LiveSnippets: Voice-based Live Authoring of Multimedia Articles about Experiences

Hyeongcheol Kim
National University of Singapore
Singapore
hckim0911@gmail.com

Shengdong Zhao
National University of Singapore
Singapore
zhaosd@comp.nus.edu.sg

Can Liu
City University of Hong Kong
China
canliu@cityu.edu.hk

Kotaro Hara
Singapore Management University
Singapore
kotarohara@smu.edu.sg

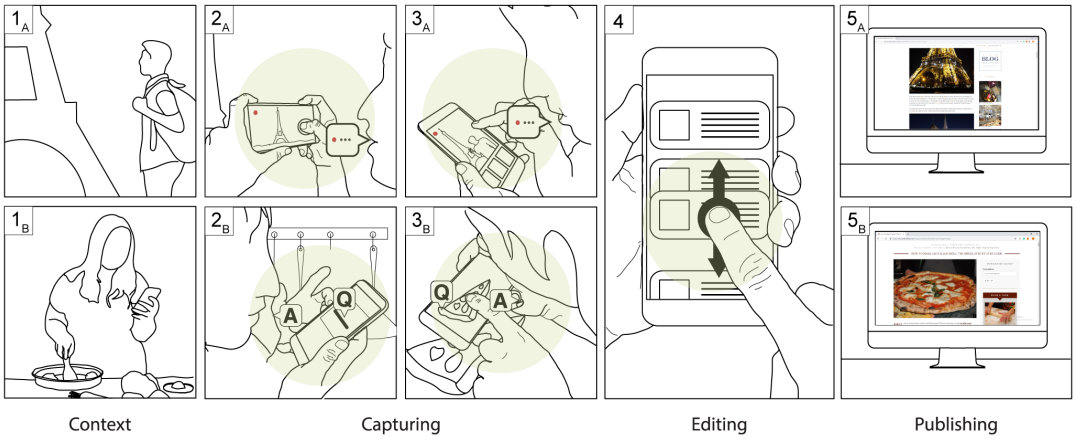

Figure 1: In-situ voice-based multimedia authoring (IVA) workflow and usage scenarios. *Top*: John loves sharing his experiences online. In his trip to Paris, John is amazed by the magnificent Eiffel Tower (1_A). He starts to describe the tower with his feelings while taking pictures of it (2_A) and narrate the moments with the taken pictures (3_A) through LiveSnippets; a mobile app that supports *IVA*. His voice is transcribed into text and saved together with the photo and contextual information (e.g. time, location, etc.) into snippets that can be edited later (4). At the end of the visit, a draft travel blog can be generated from the snippets with a simple button click (5_A). *Bottom:* Mary loves sharing her cooking experiences through recipes. While preparing a dish, she launches the recipe authoring program in LiveSnippets (1_B). The app starts to ask her a series of questions including the name, style, ingredients, steps of the dish, etc. Mary answers the questions one by one (2_B), taking photos/videos as needed while cooking (3_B). Later she rearranges and edits the snippets (4) before generating a recipe to share online (5_B).

ABSTRACT

We transform traditional experience writing into in-situ voice-based multimedia authoring. Documenting experiences digitally in blogs and journals is a common activity that allows people to socially connect with others by sharing their experiences (e.g. travelogue). However, documenting such experiences can be time-consuming and cognitively demanding as it is typically done OUT-OF-CONTEXT (after the actual experience). We propose in-situ voice-based multimedia authoring (IVA), an alternative workflow to allow IN-CONTEXT experience documentation. Unlike the traditional approach, IVA encourages in-context content creations using voice-based multimedia input and stores them in multi-modal "snippets". The snippets can be rearranged to form multimedia articles and can be published with light copy-editing. To improve the output quality from impromptu speech, Q&A scaffolding was introduced to guide the content creation. We implement the IVA workflow in an android application, LiveSnippets - and qualitatively evaluate it under three scenarios (travel writing, recipe creation, product review). Results demonstrated that IVA can effectively lower the barrier of writing with acceptable trade-offs in multitasking.

MobileHCI '20, October 5–8, 2020, Oldenburg, Germany

ACM ISBN 978-1-4503-7516-0/20/10…$15.00
https://doi.org/10.1145/3379503.3403556

CCS CONCEPTS

• **Human-centered computing → Mobile computing**; *Interaction design theory, concepts and paradigms*; Ubiquitous and mobile computing systems and tools; Natural language interfaces.

KEYWORDS

in-situ voice-based multimedia authoring; voice-based multimedia documentation; multimedia authoring in situ

ACM Reference Format:
Hyeongcheol Kim, Shengdong Zhao, Can Liu, and Kotaro Hara. 2020. LiveSnippets: Voice-based Live Authoring of Multimedia Articles about Experiences. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20), October 5–8, 2020, Oldenburg, Germany.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3379503.3403556

1 INTRODUCTION

Experiential writing is difficult because it is out-of-context. As human beings, "we are made up of, engage in, and are surrounded by stories" [24]. Writing down one's personal autobiographical narratives (experiential writing) can help to develop and maintain one's self [18, 24, 25], and has shown to provide strong mental health benefits [27]. Today the technology advancement has largely enriched the content of experiential writing with multimedia representations, as commonly seen in blogs. However, experiential writing has typically been performed out-of-context. This is partially because writing is a delicate task and requires a suitable environment. People prefer to do it while sitting in front of a desk with proper equipment so that they can type or write comfortably in front of a large display [32]. However, due to the lack of a proper writing environment during the experience, people are often unable to write in-context; instead, writing about one's experience is often performed after the experience at a separate time and place, which imposes a high memory and cognitive load to recall the experience during the writing process.

We aim to re-design the process of experiential writing by making it more in-context and less cognitively demanding. We present in-situ voice-based multimedia authoring *(IVA)*, a new approach for experience writing. In *IVA*, content creation and editing are separated, whereby, content creation is largely performed during the experience using voice-based multimedia authoring. The content is generated moment by moment in-context, and each moment is saved into a multimedia snippet that consists of visual, text, and contextual information. The captured snippets can be edited after the experience before generating an article. To overcome the lack of structure of spoken content, scaffolding questions that follow the best practices of experience sharing guide the user through the content creation process.

We envision this approach can be implemented in a variety of mobile or wearable devices including smartphones and smart glasses. As a first attempt, we instantiate it with LiveSnippets; a mobile application that supports the entire writing life cycle of experience capturing/sharing with *IVA* (Fig. 1). We present its iterative design, implementation and evaluation process involving: (i) two rounds of prototyping with four pilot users to create low-fidelity prototypes and elicit design requirements; (ii) implementation of a hi-fidelity working prototype; and (iii) a survey and interview-based empirical evaluation of the hi-fidelity prototype with twelve participants from varying backgrounds (e.g. travel bloggers and product reviewers). The results of our evaluation showed that compared with the traditional writing process of experience capturing/sharing, LiveSnippets allowed a significant amount of authoring to take place in context and the content needed only a small amount of additional editing before publishing. The participants felt writing with LiveSnippets was both simpler and more engaging compared to the traditional writing process.

Approaches of using voice for text input and Q&A for scaffolding exist in previous work; however, they are applied to relatively simple content generation tasks (i.e. voice-based annotations to support recall of ambiguous events [8]), eliciting answers from users in daily context (a.k.a. experience sampling method), or in other domains (i.e. video analysis and editing [30]). We combine and extend the previous approaches to enable authoring more complex multimedia documents in situ by introducing divide and conquer, in-context writing, minimal modal resource contention, and multifaceted data as design requirements. Unlike short texts, longer and more complex documents demand much higher mental and physical efforts. In addition, speaking uses a different mental process from writing. Thus, it was unknown whether previous approaches involving in-situ speaking for short text is suitable for authoring more complex multimedia documents. Our initial investigation for in-situ voice-based multimedia authoring with Q&A scaffolding demonstrates that it can be adapted to help the traditional writing experiences of more complex multimedia documents in three genres of experiential writing (i.e. travel blogs, product reviews, and cooking recipes). Our approach makes the writing process simplified and more engaging, pointing to a promising new way of experiential writing in the future.

The contribution of this paper is threefold:

- A novel "*in-situ voice-based multimedia authoring (IVA)*" approach to transform experiential writing using voice-based multimedia interface and scaffolding in Q&A style;
- The iterative design of a proof-of-concept application, LiveSnippets, to show the potential of this approach in three scenarios of experience recording: travel, cooking and product review;
- Initial empirical findings from evaluating the *IVA* approach in real-world settings, showing the potential of this new approach to reduce memory load and procrastination in experiential writing.

Based on our findings, we discuss how *IVA* could transform the way we write in our everyday lives, and how it may inspire us to re-think and extend how we write in the future.

2 RELATED WORK

The design of the *in-situ voice-based multimedia authoring (IVA)* workflow involves the assembly and repackaging of many of the existing techniques and approaches so that they can seamlessly work together to transform into a new, effective experiential writing experience. It has incorporated the techniques and valuable

lessons from the following domains of works: voice-based multimedia authoring, automatic recording and transcribing contextual data, life-logging experience and scaffolding for structured content.

2.1 Voice-based multimedia authoring

In *IVA* approach, voice-input/recognition is a key factor to enable in-situ authoring of multimedia articles. Voice-input has been explored as a popular modality for authoring multimedia contents due to its efficient [28], familiar [1, 3, 33], and omnipresent nature [10, 16]. Voice-input was initially explored as audio type of auxiliary method to support authoring multimedia contents. Examples include: Balabanovi *et al.* introduced a new device to facilitate both local and remote sharing of stories through a combination of photos and voice commentaries [3]; Vronay, *et al.* created PhotoStory that allows users to efficiently author and narrates an attractive photo-based story with cinematic effects [33]. Recent advances in voice recognition technology allows utilizing transcribed texts from speech, which can reduce traditional efforts for text-input. This new capacity is recently applied to reform previous text-input scenarios e.g. adding textual tags on multimedia with speech [30], composing and editing short messages on-the-go with speech [16]. Extending the body of work on voice-based multimedia authoring, we apply it to a new domain of in situ authoring of complex multimedia documents.

2.2 Automatic recording and transcribing contextual data

Automatic recording and transcribing contextual data for ease of experience sharing is an active research area involving a large body of work (e.g. [6, 9, 11, 23]). Chen, *et al.* [11] used a semi-auto tagger to tag photos with collected contextual information during experiences. These tagged photos can be used to generate sentences for quick travel blogging. Moreover, semi-automatic creation of blog posts from contextual information has been proposed by Mobilog [9]. This solution generates short mobile blogs that can leverage the automatically retrieved information by an annotator with synthesized contextual information such as: weather, location or personal data from social network service collected during the trip. Melog [23] and Braga, *et al.* [6] showed a more advanced approach for recording and utilising contextual information with automatic content generation for experience sharing. They identified specific events by considering the relationship between photos and location-based data, whereby, they used these detected events for automatic blogging in a micro-blog at a later time. Inspired by previous approaches, the *IVA* workflow also auto-captured contextual information to simplify the recording process and to preserve the rich contextual information associated with events.

2.3 Life-logging experience

Many studies in life-logging support keeping and recalling daily moments with providing external memory cues (e.g. [4, 19, 22]). We adopt this technique to facilitate the editing process of *IVA* workflow. According to Tulving [31], episodic memories are memories of *who, what, where and when*; different types of cues, e.g. information of time or location, social contexts or photos, may have a different ability to cause different parts of a memory to be reactivated. Gouveia *et al.* [17] analyzed and explained how multimedia cues trigger memories of events, and showed the effectiveness of the cues for memory recall. With reflection on the principle for memory recall, we encourage users to capture a moment in a multi-faceted way, taking pictures while speaking out their thoughts or feelings, and saving the voice-annotated visual with time and location information of the moment together, thereby, the content editing can be performed with multimedia cues for the captured moments with lower cognitive loads for memory recall.

2.4 Scaffolding for structured content capturing

To overcome one challenge of voice-based in situ content creation - distraction from structuring and planning in-situ speech - we consider scaffolding. Scaffolding was introduced as a concept for tutored instruction for education [34], but has later extended and applied to support people in creating digital contents. Chi and Lieberman proposed a system called *Raconteur*, which guides the users to develop their travel story in an attractive narration [12] so as to help users generate content that is richer and more memorable for readers [13]. Kim *et al.* [20] introduced a pattern-based scaffolding system called *Motif* to make it easier for novice users to create high-quality video content with recording patterns extracted from professional works. We use a Q&A style of scaffolding to reduce distraction for structuring in-situ speech in the workflow of *IVA*. The Q&A approach to elicit answers from users is common in diary studies, i.e. Carter and Mankoff [8] suggested that photo-elicitation combined with a specific set of questions might draw longer annotations. We scaffold the voice-based content creation with sequenced questions, whereby, with less cognitive demands for structuring in-situ, speech answering questions in sequence can generate organised snippets, resulted in more structured and higher quality drafts.

3 THE VOICE-BASED LIVE AUTHORING APPROACH - DESIGN RATIONAL (1ST ITERATION)

To design and build LiveSnippets, we employed an iterative user-centered process that involved design iterations within the research team and informal testing with four external participants. Our initial design rationale is explained below.

In-situ Voice-based Multimedia Authoring of an Experience. The existing process of experience capturing involves retrospective writing where people first encounter memorable events, then go through iterations of drafting about the events and editing afterwards. This process can be time consuming. For instance, to write about her experience of one month travel, Jurga—a popular travel blogger—spent the entire two month after the trip [14]. Our goal was to improve the process via *in-situ voice-based multimedia authoring (IVA)*, by incorporating the following characteristics.

- **Divide-and-conquer.** Writing is a complex task that involves at least two aspects: content creation and editing [26]. We can divide these two aspects into sub-tasks and conquer them separately. For writing about experiences, much of the

content creation can be pushed to the moments during the experience, leaving the majority of the editing work to be done after the experience. In addition, instead of asking users to create content in one sitting, it can be divided into small pieces and be finished moment by moment, which makes the process easier to accomplish.

- **In-context writing.** We would like to support in-context writing, in which writers write down their thoughts at the moment when the experience is happening. This allows them to use distributed cognition to describe what's happening, making writing more vivid and easier to perform. Another potential benefit of writing in-context is to save the effort of deliberate planning for a different time and place to write, so that writing can start before memory fades away.
- **Minimal modal resource contention.** We want to design a writing process that does not divert user's attention to the experience. Because humans heavily rely on vision to perceive the experience, it can be beneficial to use eyes-free modality for content creation. We turn to voice-based input to design *IVA* workflow. Voice-based input has an added benefit of being fast, which is essential for composing content in a timely manner. To make voice input more seamless during the experience, we propose combined voice input with photo and video taking activities so that speaking to write feels more natural and integrated with the experience, as previous studies showed that users often naturally give voice commentary during such activities [15, 21].
- **Multifaceted data.** One problem of the traditional experience writing process is that users do capture multimedia during the experience (such as photo, video, written or audio notes, etc.), but they are stored in separate applications. This creates a significant amount of overhead after the experience to connect all the data together. Here we want to enable a more organised way to capture the multifaceted data. That is, in addition to capturing user's speech, we also capture other types of information relevant to the experience, such as pictures, videos, location, date/time, and so on. Different types of data will be indexed by moments and stored together episodically.

Snippet as a Unit of an Event. Designing the *IVA* technology via internal design iterations and informal study led us to represent experience as a series of events. We made this design choice because it aligns with the fact that people perceive experience as a series of events [36] and share the series as a narrative [29]. Each event is represented as user-generated text and other media (e.g. pictures, videos). Thus, the *IVA* tool should (i) let people compose a small chunk of text about an event, (ii) associating the composed text with other visual media, and (iii) digitally store the composed chunks in a sequential and cohesive way to represent a narrative. We call each chunk of composed text and media as a *snippet*. We design the technology that supports piece-wise composition of media-rich snippets and enables to capture them as a sequence of events thus people can easily review and edit them. We describe the detail of this design choice in the section Snippet Model as below.

4 LIVESNIPPETS - AN EXAMPLE APPLICATION

We developed LiveSnippets, a mobile Android app that manifests the design goals described above to enable *in-situ voice-based multimedia authoring (IVA)*. The initial prototype consists of the internal data structure and voice-based multimedia authoring interface.

4.1 Snippet Model

The snippet model is the data structure to record users' experiences. We designed the snippet model to capture *who* or *what* involved in the story with textual or visual information, *where* and *when* the event happened with the objective meta-data (GPS and timestamp). In addition, the model includes *how* and *why* the event occurred and the elements that make a moment special as they tell the underlying story, to connect the readers with the author and make the story more believable. They can describe motivations, needs, feelings, inner thoughts, or reflections. As we describe below, we encourage the user to capture this information through scaffolding via voice-based text composition [12].

The snippet model encourages holistic capturing and manipulation of snippets. Whenever possible, each snippet tries to include the relevant information, 5W and 1H, together as one unit. As each snippet represents an event, the re/arrangement of snippets naturally forms a narrative in which the users share their experiences with others.

4.2 Interaction Design of In-situ Voice-based Multimedia Authoring

As illustrated in Fig. 2, after starting LiveSnippets, a user can see a view of camera just like when launching a default camera app installed in an Android phone. The initial snippet is created at this moment and any user's utterance about their experience can be added into the current snippet until the user goes for next snippet creation. Once the user has captured a photo, the system shows its preview. Retaking a photo is allowed if the user stays on the current snippet creation. Since it is common to see good multimedia articles with groups of photos, we also allow users to take multiple photos in one snippet by providing a multi-photo option.

The captured photo, the recorded voice, and other metadata (e.g. GPS location and timestamp) are saved together as a bundle of information (i.e. a snippet) when the user clicks on the photo-save button at the photo view. To indicate that the system is recording the user's utterances, the system shows a mic icon on the top left corner at the camera/preview view. Once the user finishes capturing experience, the user can click on the document icon at the left-bottom corner of the camera/photo view, and the system stops recording and generates a draft at the snippets view as a stack of snippets chronological ordered by default based on the timein which the snippet was first created. Thereby, the user can review and edit each snippet presented as a card containing a picture/video clip, transcript of the recorded utterance, and other metadata (e.g. GPS location). Once the user is satisfied with the snippet's organisation and contents, the user can simply convert the draft of snippets into a HTML format of article by clicking on the publish button at the top of the snippets view.

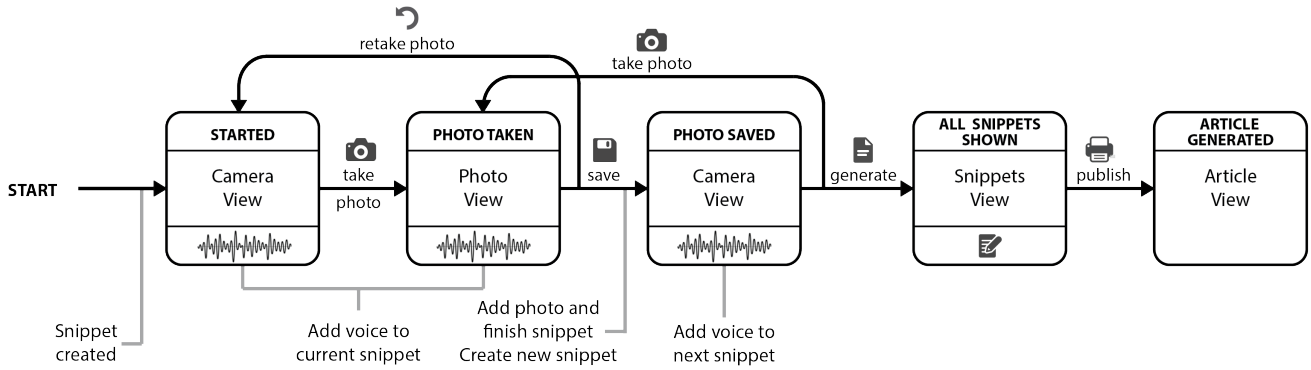

Figure 2: In *IVA* flow, the user can capture live experience as snippets with voice-based commentary over captured visuals, whereby, the snippets can be quickly reviewed, edited and published.

4.2.1 Implementation. LiveSnippets is implemented as an Android app on Oreo 8.1.0. LiveSnippets contains two main modules, ExperienceRecorder and SnippetCreator. The former responds to user's input events to capture moments for experience, including location, time, visual, and speaking. The latter creates snippets with texts, transcribed from the audio file by Google cloud Speech-to-Text API. We use Google cloud Speech-to-Text API for transcription (Java library for Android : v1-rev54-1.24.1).

5 INFORMAL TESTING

We informally evaluated the first working prototype of LiveSnippets with a small number of users. We were interested in getting initial feedback on the viability, usefulness, and potential drawbacks of the *in-situ voice-based multimedia authoring (IVA)* approach.

5.1 Study Design

5.1.1 Participants. 4 participants (2 females, mean age = 26.8, SD = 3.4) were recruited from the local university community. All participants were tech-savvy and familiar with using mobile devices. Two of them (P1, P2) rarely use social media, although P1 had run a blog for several years before discontinuing it last year. The other two participants (P3, P4) had personal blogs updated on a weekly basis.

5.1.2 Procedure. 1) *Demonstration*: participants were first introduced to the main functionalities and workflow of LiveSnippets by a facilitator (*3-minutes*) 2) *Try-out*: the participants were asked to try out the system until they self-reported being comfortable using the system (~5 minutes). 3) *Guided task*: participants were asked to generate one draft document describing the room they were located in using three or more snippets (10 minutes). 4) *Free task*: participants were asked to generate a document of another experience using three or more snippets. (10 minutes) 5) *Interview*: participants were interviewed about their experience of using LiveSnippets.

The whole process, including the post-study interview, were video recorded for further analysis.

5.2 Observation and User Feedback

Participants responded enthusiastically towards our new approach. All participants highly valued the convenience of being able to produce a relatively detailed multimedia document in a short time by speaking. It came to them as a pleasant surprise, as before using our prototype, P1 never thought he could "*get a draft of multimedia document just by speaking.*" But after using it, he was surprised by how convenient it was to write about one's experience using speech. A similar idea was expressed by P4, as she found "*it is more convenient than I thought before using it.*" In addition to the convenience, participants liked how the visual, speech and contextual information were recorded together in snippets, as it allowed moments to be described in a complementary fashion so that "*I don't forget any details*" (P4).

We had two concerns before the study; 1) Are participants fine with speaking in public areas to generate content? 2) Are they able to speak with little preparation? The feedback we received eased our concerns: they found using voice-input to author multimedia article during the experience was not obtrusive as much as they expected. Besides, "*it was not as difficult to speak as I thought because I just needed to describe what I saw in front of me*" (P1).

However, there was an unexpected concern with one design feature: in our first prototype, since we did not know when the participant would speak, we left the speech recording to be continuously recording. The participants did not find it comfortable because it gives a pressure to always try to give a speech. In addition, two participants (P3 and P4) became more conscious of what they say and try to plan what to say, and debated what was appropriate to talk about because they wanted to make a perfect and structured speech to produce quality of outcomes. With this mindset, the two participants showed frequent hesitations and pauses during speaking.

Another finding from the study was that although participants were pleased with the detailed output document, they felt the content they generated was not interesting enough to share. When asked why, they said they didn't have much time to think, and as a

result, the generated content often lacked the deliberate structure of a carefully written article. The content typically only consisted of a series of photo text blocks describing a sequence of events, but lacked emphasis and contrasts, which made it less interesting to read.

5.3 Section Summary: Additional Design Goals

Our initial test, although with a small sample, provided valuable information regarding the *IVA* approach. From users' feedback, we could see a promising potential of our approach: all participants found it useful, time saving, easier, less effortful. We also discovered that always-on audio recording was not preferred by users, and a straightforward recording of events may generate output that is mundane and less interesting. Based on the feedback, we worked on two enhancements for the initial prototype.

1) For a convenient control of audio recording, we added a simple function to pause and resume audio recording by a long press anywhere on the screen. This function allows users to initiate audio recording instantly with one-step interaction while staying focused on what they want to capture.

2) We realized that users need guidance to help organise their spoken contents so that the output can be better structured. Due to the impromptu nature of speaking, they often don't have time to think about the best structure to organize their content, especially for less experienced users. This led us to think about possible scaffolding guidance to help them generate more professional documents, which we discuss in more detail in the next section.

6 SCAFFOLDING CONTENTS CREATION - (2ND ITERATION)

To support the structured capturing process in the *in-situ voice-based multimedia authoring (IVA)* workflow, a Q&A style scaffolding approach was considered. Carter and Mankoff [8] suggested an insight for question-and-answer based annotations for capturing media to induce more information from users. From a different view, we adapt this insight to guide user's voice-based live capturing process; whereby, users can generate contents by answering a list of questions asked by the system. These questions appear while capturing either through voice or text on screen depending on the user's setting. Fig. 3 shows how users use scaffolding features in *IVA* to create recipes. This approach aims for users to attain a more structured stack of snippets without deliberate planning during the capturing process.

6.1 Scaffolding templates

The ultimate goal of designing scaffolding templates is to use questions to guide users to develop a draft that has a close-to-publishable quality. To design example scaffolding templates, we sought popular topics in everyday writing according to four general types of discourse: narration, description, exposition, and argumentation [35]. Considering that instant speech might not be suitable for argumentation discourse, we focused on the other three types and chose one example topic for each: travel (narrative), recipe (expository) and product review (descriptive). To increase the practicability of the templates, we investigated the "best practices" on the Internet for published documents in the three topics: 20 travel writings of established writers ranked by Forbes, 20 recipes posted by BBC Food, and 40 product reviews generated by CNET / Wired and written by top customer reviewers ranked by Amazon. For each set of example documents, we analyzed their most frequently appearing elements and their structures and sequences.

Note that we are more interested in how scaffolding as a technique can help *in-situ voice-based multimedia authoring (IVA)* in terms of aiding in sequencing in-situ speaking, rather than designing the best possible scaffolding templates for authoring quality of multimedia documents. We envision the best templates will be generated and shared by users. Future systems that implement the in situ author approach should support the upload, rating and sharing of these templates. However, in order to study the effect of scaffolding templates to user's motivation in *IVA* workflow, we needed to generate a few ourselves.

6.1.1 Recipe & Product review. From BBC's recipe examples, we extracted the eight most frequently appearing elements: title, photos of the dish, preparation/cooking time, serving information, introduction, ingredients, methods, and tips or tricks. To keep the template simple, elements were filtered based on their popularity among readers: informative contents, such as: photos of the dish, introduction, and tips were included, and more advanced elements such as calorie count, nutrition, or specific information for cooking supplies were excluded.

Similarly, we extracted eight elements for product reviews including: title, photos of the product, introduction, general information (such as price, target user, price and company), description with experience, pros and cons, evaluation, and conclusion for the product. We created simple questions based on these elements for each topic, see them in the supplementary material.).

6.1.2 Travel writing. Compared to the recipe and product review, travel writing is more free-form and difficult to extract a common pattern across the examples we retrieve online. Instead, we embedded five abstract headings, which have been partially observed in the best travel practices as common elements, into a stack of snippets at the top of the editing view in order to encourage users to regard the stack as a draft and edit it for publishing; these headings are: title,interests, others, information, and summary. Users were also allowed to edit the given five headings or add their own headings, texts or images at the editing view according to their style and preference.

7 EVALUATION

With the improved version of LiveSnippets, we conducted a more comprehensive evaluation. In particular, we are interested in understanding how the key design decisions associated with the *IVA* approach (divide and conquer, in-context writing, minimal modal resource contention and multifaceted data) changes writing experience of both types of authors, how they accept the changes, and what challenges and opportunities are uncovered in the *IVA* approach under different experience writing scenarios. We are also interested in understanding the roles voice-based multimedia authoring and scaffolding played in *IVA*.

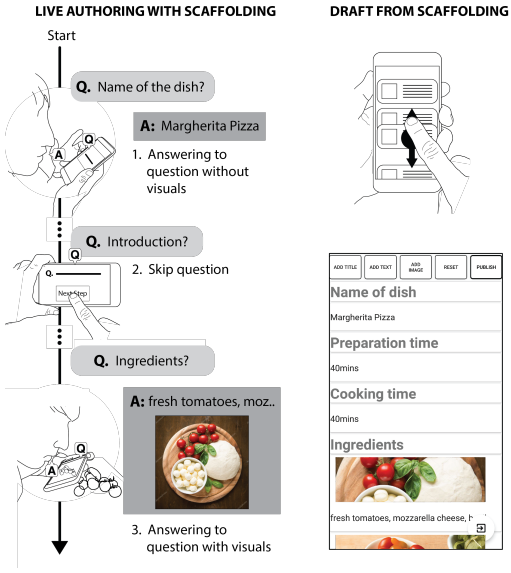

Figure 3: Q&A scaffolding for recipe in *in-situ voice-based multimedia authoring (IVA)* flow. Sequentially captured contents with three patterns of answering questions (i.e. answering only, passing or answering the given question with visuals presented in sequence on the left) result in a structured stack of snippets with subtitles (right).

7.1 Study design

To study the generalizability of our approach, we introduced two factors in our study: *Writing Scenario* and *Writing Expertise* as explained below.

- *Writing Scenario:* Experience writing covers many scenarios. Questions remain as to whether *IVA* can be used in different experience writing scenarios. To seek insights into this question, we picked three different types of experiences (travel, cooking, product usage) to test the generalizability of the *IVA* approach, covering three basic types of discourses as explained in the previous section.
- *Writing Expertise:* In HCI, it is well known that expert users focus more on efficiency (e.g. keyboard shortcut), while novice users prefer ease of use (e.g. drop-down menus). To investigate how *IVA* serves users who have different writing expertise, we divided them into two groups: expert vs. novice writers. Expert writers are the ones who publish articles actively (at least 10 articles/month in recent months, as the top 20% of online bloggers), while novice writers are the ones who publish occasionally, with less than 1 article/month (as the bottom 20% of bloggers in productivity [2].

7.1.1 Participants. Twelve participants (3M, 9F, mean age = 25.4, SD = 5.1) were recruited. Six of these participants were expert blog writers while the others were novice. The sample was divided into three groups with each group constituting two expert and two novice bloggers, to cover three different experience capturing/sharing scenarios: travel (P1-P4), recipe (P5-P8), and product review (P9-P12). Participants were paid the local equivalent of US$10 per hour, and we awarded US$35 to the participant who created the "best blog" as chosen by the research team.

7.1.2 Procedure. Initially, the experimenter give a 3-minute demonstration of the main functionalities of how to use LiveSnippets as described in section Interaction Design of Voice-based Multimedia Authoring above. Participants were then given a training session of using LiveSnippets to create a short sample article for the assigned scenario. For cooking, they were asked to select a random ingredient available in their kitchen (i.e. egg, potato, instant noodles, etc.) and taking three snippets (each snippet consisting of a photo with voice commentary) to explain part of the cooking process. For product review, participants were provided with a sample product (a Starbucks coffee mug) and asked to create three snippets to review their experience of using it. For travel, the experimenter brought the participants to a nearby garden to create three snippets from the scenes to generate a short travel article.

After the training task was completed, participants could continue to play with the system until they were comfortable with it before we proceed to the measured tasks. The training session took 20 to 30 minutes with an average of 24.8 minutes.

The participants need to perform in each scenario:

- Travel: participants were asked to take a tour of a scenic spot within or near the university campus, and write a journal about their travel experience. P1: garden, P2: museum, P3/P4: an island.
- Recipe: participants were asked to cook one of their favourite dishes in their own kitchen and write a recipe for it.
- Product review: participants were asked to bring one of their favourite portable products to a designated room in the university and write a review of the product.

For the recipe and product review writings, since we expected that participants could finish the task relatively quickly (within 1 hour), we did not constrain the time; the travel writing would be more time consuming, so to avoid taking too much time, we set a limit of two hours to complete the study.

In the recipe and product review writings, participants can proceed or skip questions if they needs to control pacing of Q&A scaffolding, however, returning to previous questions was disabled to reduce the possible tedium and anxiety associated with planing and revising. Scaffolding questions were presented to participant in both audio and text formats. When a participant proceeded to the next question, the audio was delivered by a text-to-speech engine, while simultaneously the text was also displayed on the camera-view.

After completing capturing and generating a stack of snippets, participants were asked to find the experimenter so that they could edit snippets on their mobile or a desktop with a web-based HTML editor. They could have as much time for editing as they wanted until they were satisfied with the quality of their documents, and they were asked to have comparison of estimated time-to-completion between the *IVA* workflow and their traditional ways when finishing their editing. Since the time-to-completion can be affected by various factors, i.e. familiarity with writing, editing, tools, etc., which is out of our main interests, we regard it as an individualised process and only perform personalised comparisons.

7.2 Data collection

In total, 12 detailed multimedia documents were created by our participants as drafts for given tasks. All participants were video-recorded as they used LiveSnippets for later analysis to gain a deeper understanding of their behavior. Upon completion of the given task, all participants were asked to fill out a post-study questionnaire on google forms as well as an interview. All interviews were video recorded, transcribed for analysis.

7.3 Results

7.3.1 Overall Experience and User Acceptance. All participants found the new approach easier than the traditional approach of experience writing (lower the barrier): it is quicker to have a draft (drafting, 10/12: P1-P5, P7-P9, P11-P12) and that it imposed less mental load for content gathering and post-processing (editing, 6/12: P3,P4,P6,P8,P10,P12). After participants created their draft using LiveSnippets, we ask participants to estimate how much time it takes to generate drafts of similar quality and length using the traditional approach. The average time logged for participants to complete the draft using *IVA* is 24.08 minutes (SD = 17.3), while the average of estimated time for finishing a draft with similar quality using traditional approach is 59.2 minutes (SD = 31.1). This results in a 59.3% time savings in completing the writing task. However, we suspect the actual time saving may even be greater since users tend to underestimate the time they needed to complete a task due to a well known psychological phenomenon called planning fallacy [7], and 59.2 minutes is much lower than the average time (3 hours and 28 minutes) users spend on writing a blog post [2].

The improvement, according to the participants, was due to a number of reasons. First, *IVA* helped them to generate a significant amount of content during the experience, which *"gives me something to start from"* (P2), and *"I just need to edit it, instead of writing from scratch"* (P4), which significantly lowered the *"mental weight/effort"* (P2) associated with completing a piece of writing. Second, such content was generated using the regular photo/video taking with voice comments, which *"removes the tedious typing"* (P5). Moreover, participants found the way in which the captured content was organized (snippet model) helpful. The output produced under snippet model arranged and formatted the photo, text, and context information automatically in segments, reducing the effort to *"format of the position the photos"* (P5). The grouping of multimedia information (visual, text, context information together) was also highly valued by participants. As P1 put it *"the ability to tightly couple comments to photos jogs the memory and documents the experience more closely. If I were to replace the generated text to be more succinct and add some signposts, it would make for a coherent and informative post"*. Therefore, participants believe the new approach would motivate them to *"write more"* (P7 & P8).

Six participants (P1-P4, P6 and P9) stated that voice input generated many errors in transcription, which is not ideal. This is the main reason why P9 was less satisfied with the overall experience and rated it as 3: *"language was not well processed, when I look at the transcript, it is inaccurate and I forgot what I said before. So I give it 3"*. Other participants were also bothered by the inaccuracy of speech recognition, but they thought it was still useful *"Even the text can be inaccurate, but it is fine to me because it is enough for me to remind"* (P2).

In addition, two participants (P9, P10) expressed doubts in adopting *IVA* for their formal experiential writing. One concern was not being used to write using speech as P9 mentioned *"I write slowly. Using LiveSnippets, I feel more pressure to 'write' in real time, so I am not sure whether it works for me"*.

P10 was not convinced it could be used to write publishable articles, as this required more careful thinking and efforts, but she believed it could be used for other purposes, such as *"writing a personal diary, because it gives me some drafts easily through what I talked"*.

The average satisfaction score of the draft produced was 3.91/5, with 1 being unsatisfied and 5 being satisfied. Almost all the participants rated LiveSnippets 4 out of 5 with the exception of P9 who gave a rating of 3.

7.3.2 Divide and conquer. From participants' feedback, it appears that divide and conquer offers two benefits. 1) instead of trying to write everything at once, *IVA* divides the writing tasks to individual moments, and each time, participants only need to talk about what's happening in front of them. This approach makes generating content much easier, since the content is generated little by little using speaking, "it didn't even feel like work" (P2). Another benefit of divide and conquer mentioned by the participants is the clear separation between content creation and editing. During the experience, they only need to worry about content creation, as the interface itself does not provide an easy way to edit the content on the go. The participants appreciate this separation. P10 elaborated: *"It does benefit me, as I do not need to stop at every paragraph and think about what else I left out, and/or scrutinising the types of words that I should use, etc. Thus, it does save time."*

Although participants appreciate the separation between content creation and editing, they don't want this separation to be strict. At times, they may want to perform some editing even during the experience, if there is a convenient location and time for them to do so, such as when having a break in a coffee shop. They also want to be able to create more content later, as thoughts can emerge after the experience.

7.3.3 In-context writing. Participants found in-context writing allow them to come up with content much more easily. This was true even for casual writers. Even though they were not initially sure what to say, when they actually used the application, they found it easy to describe what they had seen while taking the photos.

Being able to write in-context allowed them to provide a lot more details of what they see. As mentioned by a participant, *"When I was taking the photo, I was simply talking about the different features and steps naturally, but when I reviewed the snippets, I am surprised by some of the details I was able to provide in the snippets, as I won't remember to write it down if I was not"* (P6), similar ideas has also been expressed by P9. This finding is consistent with theories of contextual inquiry [5], in which users are more likely to provide details when they explain things in context. If they try to describe things retrospectively, they tend to summarise and forget about the details.

7.3.4 Minimal modal resource contention. The average rating of disruption of the recording activity during *IVA* is 1.16/5 (1=non-disruptive, 5=disruptive) among the 12 participants. Moreover, almost all participants rated LiveSnippets as 1 out of 5 with the exception of P8, who gave it a rating of 3.

Although the participants commonly regarded LiveSnippets as less disruptive, their reasoning differed according to their given task. All the participants for travel writing found the *IVA* approach less disruptive (P1-P4), while most participants from product review (P9, P11, and P12) found the disruption to be an acceptable trade-off.

This is because *IVA* by taking photos and providing voice comment is natural to perform during travelling and product review has little conflict with other activities during the experience.

Whilst, in cooking and product review, participants need to use their hands all the time to cook or manipulate the product, it can be difficult to perform *IVA* if their hands are busy. This conflict is more apparent in cooking than product review. P8 (rated 3 out of 5) elaborated: "I was quite concerned if I made the phone dirty by my wet hands during cooking. Also, sometimes, I needed to focus on the cooking itself to prevent it from being overcooked. Thus, I just wanted to quickly take a picture and then hoped to comment on it later". Nevertheless, all participants, except P8, reasoned that the disruption is acceptable if they consider the gain of having a draft to be prepared.

7.3.5 Multifaceted data. Participants find great value in capturing the multimedia data together at each moment. Having text and contextual information *"captured valuable parts of an experience that cannot be captured by photos alone"* (P1). Also, *"to have both visuals and text together serve as a great reminder for further modification later"* (P3), because sometimes they could take a photo, but forgot why they had taken it, or took down some notes, but forgot to capture the visual (P2).

P2: *Usually I took a bunch of photos and dumped them into Dropbox, but sometimes I forgot why I took this photo, so it is quite nice that this tool gives me some reminding for what I captured.*

7.3.6 Varied Scaffolding Experience. The rating for the scaffolding feature differs across the different experimental conditions is (Travel: 2.75/5, recipe: 4.25/5, product review: 4/5, 1=unhelpful, 5=helpful). Compared to the scaffolding template for travel, the ones for recipe and product review had more agreeable structures as they are more standard.

When participants agreed with the provided structure of the template, they were happy with the automatic arrangement for them *"I like it because it provides the structure automatically while I can still control the content"* (P5), which allow them to *"not think much about the structure"*. Another benefit of scaffolding is that it makes the writing process more engaging as *"you feel you are telling a person who is listening rather than just saying something"* (P6).

However, for more free-style writing like for travel, a rigid structure throughout the experience was shown to provide less value. An effective scaffolding template should fit the writer's writing style, or matches writers' expectations of how the content should be structured. P2, see the basic headings suggested by the travel scaffolding template, but he *"wouldn't use it"* as it does not match with what he wanted to do. P1, P3 and P4, also found the basic heading we provided unnecessary as it does not fit the story they have in mind.

This however, does not means that template can not be applied to the more free-form travel experiences. In fact, participants suggest that it can be applied retrospectively, instead of ask questions in real time, the system can ask questions at the end of the experience for the three most memorable moments, and for each moment, please tell us why you found it memorable, and provide a summary on the main take-away of the trip to conclude your experience. These guiding questions will be useful for writers to pick the most interesting aspects of the experience to share with their readers.

In addition to the usefulness of scaffolding, we are also interested in whether system initiated questions will disturb the participants. We found that participants were generally not bothered by the questions, some even mentioned that they find it more engaging. The average rating of disruption from the scaffolding feature in live mode is 4/5 (1=disruptive, 5=non-disruptive) among the 12 participants, but there are places in which participants want improvement. P10 and P11 who rated it a 3 as they found it necessary to have flexibility on the scaffolding questions for less disruption. P11 commented *"If users could select questions and decide the questioning sequence by themselves, it would be less disruptive because they could review the questions before answering them"*. For P10, the scaffolding feature was not suitable for her usual writing habit.

7.3.7 Expert vs Novice Writers. Although both groups appreciates *IVA*, they have different expectation and emphasis. This is largely due to their differences in audience, content, and quality expectation of their writing.

Audience: All novice writers, except P11 (product review), mainly shared their writing with family or friends only. Expert writers, on the other hand, share writing with the public.

Content: Novice writers favour self-centred capturing as reported by Motif [20]. Expert writers think more about content that generate public interest.

Quality expectation: Expert writers has higher expectation of the quality of the draft, including the visuals and the transcribed texts while the novice group are less picky about the quality of the writing.

Such differences certainly translate into different requirements and expectations of the *IVA* approach. For instance, P3,P4 (novice writers) both expressed the desire of capturing more about themselves (i.e. suggestion for a selfie function or a selfie stick with remote microphone). Expert writers, on the other hand, seek for features that can make their writing stands out. i.e. using a selfie stick with remote microphone to *"take a 360 degree of panorama with clear voice recording so as to deliver moments more vividly to the readers"* (P1). Moreover, they have different expectations for editing and scaffolding. Novice writers expected less editing and are more happy with the default scaffolding features, while expert writers wanted additional text editing (P1,P5 and P9) or refining on their drafts to make them more shareable (P2). Expert writers looked for more detailed levels of scaffolding according to their capturing style than novice writers.

We found scaffolding is more important for novice writers than expert writers. Almost no novice writer knew what experiences to capture in the beginning and they felt self-conscious speaking out

loud to describe their experiences. Thereby, scaffolding was important for guiding them through the initial barrier. Expert writers saw less value in the scaffolding features we provided, as they tended to already have a plan of what and how to capture, which might not be accommodated by our scaffolding templates. Some felt the questions we used were too generic. They wanted more intelligent and context specific questions that can help them to produce more interesting content. For example, when P9 was shooting the water gun, he commented that if the system initiated a dialogue (like a curious human will do) regarding the different aspects of the water gun (i.e. what's shooting range?, etc.), instead of a general question *"How to use this product"*, it would help participants to produce more interesting content. Similar ideas were expressed by P10.

8 DISCUSSION

Our evaluation shows that participants were enthusiastic for *in-situ voice-based multimedia authoring (IVA)* as a new approach for experience writing; the drafts created by participants from the study and their comments demonstrate the feasibility and acceptance of the new workflow. Participants found it natural to perform the recording task during the experience (4.08/5, 5=non-disruptive). Compared to the traditional approach of experience writing, all participants found the new approach easier: it is quicker to have a draft (drafting, 11/12) and it imposes less mental load for content gathering and post-processing (editing, 6/12). Moreover, participants valued the well formatted draft including text transcription from their speech because it significantly reduced the need/effort for memory reconstruction and was much more convenient to arrange the media resources with the draft.

In addition, participants suggested compatibility of writing in-context and out-of-context experiences with the *IVA* workflow, i.e. drafting diary entries can be done at moments with *IVA* to keep vivid thoughts, feelings and opinions, and reflection on the instant moments can be made afterwards during organising and revising the entries with enough time for thinking.

8.1 Designing for Expert vs Novice Writers

Our observation showed that professional versus novice users have very different goals and requirements. To reflect the user difference, it needs further study and additional design. Future system implementing *IVA* approach may consider subdividing the users into various types to determine what features are differently expected and how they should be differently applied. To design and develop the various versions created for different types of users, it is necessary to study when, where and how the features are utilised in real-world situations according to the different user groups.

8.2 Scaffolding Regarding Topics, Contexts and Users

The user feedback for scaffolding implies that flexibility on scaffolding is desirable to users; in addition, scaffolding needs to be designed different for different type of experiences. For more free form experiences such as travelling, scaffolding during the experience can be difficult. Also, scaffolding should accommodate different type of users. Our study shows potential usability issues and calls for further studies. For instance, scaffolding questions customised by untrained users may not match their usage contexts. In addition, in real-life environment, adjusting scaffolding could be tedious or easily forgotten as observed from participants in our study. In particular, increasing flexibility on scaffolding can negatively affect the quality of the draft and create another barrier for sharing.

8.3 Improving User Interface for advanced capturing

One obvious improvement that can benefit *IVA* is to increase the speech recognition accuracy. As explained previously, *IVA* utilises speech recognition to facilitate detailed sharing of experience via instant composition of a media-rich document. However, despite the technological advancements in this field, the text transcription still contains some errors which are highly dependent on the accent of the user as well as the noise level of the environment. Although the overall feedback is that the generated content, even with errors, is still valuable and useful, any improvement in speech recognition accuracy can significantly improve the *IVA* experience.

Another area of improvement is to design more advanced features to enable more sophisticated *IVA*. P2, an experienced travel blogger, utilised the group photo option plus the figure gesture technique to advance his storytelling—he captured a series of photos as a sub-story of his storytelling and elaborated on it with finger-pointing over photos. He expressed the desire for the enhanced accuracy of the gesture technique; and further improving the UI design towards increasing degrees of freedom; such as: allowing to draw lines over photos or decorate them with emoticons. To increase users' freedom with loose but robust UI design can improve the value of *IVA* in terms of advancing the users' experiences by reflecting users' variant patterns on capturing.

8.4 Generalisation toward various topics and other types of writing

According to the users' feedback, our core idea of *IVA*: using voice-based multimedia input during the experience, seems feasible to write many types of experiences. In addition to the three tested scenarios, participants feel that the same approach can be utilised for birthday parties, social gathering, dining, playing in the park, watching a sport game, etc. However, if the *IVA* approach is not preferred by users, or if an experience is not suitable for performing *IVA* with prohibition of speaking loudly or impossibility of carrying a mobile device, i.e. a quiet concert, swimming, etc., it will be difficult to apply the *IVA* approach.

In addition to experience writing, the *IVA* approach may also be used for fiction writers to create realistic settings about their story. Instead of sitting somewhere to imagine a setting, writers can go to a place to perform in-context writing using our approach. *IVA* can also be useful for journalists to collect stories and interviews during their trip. Instead of only having video clips, which is tedious to edit, they can obtain multimedia snippets in which text is automatically transcribed, making their editing job easier.

However, for each topic or genre, the best practice of writing can be different. Our design for the Q&A based scaffolding templates carefully extract questions from the best practices in targeting common fields necessary to generate a structured draft. However, it is impossible for us to provide all the templates to suit the diverse

needs of all users. Future systems implementing the *IVA* approach may want to create a platform which allows general users to share their own templates as a set of tuned questions to various capturing situations, but in turn poses a necessity of a tool or mode for authoring of template with ease.

9 CONCLUSION

In this paper, we introduced *in-situ voice-based multimedia authoring (IVA)*, a new approach for experience writing. By seamlessly weaving content creation into capturing, we aim to transform the traditional experience writing process. Instead of delaying the media-rich document composition after the experience occurs, with *IVA* users are naturally induced to draft a media-rich document from experience in structure by combing taking visuals with speaking under a Q&A typed guidance. This structured draft provides higher satisfaction to users in terms of quality and results in two main benefits for the post-editing process: less mental load for reminiscence and efficiency for organisation. Moreover, the user evaluation confirms that users are enthusiastic for the voice-based novel strategy for detailed sharing and want to generalise the utilisation of *IVA* in terms of topics, contexts and users with advanced UI design for increased user freedom over manipulations. Finally, we hope that our new approach of *IVA* can inspire alternative approaches beyond experience writing. In addition, by advancing the adaption of speaking into future UI design, we seek for an opportunity to enhance the ability to efficiently transform mental concepts into digital resources.

REFERENCES

[1] Morgan G Ames and Lilia Manguy. 2006. PhotoArcs: ludic tools for sharing photographs. In *Proceedings of the 14th ACM international conference on Multimedia*. ACM, Association for Computing Machinery, New York, NY, United States, 615–618.

[2] CRESTODINA ANDY. 2018. Blogging Statistics and Trends: The 2018 Survey of 1000+ Bloggers. https://www.orbitmedia.com/blog/blogging-statistics/

[3] Marko Balabanović, Lonny L Chu, and Gregory J Wolff. 2000. Storytelling with digital photographs. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, Association for Computing Machinery, New York, NY, United States, 564–571.

[4] Gordon Bell. 2001. A personal digital store. *Commun. ACM* 44, 1 (2001), 86–91.

[5] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual design: defining customer-centered systems*. Elsevier, Amsterdam, Netherlands.

[6] Reinaldo Bezerra Braga, Sócrates de Moraes Medeiros da Costa, Windson Viana de Carvalho, Rossana Maria de Castro Andrade, and Hervé Martin. 2012. A context-aware web content generator based on personal tracking. In *International Symposium on Web and Wireless Geographical Information Systems*. Springer, Springer Netherlands, Dordrecht, 134–150.

[7] Roger Buehler, Dale Griffin, and Michael Ross. 1994. Exploring the" planning fallacy": Why people underestimate their task completion times. *Journal of personality and social psychology* 67, 3 (1994), 366.

[8] Scott Carter and Jennifer Mankoff. 2005. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, Association for Computing Machinery, New York, NY, United States, 899–908.

[9] Pujianto Cemerlang, Joo-Hwee Lim, Yilun You, Jun Zhang, and Jean-Pierre Chevallet. 2006. Towards automatic mobile blogging. In *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, US, 2033–2036.

[10] Richard Chalfen. 1987. *Snapshot versions of life*. University of Wisconsin Press, 728 State Street, Suite 443, Madison wisconsin, United states.

[11] Yi-Jiu Chen, Wei-Sheng Zeng, and Shian-Hua Lin. 2013. Automatic Travel Blog Generator Based on Intelligent Web Platform and Mobile Applications. In *Information Technology Convergence*, James J Jong Hyuk Park, Leonard Barolli, Fatos Xhafa, and Hwa-Young Jeong (Eds.). Springer Netherlands, Dordrecht, 355–364.

[12] Pei-Yu Chi and Henry Lieberman. 2011. Intelligent assistance for conversational storytelling using story patterns. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, Association for Computing Machinery, New York, NY, United States, 217–226.

[13] Pei-Yu Chi and Henry Lieberman. 2011. Raconteur: Integrating authored and real-time social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Association for Computing Machinery, New York, NY, United States, 3165–3168.

[14] et al. Dawn. 2018. Starting a Blog – My First 3 Months Experience. https://fullsuitcase.com/starting-blog-experience/

[15] Jan Fortune. 2018. Why writing in place is so powerful. https://medium.com/personal-growth/why-writing-in-place-is-so-powerful-73534824247a

[16] Debjyoti Ghosh, Pin Sym Foong, Shengdong Zhao, Di Chen, and Morten Fjeld. 2018. EDITalk: Towards Designing Eyes-free Interactions for Mobile Word Processing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Association for Computing Machinery, New York, NY, United States, 403.

[17] Rúben Gouveia and Evangelos Karapanos. 2013. Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Association for Computing Machinery, New York, NY, United States, 2921–2930.

[18] S. Graham. 2006. *Writing in P. Alexander & P. Winne (Eds). Handbook of Educational Psychology (pp. 457-478). Mnawah*. NJ: Erlbaum, Mahwah, new jersey, United States.

[19] Tetsuro Hori and Kiyoharu Aizawa. 2003. Context-based video retrieval system for the life-log applications. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, Association for Computing Machinery, New York, NY, United States, 31–38.

[20] Joy Kim, Mira Dontcheva, Wilmot Li, Michael S Bernstein, and Daniela Steinsapir. 2015. Motif: Supporting novice creativity through expert patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Association for Computing Machinery, New York, NY, United States, 1211–1220.

[21] Stephen D. Krashen. 1984. *Writing, research, theory, and applications*. Pergamon, Oxford.

[22] Mik Lamming, Peter Brown, Kathleen Carter, Margery Eldridge, Mike Flynn, Gifford Louie, Peter Robinson, and Abigail Sellen. 1994. The design of a human memory prosthesis. *Comput. J.* 37, 3 (1994), 153–163.

[23] Hongzhi Li and Xian-Sheng Hua. 2010. Melog: mobile experience sharing through automatic multimedia blogging. In *Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing*. ACM, Association for Computing Machinery, New York, NY, United States, 19–24.

[24] Kate C McLean, Monisha Pasupathi, and Jennifer L Pals. 2007. Selves creating stories creating selves: A process model of self-development. *Personality and social psychology review* 11, 3 (2007), 262–278.

[25] Bonnie A Nardi, Diane J Schiano, Michelle Gumbrecht, and Luke Swartz. 2004. Why we blog. *Commun. ACM* 47, 12 (2004), 41–46.

[26] Mark. Nichol. 2019. What's the Difference Between Writing and Editing? https://www.dailywritingtips.com/whats-the-difference-between-writing-and-editing/

[27] James W Pennebaker and Janel D Seagal. 1999. Forming a story: The health benefits of narrative. *Journal of clinical psychology* 55, 10 (1999), 1243–1254.

[28] Sherry Ruan, Jacob O Wobbrock, Kenny Liou, Andrew Ng, and James Landay. 2016. Speech is 3x faster than typing for english and mandarin text entry on mobile devices. (2016).

[29] David C Rubin, Robert W Schrauf, and Daniel L Greenberg. 2003. Belief and recollection of autobiographical memories. *Memory & cognition* 31, 6 (2003), 887–901.

[30] Anh Truong, Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. 2016. Quickcut: An interactive tool for editing narrated video. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, United States, 497–507.

[31] Endel Tulving. 1984. Precis of elements of episodic memory. *Behavioral and Brain Sciences* 7, 2 (1984), 223–238.

[32] Bret Victor. 2014. Humane Representation of Thought: A Trail Map for the 21st Century. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 699–699. https://doi.org/10.1145/2642918.2642920 event-place: Honolulu, Hawaii, USA.

[33] David Vronay, Shelly Farnham, and John Davis. 2001. PhotoStory: Preserving emotion in digital photo sharing. (01 2001).

[34] David Wood, Jerome S Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Journal of child psychology and psychiatry* 17, 2 (1976), 89–100.

[35] Linda Woodson. 1979. *A handbook of modern rhetorical terms*. Natl Council of Teachers of english, Washington, United states.

[36] Jeffrey M Zacks and Khena M Swallow. 2007. EVENT SEGMENTATION. *Current directions in psychological science* 16, 2 (04 2007), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x