Ubiideas: Catalysing Divergent Thinking with Heads-Up Display

Journal:	*Transactions on Visualization and Computer Graphics*
Manuscript ID	TVCG-2022-11-0632
Manuscript Type:	Regular
Keywords:	L.1.0.c Cognition < L.1.0 Touch-based properties and capabilities of the human user < L.1 Human Haptics < L Haptics, D.2.14.a User interfaces < D.2.14 Human Factors in Software Design < D.2 Software Engineering < D Software/Software Engineering, I.2.7 Natural Language Processing < I.2 Artificial Intelligence < I Computing Methodologies
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
ubiideas (tvcg).mp4	

Ubiideas: Catalysing Divergent Thinking with Heads-Up Display

Katherine Fennedy, Chaoran Chen, Jiayi Zhang, Yixiao Kang, Shengdong Zhao and Chun Yu

Abstract—
Creative insights (or Aha! moments) occur ubiquitously, yet our mobile devices lack the compatible affordance to conveniently capture them in situ. Users often had to pause their activities just to interact with their smartphones. We propose Ubiideas, a smartglass system that leverages audio-visual channels to facilitate the generation of ideas anytime and anywhere. Users only need to verbalise their thoughts, and our Natural Language Processing pipeline will render the keywords within their direct view and in various organisational structures (e.g. chronological list, categorical list, and mindmap). We conducted two studies to reveal tradeoffs between input modalities and between visual layouts when ideas were rendered on glasses while multitasking. Results show that speaking enabled users to elaborate ideas more easily than typing. To our surprise, glasses were comparable to a phone despite the latter being a mature and more familiar platform. In addition, we recommend using Linear layout as the primary mode of display and Mindmap as the secondary mode. Lastly, we discuss the impacts of Ubiideas on both divergent thinking and beyond for future work to consider.

Index Terms—Artificial, augmented, and virtual realities, User Interfaces, Interaction techniques, Natural Language Processing, Information visualization.

1 Introduction

Archimedes discovered a way to calculate density and volume while in a bath[1]. Newton discovered the law of universal gravity while observing a falling apple[2]. What do these anecdotes have in common? Besides the groundbreaking discoveries that have been impacting generations, we are intrigued by how the *mind* materialises these insights while the *body* immerses in routine daily activities. Today, creativity is recognised as the peak of an individual's knowledge capability [1] and a highly valued outcome in today's globalised economy [2]. Many workplaces have optimised their designs and spatial layouts to support individual creative output [3]. Instead of modifying the physical space, which requires time and resources, we ask ourselves the question: how might we leverage knowledge workers' *physical* daily activities to effectively facilitate their divergent *thinking* (*i.e.,* generating and capturing ideas)?

One solution is to tap on prevalent mobile devices like smartphones (henceforth, just phones) which allow users to capture and access ideas anywhere they like. However, the convenience of phones serving as "digital notebooks" may be limited to when users can avail their hands for typing. Many daily activities *e.g.,* cooking and riding a bicycle, rely a lot on our hands, and yet users often have to pause their activities just to interact with their phones. Thus, phones do not effectively support multitasking and instead require users to adopt the undesirable heads-down posture [4]. It has motivated us to look for an alternative solution: Optical See-Through Head-Mounted Display (henceforth, just glasses).

We conducted two studies to study how the glasses could effectively facilitate more ubiquitous idea generation and capturing, and then implemented a glasses-based prototype named Ubiideas. For our first study, we compared three modalities (Phone+Typing, Phone+Speaking, and Glasses+Speaking) in their effectiveness to support multitasking between daily activities and divergent thinking. The purpose is to investigate if there is any advantage afforded by the glasses as compared to the phone. Results show that using hands-free (*i.e.,* voice) input enabled participants to generate more elaborate (and not the number of) ideas than typing. In addition, the benefits of glasses over a phone are in terms of recalling previous ideas and reminding active tasks. Since Study 1 standardised the visual layout to a simple list format, we investigated in Study 2, how the List can affect divergent thinking as compared to two alternative layouts (Grid and Mindmap) of keywords on the glasses that are commonly used. To our surprise, the simple chronological List contributed to more novel ideas than Mindmap due to the benefits of having ample space to see through. However, the meaningful categorisations offered by Mindmap were also favoured by the participants.

Integrating the insights from the first two studies, we implemented Ubiideas: a visualisation system that customises Natural Language Processing (NLP) pipeline to convert user audio recordings into various levels of structured information in real-time. Since users need visualisation assistance in gradual increments, we incorporated a total of seven levels (Transcript, Summary, Mindmap, Smart List, Full List, Recent, and Mic only), ranging from the most information dense to the least. Users could flexibly switch between these levels using a wearable ring mouse, depending on their

•
•

Manuscript received Month DD, YYYY; revised Month DD, YYYY.

1. https://ed.ted.com/lessons/the-real-story-behind-archimedes-eureka-armand-d-angour
2. https://education.nationalgeographic.org/resource/isaac-newton-who-he-was-why-apples-are-falling

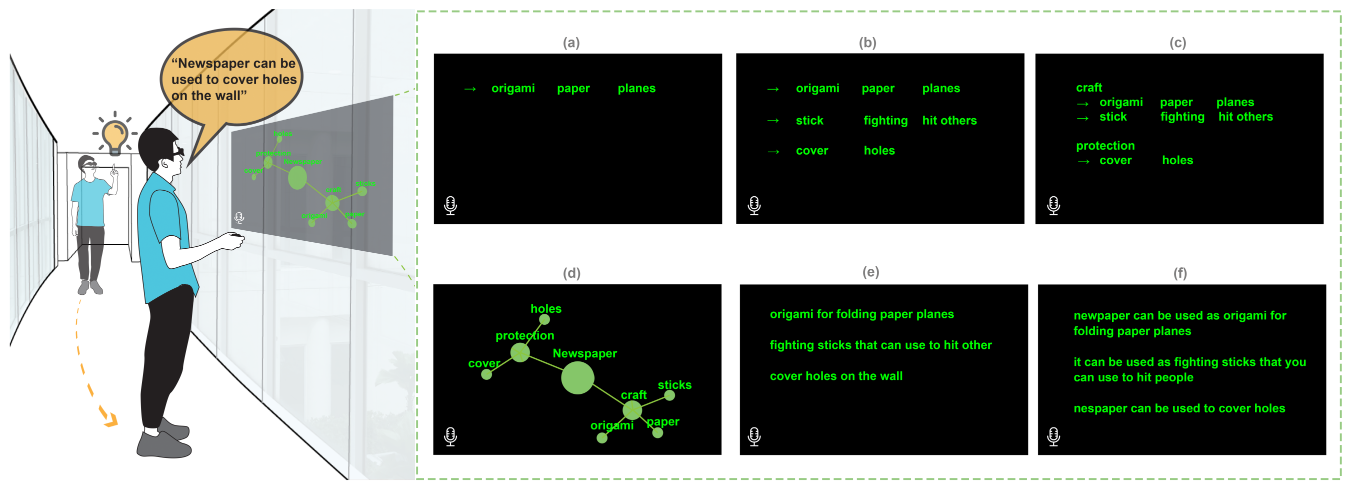

Fig. 1. Imagine a user who just had a spontaneous idea while on-the-go. Our proposed system Ubiideas allows the user to verbalise it hands-free and renders the ideas within their line of sight in real-time. There are a total of seven levels of varying information density which users could flexibly switch by clicking on a ring mouse. Here we illustrate six of them: a) Recent, b) Full List, c) Smart List, d) Mindmap, e) Summary, and f) Transcript.

multitasking familiarity and needs. While unrelated daily activities may have been traditionally seen as distractions, we hope that Ubiideas can paint an alternative picture of how these “distractions” can instead be capitalised through glasses to incubate creativity, thus offering a new source of inspiration.

Our main contributions can be summarised as follows:

- We investigated the impact of input modality on idea generation while multitasking.
- We investigated tradeoffs when adopting existing visualisations for the glasses and multitasking scenarios.
- We implemented Ubiideas, a novel smartglass system that effectively leverages audio-visual channels to facilitate ubiquitous idea generation. We also open-sourced[3] a working prototype of Ubiideas for future work to take advantage of.

2 Related Work

Our work relates to three areas in the literature: 1) what creative insight is, 2) how insightful moments can be facilitated in the course of daily activities, and 3) what support tools have been proposed thus far? These questions fundamentally relate to human cognition and form the bedrock of our work.

2.1 The Nature of Insight

Contrary to popular belief, insight is not a mysterious process of the human mind that comes from nowhere. By studying the minds of famous inventors and polymaths, Wallas [5] outlined four stages of the creative process: 1) preparation, 2) incubation, 3) illumination, and 4) verification. One explanation for why the process felt sudden is that the creative process is much more recursive and chaotic [6], than linear. As opposed to the right brain myth [7], the creative process involves a dynamic interplay of many brain regions [8], thinking styles [9], and emotions [10]. Another reason is that our minds lack consciousness when we incrementally and loosely form associations between disparate ideas [11]. In fact, Sternberg and Davidson [12] conceptualised creativity as the reorganisation of a person’s mental representations of a stimulus. Despite the similarities between insight and creativity, the former is simply one way to produce creative solutions [13]. Another way is through a conscious, deliberate analysis of the problem [14]. While there remains a body of literature we have yet to exhaust, these two reasons sufficiently motivated our work to find a plausible “method to the madness” that creative processes may entail. More specifically, our work aspires to facilitate and streamline the creative process through an AI-supported visualisation system.

2.2 Routine Multitasking is opportune for creativity

Besides the powerful human minds, let us not ignore the equally crucial role activities and environments play in generating creative insights. Seifert et al. [15] interpreted the Prepared Mind theory [16] as: creative insights do not necessarily occur to people whose minds are prepared, but rather, prepared minds are more likely to identify and take advantage of fortuitous opportunities upon external stimulation. This interpretation motivated Seifert et al. [15] to introduce the opportunistic-assimilation hypothesis, which promotes how the course of daily activities could facilitate the incubation process and may eventually lead to serendipitous solutions to an existing problem. In fact, Ohly et al. [17] were one of the first to present evidence that work routines are positively related to creativity and innovation. Empirical studies suggest that when user behaviour reaches automaticity (or becomes a habit), cognitive resources can be freed [18] to develop new ideas. A very clear example is illustrated by philosopher Friedrich Nietzsche [19], who once wrote, ‘all truly great thoughts are conceived by walking.’ It naturally begs the following question: what other activities of daily living [20] can be opportunistic for creative insights?

3. We will make it available for the public upon official publication of this manuscript.

From exercising to showering, there are plenty of anecdotes of people experiencing flashes of insight [21]. The Attention Restoration Theory [22] posits that exposure to nature encourages more effortless brain function, thereby renewing its directed attention capacity [23]. However, Oppezzo and Schwartz [24] separated the effects of outdoor stimulation and walking: the former has many cognitive and emotional benefits, while the latter explicitly improves creativity. In an adjacent research field, neuroscientists have studied how mind wandering [25] may hold the key to our creative bursts because it is likely to be the brain state that corresponds to the incubation process [26], often followed by the Aha! Moment [5]. Sawyer [8] suggested that although half of the population is not aware of their mind-wandering episodes, unrelated daily activities can provide incubation moments for creative thought [27]. We note that small ideas should not be overlooked as they can often lead to potential solutions [28]. We consider these findings as we investigate the everyday routine tasks where participants can leverage creative processes.

2.3 Creativity Support Tools

Digital technologies that facilitate creative thinking have been well-studied. However, few studies have investigated how we can facilitate divergent thinking, one of which is Idea Expander [29], a proof-of-concept tool that supports group brainstorming. It dynamically presents pictorial stimuli based on the group-like conversation. In contrast, InspirationWall [30] and IdeaWall [31] used speech-to-text techniques to monitor the groups' verbal discussion and presented suggestions using keywords. The main difference is that the former uses a non-intrusive desktop visualisation, while the latter uses a literal room wall to project the keywords. While these three related works have revealed technology's role in supporting creativity, they were mostly constrained to stationary and group settings. On the other hand, research has shown that group interaction can result in cognitive interference as much as it cognitively stimulates [32]. Therefore, we chose to focus on facilitating individual brainstorming. Future work could extend the system to multiple users in versatile environments.

In the last decade, we have witnessed how Optical Head-Mounted Display technologies have garnered relevance among the human-computer interaction research community. This is because wearing the glasses affords the user to maintain direct visual contact with both the physical surrounding and the digital content simultaneously [33]. From reading/editing texts [34, 35], to watching videos [36], and to browsing an augmented physical space [37], these examples demonstrate the emerging applications of such interactive glasses. The closest work to our on-the-go divergent thinking context would be LifeTags [38]. In an attempt to support digital lifelogging, it passively captures snapshots from life experiences and then renders a summarised word cloud on the smartphone. Although our work leverages the same life experience, we have a very different visual output (*i.e.,* keywords within direct user view) and goal (*i.e.,* maximise the individual creative potential).

3 Study 1: Investigating Input Modalities

Reviewing the literature has revealed an interesting gap: routine multitasking is opportune for creative insights, but to what extent have today's device platforms enabled us to reap the benefits from such opportunities? On the one hand, mobile devices like phones and tablets have reached mainstream adoption. Since the latter is not as readily available on-the-go due to its relatively large size (*e.g.,* cannot fit in a pocket), we chose to focus on the ubiquitous phones instead. On the other hand, wearable candidates range from the recently adopted smartwatches to the emerging heads-up display (or glasses). Due to the limited screen real estate of the watch, we chose to focus on the glasses for a more meaningful comparison with the phone.

As our first study, we evaluate three input modalities with the two platforms: Phone+Typing, Phone+Speaking, and Glasses+Speaking. The reason why we did not include Glasses+Typing and tablet-and-stylus alternatives is due to the lack of established methods that are convenient enough to perform in mobile multitasking scenarios. To be more specific, we aim to answer the following research questions:

- **RQ1.1**: How does input modality impact the objective performance of idea generation while multitasking?
- **RQ1.2**: How does input modality impact the subjective perception of idea generation while multitasking?

Among the three modalities, we hypothesise that Glasses+Speaking would facilitate better idea generation than Phone+Speaking because the former affords a heads-up posture to perceive the external environment more than the latter. In addition, since previous works have already demonstrated that speaking alone is more productive than having to type [39], we hypothesise that Phone+Typing would perform the worst and be perceived negatively for generating ideas while multitasking. In this study, we minimise technical limitations on speech recognition by using the Wizard-of-Oz approach.

3.1 Task

When coming up with thoughts regarding a particular topic, what typically happens is that the first batch of ideas comes easily and quickly, until reaching some threshold (or problem-solving impasse [15, 40]). At this point, one has difficulty generating heterogeneous ideas and gets 'stuck' in thinking mode. Since our focus is on helping users to achieve creative leaps and be more out-of-the-box, we are interested in how multitasking could help users overcome their mental impasses. Thus, we will first use a single-task scenario as a baseline to simulate a mental impasse before introducing participants to the multitask scenario. When picking the experimental task, we chose Guilford's Alternative Uses (GAU) [41] used in previous studies [24, 31, 42], which tests divergent thinking abilities with everyday objects [43]. For instance, we asked participants what the alternative uses for brick are and listed the assumed normal usage is to build a wall/house. Valid alternative usages include "to throw through a window". We also instructed each participant to elaborate and justify their chosen use case, "for example, during an emergency". Thus, we can consider the feasibility of each idea for further analysis (see results section).

We adapted GAU by simulating two distinct task scenarios for participants to complete. The first scenario (*SingleTask Scenario*) requires each participant to be seated in a room. The only goal of the participant was to generate as many distinct and relevant ideas as possible through either typing or speaking (*i.e.,* think-aloud), depending on the assigned modality. One potentially confounding factor is that people may have different experiences with the objects we pick. To minimise this factor, we paid careful attention to filtering the chosen objects as topics: 1) shirt button, 2) shoe, and 3) newspaper. First of all, the everyday objects we chose varied in size and shape: button is the smallest and of circle shape, shoe is of middle size and odd shape, while newspaper is the largest size. Secondly, based on our pilot study (of 6 participants), we did not observe any significant differences in the quantity of ideas generated between these 3 chosen test topics. As a final approach, we also ensured all 3 modalities were equally paired with 3 topics across all participants.

The second scenario (*MultiTask Scenario*) requires each participant to multitask between daily activities and idea-generation while using the assigned modality of SingleTask Scenario. The primary goal of the participant is to complete daily activities as naturally as possible, while the secondary goal is to resume and maximise the idea-generation from the same object (or topic) used in SingleTask Scenario. To simulate the incubation phase [15], we curated five daily activities ranging in legs- and hands-busy (see Table 1 for an overview). To avoid systematic biases, we followed Aaron Dollar's [20] categorisation of activities of daily living (ADLs) and ensured that each category is represented by at least one activity. While having insights in the shower is commonly reported [21], we did not include them for privacy reasons. Please refer to our supplemental file for more details.

Aaron Dollar's Categorisation [20]	Activity	Hands-busy	Legs-busy
Domestic ADL	Fold clothes (5 clothes and 2 pants)	2H	stand
	Watch video on laptop (2min)	0H	sit
Extradomestic ADL	Walk + Hand-carry 1.5kg bag	1H	walk
Physical Self-Maintenance	Leisure walking (2min)	0H	walk
	Wash hands with soap + Dry hands	2H	stand

TABLE 1
List of activities participants had to complete for MultiTask Scenario of Study 1 and 2. The number in 0H, 1H, and 2H labels represent the minimum number of hands required.

To standardise the visual format in each modality, we captured the ideas across all modalities using a linear list of keywords (see Fig. 2), where one line represents one idea. This is because a linear layout allows faster reading and more preferred over non-linear layouts [44]. For Phone+Typing, we instructed participants to concisely type the keywords of their use cases and explanations while ignoring grammar or spelling accuracy. For Phone+Speaking and Glasses+Speaking, participants verbalised their ideas and ignored pronunciation accuracy, while the experimenter simulated the automatic capturing of keywords. The standardised presentation of keywords were controlled: regardless of the modality, participants could consistently refer to their previously-generated ideas as visual feedback, which is captured by the system to show their intent or to inspire new ideas [31]. Through our pilot study of 6 participants, we learned that the persistent display of the glasses could be overwhelming or occluding the participants' vision, and half of them wished to toggle the heads-up display on and off flexibly. Therefore, we allowed participants to verbalise commands like "Hide/Show ideas", "Scroll up/down", and "Zoom in/out".

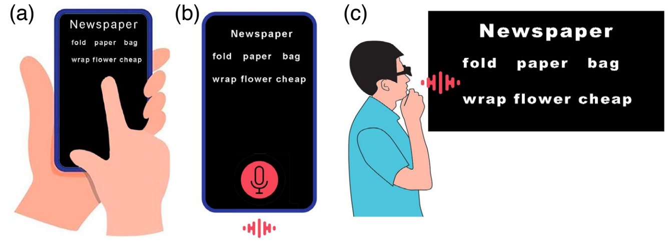

Fig. 2. Participants used one modality to generate ideas in each condition: (a) Phone+Typing, (b) Phone+Speaking, and (c) Glasses+Speaking. Since the glasses will render the black background colour as semi-transparent, we activated dark mode on both Phone conditions for consistency.

3.2 Procedure

For the training phase, participants familiarised themselves with the idea-generation task for 1 minute, followed by a guided walkthrough of the route and tasks to be completed for daily activities. For the test phase, although participants were not informed of how long the first scenario would need, the experimenter would start the timer as soon as the participant paused typing or speaking. Only if the participant did not generate any idea throughout the 30 seconds continuously would the SingleTask Scenario end, and they would proceed to the MultiTask Scenario. This procedure design is inspired by prior research, which had demonstrated that originality of ideas increases with quantity. The purpose of this design is to establish an experimental control such that only when participants have sufficiently exhausted their ideas, that they could finally proceed to the next scenario. We called this the idea *saturation threshold* and determined its duration through 6 pilot participants who spent a maximum of 25 seconds before generating a new idea. The study ended with a questionnaire and an open-ended interview. It took approximately 2 hours to complete.

3.3 Participants

We recruited a total of 18 participants (6 male), aged 19 to 25 (M=21.6, SD=1.9) from the university community. Their proficiency level with the English language was either native or full professional and they received approximately $7 USD for every hour of their participation.

3.4 Apparatus

For both phone modalities, we used a Samsung S20 phone installed with Android Version 12. For the glasses modality, we connected NReal Light glasses to the phone via USB. The glasses have customisable nose pads and prescription lenses to meet the individual needs of the participants. For all three modalities, we required participants to be wearing two items. Firstly, a wireless Bose headphone was worn to minimise external noise and maximise the quality of audio recordings. Another reason is to maximise the advantages of the Phone+Speaking modality such that users don't have to bring the phone up to transmit information. Secondly, a body-strapped iPhone X was used to record the verbal generation of ideas while multitasking. We connected the phone to an experimenter's laptop and iPad Air via a Zoom meeting so that the experimenter could remotely and simultaneously view three sources: 1) digital content (by either phone or glasses modality), 2) physical tasks, and 3) external environments engaged by each participant.

3.5 Design

We used a within-subject design with input MODALITY as one independent variable with 3 levels { PHONE+TYPING, PHONE+SPEAKING, GLASSES+SPEAKING }, and SCENARIO as another independent variable with 2 levels { SINGLE-TASK, MULTITASK }. The order of the modalities and topics were counterbalanced using Latin Square. In terms of dependent variables, we measured both objective performance and subjective perception. Please refer to our supplemental file for more details.

3.5.1 Objective Performance

Firstly, we transcribed the audio recordings and adopted best practices from prior work on how to score divergent ideas objectively. [45] evaluated three approaches: summation score (fluency + originality + flexibility), uncommon score (number of ideas given by <5% of the sample), and weighted-fluency (determined by relative frequency of ideas). Fluency refers to the number of responses, originality refers to the number of unique responses (only appear once within the sample), and flexibility refers to the number of categories in the responses. [45] reported that weighted-fluency has the highest validity and thus, we chose to focus on this measure for our analysis. However, we note that its weight calculation only takes novelty (*i.e.*, frequency of ideas) into consideration, but not flexibility, feasibility, and elaborateness, which were acknowledged as equally important by [46] and also incorporated by recent works [24, 42].

As a result, we decided to evaluate not only *Fluency* (which focuses on quantity of ideas), but also the weighted sum based on each quality of ideas: *Novelty*, *Flexibility*, *Feasibility*, and *Elaborateness*. For *Novelty* and *Flexibility*, the weight of each response is calculated based on this formula:

$$1 - \frac{frequency}{sample\ size}$$

This means that the less frequent the response or category (each response belongs to), the heavier the weight. For *Feasibility* and *Elaborateness*, the weight of each response is calculated by rating each of the transcribed responses using 1-5 Likert Scale, and then computed into a percentage (*i.e.*, the minimum value attributed to an idea is 0.2 and the maximum is 1.0).

Secondly, we measured the disruption caused by multitasking in Scenario 2. This can be done by applying the formula below to the task of carrying, washing, and folding completed by each participant:

$$\frac{time\ with\ multitasking}{time\ without\ multitasking} - 1$$

Thus, the lower the better and we call this measure *Time Increase* (%).

3.5.2 Subjective Performance

We measured subjective perception of participants using a 1-7 Likert Scale for statements relating to physical and mental demand of participants during multitasking [47]. We also asked participants to rate primary and secondary tasks separately. For the primary task, we asked how easy it was for them to complete the assigned daily activities. For the secondary task, we asked them to rate the perceived quantity and quality of ideas they generated. All statements and ratings were consistently presented such that higher rating corresponds to more desirable attributes. In addition, we asked participants their ranking preference for each modality and each primary task with regards for their future idea generation.

3.6 Results

We recorded an estimated total of 810 minutes (18 participants x 3 modalities x 15min) period of idea generation. The recordings were transcribed to a total of 867 ideas, in which 108 (12.5%) were discarded because they are either repeated (by the same participant for the same MODALITY) or not following task guidelines. Following [48], we processed the remaining valid 759 ideas in two steps. First, we abstracted similar ideas (*e.g.*, "currency for trading" and "fake money for children" for alternative uses of "shirt button") into **generalised keywords** (*e.g.*, "coin"). Then, we grouped the keywords into **generalised categories** (*e.g.*, "coin", "for magic trick", and "monopoly pieces" were grouped under "props"). This 2-step process produced a total of 230 keywords and 48 categories across all three topics.

For analysis, trials were first aggregated by participants and then by factors used in the test. When the assumption of sphericity was violated, we corrected both p-values and degrees of freedom using Greenhouse-Geisser ($\epsilon < 0.75$). We used Shapiro-Wilk and found all data residuals to be of normal distribution. The data residuals for *Time Increase* were found to be non-normal by Shapiro-Wilk, we used Aligned Rank Transform [49] to correct normality for analysis. A two-way ANOVA with pairwise t-tests (with Bonferroni corrections for post hoc comparisons) were used for objective continuous measures. Wilcoxon Signed Rank tests were used to analyse subjective ratings. To be concise, we grouped the reporting for measures that exhibit a similar pattern of results.

3.6.1 Elaborateness

As shown in Fig. 3e, participants elaborated their ideas 25.2% more while doing a single task (than multitasking[4]) and 37.0-44.8% more when they were speaking (than

4. The MultiTask Scenario starts immediately after the SingleTask Scenario, which only ends upon reaching the idea *saturation threshold*, *i.e.*, when the participant failed to generate idea in a continuous period of 30 seconds.

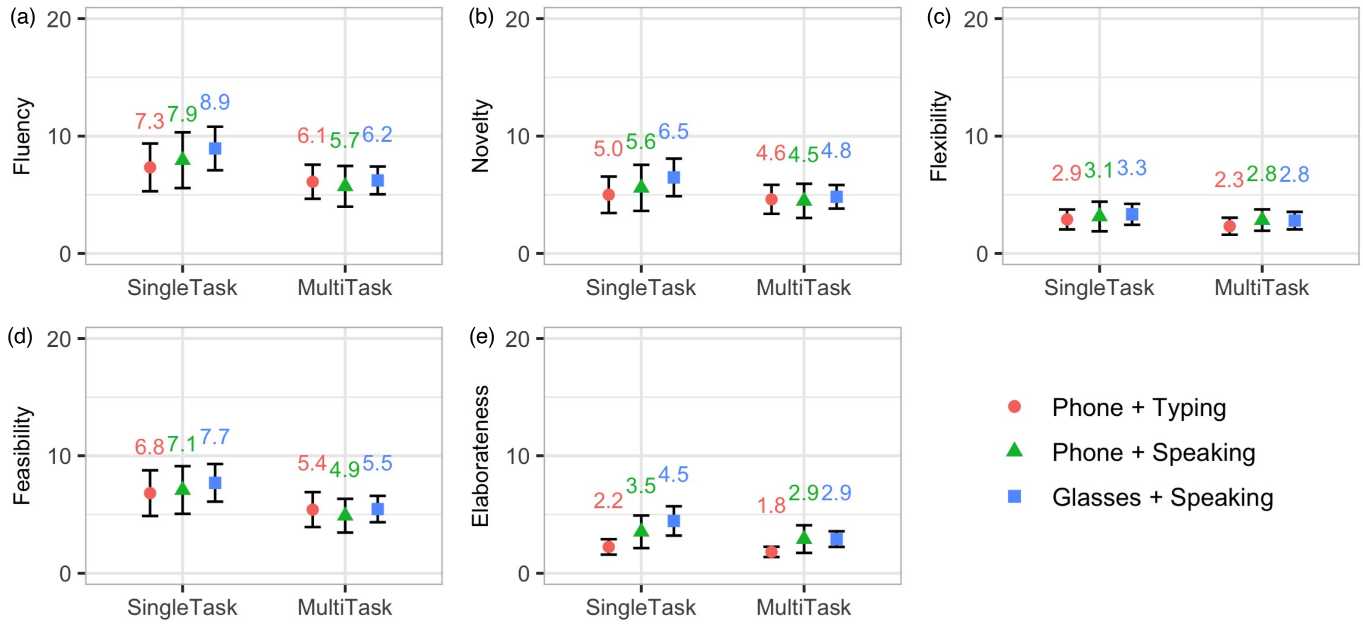

Fig. 3. Scoring of ideas by SCENARIO for each MODALITY in terms of (a) *Fluency*, (b) *Novelty*, (c) *Flexibility*, (d) *Feasibility*, and (e) *Elaborateness*. Error bars are 95% confidence.

typing). There was a significant main effect of SCENARIO ($F_{1,17} = 10.06$, $p < .01$, $\eta_G^2 = .04$) and MODALITY ($F_{2,34} = 10.41$, $p < .001$, $\eta_G^2 = .10$) on *Elaborateness*. This means that, even after reaching a saturation point, participants could still elaborate new ideas albeit to a lesser extent. Pairwise comparisons show that participants could elaborate more ideas while using GLASSES+SPEAKING ($M = 7.37$, $p < .01$) and PHONE+SPEAKING ($M = 6.44$, $p < .00001$) than while using PHONE+TYPING ($M = 4.07$). We did not find any interaction between SCENARIO and MODALITY, on *Elaborateness* of generated ideas.

3.6.2 Fluency and Feasibility

Participants were more fluent (by 34.2%) and they generated more feasible ideas (by 36.9%) while doing a single task than while multitasking (Fig. 3a and d). There was a significant main effect of SCENARIO on *Fluency* ($F_{1,17} = 9.32$, $p < .001$, $\eta_G^2 = .07$) and *Feasibility* ($F_{1,17} = 8.83$, $p < .001$, $\eta_G^2 = .07$). We did not find any main effect of MODALITY and any interaction between SCENARIO and MODALITY, on both *Fluency* and *Feasibility*. In terms of *Fluency*, participants on average scored 13.44 for PHONE+TYPING, 13.67 for PHONE+SPEAKING, and 15.17 for GLASSES+SPEAKING across both scenarios. In terms of *Feasibility*, participants on average scored 12.24 for PHONE+TYPING, 11.99 for PHONE+SPEAKING, and 13.17 for GLASSES+SPEAKING across both scenarios.

3.6.3 Novelty and Flexibility

The ideas generated were similarly novel and flexible across the three modalities and two scenarios (Fig. 3b and c). There was no significant main effect of SCENARIO on *Novelty* ($F_{1,17} = 4.34$, $p > .05$) and *Flexibility* ($F_{1,17} = 3.11$, $p > .05$). We did not find any main effect of MODALITY and any interaction between SCENARIO and MODALITY, on both *Fluency* and *Feasibility*. This means that participants performed similarly for both measures regardless of the input modality. In terms of *Novelty*, participants on average scored 9.62 for PHONE+TYPING, 10.07 for PHONE+SPEAKING, and 11.31 for GLASSES+SPEAKING across both scenarios. In terms of *Flexibility*, participants on average scored 5.23 for PHONE+TYPING, 6.00 for PHONE+SPEAKING, and 6.14 for GLASSES+SPEAKING across both scenarios.

3.6.4 Time Increase

Participants experienced a similar level of speed disruption regardless of the input modality. There was no significant main effect of MODALITY ($F_{2,31364} = 0.13$, $p > .05$) on *Time Increase*. While there was no interaction between MODALITY and TASK, there was a significant main effect of TASK ($F_{2,30} = 3.38$, $p = .037$, $\eta_G^2 = .05$) on *Time Increase* (Fig. 4b). However, pairwise comparisons did not show any significant difference between the TASKS. On average, there was a 24.9% for CARRYING a bag while walking, 21.3% for WASHING hands, and 19.3% increase in time for FOLDING clothes.

3.6.5 Subjective Rating

Speaking made it easier for participants to multitask daily activities and generate ideas simultaneously, as compared to typing (Fig. 5). There were significant main effects of MODALITY ($V = 4$, $p < .001$ and $V = 17$, $p < .01$) on *Multitasking Ease* and *Low Physical Demand*, where participants rated PHONE+SPEAKING and GLASSES+SPEAKING (both median=6 for both metrics) more favourably than PHONE+TYPING (median=3.0 and median=4.0 respectively). There was another significant main effect of MODALITY ($V = 0$, $p < .001$ and $V = 4.5$, $p < .001$) on *Low Mental Demand* and *Daily Activities Ease*, where participants rated PHONE+SPEAKING (median=6 and median=5.5) more

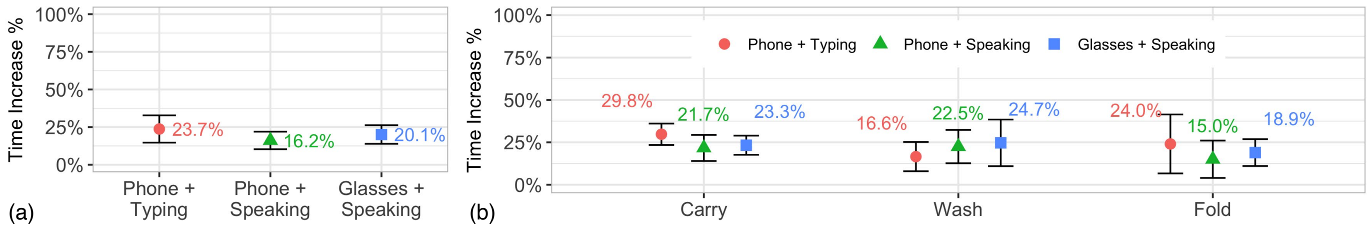

Fig. 4. (a) *Time Increase* by MODALITY for a combination of tasks: walking while carrying bag, washing hands, and folding clothes. (b) *Time Increase* by TASK for each MODALITY.

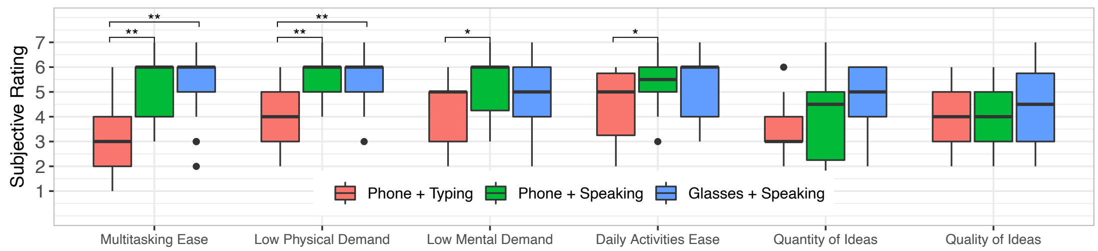

Fig. 5. Subjective rating for each MODALITY (1: Strongly Disagree, 7: Strongly Agree).

favourably than PHONE+TYPING (median=5.0 for both metrics). However, there were no significant differences between all three MODALITIES in terms of perceived *quantity* and *quality* of ideas.

3.7 Discussion

3.7.1 Typing vs.Speaking

Since speaking is less effortful than typing [39], it is reasonable to expect the former modality to generate more ideas (from the same topic) than the latter. However, our results revealed that speaking and typing were comparable in terms of fluency of ideas. While this may not be intuitive, we suggest looking at the elaborateness score to shed some light on the plausible reason. During the single task scenario, verbalising ideas (either in Phone+Speaking or Glasses+Speaking) increased the amount of elaboration by 59-105% as compared to the act of typing (*i.e.,* Phone+Typing). We also observed that while a minority of 4 participants dutifully typed out the details, the large majority chose to concisely capture a single idea by typing key phrases of 7 words or less. This is because the convenience of speaking made it easier for participants to be generous with fleshing out the details. 6 participants even enhanced their idea descriptions by gesturing their hands as they talked. For typing, both hands are instead occupied with translating thoughts into words. In other words, this tradeoff between fluency and elaborateness scores highlighted the underrated impact between typing and verbalising ideas.

Csikszentmihalyi [50] refers to elaboration as the "99% perspiration of creativity," while the rest is 1% inspiration, as suggested by Thomas Edison. One reason is that the process of elaboration allows ideas to interact with reality which then reveals potential rooms for improvement [51], thus enabling future inspiration and boosting overall output of ideas. Considering the importance of elaboration, we reflected on the visual design used in this study to render ideas. For instance, P14 and P19 suggested revealing raw transcription to supplement the already extracted keywords by the wizard. This is because there were times when the extracted keyword (*e.g.,* "decoration") was not sufficient to paint the complete picture of the intended idea (*e.g.,* "cut out newspaper articles forming interesting patterns for scrapbook decoration"). It is challenging to expect a human experimenter to manually record more relevant keywords for a single idea because some participants spoke very fast. However, we foresee this limitation could be easily overcome when we implement an automatic pipeline (in Section 5) to record and process the user ideas more systematically. In fact, we are also motivated to further investigate how the visualisation might be designed differently (from a linear list) to capture the elaborated details of an idea. Would semantic grouping (*e.g.,* mindmap) of ideas be more effective in reflecting details of an idea, as compared to a linear list? To summarise, our results highlighted not only the importance of elaborating ideas, but also how visualisations can play a role in capturing the elaboration, to ultimately support the multidimensional needs of the creative brainstorming process [52].

3.7.2 Phone vs.Glasses

Despite the similar quantitative performance between Phone+Speaking and Glasses+Speaking, we discovered qualitative differences favouring the glasses in terms of recall. Note that since the phone is a mature and more familiar platform than glasses, the fact that the performance was similarly well might suggest a potential advantage of glasses over the phone.

On the one hand, glasses offer easier access to revisit ideas by rendering the captured keywords within their line of sight. For instance, P15 liked the fact that she could "see the ideas and the real world as background simultanesously" with the glasses, and P13 shared how the glasses enabled her to "be more aware of the surroundings". On the other hand, the phone usage required P3 to frequently pause her folding clothes (*i.e.,* primary task) just to refer to her previous ideas (*i.e.,* secondary task). This switching between primary and secondary tasks, if repeated often, would incur significant cognitive resources and affect overall multitasking performance [53]. This is especially important because we witnessed the natural tendency of participants to forget previously generated ideas. P15 asked, "have I said this before?" and since the experimenter was not allowed to respond, P15 proceeded with repeating the ideas instead. Therefore, we argue that despite rendering the same consolidated information on both phone and glasses, participants struggle less to locate the information. As a result, users would need fewer cognitive resources to switch between views, thus availing them with more resources for future creative output.

Beyond recalling ideas, we observed how wearing the glasses could serve as a physical reminder of active mental tasks. The multitasking context means that user attention had to be shared between idea generation and various daily activities. For instance, we observed that P2 did not generate any new ideas while folding clothes, not because he was stuck, but because the phone was accidentally hidden under the pile of clothes to be folded. We interpret that the phone is more likely to exhibit an "out of sight, out of mind" phenomenon while wearing the glasses is kin to putting on a "thinking cap".

Overall, we conclude Study 1 by:

- **Answering RQ1.1**: Speaking on both phones and glasses led to a higher rate of elaboration, as compared to typing on the phone. However, there were no objective differences in other metrics (*i.e.,* fluency, novelty, flexibility, and feasibility) and between speaking on the phone and speaking on glasses.
- **Answering RQ1.2**: The advantage of speaking on glasses over the phone is that glasses were perceived to provide a more convenient access to recall previous ideas, and also a more salient reminder of the active idea generation task.

The key takeaway is that it is not only the user input mode of speaking that is beneficial, but also how the system readily outputs consolidated information to the user is equally important, thus warranting further investigations in the next study.

4 Study 2: Investigating Visual Layouts

Results from Study 1 revealed the potential advantages of combining verbal idea generation with a visual heads-up display (*i.e.,* glasses). Such multimodal interaction addresses users' needs by facilitating effective multitasking opportunities, whereby users can attend to their daily activities whilst engaging with the digital world. However, Study 1 uses a very simple visualisation (*i.e.,* linear list of texts) and only scratches the surface. From the literature of divergent thinking, we are inspired by more advanced visualisations (e.g. animated keywords [30], image matrices [31] and mindmaps [54] that have been rendered on desktop screens or walls of a room. Collectively, these approaches demonstrated positive impacts on cognitive group activities like brainstorming. Translating them to the individual context and glasses requires thorough exploration spanning the 2D, 2.5D, and 3D visualisation research areas [55]. Prior works have shown mixed benefits for human performance and implied task-display interactions [56, 57]. Therefore, as a preliminary attempt, Study 2 focuses on evaluating the minimal 2D visualisation variants for the glasses, and scopes out the more immersive 2.5D and 3D display modes for future studies.

Recent works in the literature of OHMD visualisations had identified 1) optimal text presentations (font styles, size, and positions) [34] and 2) hybrid image combinations [36]. While Ram and Zhao's contributions focus on navigational tasks and 2D video learning context, it remains unclear how spatial arrangement and density of the above-mentioned advanced visualisations can be adapted for the unique platform of glasses. This is especially important for our idea generation task, which needs to support divergent thinking with a smaller scope [59] than the complex educational contexts explored by Ram and Zhao. Thus, we narrowed down the gap between both literature by conducting a systematic evaluation of 2D visualisations on OHMD over a pilot and a formal study. First, we group prior works into three main visualisation strategies: 1) linear layouts L1 to L2, 2) grid layouts G1 to G4 , and 3) radial layouts R1 to R4 (see Fig. 6), based on its spatial distribution [60]. In each strategy, we enumerated the variations proposed by prior works and elaborate on their shared characteristics. The goal of the pilot study is to first determine the best variant to represent each strategy (as well as alternative strategies that users may prefer), before a more rigorous comparison between strategies in the formal study to investigate in the multitasking scenario.

Specifically, we are motivated to answer the following research questions:

- **RQ2.1**: What are the promising visual layouts of ideas for glasses to adopt?
- **RQ2.2**: How does visual layout on glasses impact the objective performance of idea generation while multitasking?
- **RQ2.3**: How does visual layout on glasses impact the subjective perception of idea generation while multitasking?

4.1 Visualisation Strategies

4.1.1 Linear Layouts

L1 was used in Study 1 and displayed 12 items within a single screen. Unlike random alignments in InspirationWall [30], we aligned all text to the left [61]. L2 is accompanied with visual cues, positioned below the text [31] and hypothetically should be preferred for idea generation [29, 31], but as depicted above, the quantity is reduced by half (from 12 to 5.5). We also conducted a quick pre-pilot to determine suitable image size relative to the text. For image size, we transformed the image height to one, two, three, and four times of the text height. Resulting in the corresponding 7.5, 5.5, 4, and 3.5 items within a single screen. We found that using one or four times the text height is respectively too small and too big. Between two and three times the text height, we chose the latter because participants found it more suitable to see when details were small (*e.g.,* wall display). The reduced quantity may or may not be a trade-off (to be investigated in the pilot) because empty digital

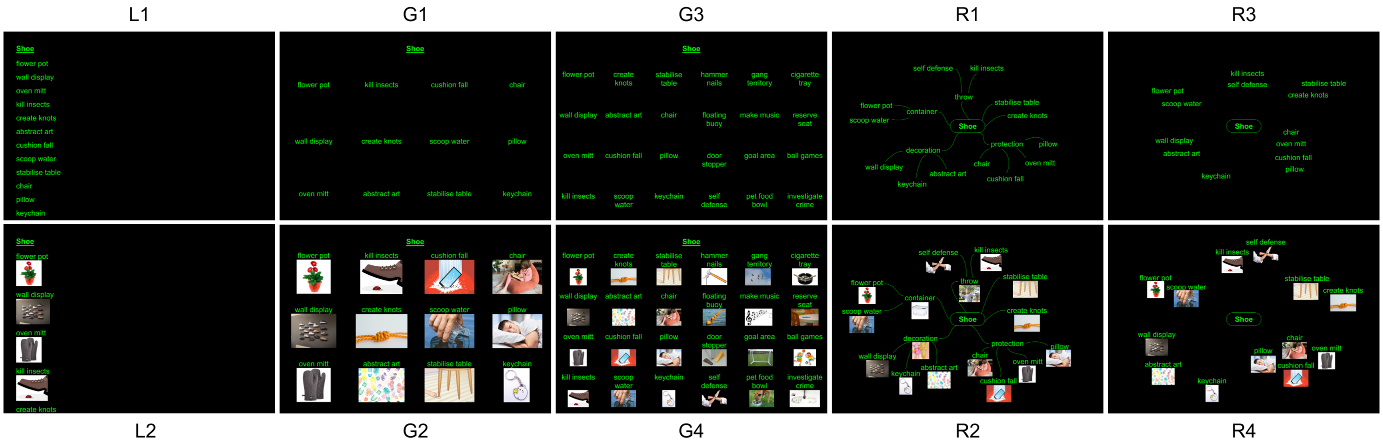

Fig. 6. The ten visualisation variants explored in the pilot study. Each variant is rendered with 1920x1080px stereoscopic display (115” diagonal), which is projected at 3m depth away from the user. All variants use Arial font size 32px [58] for text.

real estate can be used to accommodate physical tasks due to the see-through nature of the glasses. Both L1 and L2 can be scrolled vertically to reveal more ideas.

4.1.2 Grid Layouts

G1-G4 uses a grid layout where ideas are appended sequentially: vertically first, then horizontally. G1 and G2 use 3x4 layout to accommodate the same number of items as L1. G3 and G4 use 4x6 layout to accommodate double that of G1 and G2, from 12 to 24. This is similar to the visualisation used by [62] but for capturing frequent words in a meeting. The images used in G2 and G4 are all constrained to fit the uniform cell size afforded by each layout: four times the text size for G2 while two times for G4. G1-G4 can be scrolled horizontally to reveal more ideas. The grid layouts are mostly derived from contributions by [31]. Hence, our attempt in this pilot is to quickly validate which one of the variants is suitable for our OHMD display context.

4.1.3 Radial Layouts

Unlike the linear and grid layouts, radial layout explores the potential of automatic associations [54] through non-linear and networked-structured representation: Mindmap and Dynamic Cells. Radial layouts could break the traditional “linear thinking” and thus predicted to be more innate in human cognition [54]. Mindmaps (R1 and R2) involve explicit group names [31] generated by the system (*e.g.,* “flower pot” and “scoop water” under “container”). At the initial phase where ideas were few, we rendered them surrounding the prompt non-hierarchically [63]. When a meaningful group can be identified, the group name will be rendered nearest to the prompt while the respective group members are now linked to the group name instead. We included mindmap because previous works found that it supports users in internalising the content and structure of information, more effectively than word clouds and image collages [61, 64]. Dynamic cells (R3 and R4) cluster structures based on association and this was heavily used in IdeaWall [31]. The main difference with our variant is that we place the prompt in the centre [54] and spawn the item in the adjacent space, radially emanating outwards from the prompt [62]. As more items get added, and the system recognizes meaningful groupings, the relevant item will be repositioned to form a cluster. The empty space separating each item determines the relative similarity and thus clustered some items together [61]. For example, in R3 and R4, “wall display” and “abstract art” can be perceived as in the same sub-cluster while “keychain” is another sub-cluster. The images used in R2 and R4 are all constrained to two times the height of the text. R1-R4 can be panned horizontally and vertically to reveal more ideas.

4.2 Pilot Study

We recruited 6 participants (2 male, M= 24.8 years, SD = 4.2 years) and asked them to wear the same glasses used in Study 1. Then, we simulated a live capture system using the Wizard-of-Oz approach. The main difference from Study 1 is that this time, we displayed 10 visualisation variants (one at a time) and only dedicated 3 minutes for each variant. The tasks from Study 1 were simplified to only walking along a predefined loop. After all the variants belonging to a layout group (*e.g.,* G1-G4 for Grid) have been presented, we interviewed participants to extract qualitative insights and better understand which one is preferred from that specific group. Note that 3 minutes were not sufficient to populate each layout for 3 participants. Thus, after all 10 variants had been presented, we revealed the more complete/complex scenarios: when each variant is populated beyond a single screen and requires the user to instruct the system to scroll/pan to reveal more items. By viewing the greater potential of each variant, we hope to more effectively contextualise users’ preferences for usages of longer duration.

Overall results confirm that the use of images boils down to user preferences. This is consistent with [31] whose results imply that there is no objective difference between with and without images, but subjective differences may arise. In our pilot group, 5 out of 6 preferred not to use images because “images take up a lot of space” [P1,P3] and they felt that text alone is “sufficient to capture previous ideas” [P6]. Since visual information load is an important design consideration for the glasses [36], we chose L1 to represent the Linear layout and exclude the use of images for comparison within the Grid and Radial group (*i.e.,* G2, G4, R2, and R4). This is also so that we can also minimise the confounding effect of the use of images in the other groups. However, we note that the use of images can be made optional instead of system default, as suggested by P4 who wished for a "flexible switching" between L1 and L2.

For Grid layout, between G1 and G3, 5 out of 6 participants preferred G1 (3x4) over G3 (4x6). They explained that G1 offers "more empty space" [P5] to see through the glasses and into their physical world. The one participant (P6) that preferred G3 shared that seeing more items helped he track older ideas and generate newer ideas. To ensure the overwhelming display does not occlude her navigation, P6 would scroll the screen and effectively hide the ideas temporarily. Thus, we chose G1 to represent the Grid layout.

For Radial layout, R1 (mindmap) was unanimously preferred over R3 (dynamic cells). Participants mainly cited reasons like how the latter "looks more disorganised" [P1,P2,P3] while the former provides lines to help users "draw explicit connections" [P5] between ideas. In addition, Shi et al. [31] used dynamic cells and highlighted significant room for improvement. Thus, we chose R1 to represent the Radial layout. When we asked participants for alternative layouts for us to consider, all of them did not mention any new layouts.

Through this pilot study, we can now more concretely **answer RQ2.1**. Designs proposed by previous works cannot be directly transferred because there are tradeoffs when considering the unique interaction afforded by the glasses. We narrowed down the 10 visualisations for the glasses to the following 3 promising candidates for further investigations in the multitasking context.

1. **L1 for Linear**: 12 ideas vertically listed within a screen.
2. **G1 for Grid**: 12 ideas distributed across a 3x4 grid cells.
3. **R1 for Radial**: ideas clustered in a network of semantic associations. Note that we will directly refer R1 as **mindmap** layout from this point onward.

4.3 Formal Study

We used similar tasks and procedures as the previous Study 1 (see Section 3). The main difference is that Study 2 involved a new group of 18 participants (6 male, aged 19 to 28, M=22.7, SD=2.5). The two phone conditions in Study 1 were excluded and only the details for the glasses condition were retained for the three identified layouts. In addition to the ranking question asked in Study 1, we also asked participants to rank their perceived level of occlusion in each layout.

4.3.1 Design

We used a within-subject design with input LAYOUT as one independent variable with 3 levels { LINEAR, GRID, MINDMAP }, and SCENARIO as another independent variable with 2 levels { SINGLETASK, MULTITASK }. The order of the conditions and topics were counterbalanced using Latin Square to avoid any potential ordering effect. In terms of dependent variables, we analysed the same objective and subjective measures as Study 1.

4.3.2 Formal Study Results

We recorded an estimated total of 810 minutes (18 participants x 3 layouts x 15min) period of idea generation. The recordings were transcribed to a total of 1,125 ideas, in which 137 (12.2%) were discarded because they are either repeated (by the same participant for the same LAYOUT) or not following task guidelines. We repeated the same 2-step process used in Study 1 and abstracted a total of 114 keywords and 34 categories across all three topics. For analysis, we followed the same procedure as stated in Study 1.

4.3.2.1 Novelty: Participants generated more novel ideas (by 24.5%) while using the linear layout than with mindmap (Fig. 7b). There was a significant main effect of LAYOUT on *Novelty* ($F_{2,34} = 3.41$, $p = .045$, $\eta^2_G = .02$). Pairwise comparisons show that LINEAR ($M = 14.55$) layout resulted in more novel ideas than MINDMAP ($M = 11.69$, $p < .01$). Although LINEAR list also resulted in more novel ideas than GRID ($M = 12.47$) on average, the difference was not significant ($p = .364$). We did not find any main effect of SCENARIO and any interaction between LAYOUT and SCENARIO on *Novelty*.

4.3.2.2 Fluency and Feasibility: Participants were more fluent (by 22.1%) and they generated more feasible ideas (by 34.5%) while doing a single task than while multitasking (Fig. 7a and d). There was a significant main effect of LAYOUT on *Fluency* ($F_{1,17} = 5.07$, $p = 0.038$, $\eta^2_G = .03$) and *Feasibility* ($F_{1,17} = 9.88$, $p < .001$, $\eta^2_G = .08$). We did not find any main effect of SCENARIO and any interaction between SCENARIO and LAYOUT, on both *Fluency* and *Feasibility*. In terms of *Fluency*, participants on average scored 19.83 for LINEAR, 18.11 for GRID, and 16.83 for MINDMAP across both scenarios. In terms of *Feasibility*, participants on average scored 17.23 for LINEAR, 16.3 for GRID, and 15.49 for MINDMAP across both scenarios.

4.3.2.3 Flexibility and Elaborateness: The ideas generated were similarly flexible and elaborate across the three layouts and two scenarios (Fig. 7c and e). There was no significant main effect of SCENARIO on *Flexibility* ($F_{1,17} = 0.19$, $p > .05$) and *Elaborateness* ($F_{1,17} = 4.04$, $p > .05$). We did not find any main effect of LAYOUT and any interaction between LAYOUT and SCENARIO, on both *Fluency* and *Feasibility*. This means that participants performed similarly for both measures regardless of the visual layout. In terms of *Flexibility*, participants on average scored 7.49 for LINEAR, 6.58 for GRID, and 5.86 for MINDMAP across both scenarios. In terms of *Elaborateness*, participants on average scored 7.91 for LINEAR, 7.47 for GRID, and 6.91 for MINDMAP across both scenarios.

4.3.2.4 Time Increase: Participants experienced a similar level of speed disruption regardless of the layout used. There was no significant main effect of LAYOUT ($F_{2,136} = 1.50$, $p > .05$) on *Time Increase*. While there was no interaction between LAYOUT and TASK, there was a significant main effect of TASK ($F_{2,136} = 9.93$, $p < .0001$, $\eta^2_G = .13$) on *Time Increase* (Fig. 8b). Pairwise comparisons show that FOLDING ($M = 12.26\%$) was disrupted the least as compared to WASHING hands ($M = 26.36\%$) and CARRYING a bag while walking ($M = 27.84\%$).

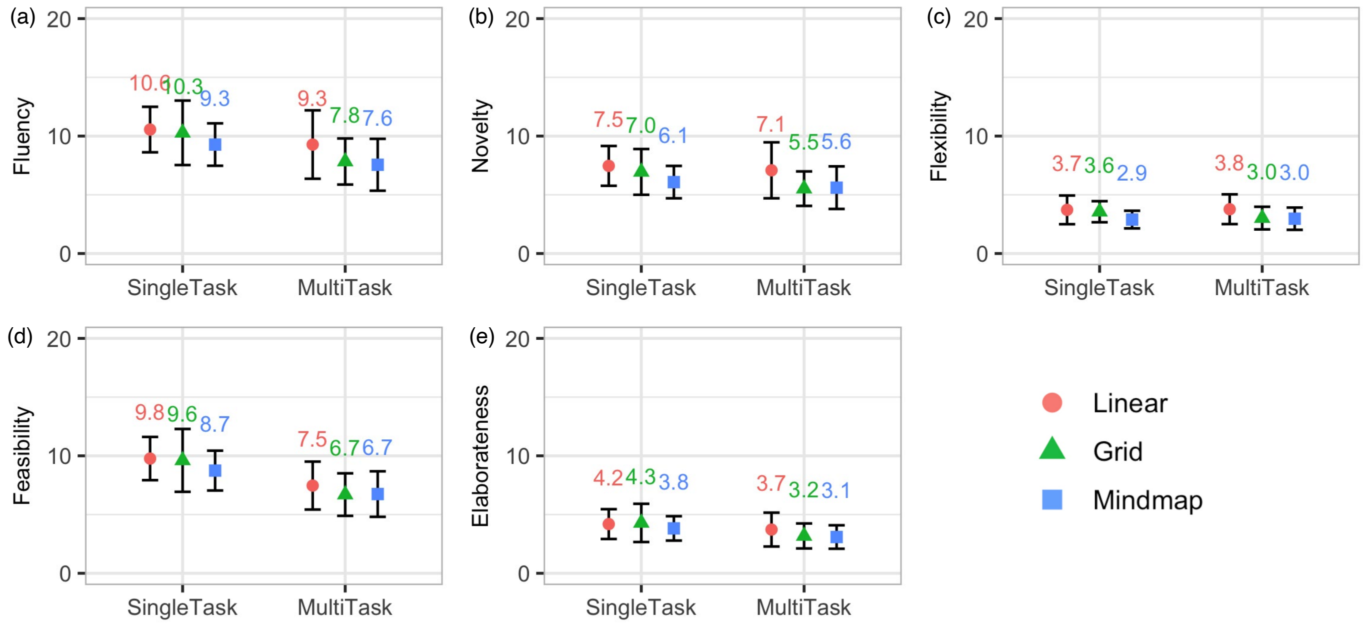

Fig. 7. Scoring of ideas by SCENARIO for each LAYOUT in terms of (a) *Fluency*, (b) *Novelty*, (c) *Flexibility*, (d) *Feasibility*, and (e) *Elaborateness*.

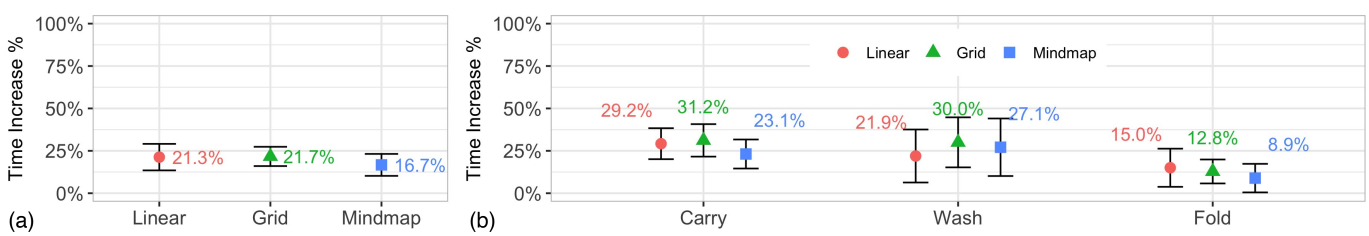

Fig. 8. (a) *Time Increase* by LAYOUT for a combination of tasks: walking while carrying bag, washing hands, and folding clothes. (b) *Time Increase* by TASK for each LAYOUT.

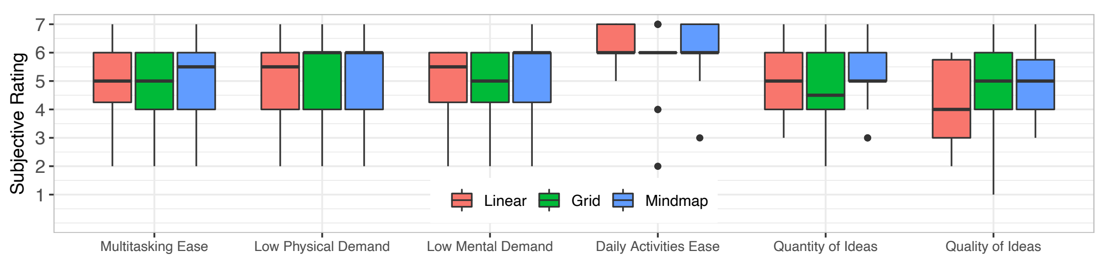

Fig. 9. Subjective rating for each LAYOUT (1: Strongly Disagree, 7: Strongly Agree).

4.3.2.5 Subjective Rating: Participants rated the three visual layouts similarly in terms of ease of multitasking, physical and mental demand, ease of daily activities, quantity and quality of ideas generated. There were no significant effects of LAYOUT in all measures.

4.3.3 Discussion

4.3.3.1 Maximise creative output by minimising perceived occlusion: Listing ideas linearly is a simple yet an effective strategy to generate novel ideas on the glasses. According to our analysis, the Linear layout delivered a statistically higher novelty score by 24.5% as compared to that of Mindmap. Reasons could be attributed to how the organisation of digital information affects the amount of visual occlusion and thus, influencing user cognitive workload and creative output [65]. The majority (16/18) of the participants perceived that Linear occludes their vision the least, followed by either Grid or Mindmap in no clear order. They explained that Linear's approach in "concentrating all ideas within a small area on the side made it easier to glance over quickly" [P13] and therefore, "suitable for multitasking" [P4]. In contrast, the text in Grid and Mindmap looks like being "rendered everywhere" [P5].

However, the distinction in perceived occlusion between Linear and Grid is not as straightforward. Both utilise the same amount of glasses screen space to render the same number of ideas, and yet Linear was perceived to occlude less than Grid. Gestalt psychology [66]'s core belief in the "whole is something else than the sum of its parts" [67] hinted that humans perceive the same quantity of objects differently, depending on how they were arranged in space [68]. So, we dig deeper into fundamental oculomotor research to help make sense of this less straightforward phenomenon. For instance, studies conducted by Glenberg et al. [69] revealed that when people try to respond to difficult questions, they avert their gaze from engaging in visual inputs. The perceived occlusion increased in Mindmap as the number and structure of ideas (and lines connecting them) grew over time, Mindmap "induced stressful feeling" [P1] and prompted P2 to request hiding the display One reason could be because empty digital space accommodates the users to render the content of their short-term memory, which in our case, imagined candidate of alternative uses of the given prompt [70]. This user behaviour of 'looking at nothing' [71] facilitates the shift of attention from external stimuli to internal thoughts [72], like mind-wandering. An empirical study reported that conditions favouring mind-wandering could enhance creativity [27] by increasing unconscious associative processing [73] or by stimulating interactions between various networks of the brain (*e.g.,* default mode and central executive) [74].

4.3.3.2 If less occlusion is desirable, then why not eliminate the ideas completely?: Salvi and Bowden [72] summarised various theories linking eye movements to improved memory, attention, and also creative insights. One important implication we need to highlight is that "it is not the movement per se but the goal of the movement" [75] that has an effect on cognitive processes [76]. Despite the available feature to hide all the keywords in any layout, we were surprised to observe that the majority (12/18) of the participants did not hide the display while watching videos. They explained that hiding the display completely is not helpful for three main reasons. First, participants may forget about the idea generation task (a similar observation in Study 1), and thus need to be reminded to generate future ideas. Second, participants need to search/compare amongst previous ideas to avoid repeating them, *i.e.,* feed-forward for future ideas. Third, participants need to verify if the system is working, *i.e.,* feedback on prior ideas. For these reasons, it is important to display the right amount of information depending on these user needs and goals. In the Linear layout, participants could take the advantage of the empty space to wander their minds, while the list of keywords (on the left side of the glasses) serves as a minimal anchor to their fleeting attention. Therefore, we recommend **using Linear layout as the default/primary mode** because its least perceived occlusion both figuratively and literally makes room for creative inspiration. Since Grid is very similar to Linear in terms of performance and the absence of hierarchy, and yet perceived to occlude more, we chose to focus on Linear and **did not consider Grid layout** when developing the prototype in the next section.

4.3.3.3 Tradeoff of using Mindmap: Despite the high perceived occlusion in Mindmap, the overall ranking shows that the first choice is shared between Linear (8/18) and Mindmap (10/18) and both user ratings were similarly positive. One reason could be that participants were familiar with Linear and Mindmap layouts as they were commonly rendered using everyday pen on paper or on desktop. Another benefit is that the grouping in Mindmap helps participants to "see if an area has been well populated with ideas" [P6] and signals them to explore neighbouring topics. In other words, Mindmap facilitates analytical solving [77]: a conscious, deliberate search through a space of potential solutions. Participants also liked the fact that Mindmap enabled them to "easily glance at the structure of ideas" [P2,P9,P16] and "avoid repeating ideas" [P15]. However, depending on what the participants generated, ideas may not have a semantic connection at all, which results in meaningless and inaccurate categorisation. Therefore, we recommend **Mindmap as the optional/secondary mode of display** to prepare for scenarios where the generated ideas can be grouped meaningfully. The benefits of Mindmap may complement the 'solving by insight' [72] (or think-out-of-the-box) facilitated by the default Linear layout.

We conclude Study 2 by:

- **Answering RQ2.2**: Linear layout contributed to more novel ideas than Mindmap layout. However, there were no objective differences in other metrics (*i.e.,* fluency, flexibility, feasibility, and elaborateness) and also between any comparison with the Grid layout.
- **Answering RQ2.3**: Despite the similar subjective ratings across the three layouts, we highlighted how participants perceived the grouping by Mindmap as offering a different benefit that could supplement the Linear layout.

5 UBIIDEAS PROTOTYPE

We consolidate the insights from both Study 1 and 2 by implementing appropriate features of our proposed system (*i.e.,* Ubiideas) and the system pipeline to realise them. Please refer to the accompanying demo video for more details on the interactive prototype.

5.1 Apparatus

In addition to the Samsung phone, Nreal glasses, and Bose headphone used in Study 1 and 2, we realised the need to disambiguate between idea input and command selection modes. While reviewing the literature on appropriate techniques to facilitate interaction with the glasses, we found the suitability of using a ring mouse [78] for hand-busy everyday scenarios. As such, the user can simultaneously verbalise their ideas through the headphone and select commands through the ring mouse. We took the advantage of Samsung's native DeX station to mimic a desktop website (programmed in Javascript) experience with the visuals projected by the glasses.

5.2 Design Features

5.2.1 Flexible Level of Details

Our participants in Study 2 requested to hide the display when visually overwhelmed, and they also desired to flexibly switch between Linear and Mindmap layouts. This "no one size fits all" observation motivated the design of Ubiideas to consider multiple visual layouts to vary information density in gradual increments [79]. We envision that for relatively challenging tasks (*e.g.,* folding clothes), users could benefit from less dense information being displayed, as compared to simpler tasks (*e.g.,* free-roaming). Switching between various information densities may enable the user to rehearse 'problem representation' [80], which could significantly improve multitasking performance [53].

As illustrated in Fig. 10, we introduce a total of 7 levels (or views) that users could flexibly switch between through a single click on the ring mouse. All levels are accompanied by minimal feedback, which indicates if the system is "listening", "processing" the ideas, or "ready" to capture new ideas. The level with the least information is 'Mic Only', in which only a small start/stop button serves two functions: 1) to record ideas verbally and 2) to remind users of the ongoing idea generation task. Suppose the user would like to see more system feedback, they could switch to the next level called 'Recent', which will render only the latest idea generated by the user. For the complete enumeration of ideas, users can switch to 'Full List' which adopts a chronological order (Fig. 11a). Then, the next two levels ('Smart List' and 'Mindmap') represent two approaches to visualising semantic associations between ideas. Although both share the same information hierarchy, Smart List adopts the linear layout, while Mindmap adopts the radial layout (as introduced in Section 4). Finally, the last two levels ('Summary' and 'Transcript') offer users more context than the remaining levels by reflecting almost every word they said (in Transcript) and also the concise version of it (in Summary).

There are two optional toggles shared between some of the levels (see Fig. 11a). Firstly, users could toggle the optional 'image' switch in Recent, Full List, and Smart List levels (Fig. 11d). This design decision is based on our participants' mixed preferences for images to accompany the extracted keyword/keyphrase. Secondly, with more ideas added over time, users would need to manually scroll the linear layout adopted by Full List, Smart List, Summary, and Transcript. As recommended by seven participants in Study 2, we added the option to automate the vertical scrolling, which when activated, would always show users their last few ideas captured by the system.

5.2.2 Support Error Recovery

To mitigate the expected limitations from our NLP implementations, we enable the users to edit the captured keywords on-demand. This editing feature (see Fig. 11e) can be activated in two simple steps. Users first need to select the keyword/keyphrase using a ring mouse before verbalising the new word(s). Editing one keyword/keyphrase in one view will synchronously update those in another view as well as the corresponding images. We constrained the editing to be executable in only Recent, Full List, or Smart List views because it is challenging for users to precisely locate keywords in nodes of a mindmap and also identify mismatches among sentences (in Summary and Transcript views).

5.3 Algorithm Implementation

Now that we have revealed how Ubiideas looks from the user point of view, we will now explain the five backend stages from the system point of view: transcription, summarization, sentence clustering, keyword extraction, and finally visualisation. Firstly, we chose Mozilla Web Speech API[5] to enable the automatic conversion of speech to text directly in the frontend without having to first transmit the voice data to the backend. Although Google Speech API[6] is a powerful cloud-based alternative, it consumes more time than Web Speech API and works better for long content transcription. Secondly, we used OpenAI[7] to summarise the transcribed text because of its training model optimised by human feedback. Summarisation has the ability to not only shorten the length, but also improve the semantic details that fail to be captured by the transcription process. This is important to address the common challenges by audio conversations, which include speaker disfluencies, informal prose styles and lack of structure [81]. Thirdly, we used Meaningcloud Text clustering API[8] to cluster the summarised sentences based on their semantics and assign a descriptive label to each cluster. Instead of the basic document grouping mode, we choose the topic modelling mode implemented with the K-means algorithm. This approach helps to discover hidden themes in sentences by providing more descriptive labels than classical clustering algorithms. With the descriptive labels of each cluster, we then build a hierarchy needed for Smart List and Mindmap. Fourth, we again used OpenAI[9] to extract the keywords from the summarised transcriptions. Instead of RAKE [10] [82], Spacy [11], and TextRank [83], we chose OpenAI as it suits the context of our work. For example, only OpenAI supports extraction of both keyword and keyphrases. Our pilot explorations reveal that the keywords extracted by OpenAi were more meaningful and closer to reflect the gist of user speech. Lastly, we visualised the mindmap using an open-source framework called Highchart.js [84]. The remaining views do not require any external library as it simply adopts a linear layout. For layers with optional image toggles, we fetched the images based on each keyword/keyphrase using Google Image Search [12].

6 Overall Discussion

Results from both Study 1 and 2 revealed the benefits of speaking with glasses and rendering ideas with either a linear or mindmap layout. In this section, we motivate future works by discussing additional insights from conducting the two studies and implementing the Ubiideas prototype.

6.1 Language as a Mode of Thinking

Research in cognitive psychology on how language shapes thinking [85], and vice versa, hinted how our findings may have only scratched the surface by using only English. According to Quinteros Baumgart and Billick [86], different languages activate different cognitive brain functions.

5. https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API
6. https://cloud.google.com/speech-to-text
7. https://beta.openai.com/examples/default-summarize
8. https://www.meaningcloud.com/
9. https://beta.openai.com/examples/default-keywords
10. https://pypi.org/project/rake-nltk/
11. https://spacy.io/
12. https://github.com/jimkang/g-i-s

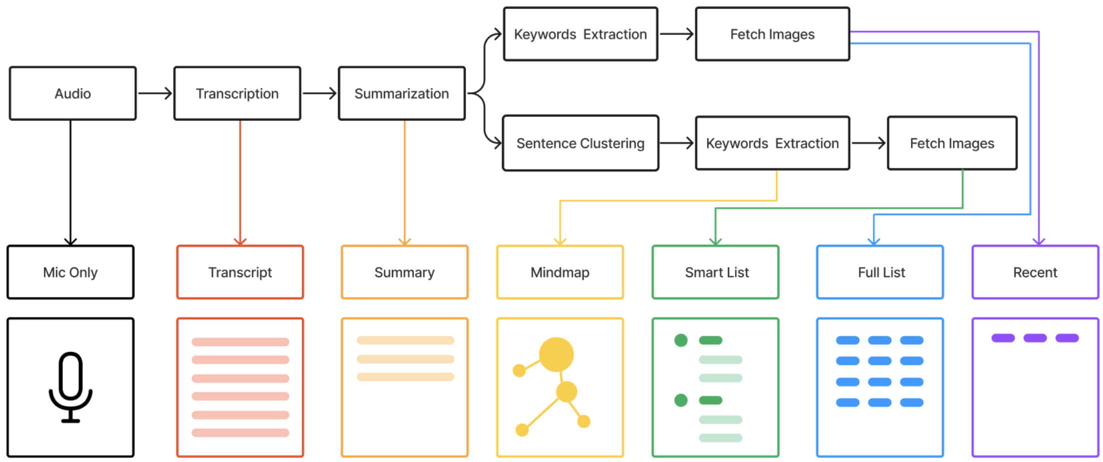

Fig. 10. Ubiideas Pipeline. Our system takes in user audio recordings and processes them through various stages to render the seven visualisations of dynamic information in real time.

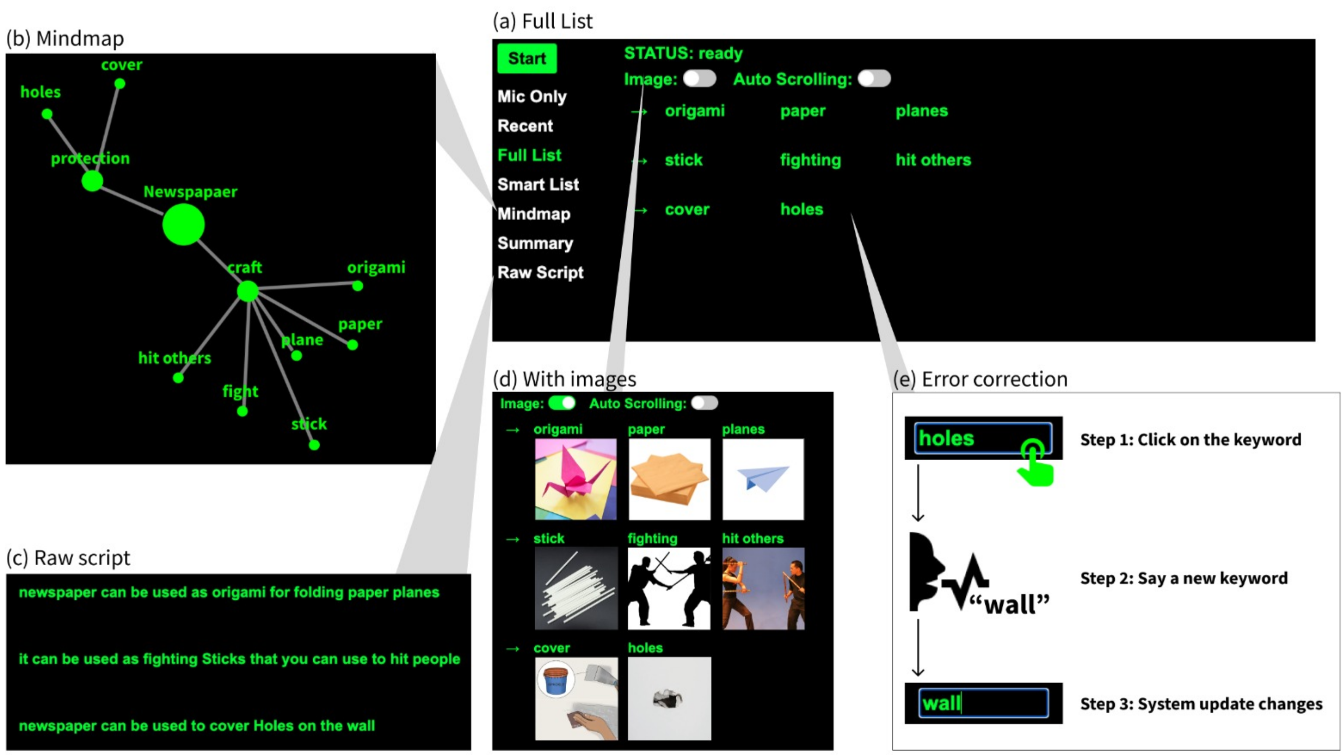

Fig. 11. The three primary levels of Ubiideas: (a) Full list, (b) Mindmap and (c) Raw script. It is optional for users to (d) toggle images and (e) correct errors.

By restricting idea expression to only one language, we are indirectly restricting the way users see the world [87] and consequently, generate ideas. In Study 1, P18 shared a similar sentiment, "sometimes I know a shorter phrase in my second language to better communicate my ideas". We hypothesise that the forbidden code-switching[13] could be why results from Study 1 shows that speaking boosts elaboration rate, instead of the expected number of ideas (*i.e.*, fluency). While future work needs to be conducted to confirm this hypothesis, it is an essential direction for NLP research to optimise language models [88] and effectively understand code-mixed sentences communicated by the user. By integrating these models, Ubiideas will not only benefit more users regardless of their mother tongues, but also capture the hidden nuances to catalyse the idea expressions.

13. Note that all our recruited participants were either native or full professional English speakers and we require them to use English only.

6.2 Beyond Divergent Thinking

While our study context is in generating ideas for Guilford's Alternative Uses tasks, we discovered how expository writing can benefit by using the Ubiideas system. Many perceive expository writing is difficult because it involves multiple cognitive processes at multiple levels [89], and sometimes in a non-linear fashion. For instance, authors often dabble between planning, generating, and revising, which could influence the next word, the next paragraph, or even the complete story. These challenges motivated recent works to divide and conquer the processes by leveraging visual diagramming [90] and generative capabilities of large language models like GPT-3 [91]. Although these contributions have made it possible for authors to collaborate with artificial intelligence in their creative and argumentative writing [92], the input modalities are limited to pen-and-paper or keyboard-typing. On the other hand, Ubiideas leverage the underutilised speaking modality to facilitate the incremental generation of ideas, and by exploring 7 views (or levels) of varying information density. As shown in Study 2, the 'Mindmap' view could offer semantic association between ideas, and thus formulate a writing structure for writers to follow. Thus, we think it is a viable next step to integrate the planning and revision processes into Ubiideas, to help users overcome writer's block and eventually write more effectively.

6.3 Prototype Limitation

The Ubiideas prototype could improve its image retrieval and hardware setup. Firstly, using Google's API to retrieve the images occasionally results in ambiguous or inaccurate reflections of the extracted keywords. As a solution, we suggest future work consider using OpenAI's Dalle2 [14], which can create more realistic images from a single description in natural language. We did not leverage Dalle2 due to its restricted public accessibility at the time of implementation. Secondly, despite choosing the lighter Nreal (106g) over HoloLens (556g), 12 out of the total 36 recruited participants (from Study 1 and 2) commented about the physical strain caused by wearing the smart-glasses, particularly the restricted field-of-view and the pressure on their ears and noses. Therefore, we hope that future hardware improvements address this issue and increase the 52-degree diagonal field-of-view to be as close as the existing reading glasses. As an extension, we plan to have a longitudinal study of our prototype to test the ecological feasibility of Ubiideas in more realistic and diverse activities of daily living. We also consider analysing users' temporal interaction patterns in the next step, such as how the users would use the built features to cope with the increasing complexity of the visualised ideas. Insights from the study could help improve the design and implementation of Ubiideas.

7 Conclusion

Creative inspiration may indeed happen anytime and anywhere and Ubiideas enabled users to generate and capture ideas more readily while multitasking. In Study 1, we learned that one benefit of speaking is on the elaborateness (and not fluency) of ideas, and that glasses performed similarly well as phones despite the former being less familiar among users. In Study 2, we learned that linearly listing ideas contributed to more novel ideas than using a mindmap, and thus adopted as the default layout in Ubiideas. However, participants perceived the semantic grouping in a mindmap as useful when the generated ideas have logical connections. Therefore, we integrated mindmap in Ubiideas as an optional layout to view the same information from a different perspective. To summarise, we not only revealed tradeoffs between input modalities and between visual layouts when ideas were rendered on the glasses, but also demonstrated how Ubiideas can be further enhanced and implemented. Our two studies have only taken the initial steps in exploring and understanding the design space for Ubiideas. There is so much more the world can offer in terms of ideas and insights, if only we start looking up more, and take advantage of, instead of hindered by, our daily activities.

14. https://openai.com/dall-e-2/

References

[1] D. H. Henard and M. A. McFadyen, "Making knowledge workers more creative," *Research-Technology Management*, vol. 51, no. 2, pp. 40–46, 2008.
[2] N. Anderson, K. Potočnik, and J. Zhou, "Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework," *Journal of management*, vol. 40, no. 5, pp. 1297–1333, 2014.
[3] J. Wineman, J. Peponis, and R. Dalton, "Exploring, engaging, understanding in museums," 2006.
[4] L. M. Straker, P. B. O'Sullivan, A. Smith, and M. Perry, "Computer use and habitual spinal posture in australian adolescents," *Public health reports*, vol. 122, no. 5, pp. 634–643, 2007.
[5] G. Wallas, *The art of thought*. Harcourt, Brace, 1926, vol. 10.
[6] M. Kuman, "Nonlinear mathematical modeling of creativity," 2019.
[7] M. C. Corballis, "Left brain, right brain: facts and fantasies," *PLoS biology*, vol. 12, no. 1, p. e1001767, 2014.
[8] K. Sawyer, "The cognitive neuroscience of creativity: A critical review," *Creativity research journal*, vol. 23, no. 2, pp. 137–154, 2011.
[9] S. C. Yang and W. C. Lin, "The relationship among creative, critical thinking and thinking styles in taiwan high school students." *Journal of instructional Psychology*, vol. 31, no. 1, 2004.
[10] S. W. Russ, *Affect and creativity: The role of affect and play in the creative process*. Routledge, 2013.
[11] S. Mednick, "The associative basis of the creative process." *Psychological review*, vol. 69, no. 3, p. 220, 1962.
[12] R. J. Sternberg and J. E. Davidson, *The nature of insight.* The MIT Press, 1995.
[13] W. Carpenter, "The aha! moment: The science behind creative insights," in *Toward Super-Creativity-Improving Creativity in Humans, Machines, and Human-Machine Collaborations*. IntechOpen, 2019.
[14] M. D. Mumford and D. L. Whetzel, "Insight, creativity, and cognition: On sternberg and davidson's the nature of insight," *Creativity Research Journal*, vol. 9, no. 1, pp. 103–107, 1996.
[15] C. M. Seifert, D. E. Meyer, N. Davidson, A. L. Patalano, and I. Yaniv, "Demystification of cognitive insight:

Opportunistic assimilation and the prepared-mind hypothesis," 1994.

[16] M. I. Posner, "Cognition: An introduction." 1973.

[17] S. Ohly, S. Sonnentag, and F. Pluntke, "Routinization, work characteristics and their relationships with creative and proactive behaviors," *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 27, no. 3, pp. 257–279, 2006.

[18] R. Kanfer and P. L. Ackerman, "Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition." *Journal of applied psychology*, vol. 74, no. 4, p. 657, 1989.

[19] D. F. Ferrer and F. Nietzsche, "Twilight of the idols or how to philosophize with a hammer," 2013.

[20] A. M. Dollar, "Classifying human hand use and the activities of daily living," in *The Human Hand as an Inspiration for Robot Hand Development*. Springer, 2014, pp. 201–216.

[21] L. A. Ovington, A. J. Saliba, C. C. Moran, J. Goldring, and J. B. MacDonald, "Do people really have insights in the shower? the when, where and who of the aha! moment," *The Journal of Creative Behavior*, vol. 52, no. 1, pp. 21–34, 2018.

[22] S. Kaplan, "The restorative benefits of nature: Toward an integrative framework," *Journal of environmental psychology*, vol. 15, no. 3, pp. 169–182, 1995.

[23] H. Ohly, M. P. White, B. W. Wheeler, A. Bethel, O. C. Ukoumunne, V. Nikolaou, and R. Garside, "Attention restoration theory: A systematic review of the attention restoration potential of exposure to natural environments," *Journal of Toxicology and Environmental Health, Part B*, vol. 19, no. 7, pp. 305–343, 2016.

[24] M. Oppezzo and D. L. Schwartz, "Give your ideas some legs: the positive effect of walking on creative thinking." *Journal of experimental psychology: learning, memory, and cognition*, vol. 40, no. 4, p. 1142, 2014.

[25] K. Christoff, Z. C. Irving, K. C. Fox, R. N. Spreng, and J. R. Andrews-Hanna, "Mind-wandering as spontaneous thought: a dynamic framework," *Nature Reviews Neuroscience*, vol. 17, no. 11, pp. 718–731, 2016.

[26] T. Tan, H. Zou, C. Chen, and J. Luo, "Mind wandering and the incubation effect in insight problem solving," *Creativity Research Journal*, vol. 27, no. 4, pp. 375–382, 2015.

[27] B. Baird, J. Smallwood, M. D. Mrazek, J. W. Kam, M. S. Franklin, and J. W. Schooler, "Inspired by distraction: Mind wandering facilitates creative incubation," *Psychological science*, vol. 23, no. 10, pp. 1117–1122, 2012.

[28] N. Vaid, "Waiting for that big idea!" *APOS Trends in Orthodontics*, vol. 6, no. 2, pp. 55–55, 2016.

[29] H.-C. Wang, D. Cosley, and S. R. Fussell, "Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 103–106. [Online]. Available: https://doi.org/10.1145/1718918.1718938

[30] S. Andolina, K. Klouche, D. Cabral, T. Ruotsalo, and G. Jacucci, "Inspirationwall: Supporting idea generation through automatic information exploration," in *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, ser. C&C '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 103–106. [Online]. Available: https://doi.org/10.1145/2757226.2757252

[31] Y. Shi, Y. Wang, Y. Qi, J. Chen, X. Xu, and K.-L. Ma, "Ideawall: Improving creative collaboration through combinatorial visual stimuli," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 594–603. [Online]. Available: https://doi.org/10.1145/2998181.2998208

[32] B. A. Nijstad and W. Stroebe, "How the group affects the mind: A cognitive model of idea generation in groups," *Personality and social psychology review*, vol. 10, no. 3, pp. 186–213, 2006.

[33] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[34] R. Rzayev, P. W. Woźniak, T. Dingler, and N. Henze, "Reading on smart glasses: The effect of text position, presentation type and walking," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–9. [Online]. Available: https://doi.org/10.1145/3173574.3173619

[35] D. Ghosh, P. S. Foong, S. Zhao, C. Liu, N. Janaka, and V. Erusu, "Eyeditor: Towards on-the-go heads-up text editing using voice and manual input," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[36] A. Ram and S. Zhao, "Lsvp: Towards effective on-the-go video learning using optical head-mounted displays," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 1, mar 2021. [Online]. Available: https://doi.org/10.1145/3448118

[37] Z. Rashid, R. Pous, J. Melià-Seguí, and E. Peig, "Cricking: Browsing physical space with smart glass," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 151–154.

[38] A. Aiordachioae and R.-D. Vatavu, "Life-tags: a smartglasses-based system for recording and abstracting life with tag clouds," *Proceedings of the ACM on human-computer interaction*, vol. 3, no. EICS, pp. 1–22, 2019.

[39] J. B. Newman and M. W. Horowitz, "Writing and speaking," *College Composition and Communication*, vol. 16, no. 3, pp. 160–164, 1965. [Online]. Available: http://www.jstor.org/stable/355730

[40] K. VanLehn, "Toward a theory of impasse-driven learning," in *Learning issues for intelligent tutoring systems*. Springer, 1988, pp. 19–41.

[41] J. P. Guilford, "Basic conceptual problems in the psychology of thinking," *Annals of the New York Academy of Sciences*, vol. 91, no. 1, pp. 6–21, 1960. [Online]. Available: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749

6632.1961.tb50913.x

[42] S. Lewis, M. Dontcheva, and E. Gerber, "Affective computational priming and creativity," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 735–744.

[43] M. A. Runco, *Divergent thinking*. Ablex Publishing Corporation Norwood, NJ, 1991.

[44] S. Di Bartolomeo, A. Pandey, A. Leventidis, D. Saffo, U. H. Syeda, E. Carstensdottir, M. Seif El-Nasr, M. A. Borkin, and C. Dunne, "Evaluating the effect of timeline shape on visualization task performance," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: https://doi.org/10.1145/3313831.3376237

[45] M. A. Runco, S. M. Okuda, and B. J. Thurston, "The psychometric properties of four systems for scoring divergent thinking tests," *Journal of Psychoeducational Assessment*, vol. 5, no. 2, pp. 149–156, 1987.

[46] J. P. Guilford, "Creativity: Yesterday, today and tomorrow," *The Journal of Creative Behavior*, vol. 1, no. 1, pp. 3–14, 1967.

[47] E. Galy, J. Paxion, and C. Berthelon, "Measuring mental workload with the nasa-tlx needs to examine each dimension rather than relying on the global score: an example with driving," *Ergonomics*, vol. 61, no. 4, pp. 517–527, 2018, pMID: 28817353. [Online]. Available: https://doi.org/10.1080/00140139.2017.1369583

[48] C. Dippo and B. Kudrowitz, "Evaluating the alternative uses test of creativity," *2013 NCUR*, 2013.

[49] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," ser. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 143–146. [Online]. Available: https://doi.org/10.1145/1978942.1978963

[50] M. Csikszentmihalyi, "Flow and the psychology of discovery and invention," *HarperPerennial, New York*, vol. 39, 1997.

[51] G. Calic, E. Mosakowski, N. Bontis, and S. Helie, "Is maximising creativity good? the importance of elaboration and internal confidence in producing creative ideas," *Knowledge Management Research & Practice*, vol. 0, no. 0, pp. 1–16, 2020. [Online]. Available: https://doi.org/10.1080/14778238.2020.1730718

[52] D. J. Hansen, G. T. Lumpkin, and G. E. Hills, "A multidimensional examination of a creativity-based opportunity recognition model," *International Journal of Entrepreneurial Behavior & Research*, 2011.

[53] D. D. Salvucci, N. A. Taatgen, and J. P. Borst, "Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 1819–1828. [Online]. Available: https://doi.org/10.1145/1518701.1518981

[54] A. J. Leeds, B. Kudrowitz, and J. Kwon, "Mapping associations: exploring divergent thinking through mind mapping," *International Journal of Design Creativity and Innovation*, vol. 7, no. 1-2, pp. 16–29, 2019. [Online]. Available: https://doi.org/10.1080/21650349.2018.1463178

[55] S. Dixon, E. Fitzhugh, and D. Aleva, "Human factors guidelines for applications of 3D perspectives: a literature review," in *Display Technologies and Applications for Defense, Security, and Avionics III*, J. T. Thomas and D. D. Desjardins, Eds., vol. 7327, International Society for Optics and Photonics. SPIE, 2009, p. 73270K. [Online]. Available: https://doi.org/10.1117/12.820853

[56] C. D. Wickens, "The when and how of using 2-d and 3-d displays for operational tasks," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 21. SAGE Publications Sage CA: Los Angeles, CA, 2000, pp. 3–403.

[57] S. M. John, H. S. Smallman, and M. B. Cowen, "Designing for the task: Sometimes 2-d is just plane better," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 21, pp. 3–407–3–410, 2000. [Online]. Available: https://doi.org/10.1177/154193120004402108

[58] C.-Y. Hsiao, M.-J. Wang, Y.-T. Lu, and C.-C. Chang, "Usability evaluation of monocular optical head-mounted displays on reading tasks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2018.

[59] N. P. Gallavan, *Developing performance-based assessments, grades K-5*. Corwin Press, 2009.

[60] L. Chen and H. Tsoi, "Users' decision behavior in recommender interfaces: Impact of layout design," *CEUR Workshop Proceedings*, vol. 811, 01 2011.

[61] M. Hearst, E. Pedersen, L. Patil, E. Lee, P. Laskowski, and S. Franconeri, "An evaluation of semantically grouped word cloud designs," *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, pp. 1–1, 03 2019.

[62] S. Chandrasegaran, C. Bryan, H. Shidara, T.-Y. Chuang, and K.-L. Ma, "Talktraces: Real-time capture and visualization of verbal content in meetings," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–14. [Online]. Available: https://doi.org/10.1145/3290605.3300807

[63] V. Kokotovich, "Problem analysis and thinking tools: an empirical study of non-hierarchical mind mapping," *Design Studies*, vol. 29, no. 1, pp. 49–69, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142694X07

[64] T. Dingler, D. Kern, K. Angerbauer, and A. Schmidt, "Text priming - effects of text visualizations on readers prior to reading," in *Human-Computer Interaction – INTERACT 2017*, R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, and M. Winckler, Eds. Cham: Springer International Publishing, 2017, pp. 345–365.

[65] E. Wästlund, T. Norlander, and T. Archer, "The effect of page layout on mental workload: A dual-task experiment," *Computers in Human Behavior*, vol. 24, no. 3, pp. 1229–1245, 2008, instructional Support for Enhancing Students' Information

Problem Solving Ability. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563210001891

[66] M. Wertheimer, "Experimentelle studien uber das sehen von bewegung," *Zeitschrift fur psychologie*, vol. 61, 1912.

[67] K. Koffka, *Principles of Gestalt psychology*. Routledge, 2013.

[68] F. Jäkel, M. Singh, F. A. Wichmann, and M. H. Herzog, "An overview of quantitative approaches in gestalt perception," *Vision research*, vol. 126, pp. 3–8, 2016.

[69] A. M. Glenberg, J. L. Schroeder, and D. A. Robertson, "Averting the gaze disengages the environment and facilitates remembering," *Memory & cognition*, vol. 26, no. 4, pp. 651–658, 1998.

[70] S. A. Brandt and L. W. Stark, "Spontaneous eye movements during visual imagery reflect the content of the visual scene," *Journal of cognitive neuroscience*, vol. 9, no. 1, pp. 27–38, 1997.

[71] F. Ferreira, J. Apel, and J. M. Henderson, "Taking a new look at looking at nothing," *Trends in cognitive sciences*, vol. 12, no. 11, pp. 405–410, 2008.

[72] C. Salvi and E. M. Bowden, "Looking for creativity: Where do we look when we look for new ideas?" *Frontiers in psychology*, vol. 7, p. 161, 2016.

[73] I. Yaniv and D. E. Meyer, "Activation and metacognition of inaccessible stored information: potential bases for incubation effects in problem solving." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, no. 2, p. 187, 1987.

[74] K. Christoff, A. M. Gordon, J. Smallwood, R. Smith, and J. W. Schooler, "Experience sampling during fmri reveals default network and executive system contributions to mind wandering," *Proceedings of the National Academy of Sciences*, vol. 106, no. 21, pp. 8719–8724, 2009.

[75] K. Werner and M. Raab, "Moving your eyes to solution: effects of movements on the perception of a problem-solving task," *Quarterly Journal of Experimental Psychology*, vol. 67, no. 8, pp. 1571–1578, 2014.

[76] M. Raab and N. Green, "Motion as input: A functional explanation of movement effects on cognitive processes," *Perceptual and Motor Skills*, vol. 100, no. 2, pp. 333–348, 2005.

[77] C. Salvi, E. Bricolo, S. L. Franconeri, J. Kounios, and M. Beeman, "Sudden insight is associated with shutting out visual inputs," *Psychonomic bulletin & review*, vol. 22, no. 6, pp. 1814–1819, 2015.

[78] S. Sapkota, A. Ram, and S. Zhao, "Ubiquitous interactions for heads-up computing: Understanding users' preferences for subtle interaction techniques in everyday settings," in *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, 2021, pp. 1–15.

[79] D. Lindlbauer, A. M. Feit, and O. Hilliges, "Context-aware online adaptation of mixed reality interfaces," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 147–160. [Online]. Available: https://doi.org/10.1145/3332165.3347945

[80] J. R. Anderson and C. Lebiere, "The atomic components of thought." 1998.

[81] T. Chen, A. Tung, and L. B. Chilton, "Hierarchical summarization for longform spoken dialog," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 582–597. [Online]. Available: https://doi.org/10.1145/3472749.3474771

[82] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic keyword extraction from individual documents," *Text mining: applications and theory*, vol. 1, no. 1-20, pp. 10–1002, 2010.

[83] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, pp. 404–411.

[84] C. Walters, "Interactive javascript charts library," 2022. [Online]. Available: https://www.highcharts.com/

[85] L. Boroditsky, "How language shapes thought," *Scientific American*, vol. 304, no. 2, pp. 62–65, 2011. [Online]. Available: http://www.jstor.org/stable/26002395

[86] C. Quinteros Baumgart and S. B. Billick, "Positive cognitive effects of bilingualism and multilingualism on cerebral function: A review," *Psychiatric Quarterly*, vol. 89, no. 2, pp. 273–283, 2018.

[87] A. V. Kharkhurin, "Multilingualism and creativity," in *Multilingualism and Creativity*. Multilingual Matters, 2012.

[88] K. Li, J. Li, G. Ye, R. Zhao, and Y. Gong, "Towards code-switching asr for end-to-end ctc models," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 6076–6080.

[89] L. Flower and J. R. Hayes, "A cognitive process theory of writing," *College Composition and Communication*, vol. 32, no. 4, pp. 365–387, 1981. [Online]. Available: http://www.jstor.org/stable/356600

[90] Z. Lu, M. Fan, Y. Wang, J. Zhao, M. Annett, and D. Wigdor, "Inkplanner: Supporting prewriting via intelligent visual diagramming," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 277–287, 2019.

[91] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[92] M. Lee, P. Liang, and Q. Yang, "Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3491102.3502030

Study Design

Both studies recruited a different set of 18 participants but shared the same procedure and order of conditions and topics. The conditions of Study 1 are Phone+Typing, Phone+Speaking, and Glasses+Speaking. The conditions of Study 2 are Linear, Grid, and Mindmap.

P1	Condi1 Topic1	Condi2 Topic2	Condi3 Topic3		P4	Condi1 Topic2	Condi2 Topic3	Condi3 Topic1		P7	Condi1 Topic3	Condi2 Topic1	Condi3 Topic2
P2	Condi3 Topic3	Condi1 Topic1	Condi2 Topic2		P5	Condi3 Topic1	Condi1 Topic2	Condi2 Topic3		P8	Condi3 Topic2	Condi1 Topic3	Condi2 Topic1
P3	Condi2 Topic2	Condi3 Topic3	Condi1 Topic1		P6	Condi2 Topic3	Condi3 Topic1	Condi1 Topic2		P9	Condi2 Topic1	Condi3 Topic2	Condi1 Topic3
P10	Condi1 Topic1	Condi2 Topic2	Condi3 Topic3		P13	Condi1 Topic2	Condi2 Topic3	Condi3 Topic1		P16	Condi1 Topic3	Condi2 Topic1	Condi3 Topic2
P11	Condi3 Topic3	Condi1 Topic1	Condi2 Topic2		P14	Condi3 Topic1	Condi1 Topic2	Condi2 Topic3		P17	Condi3 Topic2	Condi1 Topic3	Condi2 Topic1
P12	Condi2 Topic2	Condi3 Topic3	Condi1 Topic1		P15	Condi2 Topic3	Condi3 Topic1	Condi1 Topic2		P18	Condi2 Topic1	Condi3 Topic2	Condi1 Topic3

List of videos used (each clipped to feature only 2 minutes) in both studies

- Thor 4: Love and Thunder (trailer)
- Minions: The Rise of Gru (trailer)
- Jurassic World Dominion (trailer)

An example prompt of "brick" illustrated how we conducted the Phone+Speaking and Glasses+Speaking conditions

User-Spoken Words	**Experimenter**-Generated Keywords
A paper weight because paper often fly away and brick is heavy	• Paper weight ○ Fly away ○ Brick is heavy
A doorstop because the wind causes slammed door accidentally and it is annoying	• Door stopper ○ Accidental wind ○ Annoying
Since it is rectangular and has holes, perhaps use it as a mock coffin at a Barbie funeral. The holes can be used to display the head.	• Mock coffin ○ Rectangles with holes ○ Barbie
Or we could also throw the brick through a window, for example during a fire emergency or escape from a kidnapper.	• Throw to window ○ Fire emergency ○ Escape kidnapper
I might use the brick as a weapon, you know throw it to a robber's head, so i can defend myself	• Self-defense ○ Robber

Subjective Ratings [7-Likert Rating]

- I find the assigned condition supports me to multitask easily
- I find the assigned condition offers me low physical demand while multitasking
- I find the assigned condition offers me low mental demand while multitasking
- [primary task] I could complete the daily activities easily in the assigned condition
- [secondary task] I generate high quantity of ideas while multitasking in the assigned condition
- [secondary task] I generate high quality of ideas while multitasking in the assigned condition

Study 1 Results

	Input Modality		
Objective Measures	Phone + Typing	Phone + Speaking	Glasses + Speaking
Fluency Score	13.44 ± 3.18	13.67 ± 3.67	15.17 ± 2.60
Novelty Score	9.62 ± 2.57	10.07 ± 3.10	11.31 ± 2.28
Flexibility Score	5.23 ± 1.43	6.00 ± 1.94	6.14 ± 1.46
Feasibility Score	12.24 ± 3.10	11.99 ± 3.00	13.17± 2.37
Elaborateness Score	4.07 ± 1.04☆✣	6.44 ± 2.45✣	7.37 ± 1.71☆
Time Increase %	23.7± 9.1 %	16.2 ± 5.8 %	20.1 ± 6.2 %

Objective measures reported in Study 1 across both scenarios.

☆ and ✣ represent significant ($p < 0.05$) post-hoc tests.

Study 2 Results

	Visual Layout		
Objective Measures	Linear	Grid	Mindmap
Fluency Score	19.83 ± 4.37	18.11 ± 3.74	16.83 ± 3.71
Novelty Score	14.55 ± 3.78☆	12.47 ± 2.79	11.69 ± 2.89☆
Flexibility Score	7.49 ± 2.29	6.57 ± 1.62	5.86 ± 1.47
Feasibility Score	17.23 ± 3.42	16.31 ± 3.62	15.49 ± 3.38
Elaborateness Score	7.91 ± 1.86	7.47 ± 2.24	6.91 ± 1.86
Time Increase %	21.3 ± 7.8 %	21.7 ± 5.7 %	16.7 ± 6.5 %

Objective measures reported in Study 2 across both scenarios.

☆ represents significant ($p < 0.05$) post-hoc tests.