

The Effects of an Audio-Based Modality on Social Media Experience

Group 5: Addison Bright, Jinwoo Oh, Michelle Morris,
Enoch Akli, Justin Kerobo

ABSTRACT: Continuous use of social media platforms like Facebook or Instagram has been shown to have negative effects on user's mental health [1]. This is primarily through the continuous comparisons the users make to their peers' appearance and perceived status [1]. Physical appearance and perceived economic status are expressed mainly through visual media. On the other hand, activities, milestones, and announcements can be expressed both visually and auditorily.

The goal of this research project was to develop an audio-based social media platform. Our research project aimed to address the following questions: (1) Does using an audio-based social media platform lead to less anxiety than using a visual-based platform? (2) Is an audio-based social media platform as satisfying to use as a visual-based platform?

To answer these questions, we recruited 12 Virginia Tech students and performed a within-subject study comparing users' experiences while interacting with an audio-based platform, a visual-based platform, and a hybrid (both audio and visual) platform. ANOVA and Tukey's statistical tests were used to assess the differences in anxiety and other mood characteristics across all three modalities. Overall, there was no significant difference in anxiety levels across different social media modalities. However, on average, participants preferred a visual social media modality over an audio or hybrid modality.

Background

Visual-based social media platforms like Facebook negatively affect mental health since the users are constantly comparing themselves to their peers [1]. This has been shown to lead to increased levels of anxiety and body-image issues among younger adults and adolescents [1] [2]. Additionally, visual processing is associated with increased levels of anxiety in healthy populations, and this association is more severe for populations with body dysmorphic disorder (BDD) [3]. On the other hand, the use of audio-based media, such as audiobooks and podcasts, has shown promising effects on lowering anxiety levels [4].

The extensive use of audio-based media shows promise that it could also be used for social media content, which often encompasses entertainment and information delivery. Many college students have found podcasts fun and entertaining and enjoyed sharing this audio-based content with their peers [5]. Additionally, audio-based content has been used to deliver the news, which allows users to integrate audio news into their daily activities, making the news more accessible and easier to fit into a busy lifestyle compared to reading or watching the news

[6]. These examples prove that audio-based media at the minimum can be a good substitute for visual-based media, and sometimes it is even the preferred modality when compared to visual media. Furthermore, audio-based media can be adaptable to keep an engaged audience. When television became widely available, radio talk shows adopted niche-casting, using specific music formats to draw in listeners [7, 8]. Today, radio stations are embracing the digital era and beginning to integrate their content with social media [9]. Therefore, audio-based media has potential to be adaptable and accommodate the new age of social media.

In the age of social media, users can access unlimited information with very little regulation. This can be problematic, especially among young adults, who are still in their development stages and are sensitive to "fitting in" among their peers [1]. Through visual-based comparisons, users can judge themselves based on their body type, skin color, appearance, and perceived wealth [1][2]. Here, artificial intelligence (AI) can be a solution to this problem. AI already has a hand in personalized teaching and learning from intelligent tutoring systems, online content monitoring, and filtering algorithms that proactively identify and stop harmful content or contexts. [11]. Besides those aspects, their use in storytelling and creating personal experiences will allow

for a more inclusive and equitable system for those who can access it. Storytelling has been shown to increase the levels of oxytocin in the brain, which leads to a decrease of anxiety and creates a calming effect [10]. Because of these effects, using AI to weave a personalized and factual story from social media, while filtering out harmful content, could potentially decrease the levels of anxiety that users typically experience while interacting with visual-based social media.

Therefore, we will design an audio-based social media application that uses artificial intelligence (AI) to analyze visual social media feeds and create a personal auditory summary. Since emotional attachment from visual stimuli can be similarly induced from auditory stimuli [12], we will investigate if our audio-based social media app can provide the user with a satisfying and immersive social media experience while limiting anxiety. Young adults, especially those with BDD, can benefit from using audio-based social media by limiting the opportunity to make body- or status-focused comparisons while still reaping the gains of social connection.

To our knowledge, there has not yet been a comparison between an audio-based social media platform and a traditional, visual-based social media platform. Our project findings will contribute to future research by (a) comparing between visual- and audio-based social media, (b) determining how audio-based social media impacts user experience, including associated anxiety with social media use, and (c) how content filtering and storytelling generated from AI can impact user experience with social media.

Task Analysis

To devise a new method of social media to interpret visual information and convert it to an auditory modality, we chose an existing social media platform to base our task analysis on. We chose Instagram because visual information is the most important, as there are no posts purely based on text, such as with Facebook or Twitter. At the same time, it is easy to see updates from friends and family on Instagram, which is the target function of our proposed audio interface.

There are many features of Instagram including, the functions for receiving and communicating with other people include stories, feeds, IGTV, and reels. Among them, the auto-generated Instagram feed is the oldest and one of the most widely used functions. Therefore, for our task analysis, we chose scrolling through Instagram feed, which includes opening the app, scrolling through pictures, interacting with a post, and sharing posts. In our task analysis, we focused on a user scrolling through social media. The entire task analysis can be viewed in Appendix 1. For our design focus, we chose the user seeing auto-populated content from friends and family (Figure 1), as we believed this step could be most improved upon using an audio-based interface. Potential design implications and areas for improvement that were identified with our Task Analysis are included in Table 1.

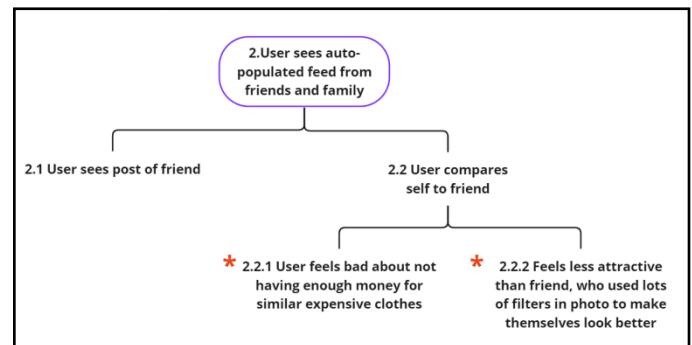


Figure 1: Excerpt from Task Analysis

Design Implications

Table 1: Design considerations derived from task analysis

Task Analysis Element to Improve	Plan for Improvement
<p>1. User feels bad about not having enough money for similar expensive clothes in Task 2.2.1</p>	<ul style="list-style-type: none"> • Change sensory modality from visual to auditory to limit emotional distress and anxiety from visual stimuli that portray income disparities. • We will train the AI algorithm used for audio-based social media to remove descriptive language concerning income, class, or perceived socioeconomic status
<p>2. User feels less attractive than friend, who used lots of filters in photo to make themselves look better in Task 2.2.2</p>	<ul style="list-style-type: none"> • Use AI algorithms to identify and describe key elements of an image without emphasizing potential emotional triggers. For instance, "A group of people gathered outside a building" rather than "A group of people partying outside a lavish mansion." • Train the AI to omit or neutralize descriptions of luxury items, brands, or other indicators of wealth which might trigger negative effects.
<p>3. User types of content they wish to comment on in Task 3.2.1.1. Tone of comment can be misinterpreted and lead to negative feelings</p>	<ul style="list-style-type: none"> • Providing an audio recording of comment would remove ambiguity of tone • We will design an option to respond to social media feed using voice recording so that comments feel more personalized and can be interpreted without tone confusion.

Methods

A convenience sample of 12 students at Virginia Tech (4 Males, 8 Females; Age = 23.3 ± 2.6) were recruited for this study. All participants reported normal vision and hearing and gave their informed consent prior to participating. This experiment was a within-subject observational study to characterize user experience when interacting with different modalities of social media. For our independent variable, we manipulated a social media experience into three distinct modalities: visual-

only, audio-only, and audio-visual. For our dependent variables, we characterized the participants' experience using each social media modality by obtaining subjective measures of stress, anxiety, and satisfaction. We also measured usage time to obtain objective measurements that relate to usage experience and stress.

Participants were exposed to a novel social media modality that was based solely on audio. This audio modality used "Wizard of Oz" technology to mimic an AI-curated audio rendition of social media feed. In reality, the

audio rendition of the social media feed was written and pre-recorded by the experimenters and played to the participants using a laptop speaker. The experimenters chose to use a human-like audio rendition of social media feed as opposed to text-to-speech, in the effort to increase the personalized experience of audio-based social media.

Each participant was exposed to 3 social media modalities: (A) visual-only, (B) audio-only, and (C) hybrid (visual and audio). Appendix 1 shows screenshots of the visual components of the social media interface and provides transcriptions of the audio-based components of the interface. For condition (A), participants scrolled through a Microsoft PowerPoint presentation using a laptop. Participants were exposed to 10 social media posts per condition and were instructed to “scroll through the social media posts at their self-selected pace.” For Condition (B), participants again scrolled through a Microsoft PowerPoint presentation using a laptop. However, the slides of the PowerPoint were left intentionally blank, and an audio file of the given social media post played at the onset of slide transition. For Condition (C), participants again scrolled through a Microsoft PowerPoint presentation using a laptop similar to Condition (A). However, this time the participant was simultaneously exposed to an audio-summary of the post, played over the laptop speakers as in Condition (B).

This study was approved by the Virginia Tech Institutional Review Board (IRB #23-1144).

Results

All data from all participants were included, as no outliers were observed. A 4-point Likert scale was employed to explore participants perceived emotional stress, as measured by the State Trait Anxiety Inventory (Appendix 3). Additionally, overall satisfaction with each type of social

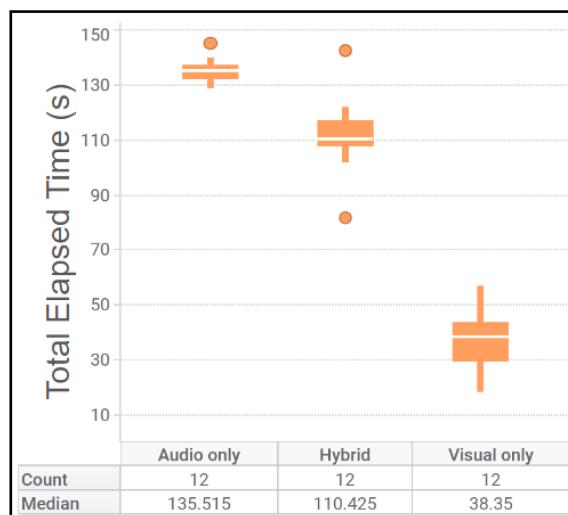
media was qualitatively assessed. The time taken to view all 10 social media posts was also recorded for each condition.

Time elapsed. Time elapsed from start of experiencing feeds to the end are presented on Table 2 and Figure 2. It was observed that elapsed times have the following ranking in ascending order: Visual only, audio only, and hybrid. Results of ANOVA showed there is a significant difference between at least one pair of stimuli, so Tukey’s HSD test was used to examine differences and significance between each set of stimuli.

Table 2: Mean, standard deviation, and ANOVA values for time elapsed.

	Visual only	Audio only	Hybrid	ANOVA p-value
Mean (s)	37.31	135.49	113.77	<0.0001
Stdev (s)	12.58	4.55	16.72	

Figure 2: Total Elapsed Time Across all Three Conditions



ANOVA table is presented on Table 3, and the result from Tukey's HSD test is presented on Table 4, and it is shown that there is a significant difference in watching time between all stimuli

Table 3. ANOVA table on elapsed time under conditions

Source	DF	F Ratio	Prob > F
Condition	2	208.8	<0.001
Error	33		
Total	35		

Table 4: Tukey's HSD test result for time elapsed

Level	- Level	Diff . (s)	Std Dr	p-Value
Audio	Visual	98.2	5.047	<0.001
Hybrid	Visual	76.5	5.047	<0.001

The results of Tukey's HSD test indicate statistically significant differences between visual-only and audio-only conditions. Consequently, it can be concluded that the participants express a greater sense of comfort with the visual-only version of social media.

Table 5: One-way ANOVA result for "How easy was it to navigate and use?"

Source	DF	SSQ	MSQ	F Ratio	Prob > F
Condition	2	5.2	2.58	7.20	0.003
Error	33	11.8	0.36		
Total	35	17.0			

Audio	Hybrid	21.7	5.047	0.0004
-------	--------	------	-------	--------

State Trait Anxiety Inventory. No statistically significant differences were reported regarding anxiety index.

Overall experience. After all types of social media were presented, a post-questionnaire for assessing the overall experience was conducted. Results from ANOVA showed a significant p-value in 'How easy was it to navigate and use?'. The ANOVA result is presented in Table 5, and Tukey's HSD test was performed for this question, with the results presented in Table 6. Overall, participants indicated that they found it easy, less mentally demanding, and less engaging, feeling the best overall experience with visual-only social media.

Level	Level	Diff for questionnaire 'How easy was it to navigate and use?' Value (1=Not at all, 4=Very Much So)	Std Dr	p-Value
Visual	Audio	0.917	0.244	0.0019
Visual	Hybrid	0.583	0.244	0.058
Hybrid	Audio	0.333	0.244	0.3713

Discussion

This study was designed to test the efficacy and overall user experience of an audio-based social media platform. It was hypothesized that users would appreciate an audio-only social media because of certain negative impacts of visual social media derived from comparison about appearance and perceived status [1].

Theoretically, the same messages could be conveyed using auditory stimuli while removing the negative visual stimuli [5]. To do this, a series of three different social media cases was employed, a visual only, an auditory-only, and a hybrid condition. Within these conditions, five “themes” were used to “Wizard of OZ” a curated AI social media feed. These themes included: friends, family, relationship, travel, and Body-focused (Male and Female) and were chosen as they mimic what one would experience in their own personal social media feed. Similarly, these are themes that can induce an increased level of negative thoughts or feelings as discussed above [1],[2]. Data was collected for 12 participants to assess user’s experience with the different social media conditions.

The data for length time spent in each condition shows a familiar trend with regards to how fast information can be conveyed to individuals. Participants spent the least amount of time on the visual-only social media suggesting that they view each post long enough to get the general idea or gist, and then move on to the next rather quickly. In contrast, the audio-only condition almost forced participants to stay for the duration of the audio to understand what the post was about. In the hybrid condition, the individual spent less time than the audio-only condition and more time than the visual condition. Given that the length of the audio clips was, on average, the same across posts in each condition, these results suggest that in the hybrid condition, participants advanced through the slide before finishing listening to the audio as they were able to retrieve relevant

information from the visual stimuli, so they advanced through the slides quicker. This data makes sense when compared to other studies evaluating information processing of visual and auditory stimuli [13].

The data for the State Trait Anxiety Inventory suggests that overall, none of our social media conditions had a negative impact on participant’s general mood at the moment. This could contribute in part to the fact that other hypotheses were not supported in the other data as well. For example, one hypothesis was that people would prefer the auditory-only condition to not experience “negative” visual stimuli [1]-[4]. As such, because none of the conditions induced any negative feelings as collected from the state trait anxiety scale, there was no effect of removing the visual component as expected. Additionally, given that participants preferred the visual-only condition the most, the overall hypothesis was not supported. This could be since an audio-only social media is a novel tool and just based on pure familiarity, individuals did not like it as much. However, it is interesting to note, that when asked what the audio-only condition would be beneficial for, participants did see the merit in this type of media while doing other tasks that require visual attention (driving, cleaning, etc.) or for those that are blind. These results corroborate what we know about multi-dimensional design and are important to consider as social media moves forward [5], [7], [9].

Limitations

Limitations of this work include the low fidelity mockup used in tests, a small sample size, posts that did not induce negative feelings, and participants recruited were not screened to see if they experience negative effects of social media. The stimuli presented was created by the researchers to mimic Instagram’s layout; however, the auditory stimuli were recorded by

the researchers and were set at one speed for the duration of the experiment. Some participants indicated that they wished the audio could be varied by speed and how much information was given at a time. Additionally, as mentioned in the discussion, the three conditions were rated by participants approximately the same on the State Trait Anxiety Inventory our goal of inducing negative feelings of BDD or perceived status. Lastly, the participants recruited were not specifically selected because they felt social media had a negative effect on their mental health, if they had been, potentially the results would differ supporting the hypotheses.

Conclusions and Future Work

Although our hypotheses were not supported fully by this experiment, the idea of an auditory social media may have promise in the future. Especially considering the user's desires to multi-task while interacting with social media. Several participants discussed how an auditory social media could benefit them while driving, cleaning, or for those who are blind like how podcasts or radio news do currently. Future research should address a higher fidelity mockup to fully gauge the appropriateness of the AI auditory-only implemented in social media. Given more ability to customize and choose when/how much auditory information to receive, users could benefit from the idea of auditory social media.

References

- [1] L. Braghieri, R. Levy, and A. Makarin, "Social Media and Mental Health," *American Economic Review*, vol. 112, no. 11, pp. 3660–93, Nov. 2022, doi: 10.1257/AER.20211218.
- [2] L. Charmaraman, A. M. Richer, C. Liu, A. D. Lynch, and M. A. Moreno, "Early adolescent social media-related body dissatisfaction: Associations with depressive symptoms, social anxiety, peers, and celebrities HHS Public Access," *J Dev Behav Pediatr*, vol. 42, no. 5, pp. 401–407, 2021, doi: 10.1097/DBP.0000000000000911.
- [3] C. Bohon, E. Hembacher, H. Moller, T. D. Moody, and J. D. Feusner, "Nonlinear relationships between anxiety and visual processing of own and others' faces in body dysmorphic disorder," 2012, doi: 10.1016/j.psychresns.2012.09.003.
- [4] L. I. Stein Duker, A. R. Schmidt, P. K. Pham, S. M. Ringold, and A. L. Nager, "Use of Audiobooks as an Environmental Distractor to Decrease State Anxiety in Children Waiting in the Pediatric Emergency Department: A Pilot and Feasibility Study," *Front Pediatr*, vol. 8, p. 556805, Jan. 2021, doi: 10.3389/FPED.2020.556805/BIBTEX.
- [5] Douglas J. Swanson, Tuning in and hanging out: A preliminary study of college students' use of podcasts for information, entertainment, and socializing, *The Social Science Journal*, Volume 49, Issue 2, 2012, Pages 183-190, ISSN 0362-3319, <https://doi.org/10.1016/j.soscij.2011.08.011>.
- [6] Loewenthal, R., & McMullan, M., (2019). "Audio opportunities for news media". INMA, International News Media Association Reports.
- [7] Em Griffin, Andrew Ledbetter, and Glenn Sparks, a First Look at Communication Theory, 10th ed., p. 562, 2019
- [8] About KBP – Kapisanan ng mga Brodkaster ng Pilipinas, 1973. [Online]. Available: <https://www.kbp.org.ph/about-kbp>
- [9] Magnaye, Robin Ceasar F., and Mary Ann E. Tarusan. "The Old and The New: Radio and Social Media Convergence."
- [10] G. Brockington et al., "Storytelling increases oxytocin and positive emotions and decreases cortisol and pain in hospitalized children," *Proc Natl Acad Sci U S A*, vol. 118, no. 22, Jun. 2021, doi: 10.1073/PNAS.2018409118.
- [11] Wang, G., Zhao, J., Van Kleek, M., & Shadbolt, N. (2022). Informing Age-Appropriate AI: Examining Principles and Practices of AI for Children. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–29. <https://doi.org/10.1145/3491102.3502057>

- [12] E. Hilton, "Differences in Visual and Auditory Short Term Memory" *IU South Bend Undergraduate Research Journal*, vol. 4, pp 47-50, 2001.
- [13] Jiang Y, Qiao R, Shi Y, Tang Y, Hou Z, Tian Y. The effects of attention in auditory-visual integration revealed by time-varying networks. *Front Neurosci*. 2023 Aug 2;17:1235480. doi: 10.3389/fnins.2023.1235480.

Appendix

Appendix 1: Task Analysis

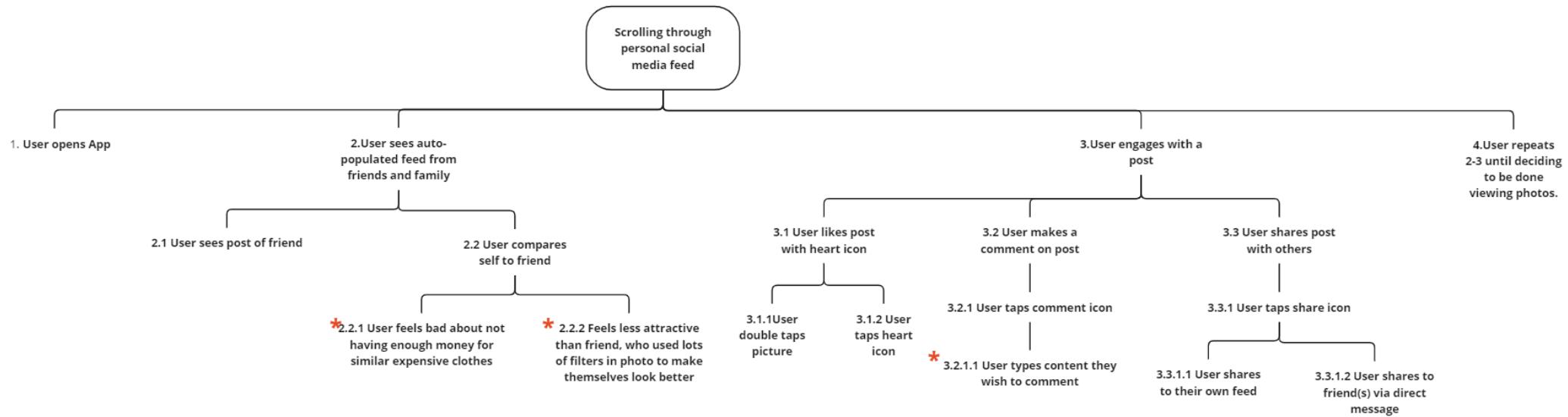
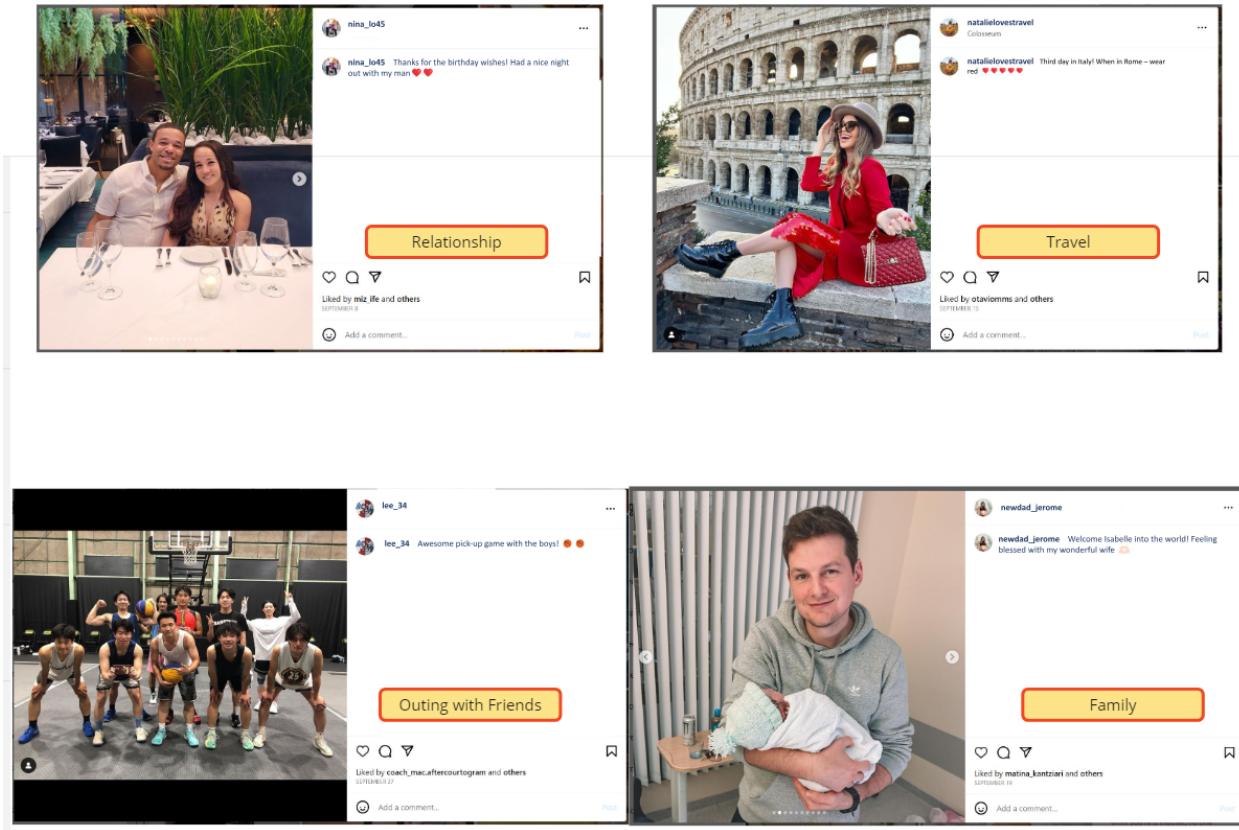
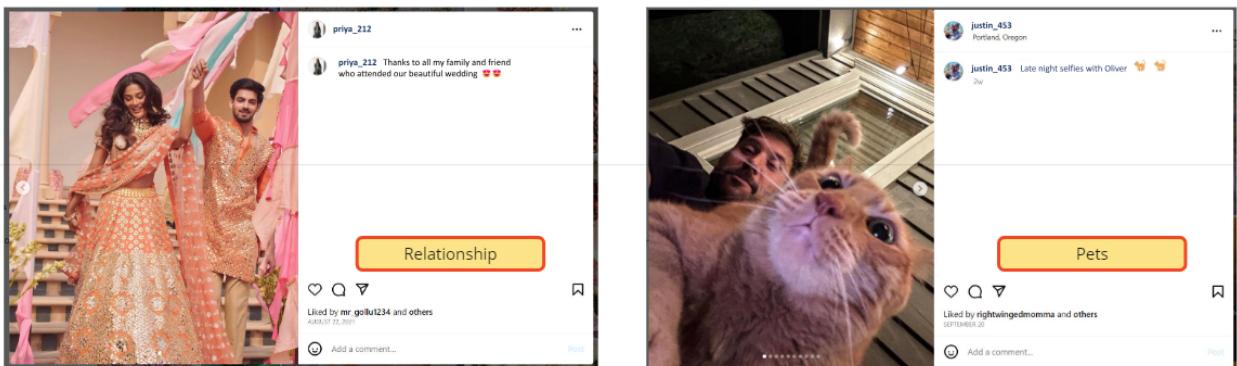
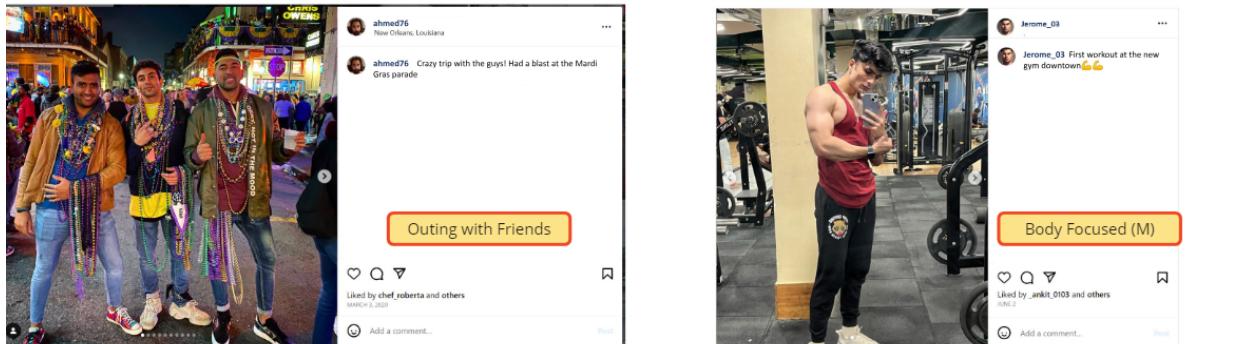
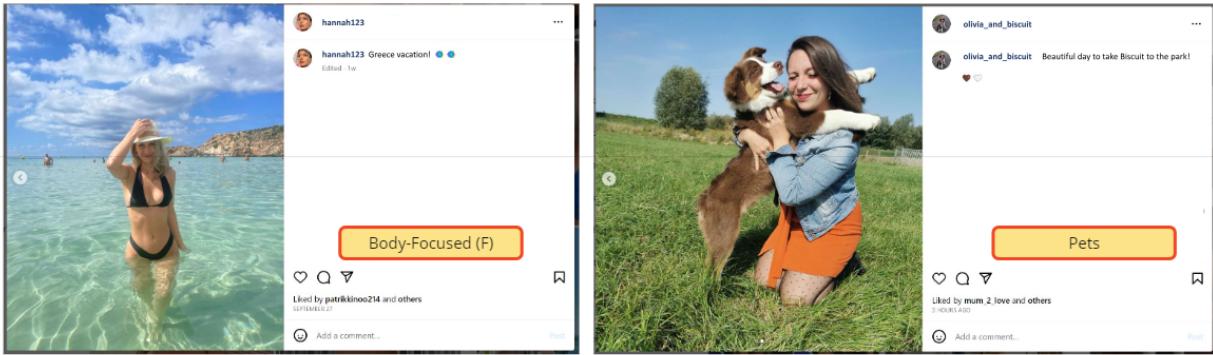


Figure 1: Task analysis flow chart for scrolling through social media

Appendix 2: Social Media Modalities

I. Visual-Only (Condition A) with Categories (Category cues were not visible to participant)





II. Audio-Only (Condition B) - Transcript

Post 1: Travel

“Your friend Emily traveled to Paris, France this week and is happily eating a chocolate croissant at a French Cafe along a cobblestone street. In her post, she says “Best croissant I have ever had!””

Post 2: Relationship

“Your colleague Mohammed proposed to Amira today and they are excited to announce their engagement. They had an engagement photo session at the Mountain Lake Lodge. Mohammed says “Here’s to a future of laughs and love with my best friend”

Post 3: Outing with Friends

“Your friend Kelly went out with her friends Shan, Diana, and Paula to have a charcuterie night at the Blacksburg Wine Lab on Sunday. Kelly is pictured smiling and writes “A much needed ladies night after a long week!””

Post 4: Pets

“Your cousin Steve just got a new corgi puppy named Cheeto. Steve is teaching Cheeto how to walk on a leash and Cheeto still has a bit of learning to do. Steve writes, “Cheeto’s first walk! Hopefully he’ll learn that the leash is not a chew toy.””

Post 5: Body Focused (M)

“Kyle is at the new gym and finished a workout. He is pictured at the gym and writes, “New gym gains!””

Post 6: Family

“Your brother Isaac is taking his daughter Nina to her first day of kindergarten. Nina is pictured in front of the school bus with a Hello Kitty backpack and looks excited to start her first day. Isaac writes, “First day of school! Can’t believe how fast this one is growing up.””

Post 7: Relationship

“Your friend Julia is celebrating her 2nd year anniversary with her boyfriend Omar. They are at the Farmhouse Restaurant for a special dinner. Julia writes, “To an amazing 2 years filled with wonderful adventures and smiles.””

Post 8: Outing with Friends

“Your friend Kyle went out to this weekend’s Virginia Tech football game with his friends Jin and Ameer. Kyle writes, “First football game with the boys!””

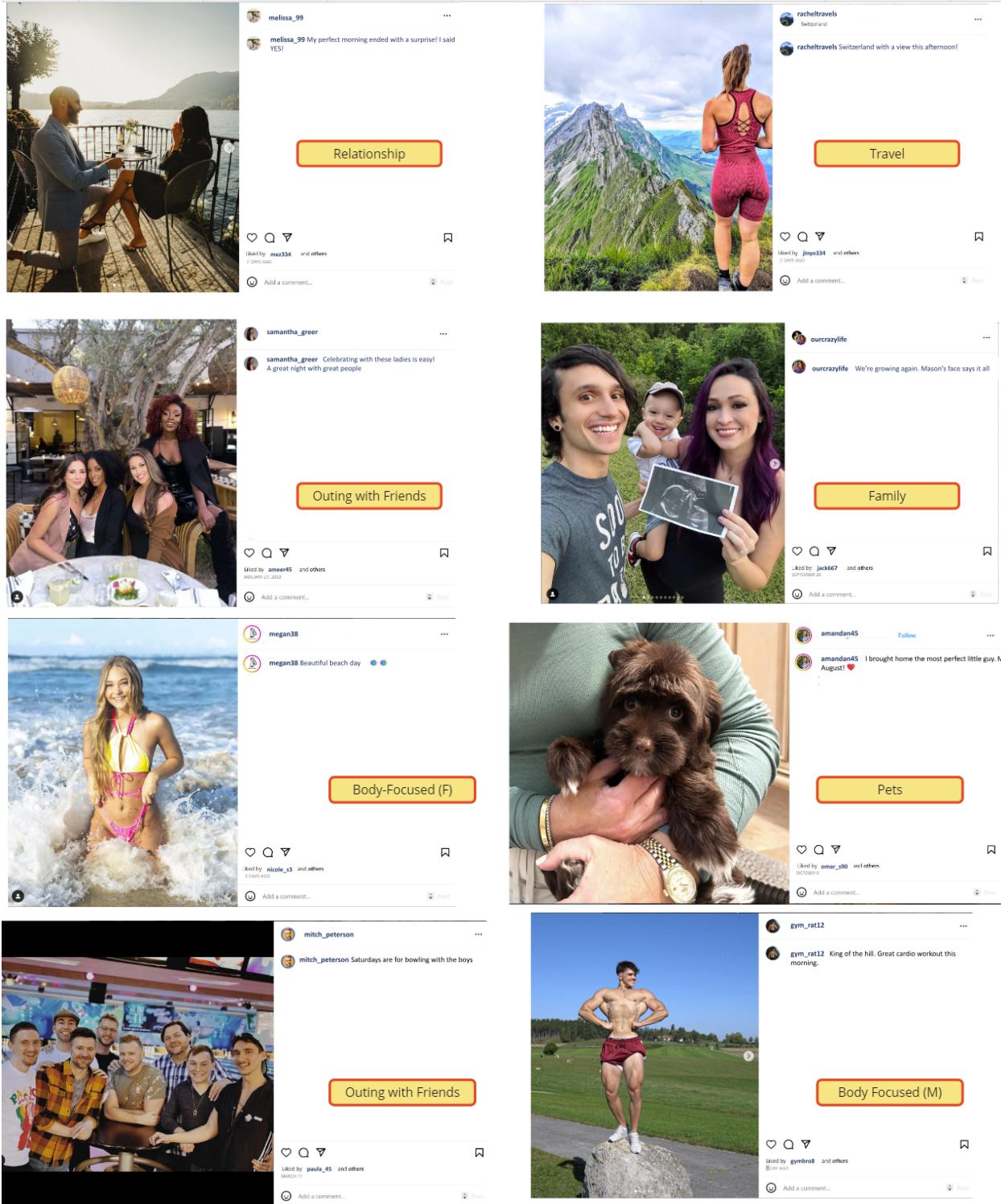
Post 9: Body Focused (F)

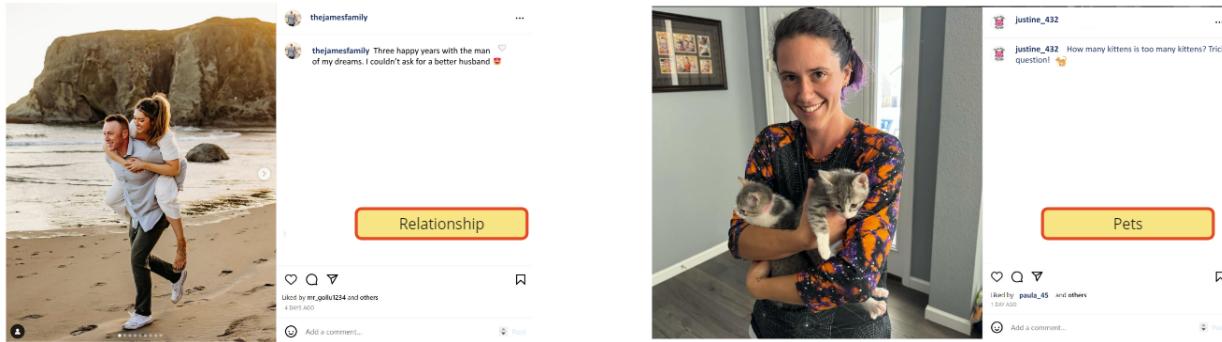
“Your colleague Sheena went to Jamaica and is loving her time on the beach. She is relaxing on the sand and writes, “Catch me on these Caribbean vibes””

Post 10: Pets

“Your friend Amina knitted a dinosaur sweater for her cat Oscar. Oscar does not appear to like the sweater and seems quite grumpy with his circumstance. Amina writes, “Dino-Oscar is not amused!””

III. Visual and Audio (Condition C) - with Categories (Category cues were not visible to participant) and corresponding Transcript





Post 1: Relationship

“Your friend Melissa’s boyfriend proposed this morning while at breakfast overlooking a beautiful lake. Melissa says “My perfect morning ended with a surprise. I said YES!””

Post 2: Travel

“Your sister Rachel traveled to Switzerland this week and is hiking nearby Mt. Matterhorn. In her post, she says “Switzerland with a view this afternoon!””

Post 3: Outing with Friends

“Your colleague Samantha went out this Saturday to a celebratory dinner with friends, she says “celebrating with these ladies is easy! A great night with great people””

Post 4: Family

“Adam and Allegra pose in their yard with their two year old, Mason, excitedly holding an ultrasound picture. The caption says “We’re growing again. Mason’s face says it all.””

Post 5: Body Focused (F)

“Your cousin Megan poses kneeled in the sand and waves in a pink and yellow bikini smiling captioned “Beautiful Beach day””

Post 6: Pets

“Your mom, Amanda, holds her new chocolate brown puppy, August. She says “I brought home the most perfect little guy. Meet August””

Post 7: Outing with Friends

“Your brother, Mitch poses with his 7 guy friends in front of the lanes at the local bowling alley. Mitch says “Saturdays are for Bowling with the Boys””

Post 8: Body Focused (M)

“Marcus stands atop a rock in the local park. Shirtless and flexing his muscles claiming “King of the hill. Great cardio workout this morning””

Post 9: Relationship

“Grace posted a picture of her husband, Graham, carrying her piggy back style across a sandy beach. She says “Three happy years with the man of my dreams. I couldn’t ask for a better husband.””

Post 10: Pets

“Justine stands smiling, wrangling two little kittens in her arms. She says “How many kittens is too many kittens? Trick Question!””

Appendix 3: State Trait Anxiety Inventory

State Trait Anxiety Inventory

Read each statement and select the appropriate response to indicate how you feel right now, that is, at this very moment. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.

	1	2	3	4
	Not at all	A little	Somewhat	Very Much So
1. I feel calm	1	2	3	4
2. I feel secure	1	2	3	4
3. I feel tense	1	2	3	4
4. I feel strained	1	2	3	4
5. I feel at ease	1	2	3	4
6. I feel upset	1	2	3	4
7. I am presently worrying over possible misfortunes	1	2	3	4
8. I feel satisfied	1	2	3	4
9. I feel frightened	1	2	3	4
10. I feel uncomfortable	1	2	3	4
11. I feel self confident	1	2	3	4
12. I feel nervous	1	2	3	4
13. I feel jittery	1	2	3	4
14. I feel indecisive	1	2	3	4
15. I am relaxed	1	2	3	4
16. I feel content	1	2	3	4
17. I am worried	1	2	3	4
18. I feel confused	1	2	3	4
19. I feel steady	1	2	3	4
20. I feel pleasant	1	2	3	4

Reference: Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory (STAII). Palo Alto, CA: Consulting Psychologists Press.