Designing CAST: A Computer-Assisted Shadowing Trainer for Self-Regulated Foreign Language Listening Practice

by

Mohi Reza

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Computer Science)

The University of British Columbia
(Vancouver)

August 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Designing CAST: a Computer-Assisted Shadowing Trainer for Self-Regulated Foreign Language Listening Practice

submitted by **Mohi Reza** in partial fulfillment of the requirements for the degree of **Master of Science** in **Computer Science**.

Examining Committee:

Dongwook Yoon, Computer Science
Supervisor

Bryan Gick, Linguistics
Examining Committee Member

Joanna McGrenere , Computer Science
Second Reader

Abstract

Speech *shadowing*, i.e., listening to some audio and simultaneously vocalizing the words, is a popular language-learning technique that is known to improve listening skills. However, despite strong evidence for its efficacy as a listening exercise, existing software tools do not adequately support listening-focused shadowing practice, especially in self-regulated learning environments with no external feedback.

To bridge this gap, we introduce **Computer-Assisted Shadowing Trainer (CAST)**, a shadowing system that makes self-regulation easy and effective through four novel interface features — *(i) contextual blurring* for inducing self-reflection on misheard portions, *(ii) in-situ annotations* for self–monitoring progress through tracking and visualization, *(iii) embedded recordings* for post-practice self-evaluation, and *(iv) adjustable pause-handles* for self-paced practice.

We base CAST on a formative user study (N=15) that provides fresh empirical grounds on the *needs* and *challenges* of those who practice shadowing using conventional software tools. We validate our design through a summative evaluation (N=12) that shows learners can successfully self-regulate their shadowing practice with CAST while retaining focus on listening.

Lay Summary

Shadowing, i.e. listening to some native speech and repeating the words at the same time, is a popular language-learning technique that is known to help foreign language learners improve their listening skills. However, existing software tools for shadowing are not well-suited for this purpose because learners cannot monitor misheard words while shadowing, nor can they assess their listening ability with these tools.

To address these problems, we built **CAST (Computer-Assisted Shadowing Trainer)**. With CAST, learners track and visualize their progress by marking over a blurred transcript. CAST reveals parts of the transcript to the learner only when reading those parts helps them notice misheard words. It creates and places shadowing recordings into the transcript so that learners can assess their listening ability by matching recordings with the text. We evaluated CAST and found that learners were successfully able to monitor misheard words and self-assess their listening ability.

Preface

This thesis is an original intellectual product of the author, Mohi Reza. The studies reported in Chapters 3 and 5 were conducted with the approval of the UBC Behavioral Research Ethics Board (certificate number H19-01380). A significant portion of this thesis will be submitted as a manuscript in a top-tier conference. I am the lead author of that manuscript. Dr. Dongwook Yoon provided supervisory assistance in formulating, framing, and ideating the problems addressed in this research, and assisted in the writing process for both this thesis and the manuscript. Dr. Bryan Gick and Dr. Strang Burton provided advice on research direction, motivation, and study design. Dr. Joanna McGrenere and Dr. Bryan Gick assisted with proofreading and editing. Dr. Strang Burton, Fatimah Mahmood, Misuzu Kazama, and Ashish Chopra assisted with recruiting research subjects. Members from the D-Lab research group led by Dr. Dongwook Yoon contributed by participating in pilot studies, prototype testing, and brainstorming sessions. Anna Offenwanger assisted with rating language proficiency levels of research subjects for the study described in Chapter 3.

Table of Contents

List of Figures

Glossary

CAST Computer-Assisted Shadowing Trainer

ESL English as a Second Language

L1 First Language

SRL Self-Regulated Learning

SRS Self-Regulated Shadowing

Acknowledgments

I thank my supervisor, Dr. Dongwook Yoon, for his constant support and guidance, and his invaluable mentorship throughout the last two years. I am also grateful to my thesis committee members, Dr. Bryan Gick and Dr. Joanna McGrenere, for their many comments that have helped me improve this work.

I thank my parents for their unconditional love, my brother and sister-in-law, for taking good care of me in Vancouver, and my dear wife, Labiba, for travelling *eleven-thousand miles* to join me in this journey.

I thank my friends, Anna Offenwanger, for introducing me to Bananagrams and preserved-peaches, Ashish Chopra for sharing those coffee breaks and walks around campus, Matthew Chun, for wide-ranging conversations, Kyle Clarkson for helping me move, Hanieh Shakeri for teaching me ice-skating, Steve Kasica for introducing me to hiking and tiny bike rides, Yelim Kim for sharing much-needed late-night snacks before a paper deadline, and Taslim Arefin Khan for being such an excellent grad-buddy. All of them have made my time at UBC so much more memorable.

Chapter 1

Introduction

"We have two ears and one mouth so that we can listen twice as much as we speak."
— Epictatus

Speech shadowing, i.e., listening to some target audio and *immediately* vocalizing the words [26], is a popular language-learning technique that is known to be effective for listening skill development [17]. Unlike written text, where the boundaries between words are clearly delineated, speech is a transient concoction of phonemes, strung together in a continuous stream of sounds. While *native speakers* can effortlessly disentangle these phonemes into words, *non-native speakers* have a much harder time. This is where shadowing helps with listening — it sharpens the phoneme perception skills of non-native speakers [19, p. 47], thereby improving their ability to disentangle sounds into words.

However, existing software tools for shadowing do not provide adequate support for *listening* practice, because mainstream usage of the technique is fixated on *speaking* skill development. The difference between shadowing for *listening* and shadowing for *speaking* is subtle but significant. The former targets *bottom-up listening skills* [20, 45], i.e., the ability to recognize words from their phonemes, whereas the latter targets aspects of *oral proficiency* such as pronunciation, accent, and intonation, that are tangential to listening skill development. A good example of a recent shadowing system from the HCI community that takes the latter approach is *WithYou* [55], a *speech-tutoring* system that automatically adjusts audio playback and difficulty level by comparing "a learner's *speech and pronunciation*" (emphasis mine), with a "speech template to determine if a learner's performance is good or not" [55]. Off-the-shelf shadowing apps (e.g., [13, 31, 37]) share a similar focus on speaking practice.

As a result, very little is known about the specific *needs and challenges* associated with listening-focused shadowing practice with software tools, and this has impeded the development of shadowing systems that can benefit many foreign language learners with weak listening skills. We can understand *why* listening hasn't received attention that is commensurate with its importance by turning

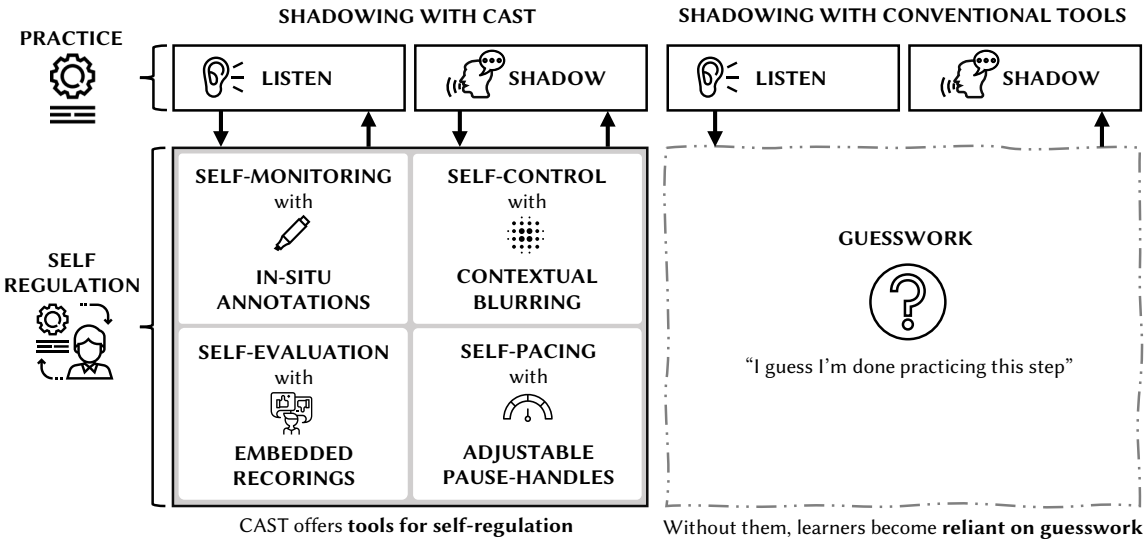


Figure 1.1: The *self-regulated shadowing* process with and without CAST

to the way it has been historically treated within English as a Second Language (ESL) pedagogy circles in relation to speaking. Listening is aptly referred to as the "Cinderella Skill" because it is often "overlooked by its elder sister — speaking" [39, p. 238]. However, the disproportionate focus on speaking within the context of shadowing is surprising because the background literature on the effectiveness of the technique for listening practice is *more substantive* than for speaking practice [20, p. 390]. This *does not* undermine the usefulness of speaking-focused shadowing systems, because speaking skills are important, and such systems may catalyze future research efforts on shadowing for speaking. However, this *does* signify a clear need for the development of shadowing systems that focus on improving listening skills.

To develop such a system, we first bridged the gap in our understanding of the specific needs and challenges of learners by conducting a formative user study with 15 ESL students, and found *self-regulation* to be the major stumbling block for listening-focused shadowing practice. We drew from a rich body of literature on Self-Regulated Learning (SRL) theory to ground our findings, and shaped the process for listening-focused shadowing around Zimmerman's SRL cycle [57]. We found that aspects of shadowing practice tied to SRL, such as, *monitoring* listening ability, *reflecting* on misheard portions of the target audio, *self-evaluating* shadowing performance, and increasing the overall *self-awareness* during practice were areas where learner's needed most support. Supporting these aspects proved to be particularly challenging for shadowing because the activity requires heavy multi-tasking (i.e. listening and vocalizing the words *at the same time*), and learners easily become overwhelmed by the sheer difficulty of the task due to high cognitive load [18].

The *transcript*, i.e. the written form of the target audio, proved to be a potential source of support to learners because we found that reading misheard words *after* listening to them enhanced the

learner's ability to reflect on their mistakes. However, we also found that using the transcript leads to what we describe as the *text-dependency problem* — firstly, when given access to the transcript, learners were tempted to read words in advance, i.e., *before* listening to them. This behaviour diminished their opportunity to reflect on their listening ability. Secondly, reading from the transcript shifted focus *away* from listening to the target audio, thereby, hampering the main learning goal of listening-skill development. This second observation is tied to prior experimental work on *selective attention* and reading while listening, which shows that our "ability to read and to listen concurrently is limited by the availability of both general and task-specific processing capacity" [28]. Our core challenge, then, was to design a system that supports the learner's ability to self-regulate their shadowing practice using the transcript, while retaining a strong focus on listening.

To address this challenge, we designed and evaluated CAST, a novel Computer-Assisted Shadowing Trainer that *enables* and *enhances* the learner's ability to self-regulate their shadowing practice in situations with no external feedback. Through iterative design, we developed four novel interface components that work together to make self-regulated shadowing easy and effective (see Figure 1.1 for an overview of how CAST improves self-regulated learning). In our approach, we use: (i) *In-situ Annotations* i.e., light-weight text-highlighting interactions over a blurred transcript, to tackle progress tracking through visualization, (ii) *Contextual Blurring*, i.e., selectively revealing portions of the transcript only when it helps with self-reflection, to resolve the text-dependency problem, (iii) *Embedded Recordings*, i.e., audio-clips of shadowing performance interlaced with the transcript, for post-practice self-evaluation, and (iv) *Pause-Handles*, i.e., strategically positioned pause markers that can be adjusted to introduce short breaks between chunks of text to reduce overwhelm during practice, without altering the target narrator's speaking rate. We validated our design through a summative evaluation study (N = 12) that provides evidence in support of the efficacy of CAST as a self-regulated shadowing tool for listening skill development.

In this work, we contribute: (i) CAST, the first shadowing system for foreign language listening practice in self regulated learning environments, (ii) fresh empirical insights on the needs and challenges of language learners for listening-focused shadowing practice, upon which we base our design, and (iii) results from a summative evaluation that validates our designs.

Chapter 2

Related Work

Our work has been informed by (i) SRL theory, (ii) previous education technologies in HCI that are concerned with self-reflection based learning (iii) the rich body of shadowing literature from cognitive psychology, simultaneous interpreter training, and language pedagogy, and (iv) interactive speech-based interfaces that use visual representations of audio to overcome its linear nature.

2.1 Enhancing Self-Regulated Learning

We set the stage by zooming out from our specific learning context of listening-focused shadowing practice, and situating CAST within a broader array of *systematic interventions* for enhancing SRL. We are motivated by previous SRL literature in favour of the notion that the "students' self-regulatory competence can be enhanced through systematic interventions"[41]. What strings together these interventions with CAST is their shared conceptual framework. These frameworks are described by various SRL models (see [36] for a review of six such models), two of which are of particular interest to us, namely, the Pintrich [38] and Zimmermann [56] models of SRL.

Pintrich's model comprises four phases: (i) forethought, planning and activation, (ii) monitoring, (iii) control, and (iv) reaction and reflection [38]. These phases are highly flexible, and only "specifies the possible range of activities" for SRL, and "does not necessitate them", nor does it "presume that the phases are linearly ordered" [40]. Therefore, when designing CAST, we have carefully considered which aspects from these phases needed most support in our specific learning context, by analyzing the empirical findings from our exploration of learner needs. For guidance on ordering, we turned to Zimmerman's model, and shaped our Self-Regulated Shadowing (SRS) process around its three cyclical phases, namely, (i) forethought, (ii) performance, and (iii) self-reflection [56].

2.2 Designing for Self-Reflection

Of the three phases in Zimmerman's model, self-reflection is of particular interest to us because we are concerned with enhancing the user's ability to reflect on misheard portions of the target audio. Supporting reflective practice through the design of new technologies has been of particular interest to Human-Computer Interaction (HCI) researchers for some time now [5, 14]. These technologies induce self-reflection through various approaches that we can broadly classify under *prompting* (e.g., [8, 42, 51], and *visualization* (e.g., [10, 16, 48]).

In the first classification, learners self-reflect by responding to prompts that concretize their thinking. For example, a learner may be asked to explain their solution to a math problem after finishing it [51], or to answer reflective questions while watching educational videos [42]. This approach works well only in situations where *interrupting* the learner during practice is okay or where we can reasonably expect the learner to *accurately recall* how their practice went after they complete the task. With shadowing, prompting is unsuitable because it induces heavy cognitive load [43]. This makes interruptions *during* practice far too obtrusive, and prompts *after* practice ineffective because cognitively loaded learners may not accurately recall how their practice went after shadowing.

In the second classification, learners use *information visualization* to glean insights from their experience, and in doing so, become self-aware of their learning process. For example, a student may reflect on how they spend their time by using a tool that charts out time-logs of their activities [16]. Visualization becomes especially helpful in situations where moving information from the learner's memory to an external form eases their cognitive burden, enabling them to see new patterns.

In the context of shadowing practice, the transcript can be used as a visual counterpart to the audio. However, as we shall discuss in Chapter 3, using the transcript for listening-focused shadowing comes with many caveats that we address through design.

2.3 Shadowing is Rooted in Listening

The background literature on shadowing reveals that the technique has important applications in two different, albeit connected, disciplines — *cognitive psychology* and *simultaneous interpretation*. While in this work, we are primarily interested in *language pedagogy*, tracing the rich history of the technique back to those two disciplines gives us useful insights on why shadowing is effective as a listening exercise.

Starting as early as the 1950s, Shadowing was used by cognitive psychologists to study selective attention [6]. A classic example of this is the application of shadowing in *auditory attention experiments* [9] to understand Moray's cocktail party effect [32] — why are we so good at tuning into a single conversation amidst a cacophony of background voices?

In those experiements, learners were given a dichotic listening task involving two different audio messages, one in each ear, and asked to shadow only one of those streams. For the stream they shadowed, participants were unable to recall the contents of the message, i.e., they focused on the *sounds* of the words, *not their meaning*. For the other stream, participants were completely oblivious to the message, and did not notice even when the language of that stream was altered mid way from English to German, i.e. they focused *solely on the shadowed audio stream*. These results hint at the power of shadowing as *focusing technique* that forces learners to pay close attention to the sounds of a single audio stream.

Shadowing found its second home among simultaneous interpreters[26][25], i.e., those who translate between languages in real-time. The technique became a precursory practice exercise that helped trainee interpreters practice timing, listening, and short-term memory skills [33]. Each of these dimensions can also benefit language learners. Here, timing refers to the ability to reproduce heard speech with little to no latency. While this skill is obviously beneficial for simultaneous translators, existing theory suggests that it can also benefit language learners by improving their phoneme-perception skills through bottom-up listening practice [17].

Building on insights from cognitive psychology and simultaneous interpretation, researchers from a Japanese EFL pedagogy context [45–47] spearheaded efforts in shaping the shadowing technique into a language learning exercise for bottom-up listening-practice. Since then, because of growing global interest in shadowing, the results of those efforts have been made accessible to a wider international audience in the form of books [19] and summary papers [17].

While there are a few examples of preliminary studies on the impact of shadowing variants on aspects of speech such as pronunciation [29], intonation [22] and oral fluency [52], the research on shadowing for listening skill development is more substantial [19, p. 390]. Our focus on listening-skill development with CAST is therefore aligned with the existing body of shadowing literature.

2.4 Existing Shadowing Systems

While the theoretical underpinnings of shadowing as a listening exercise are well-understood [19, p. 9], existing shadowing systems do not adequately address listening-focused shadowing, because popular usage of the technique remains fixated on speaking practice.

A quick search for off-the-shelf shadowing apps brings to light the imbalanced focus on speaking over listening. For example, downloadable shadowing apps such as [13, 31, 37] all focus solely on improving English *speaking* skills: [13] describes shadowing as "training for English fluency", and "the best way to improve English speaking", and [31] frames it as a technique for learning how to "speak like a native by improving your pronunciation, rhythm, and intonation."

A recent and noteworthy shadowing system stemming from the HCI community is *WithYou* [55], which uses "context-dependent speech recognition" to automatically adjust the audio playback and the difficulty of a "native speech template" when learners fail to shadow smoothly, thereby support-

ing them when they face difficulties, and helping them improve their speaking skills. We distinguish CAST from these existing systems by noting its strong focus on listening-skill development.

2.5 Visual Representation for Audio

From the perspective of interaction design, the transcript-based speech navigation features in CAST have their roots in early HCI systems such as SpeechSkimmer [4] and SCANMail [50] and more recent systems such as RichReview [53], TypeTalker [3] and Skimmer [24]. These systems overcome the transient, un-skimmable nature of audio using *visual representations of sound* such as transcripts [24, 50], threaded wave-forms [53], and captions [50, 53]. Early systems explored non-transcript based alternatives such as wave-forms and binary representations (e.g., pause vs. speech) [21] because generating transcriptions automatically was not practical. Since then, computer-generated audio transcriptions have become inexpensive and accurate, and so we opt for transcript-driven audio representation in CAST. With CAST, we build on top of these existing systems and apply what we learn to the specific context of multimedia-based language learning.

Chapter 3

Exploring Learner Needs

To gain a better understanding of our target users, i.e., language learners, we conducted a formative need-finding study with 15 ESL students, to unravel their *needs and challenges* associated with listening-focused shadowing practice. In particular, we explored how they practiced shadowing using *conventional tools*, namely, a representative general purpose media player for the audio and document viewer for the transcript, and how their needs and challenges were connected with two popular modes of shadowing instruction, namely, *video-based* and *in-person* instruction. We used conventional tools as opposed to a specialized shadowing system for our exploration because the absence of such a pre-existing representative system for listening-focused shadowing makes conventional tools a logical alternative for learners.

3.1 Method

We wanted to gain a *qualitative* understanding of learner needs and challenges, and so we conducted semi-structured interviews with our participants after providing them with two shadowing tasks, one for each mode of instruction.

3.1.1 Participants

Our target demographic consisted of 15 international students aged 18 to 24 (12 women and 3 men), who had taken formal ESL lessons within the last two years. We screened them to ensure that they spoke a variety of first languages (L1) including Chinese (Mandarin, Cantonese, and other local dialects), Korean, Russian, Ukranian, Hindi, and Arabic. Having participants with multiple L1 ensured that our findings weren't tied to specific L1 traits. English proficiency levels ranged from A2 (beginner) to C1 (advanced) on the CEFR scale [35], with B2 (intermediate) being the most common level. These levels were assigned and cross-validated by two native English speakers based on a two-minute recorded conversations at the beginning of each interview. Responses on prior familiarity with shadowing ranged from *definitely not* (13.33%) and probably not (26.66%),

to *probably yes* (50%) and *definitely yes* (20%).

3.1.2 Tasks

Participants completed the two shadowing tasks in a quiet lab environment. The first task simulated a scenario where the learner encountered shadowing in a video, and practiced on their own with no in-person guidance. The second task simulated a scenario where learners had access to one-on-one guidance from an instructor. By asking participants to practice shadowing twice, first with video-only instructions, and then with individualized guidance from an instructor, we were able to identify which of the needs and challenges were intrinsic to the technique, and which were tied to the level and mode of guidance.

3.1.3 Materials

We used a shortened version of a highly popular video on shadowing, with over 2.8 million views [34], for the first task. This video was chosen because it is representative of what learners may typically find when doing online searches on shadowing. While the title of this video mentioned speaking practice only (indicative of the fixation on speaking in popular shadowing usage), the video content covered the role of listening in shadowing. For example, it stated that shadowing "trains your ear to listen very very carefully", and that copying the target audio requires the learner to become "very good at hearing" it. For the shadowing material, we used a good quality audio narration and transcript of a standard passage, *Arthur the Rat* [2] in native British English [1]. To ensure equal difficulty level for both tasks, we divided the passage into two halves of equal length (160 words), and used one half for each shadowing task, counterbalancing the order in which they were presented to the learner.

3.1.4 Procedure

Participants completed a demographics survey before joining the study. This survey collected data on age, gender, education, first language, self-reported English proficiency level, and prior experience on shadowing. This data helped us with screening and interview-prep. First, we demonstrated how the conventional tools worked, and asked participants to try them out to make sure they were comfortable with using them. Then, participants watched the introductory video on shadowing and completed the first shadowing task. At this stage, we pointed out any mistakes that the participant made during practice (e.g. remaining quiet or not shadowing with a loud and clear voice) and provided in-person guidance on the shadowing process by going over each step with them based on a script adapted from shadowing instructions in [19]. Then, participants completed the second shadowing task. We recorded the computer-screen (with audio) during both tasks, and made observation notes from a distance. Finally, with the shadowing experience fresh in their mind, we conducted

a follow-up semi-structured interview with participants where we unpacked their needs and challenges. The entire study took approximately one hour to complete, with each task taking around fifteen minutes, and the interview around twenty minutes. We adjusted the time allocations based on results from three pilots. For more details on the procedure, see the study protocol in Appendix A.3, and the interview questions in Appendix A.4.

3.1.5 Analysis

Our data consisted of semi-structured interview transcripts, and task observation notes. Our goal was to deduce a set of requirements for a listening-focused shadowing system that addresses real-world learner needs and challenges, while remaining informed by existing shadowing theory. Therefore, we coded and analyzed the observation notes and interview transcripts using reflexive thematic analysis [11] through a hybrid, inductive-deductive lens [44], taking into account pre-existing theory on listening-focused shadowing as a central guiding theme in our interpretations, while using the empirical findings on needs and challenges to drive our selection of additional theories (i.e., self-regulated and multimedia learning theory). We chose a hybrid approach because we share the opinion that "researchers cannot free themselves of their theoretical and epistemological commitments, and data are not coded in an epistemological vacuum." [7]

3.2 Findings

In this section, we describe the findings from our exploration of learner needs and order them based on their prevalence in our data, as well as our judgement on their importance. The number after F signifies this order.

Learners want to practice alone (F1): When asked to describe the ideal environment for shadowing, 9 participants expressed a strong need to practice alone. This was due to two interrelated factors, namely, *self-consciousness* and the inability to *concentrate* on shadowing in front of others. P12 and P13 mentioned that they felt uncomfortable "speaking in front of people", and P9 felt that the "presence of others" made them "unwilling to speak". The issue of self-consciousness has been observed in related HCI sytems that require ESL students to speak, e.g., [54], but was exacerbated in our context because the shadowed speech produced by our participants was often garbled and unintelligible, especially when participants found it difficult to keep up with the native speaker's cadence (likely due to high cognitive load [18]). Isolation offered participants the *freedom to make mistakes*, which they valued. P4 noted that when alone, they were "...free to talk aloud...to make mistakes...to miss words...and to say them incorrectly". P14 noted that shadowing in front of an instructor "can be so messed up", because being observed made them "feel pressured". Therefore, the power imbalance typical in student-instructor relationships is an important sub-factor, and learners can benefit from self-regulated learning that does not require external observation by an instructor.

Certain attributes that had more to do with the nature of shadowing itself, and less to do with human relationships, made the self-consciousness issue more pronounced. For example, P16 mentioned that the act of imitating someone (as required by shadowing) was "kind of embarrassing" to do.

Even if we were to somehow overcome the issue of self-consciousness, participants still wanted to practice alone because they wanted a quiet environment where they could focus on listening. They mentioned a host of factors that hampered their ability to concentrate during shadowing, such as multitasking (since shadowing requires both listening and speaking) (P13), distraction from peers (P13, P15), and even differences in speaking style and level among peers (P8).

Learners have a hard time self-regulating their practice (F2): Practicing alone requires effective self-regulation. However, we found that self-regulation didn't come easy for learners, as they were prone to making poor strategic choices during practice, even when provided with video-instructions on how to shadow. The video instructed the participants to (i) listen very carefully to the audio, then (ii) shadow with the transcript, and then, (iii) shadow *without the transcript*. P7 and P10 remained quiet during the *entire* practice session, missing the basic requirement that the words must be vocalized during shadowing. P7 thought that vocalization wasn't necessary during shadowing and P10 quietly moved their mouth. P1 never practiced without the transcript, even though step (iii) in the video required them to do so. P3 spent considerable time reading the text *before* playing the audio, getting the order of steps wrong. Such behavioural patterns indicate that learners are unable to self-regulate their shadowing practice without added support and guidance. In the second practice session, where participants were provided with in-person guidance on steps based on their activities in the first session, they made better choices during practice. However, providing in-person guidance requires someone to observe their practice, which is in direct conflict with their need to practice alone (F1).

Reading *soon after* listening helps learners reflect on their mistakes (F4): In certain contexts, reading promoted self-reflection on mistakes because checking the transcript *soon after* listening to a difficult portion lead to *aha moments*, where participants realized that they misheard something. For example, P11 misheard "hole" as "home" during the listening step, and only realized this after shadowing that part with the text. P1, P10, and P12 had similar experiences, which P9 sums up nicely — "...I recognized so many things, so many words that I *thought* I understood, but it turned out to be a different word."

Our data offers insights on exactly *when* those aha-moments occur. Comments from P11 and P14 confirmed that participants only notice misheard words if they checked the text *soon after* the audio reaches that point in the passage. P13 felt that she could make "visual connections" between the text and audio, which wasn't possible with the audio alone. Several participants alluded to the need for timely comparisons between the text and audio by mentioning features from other systems that are familiar to them, such as subtitles (P5) and karaoke-style word highlights (P4, P14), where text accompanies audio in real time. P14 said that it took too long to "search through the text"

when looking for a misheard word. By the time the learner locates the misheard word, the audio has moved much further into passage, and this diminishes their opportunity to notice misheard words. Therefore, the timing window for self-reflection is small, and precise time-synchronized stimuli from the audio *and* the transcript is key to effective transcript-induced self-reflection. Reviewing the text *before* listening to the target audio diminishes the opportunity to mishear something, since the participant has already seen the word in writing. Reviewing the text *after* listening only works if the learner does not have to search through the text. Reviewing the text *soon after* mishearing something maximizes the chances for self-reflection — that is where the learning happens.

Learners tend to read the transcript and listen to the target audio *at the same time*(F3): Audio-only shadowing is cognitively demanding [23]. Therefore, when learners are given access to the audio transcript, they are tempted to use *reading* instead of *listening* as their primary shadowing strategy. This finding forms the bases of the *text-dependency problem* as discussed in the introduction. P9 "tried to *read the words* at the same time as the audio...", while focusing primarily on the text. P16 believed "it's so much better with the text because it's much easier" that way. P5 wished there was a step where they could read *without* the audio. P11 thought "reading the text and *then* saying things" made the exercise "kind of easy...", but with the text gone, they could only *partially* recall what was there before. We can glean two important patterns from these comments. First, we see that participants relied on the text because reading felt easier than audio-only shadowing. Second, as confirmed by additional comments from P6 and P14, participants were reluctant to remove the text because they were trying to *memorize* the the material in advance, as a shortcut. This is undesirable because the exercise requires them to rely on their *ears*, *not* on memorization when shadowing. The text-dependency problem became more pronounced in instances where participants couldn't keep up with the narrator's pace. In such instances, shadowing became a difficult game of playing catch-up. P10 felt that "...the audio [would] just keep carrying on, and you have to catch up...but that's very hard." These factors compelled learners to open the text.

Learners tend to skip difficult parts when overwhelmed, without revisiting them later (F5): When encountering difficult portions, learners tend to skip those parts. For example, P11 said "when I was lagging behind, I was like, okay, let's just skip it...and just go with the audio and then figure it out afterwards." Similarly, P8 said: "...for some really fast sentences, I just couldn't do it. If I do it, I know I'll miss that part and so I skip it." This behavioral pattern of skipping parts is problematic because learners don't always remember to return to those parts afterwards when they are overwhelmed. P10 described shadowing as "kind of stressful". It is worth noting that the sense of overwhelm is unequally distributed, because not all parts of the passage feel equally difficult. We see this in P9's comment,"in general, I don't feel that the pace is too fast, except when things became really unfamiliar." Therefore, instead of shadowing in a linear fashion, from top to bottom, participants can benefit from a more *selective* approach to practice, where in successive rounds, they focus *solely* on the parts that are still difficult for them.

Chapter 4

Designing CAST

The *overarching* design requirement (R0) for CAST is to enhance the SRS process for listening practice, because learners want to practice alone (F1), but they need support on self-regulation (F2). We expanded R0 by looking for the specific aspects of self-regulation that needed support in our learning context, and derived four supporting requirements, R1 through R4, that when fulfilled, resolves R0. Then, through iterative design, we developed four components, with each resolving an aspect of self-regulation: *self-monitoring*, *self-control*, *self-evaluation*, and *self-pacing*. We optimized these components for SRS through multiple revisions, and they serve as concrete examples of how SRS can be applied through design. Refer to Figure 4.1 to see details on how our findings, requirements.

Provide structured guidance on the self-regulated shadowing process (R0): This overarching requirement encompasses all other requirements. We manifest R0 in our design by dividing the SRS process into two simple steps: *Listen* (L) and *Shadow* (S), and structuring each step around two *interleaved modes*: *Practice* (P) and *Reflect* (R). Playing the target audio triggers P mode, whereas *pausing* triggers R, because we see pauses during practice as opportune moments for self-reflection and forethought. In L, the learner familiarizes themselves to the passage during P, and reviews difficult parts that need extra focus during shadowing in R. In S, they practice shadowing during P and self-evaluates their listening ability during R. CAST provides detailed guidance on what to do based on the current step and mode combination (see Figure 4.2).

Maximize self-reflection on mistakes without inducing text-dependency (R1): This requirement is derived from F3 and F4, and seeks to resolve the conflict between the learner's problematic inclination to depend too much on reading instead of listening when given access to the transcript, and their ability to use it to use it as a reflection device when checking misheard or mishadowed portions of the passage. We reframe this issue in SRL terms as a problem of *self-control*, i.e., learners are tempted to read from the passage even in instances where doing so inhibits their listening practice, and as such, they can benefit from more constrained access to the text. Therefore, we considered the *contexts* in which checking the text is beneficial, and tackled the text-dependency

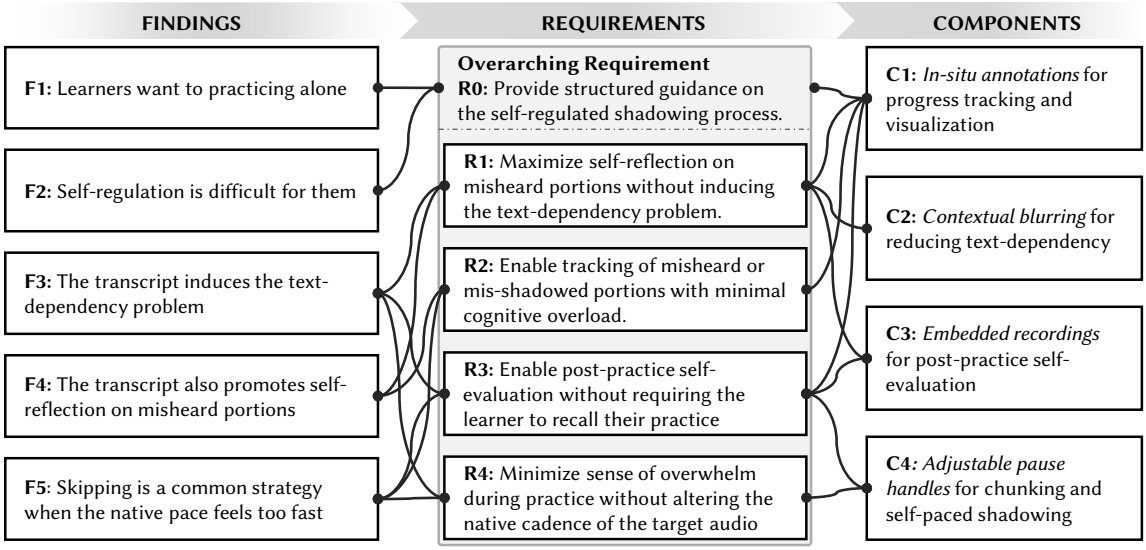


Figure 4.1: Connecting *findings*, *requirements* and *design components* in CAST

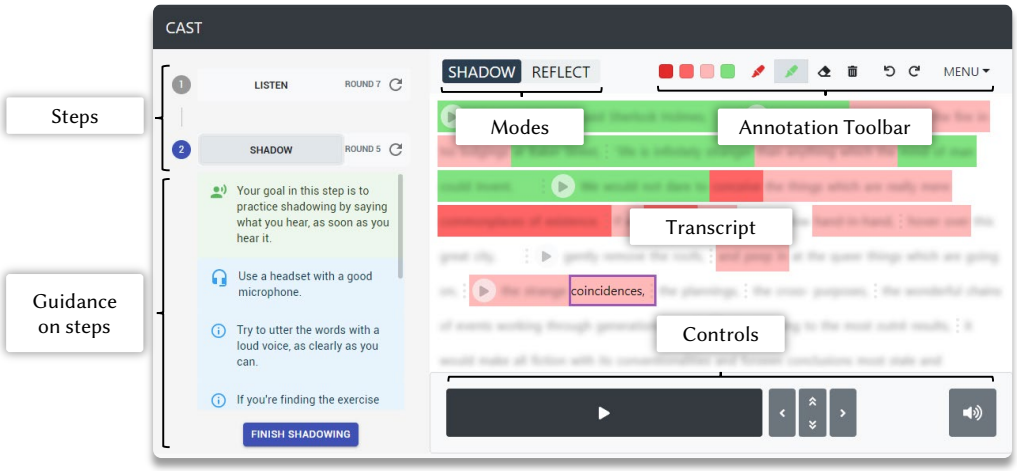


Figure 4.2: Computer-Assisted Shadowing Trainer (CAST)

problem in our design by introducing *contextual blurring* (Section 4.2), which strikes a careful balance between *revealing* parts of the text in contexts where the transcript helps with self-reflection, and keeping it *blurred* when it induces text-dependency.

Enable tracking of misheard or mishadowed words with minimal cognitive overload (R2): This requirement is derived from F4 and F5, and seeks to improve SRS through *self-monitoring*, i.e., monitoring progress over time by tracking easy and difficult portions of the passage. In F4, we found that there's a small timing window of opportunity where learners can use the transcript to notice difficult portions, i.e., misheard or mishadowed words. In F5, we found that learners

skips difficult portions, but do not always remember to return to them afterwards. If we enable learners to keep track of these portions in an external source with reasonable accuracy and minimal effort, learners become better at self-regulating because they can return to those portions during later rounds of practice even when they skip them. We support self-monitoring in CAST by introducing *in-situ annotations* (Section 4.1), which work together with contextual blurring to help learners form a *visual map* of easy and difficult portions of the target audio by using markers over the blurred transcript.

Enable post-practice self-evaluation by removing the need for practice recall (R3): In-situ progress tracking during shadowing is not feasible because the activity requires high cognitive load [23], and introducing *additional* things to do during shadowing is too overwhelming for learners. Furthermore, relying on memory for tracking is unreliable because learners do not always remember the specifics of how their practice went once they are done (F3, F5). Therefore, for the self-monitoring process to work well, learners require an easy way to self-evaluate their shadowing performance afterwards, *without* having to rely on memory or tracking during shadowing. We make this possible by combining in-situ annotations with *embedded recordings* (Section 4.3) that enable learners to self-evaluate their shadowing performance by listening to practice recordings, and comparing them with the text.

Minimize overwhelm during practice without altering target pace (R4): This requirement seeks to address the pacing issue highlighted in F5, and can be framed as a *self-pacing* problem, i.e. enabling learners to match the target audio by pacing themselves in a manner that reduces overwhelm. When learners find the narrator's pace to be too fast, they feel more inclined to depend on the text (F3). If they prevent themselves from looking at the text, they skip parts because they feel rushed (F5). Slowing things down (e.g. reducing audio speed to 0.5x) is counterproductive because shadowing improves listening by "forcing the shadower to keep pace with the audio" [6]. Therefore, we tackle the pacing issue *without* altering the target pace by introducing adjustable *pause-handles* (Section 4.4) that learners can use to break the passage into chunks.

In the next four sections, we describe in detail how each component works. We took a holistic approach to incorporating the requirements into our design, i.e., we wanted each design component to work well together as a whole, and favoured ideas that supported multiple requirements at once. As such, we describe each design idea and talk about their connection with the requirements rather than dividing them by requirement.

4.1 Self-Monitoring Using In-Situ Annotations

We solved the problem of progress monitoring by introducing light-weight in-situ annotations that can be used to visually map all the hard and easy parts in the passage using a four-tone, two-color palette (See Figure 4.3). These markers are an explicit representation of the mental mapping that learners do during practice, and removes the need for them to remember all the easy and difficult

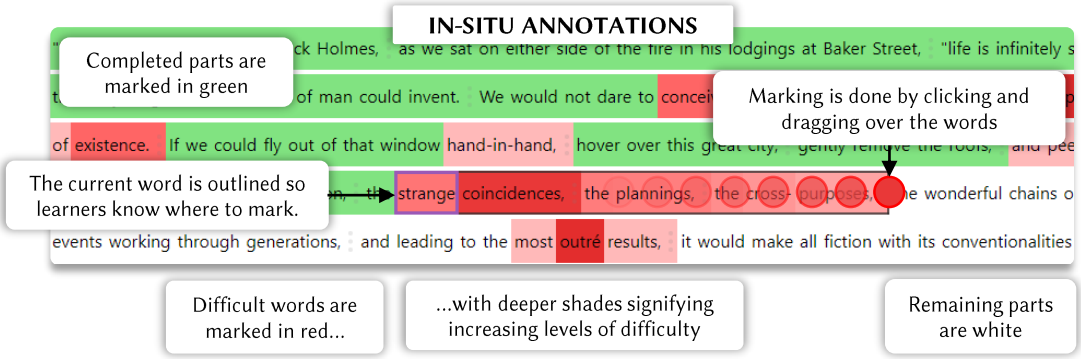


Figure 4.3: *In-situ Annotations* improve the ability of learners to self-monitor misheard portions of the target audio, and to visualize their shadowing progress using the transcript.

parts as they continue practicing. Marking is done by clicking and dragging over words. The hard parts are marked in three shades of red, with deeper shades signifying increasing levels of difficulty. Marking over the same word more than once makes it darker. Marking a difficult portion temporarily pauses the audio so that the learner can take a moment to read that part. The audio is resumed when the learner is done marking. Parts that the learner has mastered are marked in green. The red/green markers serve as a guide for which areas to focus on during practice. Using the green marker to go over a red part reduces its shade instead of completely erasing it. This allows the learner to revisit especially difficult parts multiple times.

We can draw insights from cognitive theory on multimedia learning to explain why our approach works. First, we note *dual channel assumption* in multimedia learning, which dictates that learners "possess separate channels for processing visual and auditory information" [30]. Therefore, learners are able to visualize progress by glancing at annotations over the blurred transcript, even when they are listening to the audio. Second, the pre-training principle states that students learn better when they familiarize themselves with key characteristics of the material beforehand. Going through the annotation process in the listening step helps learners with such pre-training. Third, the active processing principle states that students learn better when they actively organize the information they are given by attending to relevant portions of the passage. This is where in-situ annotations help. As the learner listens and annotates, they build a glanceable map of areas to focus that they can use and update throughout the SRS process. While the annotation process makes progress tracking easy, it does not fulfill R1 on its own because the text-dependency problem (F3) remains. To resolve this issue, we turn to contextual blurring.

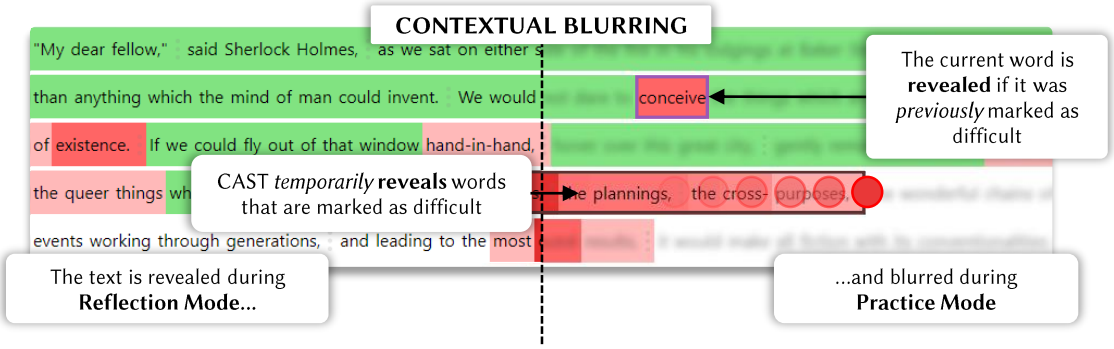


Figure 4.4: *Contextual blurring* solves the text-dependency problem by revealing parts of the transcript *only* when seeing those parts induces self-reflection on misheard words.

4.2 Self-Control using Contextual Blurring

To control attention between reading and listening, we blur the text whenever we want the learner to shift attention from reading to listening (see Figure 4.4). Smoothly transitioning between blurred and revealed text is an intuitive way for learners to know when to focus on reading, and when to focus on listening. During practice, when the learner is either listening to the audio or shadowing, we keep the text blurred. Whenever the learner pauses the audio, the text is revealed so that they can take a moment to reflect on how well they are able to listen or shadow by matching with the text. Keeping the text blurred instead of completely hiding it, and marking the current word with a box enables the learner to keep track of where they are in the passage (as indicated by a moving purple square), and to make markings even when the text is blurred.

To make quick textual references for difficult parts, we selectively reveal the text in two additional instances: (i) First, whenever the learner marks a portion as difficult, CAST temporarily reveals only that part of the text so that the learners can read it and reflect on why that part may have been difficult for them. The text progressively returns back to being blurred, to prevent learners from fixating their on that word for too long. (ii) Second, over successive rounds of listening or shadowing, whenever the audio reaches a difficult word learners get a timeline glimpse of the word. Unmarked portions, and portions marked as mastered remain blurred, so the learner gets support from the text only when necessary.

We use the signalling principle from multimedia learning theory to explain why contextual blurring works. This principle posits that multimedia-based learning is most effective when cues guide the learner's attention toward "relevant elements of the material" (in our case, the hard parts of the passage), and "highlight the organization of the material" [49] (in our case, a mental map of the hard and easy parts of the passage).

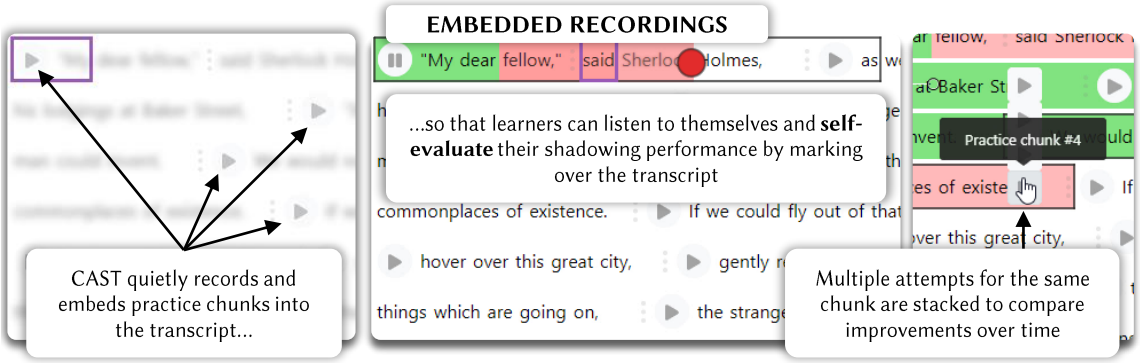


Figure 4.5: *Embedded recordings* allow for post-practice self-evaluation of listening ability through comparisons between shadowing performance and the text.

4.3 Self-Evaluation using Embedded Recordings

Making annotations over blurred text works well during listening, but not shadowing, because the latter requires higher cognitive load [23]. Early on in our prototyping phase, we tested whether learners could make markings while shadowing, but this did not work well because shadowing requires significant focus, and creating annotations during shadowing is too demanding for the learner. Therefore, any form of marking activity was viable only *after* the learner stopped shadowing, i.e., during post-practice reflection. However, annotating during post-practice reflection proved to be difficult for learners because this required them to remember clearly which parts of the passage felt easy and which parts needed more practice once they were done shadowing. So we considered creating recordings of their shadowing practice so that learners were able to listen to themselves whenever they paused to reflect, and update their annotations accordingly. This too, posed several design challenges that needed to be addressed. First, in a typical shadowing session, especially if the passage is long, it is sensible for the learner to do the shadowing in chunks (e.g. one paragraph at a time). For each chunk, the learner must create a separate recording. Second, for difficult parts, learners have to practice the same portion more than once until they get it right. This produces a large number of recordings with multiple versions that the learner has to manage and review adding to the overall workload of a task that was already demanding to begin with. Furthermore, when comparing themselves with the target audio, learners have the tendency to focus on speech attributes such as accent and style that are tangential to the overarching learning goal of listening improvement. This issue is worsened by the nature of shadowing practice, which requires learners to mimic a native speaker. Finally, because audio is linear, doing a direct comparison between two audio streams (the learners recorded audio, and the target audio) is too time consuming and tedious for self-evaluation purposes.

We solved all of these issues by embedding the practice recordings into the transcript in a way

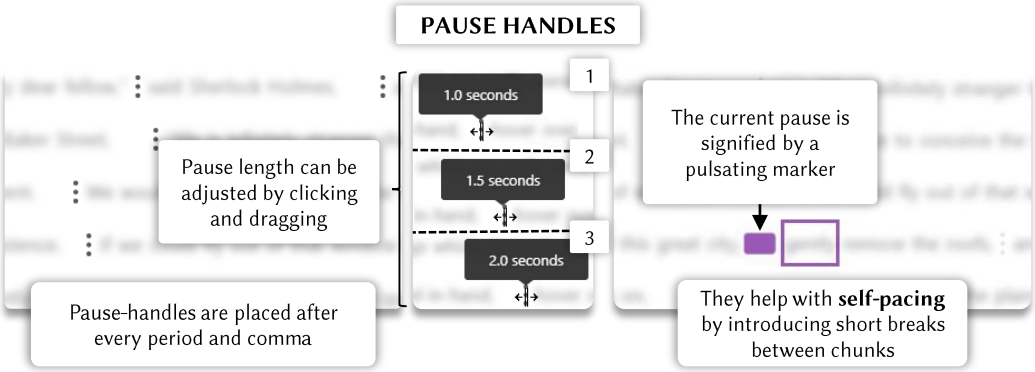


Figure 4.6: Adjustable *pause-handles* allow for self-paced shadowing without altering the native speed of the target audio.

that made comparisons with the transcript easy (see Figure 4.5). In CAST, these recordings are represented using small play buttons placed where the learner began shadowing. Hovering over a play button outlines the start and end point for that recorded chunk. Multiple recordings for chunks starting at the same point are grouped into a single play button, and revealed on hover to minimize visual clutter. The learner can revisit their recordings whenever they pause to reflect, and evaluate themselves by matching their recordings against the text, and then marking the parts where they did well, and the parts where they faced difficulty.

It may seem odd at first, that the self-evaluation process involves comparing recordings with the text and not the target audio. However, with further considerations, we argue that limiting the comparison to the text offer several benefits: First, If bottom-up listening improvement (i.e., being able to recognize words from connected speech) is the learning goal, during evaluation, the key focus should be on checking if the practice recordings indicate word recognition. This can be done more easily by matching the speech samples with the text, than by matching the speech samples with the target audio. This is because the latter process requires the learner to evaluate their listening skills *using their listening skills*, which is an oxymoron. Whereas in the former, they can make use of a different skill, i.e., reading, for self-evaluation. Second, when comparing with the audio, learners may have the tendency to focus on speech attributes such as accent and style that are tangential to bottom-up listening improvement.

4.4 Self-Paced Practice using Pause Handles

We solve the pacing issue (i.e., learners feeling overwhelmed when they are unable to keep up with the narrator's speed) by using adjustable pause handles after periods and commas in a passage (see Figure 4.6). Pausing only after punctuation marks, as opposed to a middle of a word, preserves the natural cadence of the narrator. Learners can adjust the pause length to introduce short breaks during

shadowing practice by clicking and dragging pause-handles. CAST divides embedded recordings into chunks based on these pause handles. The number of spaces between chunks corresponds to the pause duration, so that learners can *see* the size of a pause by glancing at the blurred transcript. We tested different pause lengths during prototyping, and arrived at a duration of 0 to 3 seconds with 0.5s increments based on user feedback during the piloting phase. Longer pause lengths gave the impression that the system stopped working or something broke. We use a pulsating purple box that symbolizes breathing, to signify the location of the current pause, and an earcon to inform the learner that they have reached a pause even when they have their eyes closed. Using chunking instead of altering the speed of the audio (e.g. making it 0.5x or 2x) preserves the natural cadence of the narrator, and enables the learner to practice listening to the target audio at the actual rate without feeling overwhelmed.

Chapter 5

Evaluating CAST

We wanted to test whether the inclusion of our design components positively impacted the learner's ability to self-regulate their shadowing practice using the transcript while retaining a strong focus on listening.

5.1 Method

To evaluate CAST, we conducted a summative evaluation using a baseline interface as reference. This baseline included features that are typically found in media players and document viewer, i.e., play/pause button, volume control, slider for audio navigation, and the ability to view the transcript. We removed tangential interface differences between the study conditions that could potentially confound our results, by maintaining the same overall visual layout of the common UI elements in baseline and CAST (i.e., the placement and dimension of buttons and text, font size, and color).

At the level of features, we sought answers to four specific questions: Did we really solve the text-dependency problem using contextual blurring (Section 4.2)? Do in-situ annotations (Section 4.1) induce self-reflection on misheard portions? Do embedded recordings (Section 4.1) make post-practice evaluation both possible and effective? Do adjustable pause handles (4.4) enable learners to match the narration pace?

5.1.1 Participants

The inclusion criteria for participants was the same as our need-finding study. We recruited a new batch of 12 ESL students aged 18 to 34 (7 women, 5 men) through purposive sampling, to ensure that they spoke a variety of first-languages (Arabic, Bengali, Hindi, Spanish, Turkish, Mandarin, and Cantonese). We did not include participants from our previous study to ensure that we do not bias our results.

5.1.2 Tasks

Each participant finished two shadowing tasks, one with baseline and the other with CAST. The features of CAST work in tandem. For example, the embedded recordings work on top of in-situ annotations because learners must annotate as they evaluate their recordings. Furthermore, the annotations they make during the evaluation, in turn, is taken into account in the contextual blurring feature. Therefore, we designed our tasks to give participants a sense of how the two interfaces work as a whole, instead of presenting each feature in isolation. The order of the interfaces were fully counterbalanced.

5.1.3 Materials

To provide an ecologically valid shadowing experience, we used four real-world articles as shadowing material. These articles were on two different topics (Science and Movies), and consisted of both male and female narrators to minimize the chance of domain interest and gender of the voice affecting shadowing performance. The order of the topic and gender of the article was counterbalanced. We chose these articles based on four criteria: *relevance* — (we chose news article that learners are likely to come across in real life, *neutrality* — we excluded passages that could potentially evoke strong emotions (negative/positive) from the listener, *word variety* — we chose passages with sufficient word variety so that learners were more likely to come across parts that they needed to practice, and *unfamiliarity* — we chose passages that the participants did not know in advance to prevent them from relying on memory during shadowing.

5.1.4 Procedure

We conducted the study remotely over a 1.5 hour, recorded video-conference call. Doing the study online helped us reach international ESL participants, and enabled us to simulate the experience of practicing alone as closely as possible. Since online presence could still influence shadowing performance, we kept our audio muted, and video disabled during tasks, intervening only when introducing a new feature, or when demonstrating how something works. Like before, participants completed a demographics survey before joining the study. First, we introduced the self-regulated shadowing process, and encouraged participants to reflect on their mistakes for *both* tasks for a fair comparison. For each task, participants completed a Likert-scale based questionnaire twice, once after finishing the listening step, and once after finishing shadowing. We used this questionnaire to gain insights on where the learners were focused during each step, and to see whether they were able to self-regulate their learning through self-monitoring, self-evaluation, self-reflection and self-pacing. We concluded the evaluation with a 10 minute semi-structured interview session. Each study took approximately 90 minutes to complete in total. For more details on the procedure, see the study protocol in Appendix B.3, the questionnaire in Appendix B.4, and the interview questions

in Appendix B.5.

5.1.5 Analysis

To identify whether a paired-comparison t-test was appropriate, we conducted the Shapiro–Wilk normality test on all of our Likert scale responses. These tests indicated that we were dealing with non-parametric data, and so we opted for a Wilcoxon signed-rank test. We were interested in testing whether CAST offers significant improvements over baseline, and so we chose a one-tailed test with $H_0 : B > C$. To minimize chances of committing type-1 error by accounting for multiplicity, we applied the Bonferroni correction ($\alpha = \frac{0.05}{9} = 0.00\overline{55}$ because there were 9 tests for each step).

5.2 Results

> *"The experience [with* CAST*] is pretty amazing, actually. With the first version [baseline], I had some difficulties with keeping track of where I am, and to find the hard parts. The text was making it very difficult to focus on the audio. But with the second version [*CAST*], it was very convenient...I especially liked the ability to track parts...and because the text was blurred, I could focus on the audio...also, the 'double-highlighting' feature, where I could mark difficult words in deeper shades of red, helped me practice those parts more than once...as I am not a native speaker, I couldn't keep up with the pace [with baseline], so dividing the passage into chunks [with pause handles] was pretty amazing."* — P10

The overall response to CAST, as exemplified by P10's comment, was largely positive, with 15 out of 18 indicators from our Likert-scale questionnaire (see Figure 5.1 & Appendix B.4) showing statistically significant improvements over baseline ($p < \frac{0.05}{9}$, $d > 1$) in terms of the learner's ability to focus on listening, and to self-regulate their shadowing practice. Refer to the figure 5.1 for a summary of results.

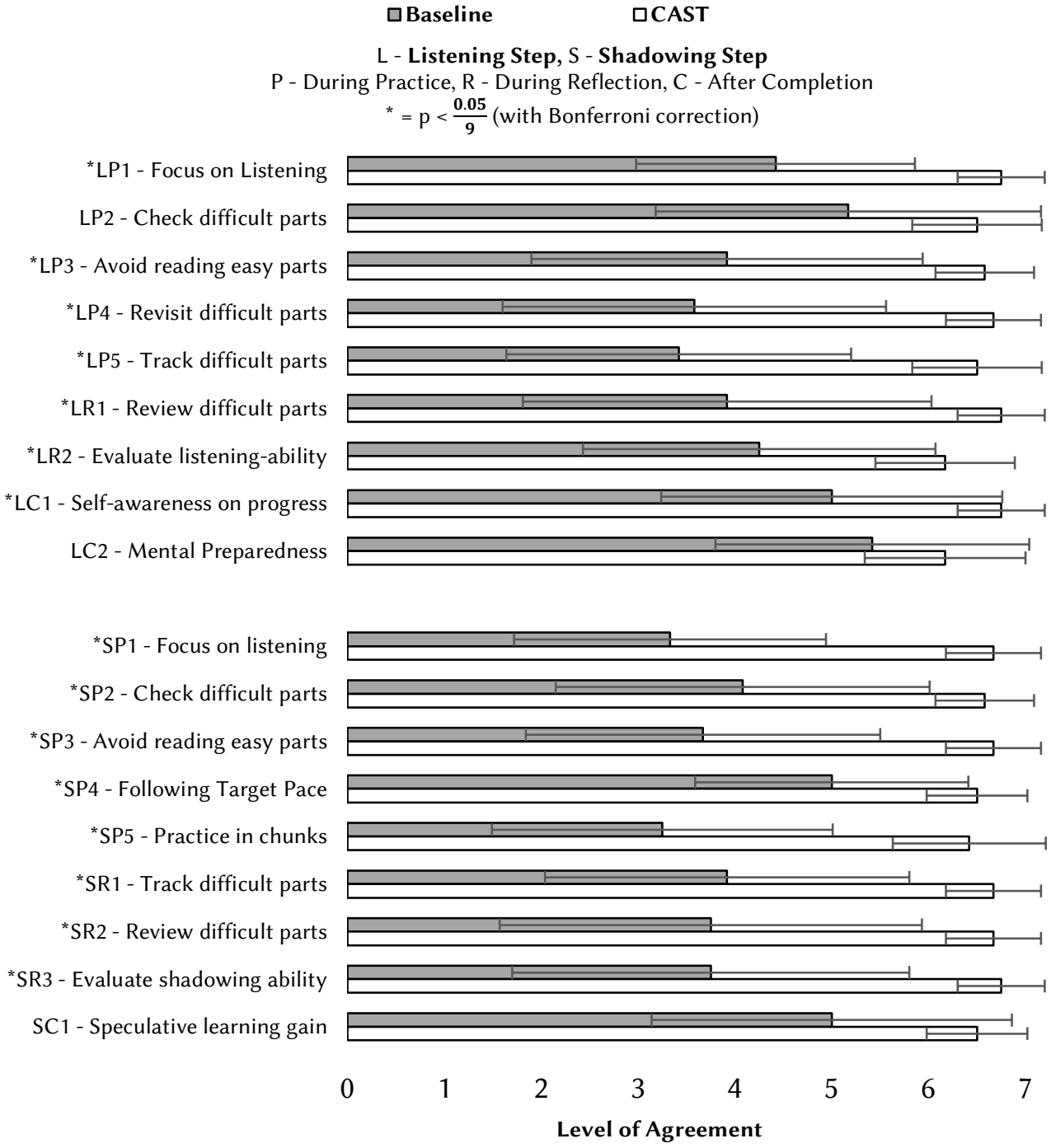


Figure 5.1: Self-regulated shadowing for listening practice is more effective with CAST

In the following sections, L and S refers to the Listening and the Shadowing step, whereas P and R refers to the Practice and Reflect mode. For example, LP1 refers to the first question about the learner's experience during listening practice, whereas SR2 refers to the second question about shadowing the learner's experience during the reflection stage in the shadowing step.

CAST improves the learner's ability to focus on listening (LP1, SP1): We note a significant improvement in the learner's ability to focus on the audio during both listening (LP1: $p < 0.001$, $d = 1.896$) and shadowing (SP1: $p < 0.001$, $d = 1.996$). This is because contextual blurring in CAST was very well received, and learners appreciated the ability to use the transcript without feeling distracted by the text, which was a recurring issue with baseline.

> *"I think the first one is much better because I can focus on listening more than reading. In the second one, I feel like I am reading the text but I am not hearing what the speaker is saying"* — P14

> *"...when the text becomes blurred you're not distracted with the other words"* — P9

While checking difficult parts during listening was easy with both versions, CAST improved the ability of learners to check difficult parts during shadowing, and enabled them to avoid unintentional glances at surrounding text (LP2, LP3, SP2, SP3): Since in baseline, the transcript is always visible, and the listening step does not require too much effort, checking difficult parts while listening was doable with both versions (LP2: $p = 0.044$, $d = 0.542$).

However, without a moving word marker and contextual blurring, participants had to rely on skimming to find difficult parts with baseline. This was too costly during shadowing because participants did not have enough cognitive resources to spare. In CAST, such skimming is not necessary, and hence we see a notable improvement in the learner's ability to check difficult parts while shadowing (SP2: $p = 0.001$, $d = 1.161$).

Furthermore, without contextual blurring, checking difficult parts forced participants to make unintentional glances at surrounding portions of the passage, even when they wanted to avoid reading those parts, and to focus on listening. Once again, CAST resolved this issue with contextual blurring (LP3, SP3: $p < 0.001$, $d > 1$)

CAST makes it easy to track, review, and read difficult parts (LP4, LP5, LR1, SR1, SR2): In-situ annotations, made tracking difficult parts during listening practice (LP5: $p < 0.001$, d = 1.526) and shadowing reflection (SR1: $p < 0.001$, d = 1.287) very effective. With all the difficult parts highlighted over the blurred transcript, participants could easily use the moving word marker and transcript-driven audio navigation features to revisit (LP4: $p < 0.001$, $d = 1.38$), review (LR1: $p < 0.001$, $d = 1.259$), and redo (SR2: $p < 0.001$, $d = 1.26$) those parts till they mastered them.

CAST enables and enhances post-practice self-evaluation (LR2, SR3): When learners pause to reflect on their listening and shadowing ability, having a visual map of areas of focus significantly improves their ability to evaluate how well they were able to listen (LR2: $p < 0.001$, $d = 1.108$).

Comments from our participants confirm that they cannot easily remember how well they were able to shadow, nor can they do in-situ annotations during shadowing.

> *"You can't remember what you spoke...that's why [self-evaluating after shadowing with baseline] wasn't good."* — P5

> *"It is too much to mark and shadow at the same time."* — P13

Therefore, our evaluation results confirm that combining in-situ annotations with embedded recordings makes post-practice self-evaluation possible and effective (SR3: $p < 0.001$, $d = 1.467$).

Pause handles make chunking significantly easier, and in turn, enables learners to match the target pace and to stop fixating on hard words (SP5, SP4): By adjusting the pause handles, learners found it significantly easier to break down the passage into meaningful chunks with CAST (SP5: $p < 0.001$, $d = 1.627$). One of the reasons why we designed these pause handles was to enable learners to match the target pace without altering the native speed of audio, and we can confirm that pause handles achieve this purpose. (SP4: $p = 0.001$, $d = 1.141$) In addition to matching the target pace, comments from P4 and P9 indicated that the pause handles provided an unexpected additional benefit — it stopped them from fixating on difficult words. P4 P4 explains this well: "...I think it's pretty good to pause while practicing...because sometimes, you're just thinking about whether you did something well in an earlier part, and maybe you realize 'oh, I didn't say that right' and so now you have that part in your mind, and you forget to say the later parts because you're still thinking about that part..." P9 is in alignment with P4's thinking and mentions that pauses are helpful for the same reason. While shadowing, when learners come across a difficult word, thinking too hard about their past shadowing performance can adversely impact their future performance and learning. The pause handles introduce small breaks that give learners a moment to reflect on past performance and move on.

CAST heightens self-awareness on progress but does not impact learner's confidence level before shadowing (LC1, LC2): The visual mapping process supported by in-situ annotations give learners a clear and complete idea of all the hard and easy parts of the passage ($LC1 : p < 0.002$, $d = 1.055$), thereby heighteing their self-awareness on progress.

However, quite interestingly, this did not impact how mentally prepared they felt to begin shadowing after completing the listening step (LC2: $p = 0.010$, $d = 0.777$). While we do not have data on the specifics of why mental preparedness was not impacted, we can say that knowing which areas need more work may not make learners feel better about their shadowing ability, but it certainly does offer them more clarity and self-awareness on their ability to listen.

Learning gain remains an open question (SC1): For the given duration of practice (approximately 15 minutes for each task), and the single session over which participants used the two interfaces, the difference between self-reported pre and post-task learning gains was not statistically

significant. The original p value for the speculative learning gain showed only a weak trend (SC1, $p = 0.009$, d = 0.797), and there's the possibility that Bonferroni adjustment may have induced a Type II error.

Comments from participants showed the promise of longer-term learning gain with CAST over baseline:

> *"[With* CAST*], I know where I am not doing well...If I can clearly identify where I'm struggling with, I can repeat it to make sure I can do it better next time."* — P1

> *"[With baseline], it's very useful if you just want to hear a story...but it won't help me learn English or practice my listening...[*CAST*] is very good for self-study"* — P9

> *"I prefer [*CAST*] because it is a great experience to hear my voice and to learn from my mistakes"* — P2

> *"[With baseline], it is pretty hard to know what you're saying and if you're doing it right or not because you don't have the recordings"* — P4

It is also worth noting that previous shadowing studies concerned with learning gain typically span multiple sessions and involve a large number of participants (see examples of shadowing research on page 25 of [19]). Therefore, long-term learning gain with CAST remains an open question, and can form the basis of a future study of that nature.

Chapter 6

Discussion

In this section, we reflect on the generalizability of our design concepts, consider the unexpected ways in which culture can influence learner behaviour, and differentiate between tools for learning and tools for consumption.

6.1 Beyond English

We chose English because it is of interest to a very large group of language learners. However, because neither shadowing nor self-regulated learning are exclusive to English pedagogy, the overarching design concepts embodied within CAST can be generalized for the acquisition of other foreign languages.

Most of the features can be used as-is, with little to no modification. For example, the notion behind contextual blurring and deblurring is applicable as long as the language in question has a written script that is supported by the computer. The same can be said about the process of self-evaluation through comparisons between the embedded recordings and the transcript.

Some of the features require additional forethought. For example, if pause handles are to be used, we must consider what constitutes a meaningful chunk in the target language because punctuation marks such as commas and fullstops are not universal. Once this has been identified, a simple regular expression which determines the placement of the pause handles within the transcript needs to be modified.

6.2 Language and Culture

> *"I'm Chinese and for us, we don't get praised for doing it well, we just want to correct all of our mistakes." — P1*

Culture can influence design in unexpected albeit significant ways. For example, P1 from our evaluation study avoided using the green marker, and focused solely on identifying and marking

all the parts that she couldn't do well in red. When asked why, she noted that in her culture, it is commendable to focus on areas of improvement rather than areas of achievement. Her cultural lens shaped how she used the in-situ annotation features of CAST. Because language and culture are inextricably linked, the design of language learning systems such as CAST must factor in cultural influences.

6.3 Consumption and Learning

> *"[With baseline], it's very useful if you just want to hear a story or say you are on the bus...but it won't help you to learn English or practice my listening..."*—P9

We can understand why CAST helped with self-regulation but baseline didn't by considering a fundamental difference between their designs - both systems can play audio and show a transcript, but the former is a tool for learning whereas the latter is *primarily* a tool for consumption. The role of the content consumer is inherently passive, whereas the role of the learner is inherently active. When learners use a media player and document viewer for shadowing, it is easy for them to consume the audio and the text, but it is not easy for them to engage with the content in a manner that makes them reflect on their consumption. From this perspective, we can view the features offered by CAST as mechanisms for engaging with the material in a structured manner, as opportunities to break free from the role of the consumer and transition into the role of a learner by becoming more involved in the learning process. This is reflected in the comments from our participants on why they felt like practicing longer with CAST compared to baseline.

> *"...you are involved in the activity [with CAST], and it's interactive, and you're spending more time, and you are involved in learning..." — P9*

> *...[CAST] was more interactive...I mean going through all parts until I got all the greens...I was feeling like going over it again and again...but with the other one [i.e. baseline], there's no progress to be made..."—P11*

6.4 Beyond Language Learning

Considering how the concepts we present in CAST can be applied to other learning activities can help us generalize our design ideas. We designed CAST with listening practice in mind. However, the notion of making annotations over a blurred document can be applied to other learning activities, as long as we can clearly define the contexts in which blurring and deblurring becomes useful. Say we want to learn the Iliad by heart (a famous epic poem with 15,693 lines [27]). The learning goal between shadowing and memorization are flipped - in the former, memorization is a vice because it removes the need for learners to rely on listening to decode words, in the latter, memorization is

the goal. Therefore, instead of beginning with a blurred document, we may begin reading from an unblurred document, and highlight parts where we feel confident to blur them.

Using what we know about memory retention over time, we can figure out the contexts where revealing blurred parts can help the learner. For example, we can apply an algorithm that uses the classic Ebbinghaus forgetting curve [12] as a basis. This curve suggests that we tend to continually halve our "memory of newly learned knowledge in a matter of days or weeks" unless we "actively review the learned material" [15]. Therefore, revealing portions that require review based on that curve can help us define the contexts in contextual blurring.

Chapter 7

Conclusion

In this work, we introduced CAST, a novel shadowing-based language learning system for self-regulated listening practice. We explored the needs and challenges of learners associated with listening-focused shadowing through a formative user study with ESL students (N=15), and found that learners want to practice alone, but doing so requires a level of self-regulation that is hard for them to attain without support. We also found that the transcript could form the basis of such support because it induces self-reflection on misheard words in the target audio, but using it *as-is* poses the text-dependency problem.

In our design approach, we used the transcript as a self-reflection device and addressed the text-dependency problem through an ensemble of design solutions in the form of contextual blurring, in-situ annotations, embedded recordings, and adjustable pause handles. We validated our design through a summative evaluation study (N=12), that showed learners were successfully able to track their progress, reflect on misheard words, and self-evaluate their listening ability with CAST.

Bibliography

[1] The complete story of arthur the rat in a british accent. URL http://www.phonetics.ucla.edu/course/transcription%20exercises/peter.htm. → page 9

[2] D. Abercrombie. Elements of general phonetics (univ. press, edinburgh). 1967. → page 9

[3] I. Arawjo, D. Yoon, and F. Guimbretière. Typetalker: A speech synthesis-based multi-modal commenting system. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1970–1981, 2017. → page 7

[4] B. Arons. Speechskimmer: a system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 4(1):3–38, 1997. → page 7

[5] E. P. Baumer, V. Khovanskaya, M. Matthews, L. Reynolds, V. Schwanda Sosik, and G. Gay. Reviewing reflection: on the use of reflection in interactive system design. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 93–102, 2014. → page 5

[6] N. Bovee and J. Stewart. The utility of shadowing. In *JALT 2008 Conference Proceedings. Tokyo: JALT*, pages 888–900, 2009. → pages 5, 15

[7] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006. → page 10

[8] H. Chen, A. Ciborowska, and K. Damevski. Using automated prompts for student reflection on computer security concepts. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*, pages 506–512, 2019. → page 5

[9] E. C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979, 1953. → page 5

[10] E. K. Choe, B. Lee, H. Zhu, N. H. Riche, and D. Baur. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 173–182, 2017. → page 5

[11] V. Clarke, V. Braun, and N. Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, pages 222–248, 2015. → page 10

[12] H. Ebbinghaus. Memory: A contribution to experimental psychology. *Annals of neurosciences*, 20(4):155, 2013. → page 30

[13] Elfiz Media. Shadowing - english speaking exercise. URL https://play.google.com/store/apps/details?id=com.pinholesoftware.shadowing&hl=en. → pages 1, 6

[14] R. Fleck and G. Fitzpatrick. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, pages 216–223, 2010. → page 5

[15] S. Gay, M. Bishop, and S. Sutherland. Teaching genetics and genomics for social and lay professionals. In *Genomics and Society*, pages 147–164. Elsevier, 2016. → page 30

[16] S. Govaerts, K. Verbert, E. Duval, and A. Pardo. The student activity meter for awareness and self-reflection. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 869–884. 2012. → page 5

[17] Y. Hamada. An effective way to improve listening skills through shadowing. *The language teacher*, 36(1):3–10, 2012. → pages 1, 6

[18] Y. Hamada. The effectiveness of pre-and post-shadowing in improving listening comprehension skills. *The Language Teacher*, 38(1):3–10, 2014. → pages 2, 10

[19] Y. Hamada. *Teaching EFL Learners Shadowing for Listening: Developing learners' bottom-up skills*. Routledge, 2016. → pages 1, 6, 9, 27

[20] Y. Hamada. Shadowing: What is it? how to use it. where will it go? *RELC Journal*, 50(3): 386–393, 2019. → pages 1, 2

[21] D. Hindus, C. Schmandt, and C. Horner. Capturing, structuring, and representing ubiquitous audio. *ACM Transactions on Information Systems (TOIS)*, 11(4):376–400, 1993. → page 7

[22] K.-T. Hsieh, D.-H. Dong, and L.-Y. Wang. A preliminary study of applying shadowing technique to english intonation instruction. *Taiwan Journal of Linguistics*, 11(2):43–65, 2013. → page 6

[23] J. Hyönä, J. Tommola, and A.-M. Alaja. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48(3):598–612, 1995. → pages 12, 15, 18

[24] T. A. Khan, D. Yoon, and J. McGrenere. Designing an eyes-reduced document skimming app for situational impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020. → page 7

[25] I. Kurz. 'shadowing'exercises in interpreter training. In *Teaching translation and interpreting*, page 245. John Benjamins, 1992. → page 6

[26] S. Lambert. Shadowing. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 37(2): 263–273, 1992. → pages 1, 6

[27] R. Lattimore, L. Baskin, et al. *The Iliad of Homer*. CUP Archive, 1962. → page 29

[28] M. Martin. Reading while listening: A linear model of selective attention. *Journal of Verbal Learning and Verbal Behavior*, 16(4):453–463, 1977. → page 3

[29] R. Martinsen, C. Montgomery, and V. Willardson. The effectiveness of video-based shadowing and tracking pronunciation exercises for foreign language learners. *Foreign Language Annals*, 50(4):661–680, 2017. → page 6

[30] R. E. Mayer. *Multimedia Learning*. Cambridge University Press, 2 edition, 2009. doi:10.1017/CBO9780511811678. → page 16

[31] MIRI KIM. English shadowing: Tedict. URL https://apps.apple.com/us/app/english-shadowing-tedict/id1455961007. → pages 1, 6

[32] N. Moray. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly journal of experimental psychology*, 11(1):56–60, 1959. → page 5

[33] T. Murphey. Exploring conversational shadowing. *Language teaching research*, 5(2): 128–155, 2001. → page 6

[34] J. Northbrook. English speaking practice — how to improve your english speaking and fluency: Shadowing, 2013. URL https://www.youtube.com/watch?v=GVWFGIyNswI. → page 9

[35] C. of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press, 2001. → page 8

[36] E. Panadero. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, 8:422, 2017. → page 4

[37] Picup Inc. Shadowing - english speaking exercise. URL https://apps.apple.com/us/app/shadowing-english-speaking-exercise/id1182789540. → pages 1, 6

[38] P. R. Pintrich and M. Zeidner. *Handbook of self-regulation*. Elsevier Science & Technology, 2000. → page 4

[39] J. C. Richards and W. A. Renandya. *Methodology in language teaching: An anthology of current practice*. Cambridge university press, 2002. → page 2

[40] D. H. Schunk. Self-regulated learning: The educational legacy of paul r. pintrich. *Educational psychologist*, 40(2):85–94, 2005. → page 4

[41] D. H. Schunk and P. A. Ertmer. Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation*, pages 631–649. Elsevier, 2000. → page 4

[42] H. Shin, E.-Y. Ko, J. J. Williams, and J. Kim. Understanding the effect of in-video prompting on learners and instructors. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–12, 2018. → page 5

[43] H. Sumiyoshi. The effect of shadowing: Exploring the speed variety of model audio and sound recognition ability in the japanese as a foreign language context. *Electronic Journal of Foreign Language Teaching*, 16(1):8, 2019. → page 5

[44] J. Swain. *A hybrid approach to thematic analysis in qualitative research: Using a practical example*. SAGE Publications Ltd, 2018. → page 10

[45] K. Tamai. *The Effect of" shadowing" on Listening Comprehension*. PhD thesis, School for International Training, 1992. → pages 1, 6

[46] K. Tamai. Shadowing no koka to chokai process ni okeru ichizuke [the effectiveness of shadowing and its position in the listening process]. *Current English Studies*, 36:105–116, 1997.

[47] K. Tamai. Listening shidoho to shite no shadowing no koka ni kansuru kenkyu [research on the effect of shadowing as a listening instruction method]. *Japan: Kazama*, 2005. → page 6

[48] A. Thudt, U. Hinrichs, S. Huron, and S. Carpendale. Self-reflection and personal physicalization construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018. → page 5

[49] T. van Gog. *The Signaling (or Cueing) Principle in Multimedia Learning*, page 263–278. Cambridge Handbooks in Psychology. Cambridge University Press, 2 edition, 2014. doi:10.1017/CBO9781139547369.014. → page 17

[50] S. Whittaker, J. Hirschberg, B. Amento, L. Stark, M. Bacchiani, P. Isenhour, L. Stead, G. Zamchick, and A. Rosenberg. Scanmail: a voicemail interface that makes speech browsable, readable and searchable. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 275–282, 2002. → page 7

[51] J. J. Williams, T. Lombrozo, A. Hsu, B. Huber, and J. Kim. Revising learner misconceptions without feedback: Prompting for reflection on anomalies. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 470–474, 2016. → page 5

[52] F. Yavari and S. Shafiee. Effects of shadowing and tracking on intermediate efl learners' oral fluency. *International Journal of Instruction*, 12(1):869–884, 2019. → page 6

[53] D. Yoon. Enhancing expressivity of document-centered collaboration with multimodal annotations. 2017. → page 7

[54] D. Yoon, N. Chen, B. Randles, A. Cheatle, C. E. Löckenhoff, S. J. Jackson, A. Sellen, and F. Guimbretière. Richreview++ deployment of a collaborative multi-modal annotation system for instructor feedback and peer discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 195–205, 2016. → page 10

[55] X. Zhang, T. Miyaki, and J. Rekimoto. Withyou: Automated adaptive speech tutoring with context-dependent speech recognition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020. → pages 1, 6

[56] B. J. Zimmerman. Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation*, pages 13–39. Elsevier, 2000. → page 4

[57] B. J. Zimmerman. Becoming a self-regulated learner: An overview. *Theory into practice*, 41 (2):67, 2002. → page 2

Appendix A

Supporting Material for Study on *Exploring Learner Needs*

This section contains copies of supporting material for the study described in Chapter 3:

1. Recruitment Flyer
2. Consent Form
3. Study Protocol
4. Interview questions

A.1 Recruitment Flyer

THE UNIVERSITY OF BRITISH COLUMBIA
Computer Science
Faculty of Science

Department of Computer Science
2366 Main Mall
Vancouver, B.C. V6T 1Z4

*You will receive a **$15/hour** honorarium for this study*

CALL FOR PARTICIPATION

Designing a Learner-Centered Computer-Assisted Language Learning Tool

Do you speak English as a second language?

Have you taken **ESL lessons** within 2 years?

Are you interested in using
technology for language learning?

Participate in our study!

Who are we?

We are Human-Computer Interaction (HCI) researchers from the Department of Computer Science at UBC, *Dr. Dongwook Yoon* and *Mohi Reza*.

Who are we looking for?

To participate in this study:

- ✓ you must have taken **ESL lessons** (e.g. an English language course) **within the last two years**.
- ✓ you must be a **non-native** speaker of **English**.
- ✓ you must be **at least 18** years old
- ✓ you must not have any significant hearing loss.

Interested in participating?

If you are interested in participating or would like more information, please send an email at [redacted] or text [redacted].

How long will it take?

The study will consist of a single session of approximately **1 hour**.

Call for participation (Version 1.1 / August 28, 2019)

A.2 Consent Form

Consent Form
[Student Participants]

Designing a Learner-Centred Computer-Assisted Shadowing Trainer (CAST)

Principal Investigator

Dongwook Yoon, Computer Science Department, University of British Columbia.

Phone: [redacted]; Email: [redacted]

Co-Investigator

Mohi Reza, Computer Science Department, University of British Columbia.

Phone: [redacted]; Email: [redacted]

Introduction

Thank you for participating in this study! The purpose of our research is to inform the design of a learner-centred language learning software tool called *Computer-Assisted Shadowing Trainer (CAST)*. This tool is meant for adults who are learning English as a second language (ESL). This research is being conducted as part of a Computer Science Master's graduate degree requirement at the University of British Columbia.

What You Will Be Asked to Do

After you have read this document, please do not hesitate to ask any questions or concerns that you may have. Once you have signed this consent form, you may be asked to:

- Fill out an online survey for basic demographic data
  (i.e., gender, age, occupation, language, academic background, etc.)
- Participate in a task-based experiment
  (using common software tools such as audio-players and text readers, and/or a prototype of CAST.)
- Answer interview questions
  (related to the prototype and/or your language learning experience.)

The demographics survey should take approximately 5~10 minutes to fill out, and the tasks and interviews should take approximately one to one and a half hours. The entire study will be completed in a single session.

Please note that during the interview and tasks, **audio recording is required and that the computer screen will be recorded while during the tasks**.

Participants are free to withdraw without giving any reason. If you withdraw from the study, all data obtained from you will be permanently discarded on the date of your withdrawal and will not be used for this study. All electronic files will be permanently deleted and all of the physical copy of documents (e.g., handwritten notes, paper transcriptions and the consent form) obtained from you will be shredded. Your name will be also removed from the code assignment file.

Project Outcomes

Although the project outcomes will be determined by the research findings, possible research products will include journal articles, reports, software prototypes and plain language summaries.

Potential Benefits

There are no explicit benefits to you by taking part in this study. However, the purpose of the study is to improve research, and as an extension of the experience of participating in research.

Potential Risks

This study is expected to take one to one and a half hours of your time, and you may feel tired during the session. If you need any break at any point of the time, you can always ask the researcher to have a 5 to 15-minute break. You can also withdraw your participation at any time.

Confidentiality

The demographic survey and questionnaires will be conducted through a UBC survey tool called Qualtrics. All hard copy documents will only be identified by an assigned code. Any electronic file names will not contain any identifiable data such as names. You will not be identified by name in the survey data or interview transcript. The link between the code associated and the actual names will be stored in a master code file under lock and key. The only documents containing your real name will be the master code file and this consent form.

The master code file will be stored on an encrypted hard drive of a password-protected laptop. During the study, any handwritten notes and paper transcripts will be kept in a locked cabinet with controlled access at UBC.

Open Access: In the future, we may be required to make the data we collect publicly available at the time of publication. Please note that once the data is made publicly available, participants will not be able to withdraw their data. In any published material (i.e., any reports, research papers, thesis documents, and presentations), participants will be named only by assigned code to preserve anonymity. There will be no identifiable data published, and transcription of the audio files will be modified to remove any personally identifiable information.

Remuneration/Compensation

In order to acknowledge the time you have taken to be involved in this project, each participant will receive $15 dollar per hour.

Contact for Information About the Study

If you have any questions or desire further information with respect to this study, you may contact Mohi Reza ([redacted] email: [redacted]).

Contact for Concerns or Complaints About the Study

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Consent

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time. Your signature below indicates that you have received a copy of this consent form for your own records. Your signature indicates that you consent to participate in this study.

Please Tick <u>One</u> of the Following

☐ I consent to be audio recorded in this study.
☐ I do not consent to be audio recorded in this study.

Please Tick <u>One</u> of the Following

☐ I consent to the computer screen being recorded during the study.
☐ I do not consent to the computer screen being recorded during the study.

I, _____, have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in this study. However, I realize that my participation is voluntary and that I am free to withdraw at any time.

_____
Participant Signature and Date

A.3 Study Protocol

Pre-Study Checklist

- ✓ Clean and sanitize headphones for the next participant.
  Sanitary Disposable Earpad Covers
- ✓ Ensure that the laptop is charged or connected to the power cable.
- ✓ Create a new folder to store recordings titled PX where X is the participant number.
- ✓ Test out screencasting software (Camtasia), media player (VLC[3]).
- ✓ Go through demographics survey responses for interview prep.

Introduction [~2 minutes]

Hello, my name is _____ and I am working with _____.
Thank you for choosing to participate in our study.

Please fill out the consent form before we begin. This study will take approximately 1 hour, and you will be compensated $15 at the end of it.

- ✓ Have participants sign the consent form.
- ✓ Complete demographics questionnaire, in case they haven't done it in advance.

Task Briefing [~7 minutes]

Shadowing is a language learning technique where learners listen to speech recordings from native speakers and *simultaneously* utter what they hear as accurately as possible.

The purpose of this study is to learn more about how language learners like you can benefit from the technique. My goal is to understand your needs and the challenges during shadowing practice. In this study, you will be asked to complete two shadowing tasks. For each task, you will be provided instructions on shadowing, and one of two halves of a short passage, *Arthur the Rat*, as shadowing material.

Please let me know at any time during the study if you ever need to take a break, or are feeling uncomfortable in any way. There will be a scheduled 3-minute break at the end of the first task.

- ✓ Demonstrate how the software tools (VLC media player and Notepad) work.
  - Show them how to use the play/pause/rewind/forward and text-resizing features using both the mouse and the keyboard.
  - Let them use whichever input mode they prefer.
- ✓ Make sure they are comfortable with using the tools before moving on.

[3] VLC is a good choice for the "typical" audio player interface because it's widely used and cross-platform.

First Task - Half of *Arthur and the Rat*: [~15 minutes]

Introduction

For the first shadowing exercise, imagine that you came across a video on shadowing, and decided to give the technique a try. I will provide you with such a video.

Video instructions [6.5 minutes]

- ✓ Open the introductory video on shadowing and ask the participant to watch it.

Shadowing Exercise [~8 minutes]

Now that you've watched the video, I want you to practice some shadowing. At this stage, I will not answer any questions, as I want you to practice based solely on whatever you have learned from the video.

- ✓ Ask the participant to begin the exercise.
- ✓ Remember to record an audio screencast of the learner while they practice using Camtasia.
- ✓ Closely observe the participant and take notes.
    - ○ Keep an eye on any for the challenges and struggles they face with regards to both shadowing itself and in using the audio player/document viewer interface.

After the participant finishes their first task, offer them an optional break.

- ✓ Optional 3-minute break.

Second Task - The Other Half of *Arthur and the Rat*: [~15 minutes]

Introduction

For this second task, imagine that you heard about shadowing from an instructor who assigned you a homework exercise. I will first give you in-person instructions about Shadowing, and then, like before, you will do some shadowing.

- ✓ Answer any questions on shadowing that the participants may have.
- ✓ Provide in-person step-by-step guidance .

In-person instructions [~6.5 minutes]

Do a step-by-step[4] walkthrough of the shadowing process with the participant, simulating a scenario where you are instructing them on the technique, and then assigning a shadowing task as homework.

- *Step 1 (Warm-up, ~3 mins):* First, listen closely to the passage to get yourself ready for shadowing. You can start shadowing what you hear in your mind without vocalizing the words.
- *Step 2 (Mumbling, ~6 mins):* Second, replay the passage and shadow with a small voice as if you were mumbling. It is okay for you to not be able to shadow perfectly. Do this twice.
- *Step 3 (Parallel reading, ~9 mins):* Third, practice shadowing using the transcript. Focus on sounds and not meaning. Do this thrice. Each time, identify portions of the passage that you find challenging and underline them with a pencil. Focus on them on your next attempt.

✓ Ask the participant to close the transcript before beginning step 4.

- *Step 4 (Regular Shadowing, ~9 mins):* Now, practice shadowing without the transcript. Focus on sounds and not meaning. Do this thrice. Each time, identify and keep in mind the portions of the passage that you find challenging. Focus on them on your next attempt.

✓ Tell the participant why they should practice shadowing multiple times - the practice might feel redundant to you but researchers recommend shadowing thrice for beginners and twice for intermediate and advanced learners. If you're struggling with your first attempt, you might find the exercise easier on your next attempts

Shadowing Exercise [~8 minutes]

Now that I've given you instructions, I will assign you a shadowing exercise. Feel free to reach out to me if you face any difficulties during practice.

✓ Ask the participant to begin the second exercise. ✓ Remember to record an audio screencast of the learner while they practice using Camtasia. ✓ Closely observe the participant and take notes.

[4] *The steps I used were adapted from Chapter 2, page 21 of Yo Hamada's wonderful textbook on Shadowing - "Teaching EFL Learners Shadowing for Listening". The optional comprehension questions from step 1 and 8 have been omitted due to time considerations. Since step 7 does not focus on phoneme perception and is there to help with 8, that has been omitted as well. What remains is a good exemplar of what instructors may use to teach about shadowing, and is, therefore, suitable for this study.*

- Keep an eye on any for the challenges and struggles they face with regards to both shadowing itself and in using the audio player/document viewer interface.

Follow-up Interview [20 minutes]

Now that you've completed both tasks, let's talk about your experience with shadowing.

- ✓ Double-check that the audio recording is on.
- ✓ Do the follow-up semi-structured interview.

Study Wrap-up [<1 minutes]

That concludes our study. Thank you for your valuable input and time! Do you have any further questions or comments? Otherwise, here is compensation for your participation in this study. Please also sign this summary sheet indicating that you have received the money.

- ✓ Have participants sign the compensation sheet.
- ✓ Give them the compensation.
- ✓ Stop and save the screencast and audio recordings.
- ✓ Store the final created recording.

Post-study Checklist

- ✓ Stop and save recording.
- ✓ Double-check to make sure the recordings have been properly saved.
- ✓ Prepare for the next participant.

A.4 Interview Questions

1. Overall, what did you think of shadowing?
2. Think of the tools that you used, i.e., the audio player and text-viewer.
   (a) What was *easy* and what was *difficult* when shadowing with these tools?
   (b) What did you *like* or *dislike* about them?
   (c) What worked *well* and what *did not work well*?
3. During the two shadowing exercises, think about moments where you felt frustrated or confused, and moments where you felt satisfied or effective. Can you give me some examples of those moments?
4. These questions are about the four steps that you tried in the second shadowing task, i.e., warm-up, mumbling, parallel reading and shadowing.
   (a) What are some specific things that made each of these steps *easy* or *difficult* for you?
   (b) Tell me more about the challenges you faced during each step.
   (c) Please rank them in order, from *most* to *least* challenging.
   (d) Could you help me understand your rankings?
5. Let's talk about your English language learning experience.
   (a) How did you learn English?
   (b) How long have you have you been learning?
   (c) Do you use any specific strategies to practice listening and speaking? What are they?
   (d) If you have experienced shadowing before, please describe your experience.
6. Is there anything else that you would like to share about your experience with shadowing?

Appendix B

Supporting Material for Study on *Evaluating CAST*

This section contains copies of supporting material for the study described in Chapter 5:

1. Recruitment Flyer
2. Consent Form
3. Study Protocol
4. Likert Questionnaire
5. Interview questions

B.1 Recruitment Flyer

THE UNIVERSITY OF BRITISH COLUMBIA
Computer Science
Faculty of Science

Department of Computer Science
2366 Main Mall
Vancouver, B.C. V6T 1Z4

*You will receive a **$15/hour** honorarium for this study*

CALL FOR PARTICIPATION

*Evaluating **CAST** - Computer-Assisted Shadowing Trainer*

Do you speak English as a second language?

Have you taken **ESL lessons** within 2 years?

Are you interested in using
technology for language learning?

Participate in our online study!

The study will be conducted remotely, over Zoom.

Who are we?

We are Human-Computer Interaction (HCI) researchers from the Department of Computer Science at UBC, *Dr. Dongwook Yoon* and *Mohi Reza*.

Who are we looking for?

To participate in this study:

- ✓ you must have taken **ESL lessons** (e.g. an English language course) **within the last two years**.
- ✓ you must speak **English as a foreign or second language**.
- ✓ you must be **at least 18** years old
- ✓ you must not have any significant hearing loss.

Interested in participating?

If you are interested in participating or would like more information, please send an email at [redacted] or text [redacted].

How long will it take?

The study will consist of a single session of approximately **1.5 hours.**

Call for participation (Version 1.2 / June 22, 2020)

B.2 Consent Form

Consent Form

Designing a Learner-Centred
Computer-Assisted Shadowing Trainer (CAST)

Principal Investigator

Dongwook Yoon, Computer Science Department, University of British Columbia.

Phone: [redacted]; Email: [redacted]

Co-Investigator

Mohi Reza, Computer Science Department, University of British Columbia.

Phone: [redacted]; Email: [redacted]

Introduction

Thank you for participating in this study! The purpose of our research is to inform the design of a learner-centred language learning software tool called *Computer-Assisted Shadowing Trainer (CAST)*. This tool is meant for adults who are learning English as a second language (ESL). This research is being conducted as part of a Computer Science Master's graduate degree requirement at the University of British Columbia.

What You Will Be Asked to Do

After you have read this document, please do not hesitate to ask any questions or concerns that you may have. Once you have signed this consent form, you may be asked to:

- Fill out an online survey for basic demographic data
  (i.e., gender, age, occupation, language, academic background, etc.)
- Participate in a task-based experiment
  (using common software tools such as audio-players and text readers, and/or a prototype of CAST.)
- Answer interview questions
  (related to the prototype and/or your language learning experience.)
- Fill out post-study questionnaires
  (related to the prototype and/or your language learning experience.)

The demographics survey and the post study questionnaires should take approximately 5~10 minutes to fill out, and the tasks and interviews should take approximately 1 hour 20 minutes. The entire study will be completed in a single session.

The entire study will be conducted remotely via online video-call and screen share using Zoom.

Please note that during the study, the video-call and screen will be recorded locally.

Version 1.3 / June 16 2020 Ethics ID - H19-01380 Page 1 of 3

We would like you to be aware of the following information and best practices when using Zoom:

- Zoom servers are located outside of Canada, and Zoom stores your name and information regarding your use of the site outside Canada.
- You can protect your identity and increase the protection of your personal information if you do not use your actual name in Zoom. You can do this by:
    - Using only a nickname or a substitute name or your participant ID
    - Turning off your camera
      (when not engaged in a study task or interview and you would like to do this)
    - Muting your microphone
      (when not engaged in a study task or interview and you would like to do this)

Participants are free to withdraw without giving any reason. If you withdraw from the study, all data obtained from you will be permanently discarded on the date of your withdrawal and will not be used for this study. All electronic files will be permanently deleted and all of the physical copy of documents (e.g., handwritten notes, paper transcriptions and the consent form) obtained from you will be shredded. Your name will be also removed from the code assignment file.

Project Outcomes

Although the project outcomes will be determined by the research findings, possible research products will include journal articles, reports, software prototypes and plain language summaries.

Potential Benefits

There are no explicit benefits to you by taking part in this study. However, the purpose of the study is to improve research, and as an extension of the experience of participating in research.

Potential Risks

This study is expected to take one and a half hours of your time, and you may feel tired during the session. If you need any break at any point of the time, you can always ask the researcher to have a 5 to 15-minute break. You can also withdraw your participation at any time.

Confidentiality

The demographic survey and questionnaires will be conducted through a UBC survey tool called Qualtrics. All hard copy documents will only be identified by an assigned code. Any of electronic file names will not contain any identifiable data such as names. You will not be identified by name in the survey data or interview transcript. The link between the code associated and the actual names will be stored in a master code file under lock and key. The only documents containing your real name will be the master code file and this consent form.

The master code file will be stored on an encrypted hard drive of a password-protected laptop. During the study, any handwritten notes and paper transcripts will be kept in a locked cabinet with controlled access at UBC.

Open Access: In the future, we may be required to make the data we collect publicly available at the time of publication. Please note that once the data is made publicly available, participants will not be able to withdraw their data. In any published material (i.e., any reports, research papers, thesis documents, and presentations), participants will be named only by assigned code to preserve anonymity. There will be no identifiable data published, and transcription of the audio files will be modified to remove any personally identifiable information.

Remuneration/Compensation

In order to acknowledge the time you have taken to be involved in this project, each participant will receive $15 dollar per hour.

Contact for Information About the Study

If you have any questions or desire further information with respect to this study, you may contact Mohi Reza ([redacted]; email: [redacted]).

Contact for Concerns or Complaints About the Study

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance e-mail RSIL@ors.ubc.ca or call toll free 1-877-822-8598.

Consent

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time. Your signature below indicates that you have received a copy of this consent form for your own records. Your signature indicates that you consent to participate in this study.

Please Tick <u>One</u> of the Following

☐ I consent to the video call (audio, video and screen share) being recorded in this study.
☐ I do not consent to the video call (audio, video and screen share) being recorded in this study.

I, _____, have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in this study. However, I realize that my participation is voluntary and that I am free to withdraw at any time.

_____

Participant Signature/Online Acknowledgement and Date

B.3 Study Protocol

Pre-study Checklist

- ✓ Start Zoom Meeting
- ✓ Check microphone and camera
- ✓ Ensure that the laptop is charged or connected to the power cable.
- ✓ Create a new folder PX where X is the participant number to store recordings.

Introduction [~10 minutes]

Hello, my name is _____ and I am working with professor _____ in the department of Computer Science at UBC.

Thank you for choosing to participate in our study.

I hope you had a chance to go through the consent form and the initial survey already. I'd be happy to answer any questions you may have on those. This study will take approximately 1 hour and 30 minutes to complete, and you will be compensated at a rate of $15/hour, so that's $23 dollars at the end of it.

- ✓ Enable local Zoom recording after getting participant consent

In this study, we are going to evaluate some tools for listening-focused Shadowing practice.

- ✓ Use orientation slides to ensure that your intro is consistent across participants.

Gist: *Shadowing is a language learning exercise where learners listen to a target audio and say what they hear, as soon as they hear it. The main goal of shadowing is to exercise your listening skills. As you shadow, you train your ears to recognize the words you hear in the target audio more quickly, because in order to shadow, you have to be quick enough to be able to say the words soon after hearing it.*

The goal isn't to read and repeat. The goal isn't to memorize and repeat. The goal is to listen and repeat, with as little delay as possible.

When shadowing, you will be provided with the transcript so that you may reflect on your practice by referring to the text.

- ✓ Before moving on, ask participants some basic questions about shadowing to see if they understood your introduction.

Now, I will introduce you to one version of a shadowing tool. You will use it to complete your first shadowing exercise.

- ✓ Launch CAST/Baseline and share your screen with the participant.
- ✓ Make sure that you tick the option to share computer audio.
- ✓ Demonstrate how CAST/Baseline works.
- ✓ Ask them to share their screen with you, and then to try it out themselves.
- ✓ Make sure they know how to use the system.

- ✓ Make sure that the participant has:
  - Shared their screen.
  - Enabled microphone access in the browser.
  - Loaded CAST/Baseline on a recent version of Chrome.
- ✓ Ask the participant to load the first passage based on the counterbalanced order list.

	11		12		21	
BM → CS	$BM_1 \rightarrow CS_1$	✓	$BM_1 \rightarrow CS_2$	✓	$BM_2 \rightarrow CS_1$	✓
BS → CM	$BS_1 \rightarrow CM_1$	✓	$BS_1 \rightarrow CM_2$	✓	$BS_2 \rightarrow CM_1$	✓
CM → BS	$CM_1 \rightarrow BS_1$	✓	$CM_1 \rightarrow BS_2$	✓	$CM_2 \rightarrow BS_1$	✓
CS → BM	$CS_1 \rightarrow BM_1$	✓	$CS_1 \rightarrow BM_2$	✓	$CS_2 \rightarrow BM_1$	✓

Tool B = Baseline
C = CAST

Genre M = Movie
S = Science

Passage M_1 = *Parasite*
M_2 = *Extraction*
S_1 = *Galaxies*
S_2 = *Koalas*

CAST Evaluation Study Orientation

1

Duration, Tools and Compensation

- The study will take 1.5 hours to complete
- You will need: chrome, headphones/earphones and a microphone.
- Upon completion, you will receive a 23$ honorarium as compensation.

2

Study Structure

- There are **two** main tasks. For each task, you will be using a different version of CAST (Computer-Assisted Shadowing Trainer) to do some *shadowing*.
- The study is structured as follows:
  a. Orientation
  b. Tool Demonstration
  c. First Task
  d. Second Task
  e. Interview
- Each Task consists of:
  a. Pre-task activity
  b. Listening Step ➡ Questionnaire
  c. Shadowing Step ➡ Questionnaire
  d. Post-task activity

3

Links

1. [redacted]
2. [redacted]
3. [redacted]

4

WHAT IS SHADOWING?

Shadowing is a language learning technique that can help you exercise your **listening skills**, and improve your ability to identify words in fluent, connected speech more quickly.

You can practice shadowing by listening to a target audio, and ***saying what you hear*, as *soon as you hear it***

5

SHADOWING IS NOT REPETITION

TARGET AUDIO: *THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.*

YOUR SPEECH: *THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.*

TARGET AUDIO: *THE QUICK BROWN FOX* *JUMPS OVER THE LAZY DOG.*

YOUR SPEECH: *THE QUICK BROWN FOX* *JUMPS OVER THE LAZY DOG.*

6

SHADOWING IS NOT ABOUT MEMORIZATION

Listening and repeating
by relying on ears

Repeating from memory

7

SHADOWING IS NOT READING ALOUD

Listening and repeating
by relying on ears

Reading aloud

8

THE OVERARCHING GOAL OF SHADOWING

To train your **ears** to recognize words in fluent, flowing speech, quickly, and automatically.

9

THE ROLE OF THE TRANSCRIPT
SELF REFLECTION

After practicing listening or shadowing, you can use the transcript to **reflect on** how well your practice went.

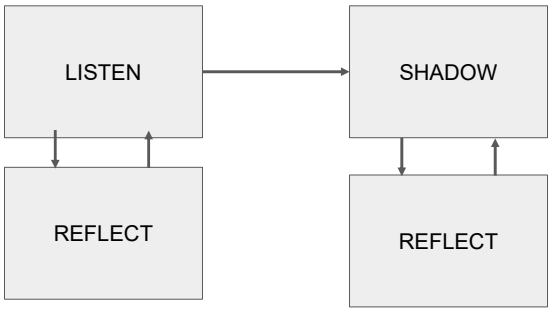
10

THE ROLE OF THE TRANSCRIPT
SELF REGULATION

As you practice, try to evaluate yourself using the transcript, and use your evaluation to decide whether or not you need to practice more (i.e, shadow more or listen more)

11

STEPS



12

[Begin first task] [15 minutes]

- ✓ Keep your camera and microphone muted while they do the task.
- ✓ Only Intervene if you need to explain how something works or if they are facing trouble .

[Ask them to fill out the post-task questionnaire] [5 minutes]

- ✓ Switch to the other version and demonstrate how it works.

[Begin second task] [15 minutes]

[Ask them to fill out the post-task questionnaire] [5 minutes]

[Conduct a short follow-up Interview] [10 minutes]

That concludes our study. Thank you for your valuable input and time! Do you have any further questions or comments?

Post-study Checklist

- ✓ End Zoom Meeting
- ✓ Send study compensation
- ✓ Save video recording of the meeting locally.
- ✓ Double check to ensure that the meeting has been successfully recorded and saved.

B.4 Likert Questionnaire

Questions on the Listening Step

Responses were collected using the following 7-point Likert Scale:

Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree

LP1 *Focus on Listening*
While listening, I could easily focus on hearing the audio rather than reading the text.

When I came across difficult parts, i.e. words or phrases that I couldn't immediately recognize by ear:

LP2 *Checking Difficult Portions*
I could easily read those difficult parts by checking the text.

LP3 *Not Reading Easy Portions*
I could easily prevent myself from also reading the easy parts when checking the text.

LP4 *Revisiting Difficult Portions*
I could easily jump back to those difficult parts as I continued listening.

LP5 *Tracking Difficult Portions*
I could easily remember or keep track of those difficult parts as I continued listening.

When pausing to reflect on my listening practice:

LR1 *Reviewing Difficult Portions*
I felt like I could easily review the difficult parts using the text.

LR2 *Evaluating Ability to Listen*
I felt like I could easily evaluate myself on how well I was able to listen.

After completing the listening step:

LC1 *Mapping Easy/Difficult Portions*
I felt like I had a clear and complete idea of all the parts that I could and couldn't identify by ear.

LC2 *Mental Preparedness*
I felt mentally prepared to begin shadowing.

Questions on the Shadowing Step

Responses were collected using the following 7-point Likert Scale:

Strongly Disagree	***Disagree***	***Somewhat Disagree***	***Neutral***	***Somewhat Agree***	***Agree***	***Strongly Agree***

While shadowing:

SP1 *Focus on Listening*
I could easily focus on hearing the audio rather than reading the text.

SP2 *Checking Difficult Portions During Practice*
I could easily check those difficult parts by reading the text.

SP3 *Not Reading Easy Portions*
I could easily prevent myself from also reading the easy parts when checking the text.

SP4 *Following Target Pace*
I could easily keep up with the narration pace (i.e. shadow at the same speed as the audio without feeling overwhelmed).

SP5 *Breaking Passage into Chunks*
I could easily divide the passage into manageable chunks and practice and reflect piece by piece rather than doing the entire exercise all at once.

When I came across difficult parts, i.e words or phrases that I couldn't say clearly, or completely missed during shadowing:

SR1 *Tracking Difficult Portions*
I could easily remember or keep track of those difficult parts as I continued shadowing.

When pausing to reflect on my shadowing practice:

SR2 *Reviewing Difficult Portions During Reflection*
I felt like I could easily review the difficult parts using the text.

SR3 *Self-Evaluation*
I felt like I could easily evaluate how well I was able to shadow.

SC1 *Speculative Listening Improvement*
After completing the shadowing step, I felt like this interface has helped me improve my ability to listen to the passage.

B.5 Interview Questions

1. Overall, what was your experience like with using the full-featured version of CAST? Was it easy to use? If not, could you describe what was hard?

2. Did the additional features feel *helpful* or *unhelpful* for your shadowing exercise? Could you describe why?

3. Let's talk about the self-regulated shadowing process that you just experienced, involving rounds of reflection and practice. Did this process make sense to you? What did you like about it, and what did you dislike?

4. Finally, let's unpack your responses to the questionnaire - help me understand your reasoning behind the ratings.

5. Is there anything else you would like to mention about your experience with the two versions that you tried?