made available by participants to make their activity accountable and understandable to others. Seeing that users made pervasive use of pausing, we first annotated the video recordings with Elan video annotation software, to specify when a video is playing, when it is paused, and when users are scrubbing. This gave an overview of interactions with the video as a first step in the research process – when they paused, moved the timeline, and pressed play. A second step was to analyse in detail a series of instances of pausing. With this incremental method, we were able to identify patterns while preserving the unique character of each instance.

From an initial total number of 150 instances of pausing in the recordings, we extracted pauses which are used to achieve specific actions, as we explain in the results section. This final collection was then analysed and transcribed in detail, drawing on studies of human talk in interaction (notably conversation analysis [37,38]). The figures representing the cases included in this article are simplified versions of more complex multimodal transcripts, including verbal (inter)actions and snapshots of important actions for the phenomenon at hand [31]. The asterisks in the text indicate the precise location of each snapshot. The participants gave their consent for their images to be displayed in scientific publications. The names used are pseudonyms.

Our setting and data have some limitations. A first limitation is that, whereas our exploratory interviews suggested that users massively use instructional videos on their own, we only have pairs of users. This comes with the risk of emphasizing the constructed character of the situation. Users sometimes divided the task between them, one of them taking charge of the video and the other performing the task, but they also easily swapped roles or jointly manipulated the laptop. In the end we can assume that dyad- or single-person use share many similarities. A second limitation is that we provided the machine on which the users viewed the video, a MacBook laptop, rather than make them use their own devices. While our users all seemed familiar with the YouTube interface, and did not have any interactional problems on that level, we do not have any data on how, for example, a phone might be manipulated to bring the video into correspondence with objects, or how the task and tool might be physically distant (e.g. in plumbing or household DIY). Nonetheless, considering the obvious consistency and regularities in how pausing is used and the results we obtained, our data proved robust.

RESULTS

In our data, we have a total of over 150 instances of pausing. Whilst one and the same technical action, it demonstrably achieves a broad variety of actions within the work of following video instructions: users would pause to discuss something that has just happened, to carry out some part of the activity, to compare objects being manipulated in the video, and so on. We found more regularity by narrowing down to instances where participants play and attend a video instruction for the first time in order to achieve the task, as opposed to no less common situations where participants replay a section of the video after they either failed to understand the instruction, or encountered a problem while attempting to do what the video instructed. We isolated and analysed systematically a collection of 52 instances from the first category, in which we identified three types of pausing roughly equally distributed across the collection: pausing to find the right things, to turn to action, and to catch up. Our first pause – pausing to *find the right thing* – concerns trying to find an object shown in the video, often before what to do

with it has been instructed. The second and third types – pausing to *turn to action* and pausing *to catch up* – relate to two different ways of coordinating video and task: users either alternate between attending the video only and achieving the task only, or they can do both at the same time, with their gaze moving from one to the other. Detailed analysis showed that in either organization, pausing has markedly different characteristics: it is prepared or not, happens at different junctions of the video instructions and the task; and more or less precisely parses the task, among others. Then, from the rest of the corpus we extracted a fourth type of pausing which reveals equally common and essential problems of following instructions, dealing with the issue of repairing when something went wrong earlier: pausing to fix a problem.

In what follows, we unpack these four types of pausing. Each is characterized by different actions from users before and after pausing, and each occurs at a different moment in the video, relative to users' needs. Rather than fundamental, this typology is a data-driven textual device, helpful in shedding lights on what pausing can typically achieve, and on purposes and actions recurrently involved in pausing to coordinate the temporality of the video tutorial with that of the physical task.

Finding the right things

Our first form of pausing is the most straightforward – pausing to find the right thing. Instructions in general involve the introduction of objects at some point – tools, products, ingredients – and in instructional videos, this is often done before the actual instruction. To be able to start at a task, users need to find and map between objects in the video and objects in the physical world. As discussed at length by both Suchman and Garfinkel [12,40], one of the biggest challenges of following instructions is that in any actual case of following, there are significant (and potentially problematic) differences between 'instruction world' and 'following world'. Users need to map between the video and what they have at hand to attempt the activity. This can be surprisingly difficult, as we can find we don't have exactly the same tool, or our device isn't exactly the same as the one on the video. When using instructional videos, users routinely pause just after the introduction of a new object. This type of pausing occurs between the introduction of the important object and the instruction of related actions (or what is then to be done with that tool).

In Figure 1, Clara and Emma are trying a new eyebrow make-up technique. The collaborative aspect of action does not have noteworthy consequences since Emma is currently leading the action, and therefore the analysis focuses on her. The instructor has just finished brushing her brow upwards, and when the clip starts, she comments on this action (*and this just shows me where my shape is*.). Meanwhile Emma is about to complete that step, looking at her image in the mirror. While the instructor says *and this just shows me where my shape is*, Emma turns her gaze to the video and acknowledges with *okay*. Acknowledgements and displays of understanding [18] such as "okay" are pervasive in our data. Saying "okay" responds to what has been said, but it also sets up a transition to a next matter, acting as Beach calls it as a *transition marker* [1].

Emma then hands over the brush to the other user,

V: And this just shows me where my shape is.
P: Okay. (1.4) (*Emma hands over brush to Clara*)
V: So: I like to use my: Anastasia Beverly hills dip brow- *(Emma pauses video)*
P: Okay so she has (.) the pomade (.) thing.

Figure 1: Pomade. (Numbers in brackets) indicate pauses, (.) small pauses, and co:lons extended phrases.

acknowledging and marking that she has finished this step. She looks at the video again shortly while the instructor initiates a new step with *so:,* and while the instructor continues with *I like to use my Anastasia Beverly hills dip bro-*, she also brings a product in the video frame (left image). In the course of this turn, Emma moves her right hand briefly to her lap and then to the computer, and she pauses even before the instructor has completed her turn-at-talk. This quick arm movement shows that the pause at this moment is not prepared, that it is instead responsive to what just happened in the video. After pausing, Emma points to the product on the screen and at the same time turns her head to where the products are located in her physical environment (right image). She then registers and identifies the product in the video with *okay so she has (.) the pomade thing*. The emphasis on *she* projects a need to find something 'they' are going to use, the same or a similar product, one which is available to them, here. They pause the video to try and get hold of a similar product as the pomade, before the instructor has given any indication as to what will be done with it. They will resume the video only after deciding which product to use, with the "pomade thing" identified.

In this video, the instructor clearly introduces the pomade before moving on to what she is going to do with it, so that users can pause in between in order to get a hold of the product before attending the instruction. But videos do not always separate instructions and objects so clearly. In cases where they instead intricately interweave them, users need to pause the video 'in flight' to get hold of their objects, and keep in mind the part of the instruction already provided.

Clip 2 is a case in point. Molly (P1) and Ali (P2) are replacing brakes on a bicycle. Despite a clear division of labour whereby Molly manipulates the video interface and

V: undo the brake pa:ds,
P1: *(brings finger closer to space bar)* **[*1& 2]**
P2: pause
V: with an Allen k-
P1: *(pushes space bar and pauses video)*
P1: an Allen key?
P2: Allen key.

Figure 2: bike repair, the Allen key

Ali gets a hold of the object, they are clearly aligned and coincide on when to pause and what for, as we will see. They have opened the callipers on the front wheel, the first step, and as the clip starts, Molly has just pressed play in order to attend the next instruction.

As the instructor initiates the next step with *undo the brake pa:ds,*, Molly gets ready to pause by bringing her fingers closer to the space bar (left image), and after Ali says *pause*, she waits a little and then pauses while the instructor names the new tool, an Allen key: *with an Allen k-* (she seems to pause independently from Ali's request). She immediately turns her head, right hand and body, to the left of the table where the tools are available. Then, she and Ali name the Allen key, which shows that they are jointly involved in looking for it, before they locate it on the table. Once Ali has taken hold of the Allen key, and as he starts opening it, he turns his gaze to the computer again, then closes the search with *okay*, and kneels close to the brakes. Thus, he is now ready to attend the video instruction of what to do with this tool, which he can now physically experience. Molly presses the space bar, and the video resumes. The instruction "undo the brake pads" will not be repeated but Ali and Molly do not rewind the video for all that, they resume it to attend to the ongoing visual demonstration of it. By pausing at this precise point and in order to get a hold of the tool, they clearly separate getting 'hold of the object' from 'performing the action' with it. And as can be seen by Ali saying aloud 'pause', they are both involved in manipulating the video. In all cases of this type of pausing, locating and *getting hold of the object* is users' first concern as they pause the video, whereas action takes second place.

While finding the right thing can be straightforward, at times it can take more work, such as identifying the object in the video, searching and locating a similar object in the local environment, grabbing or fetching it, and sometimes comparing, considering and discussing whether a similar object is similar enough to be used in place of the object in the video. In order to do this and while doing it, users have to adjust the timing of the video, which often shows the action to be performed with the object immediately after. In some cases, like Figure 1, the demarcation of the object is done in its own right, with a short section that introduces the tool. In other cases, like Figure 2, introduction of the tool and instruction are interwoven. Users then need to pause either before the instruction or in the course of the instruction to attend what is left of it later. In any case, 'pausing to find the right thing' is not planned in advance, rather it is responsive to the object's introduction in the video, which often comes unannounced.

This recurrent form of pausing tells us something about following video instructions. That users tend to get a hold of the object before attending the instruction, and thereby tend to make sure to attend an instruction with the corresponding object in hand, suggests that physical contact with the object helps to understand the instruction and to project its reproduction in the physical world. Attending the video, thus, is neither a passive nor an intellectual activity, it is already an active, embodied engagement, even though users don't necessarily move: they feel objects.

Turning to action

Our second type of pausing is characteristic of when users alternate between attending the video only, and achieving the task only with the video on pause. The focus is on breaking up the video into manageable parts – parsing the video [29,36]: they pause the video to turn to their task when they have a sufficient understanding of one or several steps. As we will show, users prepare pausing by getting physically ready, relying on cues embedded in the video and in the task's internal organization, then pause at the next relevant transition space in the video instructions.

Figure 3 involves Molly and Ali again. As the clip starts, they are fully turned towards and attending to the video with the bicycle behind them. The analysis focuses on Ali who both interacts with the video device and takes the lead to organise and turn to action. Just before this clip, the instructor gave them an overall description of the task (*Let's replace the pads on these brakes.*) and now he explains and unpacks this into smaller actions (*in order to do that*).

V: We need to release the quick release mechanism on the cable, **[*1]** Which is done by opening up the little rubber boot, a:nd (.) pulling (0.5) the calipers together, (0.8) a:nd unhook (1.5) this **[*2]**
P: *(moves finger up & down above space bar)*
V: L-shaped (0.7) tube, (1.0) called the noodle.
P: *(pushes space bar and pauses video)*
P: okay (.) let's do that

Figure 3: bicycle repair, "Let's do that"

As the voice over completes what could be heard as a first step of the task (*we need to (.) release, the quick release mechanism on the cable,*) and on the corresponding image, Ali bends towards the computer and gets ready to pause by placing two fingers above the space bar (left image). He stays in this position while the instructor unpacks the step into even smaller actions and demonstrates them: *which is done by opening up the little rubber boot, a:nd (.) pulling (0.5) the callipers together (0.8) a:nd unhook (1.5) this L-shaped (0.7) tube*. After the instructor says *unhook*, the action of unhooking is shown, and on "*L-shaped"*, Ali moves his finger up and down above the space bar (right image). With this aborted attempt to pause, Ali shows that he has almost seen and heard enough to be able to turn to action. But while the instructor adjusts his talk to the demonstrations through pauses and elongations [20], the rising intonation at the end of each short utterance also projects more to come, so that Ali can expect that what is to come is still relevant and worthy of attending to – it is part of something bigger that they should attend and prepare to achieve in one go. In addition to these intonational cues, the video also shows how one needs to maintain the pressure from 'pulling the callipers together' in order to be able to 'unhook this L-shaped tube' – the callipers are by default mechanically pulled apart by a spring. Together this can be seen as an indication for when *not* to pause. Ali then pauses exactly as the instructor ends on a falling intonation: *called the noodle*. He immediately pauses, turns away from the computer and to the bicycle, and as his proposal *okay (.) let's do that* makes explicit, they will now accomplish the same actions.

In other words, the instructor here, after giving a broad description of the task, splits it into discrete and smaller steps, incrementally increasing the granularity. In the course of this unpacking, Ali is getting ready to pause, and demonstrably relies on the instructor's intonation and on the task's internal organization to pause at a relevant transition point. He prepares to pause, waits until he sees the video instructions as complete enough to be practically followed, and pauses at the first relevant opportunity to turn to action.

This type of pausing is part of what we call the *alternate* task organization, where users alternate between the video and the task, that is, between watching the video only without doing anything, and focusing on the task only, with the video on pause. Alternate organization makes use of pausing to turn to action, with each pause also parsing the task into steps which they feel they are able to achieve at this specific moment. They can then plan to achieve the whole action, all of the actions the instructions for which they have just attended, or only a first part of them (see Clip 4). Indeed, in some cases viewing the following action is important to understand an earlier action. It has been shown that we tend to "look to what comes next to find what we're to do now, and we see more clearly what was required to be done earlier in terms of what we're currently doing" [14: p.106]. Even when one has seen and heard enough to be able to copy an action, it may be relevant to play the video longer in case what comes next is more or less directly relevant to this action.

The existing literature on parsing in co-present instructional activities [29] suggests that parsing a task into 'sub-actions' or steps is constitutive of producing instructions, and therefore mainly achieved by instructors. For example, Rauniomaa et al. [36] show how driving instructors can parse the task after they have provided a general formulation of the task, while the novice is achieving the manoeuvre. Any particular 'parsing' or breakdown of a task into subtasks needs to be done with a particular granularity – the same task can be parsed with varying granularity, for example in large 'chunks' or in more numerous, smaller units of action.

While with video they have no access to the actual task being carried out, instructors can adjust the granularity of how they make the video, parsing it to fit with their assumptions about novices' assumed skills. And as the video progresses, they can make use of the completion of previous steps as a resource in how they organise future actions. As the instructions progress, building on earlier actions, the granularity of the instruction can change as the instructor relies on the experience of previously completed actions. So, even though the difference in time and place between the production of the videos and the task obviously precludes any real-time adjustment to novices' actions and to the evolving situation on the part of the instructor, the videos themselves provide an initial parsing of the task.

In many cases, however, where to pause after an action is less clear. That is, users cannot rely on obvious transition cues in the video, between a first action and the following ones, to pause at a specific moment. In Figure 4, Ann (P2) and Jon (P1) are near the beginning of replacing the chain on their bicycle. They tacitly established a division of labour where Jon is close to the bicycle and doing most of the manual work, while Ann, with the computer on her lap, controls the

```
V:  now what you're looking for, is a link
    (.) a little bit like this one. [*1]
P1: m:okay.
V:  now this is called a quick link (.)
    and technically you can remove the
    chain (P1 turns to bicycle) without
    using any tools altogether. [*2] you
    just push the two plates together
    slightly and then slide them apart.
    (.) In reality, they're normally quite
    stiff and so you at least need a pair
    of pliers.
P2: (pauses video, turns to bicycle)
P2: do you find one?
```

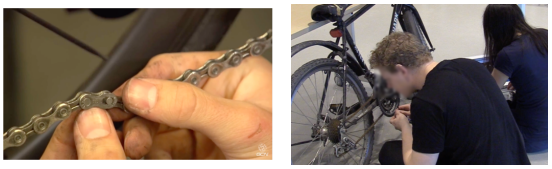

Figure 4: bicycle repair, "A quick link"

video and assists Jon. When she plays, pauses and turns to action, she visibly takes into consideration and aligns with Jon's bodily orientation showing whether he is involved in watching the video or in doing the task.

As the clip starts, they are both attending the video and the instructor says *Now what you're looking for, is a link (.) a little bit like this one*. Thus, the video not only draws attention to a particular element of the chain, it also tells to look for it, especially with the image focusing on it (left image, a close-up of the instructional video). Right after Jon acknowledges the instruction with *m:okay*, the instructor names the *quick link* and moves on to the next step: removing the chain. During the instruction for this next step, Jon turns away from the computer to the bicycle, thus to action. But Ann lets the video play and the instructor expand on the next instruction. She pauses only at the end of it, and right after pausing, she asks Jon *Do you find one?*. By engaging in action with Jon through this question, she clearly addresses a first part of the of the section of the video she has just attended, and disregards the latter. Thus, she parses the task into a smaller granularity than through pausing the video, into at least two different, successive actions which she attended in one go.

As we saw with Clips 3 and 4, users can pause in order to turn to action orienting both to the task' internal organization, and to their understanding of the video instructions so far. To do so, they often rely on transition cues embedded in the video, especially grammatical and intonational completion in the instructor's talk, which indicate a certain parsing of the task. Pausing indicates the point at which they have heard and seen enough to be able to re-do the same action, and no more. By preparing to pause and aiming for a precise moment, users aim to pause as soon as possible after an action has been carried out: the 'right' moment to pause to turn to action.

But as we saw with Figure 4, depending on how clearly transitions are emphasized in the video, users can also adjust the granularity themselves after pausing by extracting a first step from the section of the video they have just attended. This additional parsing work enables them to proceed stepwise with their task by focusing on what they need to achieve first, and leaving the rest for later. In these cases, they can be said to pause 'late' in the video.

So, users can pause during transitions to turn to action, but also 'late' compared to the additional parsing they subsequently do. These two possibilities show one and the same feature of the work of following (video) instructions: reproducing an action is easier when its overall shape is still fresh from the video. Its visible and audible details can thus be immediately re-embodied, directly translated in the physical world. As a consequence, a 'right' moment to pause to turn to action is one far enough so that one step is complete and be accomplished; early enough so that the action is fresh, and potentially even visible on the screen in the frame paused at; and it is also early enough to preserve the next action's integrity for later viewing.

Pausing to catch up

Alternating between the video and the ongoing activity is one powerful way of arranging the activity. But it is also possible to attempt both in parallel. This was the second way in which our participants arranged their instruction following - watching the video and following the instructions at the same time. Since video instructions usually go a little faster than participants first attempts at an action, the video instructions advance a little ahead of users' physical tasks, if both are ongoing at the same time. This requires users – at some point – to pause to give themselves time 'to catch up'. As we will show, pausing to catch up is not pre-prepared, in that it occurs when users are falling behind, and unlike pausing to turn to action, it tends to occur after transitions in the video, when the instructor is already progressing through the next step.

In Figure 5, we join Ali and Molly later in their task of replacing the brakes on a bicycle. They are now installing the new pads, each kneeling on one side of the front wheel, looking at their manual actions and away from the video, while it is playing. The instructor has just verbalized three consecutive actions – installing the spacers, tightening the bolts, and aligning the pad with the wheel curvature. Figure 5 begins with the completion of the last one.

As the instructor says *a:nd just very approximately (.) line it up.*, Ali removes his hands from the brake to his laps: he has completed this step of the task, installing the new pad, whereas Molly is still on it. Ali turns his head to the video (left image), shortly after the image in the video changes to a close-up on the instructor holding a new tool, and as the latter introduces this tool (*this little brake shoe tuner*), moving on to the next action, Ali bends over to reach the space bar with his left hand and pauses the video. He brings his hand back to his lap, turns his upper body to the bicycle again, and marks a transition with °*okay*°. Then he bends over the wheel to look at Molly's manipulations (right image), and thus shows that he is now waiting for her to finish installing the new pad, letting her catch up with the video.

Ali pauses when the instructor is moving to the next step. His rapid arm movement to the space bar suggests that he pauses

```
V:  a:nd just very approximately (.) [*1]
    line it up. (1.0) this little brake
    shoe tuner (.) is a-
P2: (pauses video, turns to P1's actions)
P2: °okay° [*2]
```

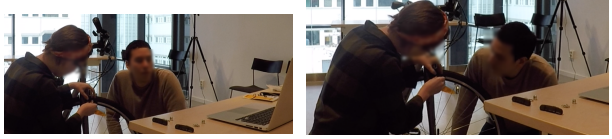

Figure 5: bicycle repair, "Line it up"

V: And I like bringing it in a little bit further than you think **[*1]** you would, because (0.4) once you do that, >I now take my foundation brush where I applied my foundation with< and then just lightly with that foundation brush just tap the front
P: mmm: *(brings hand to space bar)*
P: [that looks good]
V: [of the brow,] **[*2]**
V: >and that will< softe-
P: *(pauses video, resume tapping brow with brush)* °okay° (.) so: (.) we could use

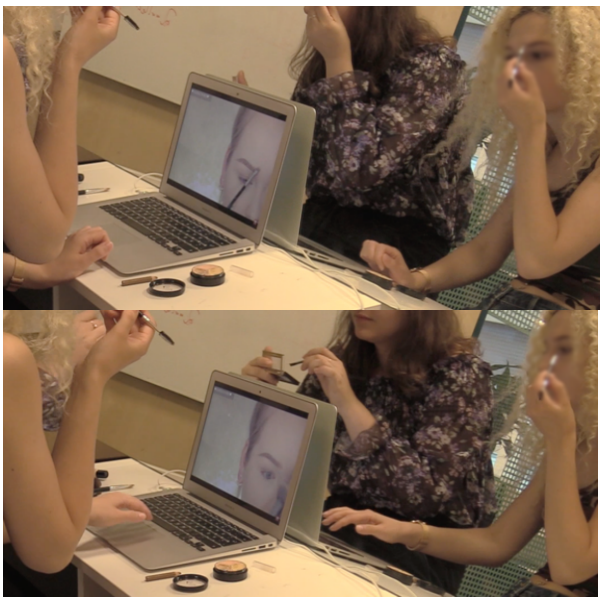

Figure 6: Applying make-up.
>quiet talk< and [overlapping talk]

in response to this initiation of a new step, in relation to Molly still achieving the previous one, and for her to catch up, and from the point of view of the whole task, for them to keep up with the video instructions.

These are the constitutive features of 'pausing to catch up', but it can take more complex forms. In Figure 6, our participants not only interweave achieving the task and attending the video instructions, they also follow the instruction for an action while still achieving the previous one. This clip takes place 7 minutes after Emma and Clara have started to use an eye-brow make-up video (a different one than in Clip 1). Like in Clip 1, our analysis focuses on Emma's actions. As the clip starts, Emma is applying powder from the tip to the inside of her brow with a small brush, looking at her actions in the mirror while the instructor starts describing the action Emma is currently achieving.

During the instructor's description *and I like bringing it in a little bit further than you think you would, because (0.4)*, Emma suspends her tapping actions with the brush and turns her gaze down to the video (first image). She turns her gaze back to the mirror before the instructor says *once you do that.* While the instructor initiates a new step, by introducing a new tool and what to do with it (*>I now take my foundation brush where I applied my foundation with< and then just lightly with that foundation brush just tap the front of the brow,*), Emma suspends her actions and turns to the video again, positively assesses this action with *mmm:*, and then looks to her left, most probably in search for a similar brush as the one the instructor is using. She brings her hand to the keyboard (second image), looks back at the video and pauses while the instructor expands on this action with the big brush (*and that will softe-*). Just after pausing, Emma turns her gaze to the mirror, marking completion and acknowledgment with *°okay°*, and resumes her tapping actions with the small brush, that is, the action preceding the one they have just attended in the video. Then, she orients to the next step with "*so:"* and a proposal as to what tool they could use in place of this big brush (*°okay° (.) so: (.) we could use*).

Here, while the video plays, Emma is alternating attending to the video and achieving the task by shifting her gaze and suspending her actions. It would be physically quite difficult to both attend the video and apply make-up on one's brows at the same time, but through this rapid alternation, she interweaves different steps of the task: the one she is currently achieving and the next two, searching for a big brush and tapping the front of the brow with it. In this way, she also manages to keep up with the video for some time, until this moment where she pauses, visibly in need to catch up before the video goes any further, too far.

Attending the video and achieving the task in parallel can be challenging, not only because of the competition for the same physical resources, but also because videos tend to proceed faster than the physical task. The physical task is always a little behind, but before it gets *too* far behind, users pause and make time to catch up with the instructions.

Fixing a problem

In the three types of pausing we have looked at so far, our participants successfully proceed with their task, they manage to move between the instructions in the video and carry out the physical task. But of course, as has been reported at length in the study of instruction following [26], things often, and nearly always, go wrong. Even with the support of videos and instructions, first attempts are often unsuccessful and require retry. Or, one may become confused during an attempt, or realize one may have taken an incorrect step. As mentioned above, we should emphasize that following instructions is not simply 'copying' or replicating what is done in the instructions, it is re-embodying and translating. The novice needs to adapt what is being done to the actual situation found in some way. Bodily movements (such as twists or manipulations) need be done by the novice themselves, not the instructor, relying upon the physical re-enactment, not necessarily a simple repeat. In our final 'pause' participants hit a problem or issue with the task, go to watch the video again, and then pause when they find a likely solution, or the information they were

V: then take the other part of the chain here feed it over (0.5) the: (0.6) smallest co:g (0.4)
P1: mm mm? (1.0)
V: on the c- the- the free wheel there, (.) bring this arou:nd, **[*1]**
P1: *(pauses video, turns to bicycle)*
P1: [we need to bring it around?
P2: [oh yeah so he *(turns to bicycle)* just brings it around,**[*2]**

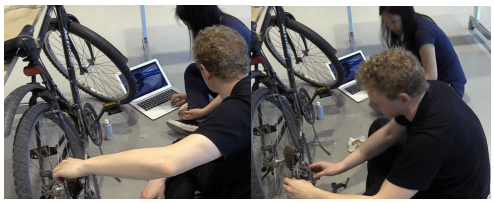

Figure 7: replacing bicycle chain

needing. Pausing is prepared – often users keep their hand ready to pause from the moment they started the re-play; but it is also responsive to the video introducing the specific information users were looking for; and it occurs outside transition spaces, that is, it is independent from both the task's internal organization and the parsing cues embedded in the video.

In this final clip (figure 7), Jon (P2) and Ann (P1) are now installing the new chain on their bicycle. One minute or so before, Jon raised a trouble: he could not establish the correspondence between the "shifter" on this bicycle and the one in the video, and was therefore stuck in the task. They moved backwards in the video and started re-playing from when the instructor shows how to install the new chain. Jon is holding the chain above the back-wheel cog.

Just as the instructor brings the chain around the shifter and says "bring this around", Ann pauses the video (left image): they were just given a candidate solution, and they concur on a similar version of it, on overlap: Ann with *we need to bring it around?*, and Jon with *oh yeah so he just brings it around,*. "Oh" marks that there is some new information - something new has emerged from this re-watching [17]. They turn to the bicycle again, re-engaging in a course of action that will address where Jon was stuck. This type of pausing is prepared: the user controlling the video is physically ready to pause, in order to pause exactly when they hear and/or see the specific element they have a problem with. Unlike in pausing to turn to action, this point is independent from parsing: users are not looking for transition points between steps. The right moment to pause has to do with users' own, emergent understanding of the task and the materials, and just after pausing they tend to verbalise this new piece of information or understanding.

DISCUSSION

Each of these four different 'pauses' shows how this simple interface feature – just pushing the space bar in most cases – can have a range of effective uses when it is part of following instructions. In this discussion, we develop these observations in two directions. First, we discuss how we could contribute to video tools for this task, covering how instructional videos themselves are structured, and how they could better support the four types of pause we have outlined. Second, we discuss how a basic user interface function can have considerable complexity *in use*, when it is incorporated into, and part and parcel of, complex activities. We go on to discuss the differences between HCI studies of multitasking, and what our findings show about the very nature of following video instructions, that is, weaving together video and physical ask as one and the same course of action.

Better supporting video instructions

As we mentioned in the introduction, within HCI there have been a number of efforts proposing how software could better support giving and following instructions. Some systems automatically identify the introduction of objects and transitions between steps based on the instructional video's content [33], and automatically pause at relevant points to assist users while they achieve the task [35]. However, our findings suggest that there are advantages to letting users control their pausing. For example, they may deliberately fall behind, as part of a method to efficiently progress the task with an eye on what comes next, and while being pushed by the video. Or, they may be proficient enough at a task to not want to pause between each step. Because pausing is based on user activation it allows for this flexibility. Accordingly, we focused our attention on *non-intrusive* ways of assisting users that don't 'break' the existing pause functionality. For each of the four 'pauses' we unpacked, we suggest a potential way in which instructional media might work.

The first type of pause shows that it is important for users to identify a new object, to see it clearly and long enough, and sometimes to be able to compare it with the objects they have at hand. One way of supporting this would be that the video announces in advance, and highlights, the introduction of new objects; and maintains objects in view for some time afterwards. We suggest the creation of tools for creators to nominate particular frames in their video that demonstrate or highlight such important elements. When playing a video, this frame could be displayed (picture in picture) alongside the actual frame that is being paused at. This functionality would support getting a better view on the specific object being introduced, without getting in the way of situations where the actual frame stopped at is important for another reason.

With our second type of pausing, to turn to action, we showed that users rely on cues embedded in the video to parse the task into reproducible steps. One way of supporting this would be to actually display transition points on the timeline, before and as they occur. This would also support our third type of pausing, to catch up, by informing users of transitions in advance and helping them anticipate when they might fall behind. On most video players, the timeline,

visible when users move their mouse, pause or start the video, displays the playhead on a line visualizing the length of the video. It could be much more informative and helpful if video creators could annotate the timeline itself. Visualizing the stepwise organization of the task, displaying not only future but also previous important sections of the video, would also support participants' endeavours to diagnose a problem or misunderstanding in achieving the task, our last type of pausing.

Embedding technology use in complex activities

While our focus here has been in following instructions, the paired nature of screen-located activity and physical activity is a very common feature of apps, and technology use broadly. From webpages held up in discussions [2], to navigation with smartphone maps [25], the juggling of device and world is commonplace in how we use smartphones and apps in our digitally-mediated lives. This paired nature is perhaps one of the most intriguing aspects of our data here.

We are not the first to engage with this – Tolmie et al. [41] outline aspects of how others can make demands upon when, where and what can be done with our mobile devices; Nardi and Whittaker [32] discuss 'outeraction'; and more recently Brown et al. [3] analyse co-operative text messaging. But while these studies tend to focus on the demands of co-present others (such as waiting for an assessment of a possible text message to send), ours focused instead on the demands of practical task and materials. What we have tried to do here is to tease out the complex relationships between what is done at the interface, and what is done around it, with a physical task proper.

In practice, this can be seen as threading together a computer interface, the instructional videos, and the task being carried out. We think of this as a threading of video and task - between what is done on the screen and what is done physically. This threading is made possible by either simultaneous or alternative task organizations. An alternate task organization requires careful attention to when and where transitions are made, and when it will be necessary to pause the video to move to the task. In contrast, a simultaneous organization will demand a user to manage both the video and task at the same time, drawing particularly on the audio commentary to the video, but also on the ability to physically arrange the task such that moving between the two will not be physically too difficult.

We have emphasized in our implications, and in our results, the importance of the pause button (and of the navigation interface more broadly) to what is done here. The screen is just as much the site where instructions are followed as in the hands of our participants. This is an interesting contribution of our method – by video-recording users around the screen, echoing classic work by Heath et al [16], we have documented *both* the interface and physical actions as our users' worksites. Action is not onscreen *or* physical, but at the screen *and* in the hands [19:38]. HCI studies of multitasking [7,15,21] emphasize the cognitive and practical demands put on participants involved in several tasks at the same time. Relatedly, in interactional research, studies of multiactivity [14] investigate the competing, social and practical demands put on participants involved in several social activities. While similar actions can be involved in using video instructions and achieving a practical task "at the same time", that users distribute their time and resources between onscreen activity and physical task does not necessarily mean that they are involved in several tasks or activities at the same time, nor that they consider onscreen and physical activities as such. Indeed, as we showed, users weave together attending the video and achieving the task in order to make it one and the same coherent course of action: following video instructions. Watching and doing are a shared concern, they are interrelated, and interdependent.

Finally, it is important to emphasize that different points of the task could require different approaches, and users could have different preferences for how to approach the task. The 'temporal order' – at its simplest when and how long it takes to do different things – is something that makes demands on how the interface or video is used. But also, either organization indicated and enacted different ways of distributing agency between the video and the users' physical tasks, while formally taking similar uses of the interface and interactions with the medium.

CONCLUSION

This paper has engaged with the massively common, educational use of online video. Through analysis of users attempting to complete different practical tasks we outlined four different ways in which the simple pause button is used. Our findings build on concurring hints from existing HCI studies around pausing videos [5,10,24,34]: across various activities, suspending the video flow can be particularly useful to make sense of the content, translate it to local circumstances, compare it to other sources, or expand on it independently from what comes next.

At the heart of our analysis is a concern for how users must manage two different temporalities. First, the video timeline – which plays until stopped and resumes when pause is hit a second time – is structured by those who make the videos, but also by users in when and where they hit pause, or navigate through the video by dragging the playhead. But when following instructions users must balance this with a second timeline – their own carrying out of the task – the order things should be done and how long it takes them to understand and follow out the task. Users adopt the pause button to bring together what they do with how the video instructs them. In this way a simple user interface feature becomes a dynamic and flexible tool – enabling an exciting new site of instruction and education.

ACKNOWLEDGMENTS

We thank our participants and our group members for comments on earlier drafts. This work was supported by the Marcus and Amelia Wallenberg grant 2015.0075.

REFERENCES

[1] Wayne A. Beach. 1993. Transitional regularities for 'casual'"Okay" usages. J. Pragmat. 19, 4 (1993), 325–352.

[2] Barry Brown, Moira McGregor, and Donald McMillan. 2015. Searchable Objects: Search in Everyday Conversation. ACM Press, In Press.

[3] Barry Brown, Kenton O'hara, Moira Mcgregor, and Donald Mcmillan. 2018. Text in Talk: Lightweight Messages in Co-Present Interaction. ACM Trans Comput-Hum Interact 24, 6 (January 2018), 42:1–42:25. DOI:https://doi.org/10.1145/3152419

[4] Graham Button (Ed.). 1993. Technology in working order: studies of work, interaction and technology. Routledge, London.

[5] Minsuk Chang, Anh Truong, Oliver Wang, Maneesh Agrawala, and Juho Kim. 2019. How to design voice based navigation for how-to videos. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 701.

[6] Kai-Yin Cheng, Sheng-Jie Luo, Bing-Yu Chen, and Hao-Hua Chu. 2009. SmartPlayer: user-centric video fast-forwarding. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 789–798.

[7] Giovanni Circella, Patricia L. Mokhtarian, and Laura K. Poff. 2012. A conceptual typology of multitasking behavior and polychronicity preferences. Electron. Int. J. Time Use Res. 9, 1 (November 2012), 59–107. DOI:https://doi.org/10.13085/eIJTUR.9.1.59-107

[8] Chris Crockford and Harry Agius. 2006. An empirical investigation into user navigation of digital video using the VCR-like control set. Int. J. Hum.-Comput. Stud. 64, 4 (2006), 340–355.

[9] Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowitcz, and Derek Nowrouzezahrai. 2008. Video browsing by direct manipulation. Proc. SIGCHI Conf. Hum. Factors Comput. Syst. ACM (2008), 237–246.

[10] Benoît Encelle, Magali Ollagnier Beldame, and Yannick Prié. 2013. Towards the usage of pauses in audio-described videos. In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A '13, ACM Press, Rio de Janeiro, Brazil, 1. DOI:https://doi.org/10.1145/2461121.2461130

[11] Harold Garfinkel. 1967. Studies in Ethnomethodology. Prentice Hall.

[12] Harold Garfinkel and Anne Rawls. 2002. Ethnomethodology's Program. Rowman & Littlefield Publishers, Inc.

[13] Harold Garfinkel and Harvey Sacks. 1970. On formal structures of practical actions. In Theoretical Sociology: Perspectives and development (J. McKinney & E. Tiryakian). Appleton Century Crofts, New York, 337–366.

[14] Pentti Haddington, Tiina Keisanen, Lorenza Mondada, and Maurice Nevile (Eds.). 2014. Multiactivity in Social Interaction: Beyond multitasking. John Benjamins Publishing Company, Amsterdam. DOI:https://doi.org/10.1075/z.187

[15] Mariam Hassib, Mohamed Khamis, Susanne Friedl, Stefan Schneegass, and Florian Alt. 2017. Brainatwork: logging cognitive engagement and tasks in the workplace using electroencephalography. In Proceedings of the 16th international conference on mobile and ubiquitous multimedia, ACM, 305–310.

[16] Christian Heath and Paul Luff. 2000. Technology in Action. Cambridge University Press.

[17] John Heritage. 1984. A change-of-state token and aspects of its sequential placement. In Structures of social action (J. M. Atkinson & J. Heritage (Eds.)). 299—345.

[18] Jon Hindmarsh, Patricia Reynolds, and Stephen Dunne. 2011. Exhibiting understanding: The body in apprenticeship. J. Pragmat. 43, 2 (2011), 489–503.

[19] Kristina Höök. 2018. Designing with the body: somaesthetic interaction design. MIT Press.

[20] Leelo Keevallik. 2015. Coordinating the temporalities of talk and dance. Temporality Interact. (2015), 309–336.

[21] Susan Kenyon. 2010. What do we mean by multitasking? - Exploring the need for methodological clarification in time use research. Electron. Int. J. Time Use Res. 7, 1 (November 2010), 42–60. DOI:https://doi.org/10.13085/eIJTUR.7.1.42-60

[22] Seungwon Kim, Sasa Junuzovic, and Kori Inkpen. 2014. The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In Proceedings of the 18th International Conference on Supporting Group Work (GROUP '14), ACM, New York, NY, USA, 167–177. DOI:https://doi.org/10.1145/2660398.2660409

[23] Ben Lafreniere, Andrea Bunt, Matthew Lount, and Michael Terry. 2013. Understanding the roles and uses of web tutorials. In Seventh International AAAI Conference on Weblogs and Social Media.

[24] Walter S. Lasecki, Raja Kushalnagar, and Jeffrey P. Bigham. 2014. Helping students keep up with real-time captions by pausing and highlighting. In Proceedings of the 11th Web for All Conference on - W4A '14, ACM Press, Seoul, Korea, 1–8. DOI:https://doi.org/10.1145/2596695.2596701

[25] Eric Laurier, Barry Brown, and Moira McGregor. 2015. Mediated Pedestrian Mobility: Walking and the Map App. Mobilities (2015), 1–18.

[26] Eric Livingston. 2008. Ethnographies of reason. Routledge.

[27] Michael Lynch, Eric Livingston, and Harold Garfinkel. 1983. Temporal Order in Laboratory Life. S. 205–238 in: KD Knorr Cetina & M. Mulkay (Hrsg.), Science Observed: Perspectives on the Social Study of Science. London: Sage.

[28] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2013. Swifter: improved online video scrubbing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 1159–1168.

[29] Helen Melander and Fritjof Sahlström. 2009. Learning to fly—The progressive development of situation awareness. Scand. J. Educ. Res. 53, 2 (2009), 151–166.

[30] David Mogensen. 2015. I Want-to-Do Moments: From Home to Beauty. Think with Google. Retrieved September 18, 2019 from https://www.thinkwithgoogle.com/marketing-resources/micro-moments/i-want-to-do-micro-moments/

[31] Lorenza Mondada. 2016. Challenges of multimodality: Language and the body in social interaction. J. Socioling. 20, 3 (June 2016), 336–366. DOI:https://doi.org/10.1111/josl.1_12177

[32] Bonnie A. Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and Outeraction: Instant Messaging in Action. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00), ACM, New York, NY, USA, 79–88. DOI:https://doi.org/10.1145/358916.358975

[33] Cuong Nguyen and Feng Liu. 2015. Making software tutorial video responsive. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 1565–1568.

[34] Amy Ogan, Vincent Aleven, and Christopher Jones. 2008. Pause, predict, and ponder: use of narrative videos to improve cultural discussion and learning. In Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08, ACM Press, Florence, Italy, 155. DOI:https://doi.org/10.1145/1357054.1357081

[35] Suporn Pongnumkul, Mira Dontcheva, Wilmot Li, Jue Wang, Lubomir Bourdev, Shai Avidan, and Michael F. Cohen. 2011. Pause-and-play: automatically linking screencast video tutorials with applications. In Proceedings of the 24th annual ACM symposium on User interface software and technology, ACM, 135–144.

[36] Mirka Rauniomaa, Pentti Haddington, Helen Melander, Anne-Danièle Gazin, Mathias Broth, Jakob Cromdal, Lena Levin, and Paul McIlvenny. 2018. Parsing tasks for the mobile novice in real time: Orientation to the learner's actions and to spatial and temporal constraints in instructing-on-the-move. J. Pragmat. 128, (2018), 30–52.

[37] Harvey Sacks. 1995. Lectures on conversation: vol 1 & 2. Basil Blackwell, Oxford.

[38] Emanuel A. Schegloff. 2007. Sequence organization in interaction: Volume 1: A primer in conversation analysis. Cambridge University Press. Retrieved September 20, 2016 from https://www.google.com/books?hl=en&lr=&id=5XbJRFQ4dhsC&oi=fnd&pg=PR11&dq=sequence+organization&ots=MjA2ETTY1r&sig=p1-_6UkzJWSJrDeDKftpm3nnn-A

[39] Emanuel Schegloff, Gail Jefferson, and Harvey Sacks. 1974. A simplest systematics for the organization of turn-taking for conversation. Language 50, 4 (1974), 696–735.

[40] Lucille Alice Suchman. 2007. Human-machine reconfigurations: plans and situated actions. Cambridge University Press, Cambridge; New York.

[41] Peter Tolmie, Andy Crabtree, Tom Rodden, and Steve Benford. 2008. "Are You Watching This Film or What?": Interruption and the Juggling of Cohorts. In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08), ACM, New York, NY, USA, 257–266. DOI:https://doi.org/10.1145/1460563.1460605

[42] Wojciech H. Zurek (Ed.). 1990. Complexity, Entropy and the Physics of Information (1 edition ed.). Westview Press, Redwood City, Calif.