

Close-up and Whispering: An Understanding of Multimodal and Parasocial Interactions in YouTube ASMR videos

Shuo Niu

shniu@clarku.edu

Clark University

Worcester, MA, USA

Hugh S. Manon

hmanon@clarku.edu

Clark University

Worcester, MA, USA

Ava Bartolome

abartolome@clarku.edu

Clark University

Worcester, MA, USA

Nguyen B. Ha

joha@clarku.edu

Clark University

Worcester, MA, USA

Keegan Veazey

kveazey@clarku.edu

Clark University

Worcester, MA, USA

ABSTRACT

ASMR (Autonomous Sensory Meridian Response) has grown to immense popularity on YouTube and drawn HCI designers' attention to its effects and applications in design. YouTube ASMR creators incorporate visual elements, sounds, motifs of touching and tasting, and other scenarios in multisensory video interactions to deliver enjoyable and relaxing experiences to their viewers. ASMRtists engage viewers by social, physical, and task attractions. Research has identified the benefits of ASMR in mental wellbeing. However, ASMR remains an understudied phenomenon in the HCI community, constraining designers' ability to incorporate ASMR in video-based designs. This work annotates and analyzes the interaction modalities and parasocial attractions of 2663 videos to identify unique experiences. YouTube comment sections are also analyzed to compare viewers' responses to different ASMR interactions. We find that ASMR videos are experiences of multimodal social connection, relaxing physical intimacy, and sensory-rich activity observation. Design implications are discussed to foster future ASMR-augmented video interactions.

CCS CONCEPTS

- Human-centered computing → Empirical studies in collaborative and social computing.

KEYWORDS

ASMR; YouTube; video; multimodal; parasocial; experience

ACM Reference Format:

Shuo Niu, Hugh S. Manon, Ava Bartolome, Nguyen B. Ha, and Keegan Veazey. 2022. Close-up and Whispering: An Understanding of Multimodal and Parasocial Interactions in YouTube ASMR videos. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA*. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3491102.3517563>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9157-3/22/04...\$15.00

<https://doi.org/10.1145/3491102.3517563>

1 INTRODUCTION

Autonomous Sensory Meridian Response (ASMR) is a phenomenon usually experienced as tingling sensations in the crown of the head in response to a range of audio-visual triggers such as whispering, tapping, and hand movements [54]. ASMR videos incorporate audio, touch, taste, observation, and roleplay effects to deliver enjoyable and relaxing feelings. Over the past decade, the creation culture on YouTube has attracted numerous ASMR creators (known colloquially by users as "ASMRtists") to design a wide array of tingle-inducing sounds and actions to intentionally induce ASMR feelings [3, 41]. ASMRtists have also leveraged ASMR videos to connect to the viewers and build online ASMR communities [3, 62]. A typical YouTube ASMR video may feature an ASMRtist whispering to the viewer, roleplaying personal attention such as massages or haircuts, making crisp sounds, or engaging in various slow and repetitive movements [5]. YouTube hosted more than 5.2 million ASMR videos in 2016 and 13 million in 2019, and the searches for ASMR grew over 200% in 2015 and are consistently increasing [42, 66]. Remarkably "ASMR" is among the top five YouTube search queries globally and in the US, with a search volume of more than 14 million¹.

In Human-Computer Interaction (HCI), experience-centered design requires researchers to capture and analyze the experiences generated from interaction and adopt the understanding of these experiences in design practices [25]. ASMR is a unique experience insofar as only some users experience the "tingles" as a response to particular triggers, and the same trigger may have different effects on different people [18, 28, 41, 54]. Over the years, ASMRtists developed highly stylized and conventionally patterned ASMR videos to engage their viewers, and as a way to enhance affect and intimacy [3, 71]. Prior research on ASMR has focused on characterizing ASMR triggers [5, 19] or understanding ASMR interactions through qualitative video analysis [3, 62, 71], user surveys [38, 54], and brain imaging [64]. Most studies described YouTube ASMR videos primarily as roleplays [1, 65, 71] or as a single video type with a mixture of ASMR triggers [54, 62]. However, little data-driven research has been conducted to categorize the wide variety of ASMR experiences developed by YouTube ASMRtists. YouTubers create ASMR videos with or without elements of social interaction, using roleplays or simply manipulating objects, and position themselves up-close or distant from the viewer. A macro understanding of the delivery

¹<https://ahrefs.com/blog/top-youtube-searches/>

mechanism and experience patterns in YouTube ASMR videos will help technology and service designers explore ways to integrate ASMR and assess its effects on the user experience. Since experience seekers need different triggers to acquire ASMR sensations, a quantitative overview of common ASMR interaction modalities will indicate what ASMR interactions may work for more users.

In this study, we collect a large number of ASMR videos and perform a mixed-method analysis to obtain an overview of ASMR interactions and experiences. This work analyzes 2663 ASMR videos collected from YouTube to examine the multimodal interactions and the ways ASMR performers para-socially attract the viewers. We focus on intentional ASMR videos – videos with “ASMR” labels in which a variety of triggers are purposefully displayed by the performer – to understand ASMRtists’ common approaches to trigger ASMR experiences. Prior work identified visual, audio, touch, taste, and scenario-based ASMR triggers [18, 56, 63, 71]. By interacting with ASMR videos, viewers are able to experience a simulation of intimacy with the video performer through “parasocial interactions” [69] – a one-sided intimacy experienced by a viewer through repeated encounters with a figure on screen. In parasocial relationships, video performers develop and manage three types of attractiveness – *social attraction*, *physical attraction*, and *task attraction* [58]. We quantify the manners in which the ASMRtists socialize with the viewer (social attraction), the camera proximity of the ASMRtists in the videos (as an alternative for physical attraction), and purposeful activities performed by the ASMRtists (task attraction). This work addresses three main research questions:

- RQ1: How are various interaction modalities employed in YouTube ASMR videos?
- RQ2: How do YouTube ASMRtists design parasocial attractiveness through multimodal interactions?
- RQ3: How do different multimodal interactions and parasocial attractions affect the expression of viewers’ feelings in the comments?

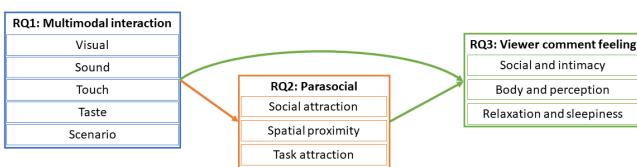


Figure 1: The structure of the research questions

Figure 1 illustrates the structure of the research questions. RQ1 provides an overview of multimodal interactions in YouTube ASMR videos to inform designs with common ASMR performing methods. RQ2 focuses on understanding the patterns of parasocial attractiveness through multimodal interactions. We summarize the experiences delivered by YouTube ASMR videos and identify the associated interaction modalities. RQ3 utilizes viewers’ comments to infer how different multimodal interactions and parasocial attractions affect viewers’ social, perceptual, and relaxation feelings. We first use grounded-theory approaches to identify subcategories of interaction modalities and parasocial attractions. Then the codebook is translated into a questionnaire task. The annotation tasks

were completed by participants recruited from Amazon Mechanical Turk (MTurk). We perform statistical analysis to address the research questions.

The development of multimodal interactions depends on the natural integration patterns that typify the combined use of different input modes [49]. Understanding diverse interaction modalities through analyzing extensive video data inform different ways to incorporate ASMR in technology design. Our results indicate social attractions are enhanced by combining multiple ASMR interaction modalities. Most ASMRtists use the closeup camera proximity as a means of building physical attractiveness. ASMRtists emulate physical closeness through microphonically-amplified whispering, manipulating objects, virtually “touching” the viewer, and making mouth noises and microphone-jostling sounds near the camera or the microphone. Many ASMR videos do not involve purposeful tasks and are not roleplays. Tasks used in non-roleplay videos include soft and routine activities such as performing medical or cosmetic treatments, eating and drinking, and demonstrating mundane daily activities. The ASMR experiences delivered by YouTube ASMRtists can be described as three experience patterns: multimodal social connection, relaxing physical intimacy, and observation of sensory-rich activities. This work aims to inspire future technologies and services to incorporate ASMR triggers to design ASMR-augmented relaxing or intimate experiences.

2 BACKGROUND

2.1 ASMR Videos on YouTube

The now widely-adopted term “Autonomous Sensory Meridian Response” (ASMR) was coined in 2010 to describe a sensory phenomenon that usually involves the sensation of tingling as a response to certain audio-visual stimuli [6]. Common ASMR videos show intentional or unintentional gentle interactions such as speaking softly, playing or brushing hair, moving hands, and tapping or scratching surfaces [18, 54], which may trigger a low-grade euphoria response and tingling sensations on the viewer’s head and spine [1]. The ASMR trend on social media began with a Yahoo group sharing personal experiences of head tingles when watching specific kinds of videos [3, 13]. Those original videos were dubbed “unintentional” ASMR, and involved real-world scenarios such as doctor’s office examinations and suit fittings, captured for some non-ASMR purpose and uploaded, but subsequently re-contextualized for their ASMR tingle-triggering properties [20, 41]. Afterward, creators made numerous “intentional” ASMR videos on YouTube in which ASMRtists purposefully use visual and sound stimuli and scripted roleplays to induce ASMR experiences [1, 40]. In 2019, there were 13 million ASMR videos on YouTube [66]. Popular ASMRtists such as GentleWhispering ASMR, SAS-ASMR, and Gibi ASMR, have millions of subscribers, and their videos attracted millions of views and generate considerable revenue for the creators [66, 71].

Despite the popularity of this emerging video genre on YouTube, studies found ASMR triggers do not work for everyone, and some individuals only experience the tingles with very precise, idiosyncratic triggers [5, 54]. ASMR was found to be associated with specific personality traits of individual viewers and to vary from person to person [18, 28]. Users’ diverse needs triggering effects drive ASMR consumers to constantly search for videos with the keyword

“ASMR.” [3] In turn, the YouTube culture of creativity and participation [10] encourages ASMRtists to make numerous ASMR videos to satisfy ASMR experience-seekers’ diverse needs. The growing trend of ASMR creation and consumption drew researchers’ attention. Prior studies focused on understanding the sensational effects through interviewing ASMR viewers (e.g., [5, 38, 54]) or scanning brain images (e.g., [64]). Some studies examined the creator-viewer interactions and the digital intimacy through qualitative analyses of a few viral videos (e.g., [40, 62, 71]).

Although the diversity of ASMR triggers is widely noted, there is little analysis of large video data sets to explain the creation practices employed by ASMRtists. ASMR can be induced with virtual face-to-face interactions [5, 65] or simply by manipulating objects without showing performers’ faces [45]. Some ASMR videos consist of constant soft speaking while others involve object manipulation without talking [36]. Some ASMR videos pretend to touch the viewers in roleplays while others perform massages on a second person who is also visible in the video [71]. ASMR sensations can emerge both in response to food consumption videos [4] and to videos showing a person studying quietly [37]. A quantitative analysis of extensive videos will help HCI designers discern the significance of different ASMR interactions and experiences. First, since an ASMR trigger may or may not induce ASMR experiences, an overview of common ASMR interaction modalities and experiential patterns will help technology designers to incorporate ASMR triggers that are effective for a broader range of users. Second, recent research noted ASMR is not just a sensory experience; it is also a kind of mediated intimacy offered by ASMRtists to deliver a sense of social connection [3, 62, 71]. However, there is limited understanding of how such social experiences are commonly constructed and their relationships with trigger interactions and social settings. Last, understanding viewers’ social, perceptual, and relaxation feelings will help technology designers understand the possible effects of different ASMR experiences.

2.2 Multimodal Interactions in ASMR videos

Researchers have examined various triggers in ASMR videos to understand this emerging media form and its physiological effects. Richard summarizes ASMR stimuli as audio, touch, visual, and scenario triggers [56]. Smith et al. examined people’s responses to five trigger types, including watching, touching, repetitive sounds, simulations, and mouth sounds [63]. Zappavigna explored ASMR roleplays and found linguistic, visual, and aural resources are used to create a sense of co-presence with the performer [71]. This study explores common interactions performed by YouTube ASMRtists. HCI researchers have identified vision, hearing, touch, smell, and taste as five sensing modalities to embody interactions [60, 70]. Grounded in multimodal interaction theories [60, 70] and trigger modalities identified in the literature [56], RQ1 explores *visual, sound, touch, taste, and scenario* as five interaction modalities in YouTube ASMR videos. *Visual interactions* describe how the performers present themselves and the trigger objects. We look into visual settings such as ASMRtists showing themselves in front of the camera, performing slow activities, or simply showing hands manipulating ASMR trigger objects. *Sound interactions* refer to what types of hearing triggers are made by the ASMRtists. This modality

seeks to capture sounds like human speaking, tapping or scratching objects, or various sound effects produced by interacting with the microphone. *Touch interactions* examine how ASMRtists stimulate haptic feelings. For this modality, we observe how ASMRtists use their hands to interact with themselves, physical objects, the camera, or another person in the video. *Taste interactions* investigate whether or not the ASMRtists eat food in the video. And finally, *scenario triggers* describe the simulated situation and environment in roleplay videos, such as haircuts, eye exams, or other dramatized forms of assistance from the figure on screen [56].

2.3 Parasocial Attractions in ASMR Videos

A large community of YouTube creators designate themselves as “ASMRtists” by regularly creating and uploading ASMR videos [3]. The pseudo-interactive nature of the videos engenders a sort of intimacy with the video creator [11, 27, 46]. The essentially one-sided intimacy with the video performers, generated by a “*conversational give-and-take*,” is defined as a parasocial relationship, and the interactions users have with videos that generate parasocial relationships are called “*parasocial interactions*.” [24, 26] Parasocial relationships and interactions have been widely found in the interactive responses of viewers to a TV or social media figure, which affect viewers socially and emotionally. Video-watching may lead some viewers to imagine themselves interacting with the performer [22]. Studies found parasocial relationships can provide social support and shield against the effects of exclusion and loneliness [24, 46]. YouTube users mostly focus their investments (of time, of energy) in parasocial interactions with the video creator, rather than building a friend and community network like other social media [47, 57]. ASMR videos can be seen as a unique form of parasocial interactions offered by ASMRtists [10]. Klausen described ASMR as a “para-haptic interactional” relation with the ASMRtists while obtaining a form of presence and intimacy [31]. Zappavigna also considered ASMR videos as a construction of the interactive context in which viewers feel co-presence with the performer [71]. Smith argued that in ASMR videos, affective experiences are intentionally construed and strategically heightened [62]. However, due to a dearth of quantitative analysis of parasocial interactions in ASMR videos, it is unknown how prevalent the social experience is, or what general approaches are best used to deliver such one-sided intimacy.

This work analyzes the associations between parasocial interactions and interaction modalities to explore the patterns of social and intimate experiences designed by ASMRtists. The parasocial interaction theory suggests that video performers develop *social, physical, and task attractions* to engage viewers and establish parasocial relationships [34, 58]. *Social attraction* refers to the degree to which one feels they would like to befriend the television or media persona [34]. In ASMR, performers may simulate conversational scenarios and socialize with the viewers through their soft vocal and bodily interactions [31, 71]. Creators use ASMR videos to attract patrons and build network of fans [40]. *Physical attraction* refers to how video performers appeal to the viewer physically [34]. In this study, we measure camera spatial proximity as a vector of physical attraction since it is difficult to quantify the attractiveness of performers’ physical appearance. ASMRtists tend to perform body

and hand movements very near the camera [65, 71]. ASMRtists may themselves appear in the video either in close up to show intimacy [3, 23, 62], or they may exclusively show trigger objects on screen without showing their faces [56]. *Task attraction* describes how ably, credibly, or reliably a performer can complete a task [34]. ASMRtists perform different tasks such as professional treatments [1, 35, 71], mundane activities like putting on makeup or sorting cards [18, 32, 37], or less meaningful activities like cutting soap or tapping their fingernails on objects [20]. RQ2 uses the parasocial attraction framework identified in [58] to examine the patterns of social, proximity, or task-observing experiences.

2.4 ASMR Experiences and Benefits

ASMR experiences are touted by many as promoting calm and relaxed feelings [33, 43, 52] and are associated with positive affect and a sense of interpersonal connection [31, 54, 62]. Barratt and Davis found that ASMR combines positive feelings, relaxation, and tingling sensation of the skin and provides temporary relief from depression [5]. Smith et al. analyzed neuroimages during ASMR tingles and found ASMR was associated with a blending of multiple resting-state networks [64]. Kovacevich examined comments to ASMR videos and found positive comments appreciated the calming or relaxing effects [33]. For social and intimate experiences, Klausen argued that ASMRtists leverage binaural sounds and haptic interactions to create a form of embodied presence and distant intimacy with the viewers [31]. Smith and Snider also suggested ASMR performers intentionally express feelings of intimacy and affection to the viewers [62]. Recent HCI research explored the use of ASMR effects in wearable technologies for enchantment and slow experiences [32]. Studies on food-eating videos (colloquially known as “Mukbang”) found ASMR a key motivator for video watching [4, 68]. YouTube study-with-me videos also use ASMR effects [37].

RQ3 seeks to obtain an initial understanding of viewers’ responses to different ASMR experiences through comment analysis. Prior work found that people can have different feelings with the same ASMR trigger [19, 53] and don’t publicly share ASMR experiences with others [4, 5]. Comments represent immediate and direct user reactions to a video and analyzing comments is a more straightforward way to capture viewers’ feelings than rating by external participants [33]. Word analysis of YouTube comments is a common approach to infer the influence of videos on viewers [2, 59, 61]. In RQ3, we measure how ASMR viewers comment on three common feelings of ASMR identified in prior research: social connection and intimacy, sensory perception, and relaxation and sleepiness. Considering the difficulties of manually annotating a large number of comments and subjectively rating viewers’ feelings, a mixed-method of Linguistic Inquiry and Word Count (LIWC) software [51] and pointwise mutual information (PMI) [8] is used. LIWC has been scientifically validated to analyze people’s social and emotional expression on social media [14, 17]. PMI is also a lexicon-based method to identify topically important keywords [14, 55]. Both methods are widely used in prior HCI research to imply psychological processes from social media data (e.g., [14, 17, 55]). Then we compare the word frequencies in comments of videos with different interaction modalities and parasocial attractions.

3 ASMR VIDEO DATA

We collected recent ASMR videos from active video creators to analyze ASMR videos on YouTube. The ASMR videos were crawled using the YouTube Data API² with the search seed “ASMR.” In the first step, we searched ASMR videos on Jun 11, 2020, and Oct 20, 2020, to collect a list of videos posted in the prior three months, respective to each search date. This step lets us identify active channels that were recently posting ASMR videos. Then the crawler collected all available videos belonging to those active channels to form a raw dataset. 227,133 videos were returned from YouTube. For each video, we requested the YouTube API to return 300 top-level comments³. Since comments belonging to a video are analyzed together, and popular videos may have numerous comments, collecting up to 300 per video ensures all videos have a similar amount of comments.

Titles and tags are processed to filter out non-English videos. We exclude non-English videos due to difficulties in the data tagging and categorization. Videos without “ASMR” in the titles are removed since this work focuses on intentional ASMR videos with a clear ASMR theme and is designed for this experience (an ASMRtist may post non-ASMR videos). We exclude videos shorter than 5 minutes ($N = 9676, 4.26\%$) due to many of them being previews of full videos and compilations of short video clips from multiple ASMR videos. We also only keep videos posted between Jan 1, 2020, and Jun 01, 2020, to ensure the videos reflect the latest creation styles and have enough time to receive comments. We remove videos with fewer than 50 comments (31.39% of videos) to ensure videos had enough comments for word analysis.

After filtering, 85,734 videos are kept for data sampling. These videos come from 697 different channels. Then we randomly sample 200 videos for grounded theory analysis. We sample up to 10 videos per channel for the final data analysis. There are many channels with less than ten videos in our dataset – the eventual sampling results in 2830 videos for data annotation. The data collection overlaps with the emergence of COVID-19, but through a rough examination, we didn’t notice a significant mention of COVID-19 in the videos. The IRB (Institutional Review Board) office at the authors’ institute has reviewed the entire research process and exempted this research from ethics board review (see Appendix B for more information).

4 METHODS

4.1 Grounded Analysis

Analysis in prior studies identified triggers from a small video sample or a few popular ASMR roleplays. Therefore, we choose to conduct a grounded theory analysis to extract common interaction modalities and parasocial attraction techniques. Grounded theory data analysis has been widely used to inductively derive models of social processes [12, 16]. This work follows open, axial, and selective coding procedures to generate and verify modality and attractiveness subcategories. We randomly sample 200 videos from 166 ASMRtists for the grounded analysis [7].

²<https://developers.google.com/youtube/v3>

³YouTube Data API does not specify how the returned comments are selected and ordered. <https://developers.google.com/youtube/v3/docs/comments/list>

Table 1: Subcategories and definition of the five multimodal interactions in ASMR videos

Category	Definition
Visual	Face-to-face The ASMRtist looks at the camera to mimic face-to-face interactions with the viewer.
	Mukbang The ASMRtist presents and consumes large quantities of food (Mukbang)
	Object only The ASMRtist interacts with physical objects without showing their faces
	Serve people The ASMRtist performs a treatment/service on another person
	Images Static image(s) or black screen
	Gaming Video shows clip(s) of gaming, with or without the ASMRtist in view
	Animals The video has animals as the characters
	Object Sounds made by interacting with a physical or liquid object by tapping, scratching, pouring, spraying, etc.
	Whispering Whispering or talking in a low volume
	Mouth Sounds made with mouth by eating, drinking, lip smacking, tongue clicking, kissing, licking, or sucking
Sound	Body&cloth Sounds made by touching/brushing/scratching themselves, another person, or a fake/silicon body in the video
	Ambience Ambient and background sounds emitted from a real or fake environment
	Mic Sounds made by interacting with the microphone
	Viewer The ASMRtist reaches to the viewer with their hands or tools in front of the camera
	Objects The ASMRtist clicks, taps, scratches, squeezes, or rubs physical objects
	Own body The ASMRtist touches their own head, body, clothes by rubbing, scratching, combing, applying makeup, etc.
	Real person The ASMRtist uses their hands or tools to interact with another real person in the video
Touch	Taste The ASMRtist eats or drinks for more than half of the video
	Service The video is a treatment or service roleplay in which the ASMRtist acts as a service provider and the viewer acts as a customer/patient (e.g., massage, haircut, makeup application, clinical exam, interview, customer service).
	Fantasy The video is a roleplay in which the ASMRtist acts as a character in a fantasy, surreal, or otherwise unrealistic scenario (e.g., historical/anime/comics character)
	Romance The video is a roleplay in which the ASMRtist acts as an intimate partner and directly interacts with the viewer intimately or romantically.
Scenario	

Table 2: Subcategories and definition of the three parasocial attractiveness in ASMR videos

Category	Definition
Social	Talk To The ASMRtist talks to or reads to the viewer
	Talk with The ASMRtist pretends to talk with or chat with the viewer, pretending the viewer responds to the ASMRtist
	Gesture and text The ASMRtist makes eye contact with the viewer and uses body language/closed captions/texts to communicate with the viewer
Proximity	Closeup One of the 3 camera shot scales (Extreme closeup, Closeup, Medium closeup) in which ASMRtists placing themselves close to the camera
	Medium One of the 2 camera shot scales (Medium shot and Medium-full shot) in which ASMRtists placing themselves in medium distance to the camera
	Fullshot The ASMRtists show full body in the camera
	No face Static image(s), black screen, or no human face in the video
	Partial face Showing half-face (upper or lower half face)
Task	Treatment and service The ASMRtist performs treatment/service on the viewer or another person in the video (e.g., massage, makeup application, interview, office visit, hypnosis, Reiki, etc.)
	Common activity The ASMRtist engages in common daily activity(s) such as painting, writing, folding clothes, preparing food, or applying makeup to themselves.
	Eat and drink The ASMRtist eats and/or drinks in the video

In the open coding phase, two of the authors of this research watched 50 videos each and take notes on the visual, sound, touch, taste, and scenario triggers described in [56]. The example multimodal interactions can be found in Figure 2. For parasocial attractiveness, the authors annotated how the ASMRtists simulate communication with their viewers (social attraction), where the ASMRtist is situated in proximity to the camera (spatial proximity), and the tasks the ASMRtists perform (task attraction).

For axial coding, the two authors used the affinity diagramming approach to summarize these notes and develop subcategories of interaction modalities and parasocial attractions (see Table 1 and Table 2 for the codebook). The categorization of social attraction identifies that ASMRtists may communicate with the viewer in the form of a one-sided talk, or may chat with the viewers as in a back-and-forth conversation by pausing and waiting for the viewer to reply. Some other ASMRtists make ASMR videos without any human voice on the audio track, instead using gestures and text to communicate. For spatial proximity, we focus on categorizing the proximity with which ASMRtists positioned themselves in relation to the camera. We adopt the shot scales used in film and TV⁴ (Figure 3) to annotate how the ASMRtist is displayed within the video frame (including *extreme closeup*, *closeup*, and *medium closeup*), medium distance (including *medium shot* and *medium-full shot*), or showing the whole body from head to toe (*full shot*). The annotation of the tasks performed by the ASMRtists finds three main categories of activities with clear goals. Treatment and service tasks seek to perform actions such as massage or haircut on the viewer. Some videos perform everyday tasks such as painting, writing, or applying makeup. Other videos demonstrate eating or drinking a large quantity of food (called Mukbang videos on YouTube [4]).

In selective coding, two authors annotated the remaining 100 videos using the codebook to validate the subcategories and obtain the inter-rater agreements between experts. Audio and touch were annotated as multi-categorical values. Visual, taste, scenario, social, physical, and task were annotated as single-categorical values. After annotation, 12 of 100 videos were removed due to unavailability (e.g., deleted, private, age-restricted, or non-English). Fleiss Kappa with Jaccard distance was used to calculate between expert agreement. For the 88 videos, all multimodal and parasocial categories reach substantial agreements with kappa scores between 0.62 and 0.88 (Table 3). Social and task have relatively lower agreement due to the differences in deciding if the ASMRtist talks to or talks with the viewer (e.g., a video is tagged differently because the ASMRtist mostly whispers by herself but also greets the viewer) and whether an activity is considered common and daily (e.g., one rater feels mixing makeup slime is a common activity while the other rater annotate it as a non-task video). Sound and touch have a lower agreement because they are multi-choice categories. Then the third author annotated disagreed answers independently to solve discrepancies and generate 88 expert annotations. The expert annotations were used to assess the accuracy of annotations completed on Amazon Mechanical Turk.

Table 3: Agreement scores between experts and between experts and MTurk participants. Calculated using Fleiss' Kappa with Jaccard method.

	Social	Proximity	Task	Visual	Sound	Touch	Taste	Scena.
Between experts	0.65	0.83	0.62	0.84	0.68	0.64	0.88	0.78
Between experts and MTurk	0.73	0.79	0.72	0.80	0.68	0.67	0.93	0.76

⁴<https://www.studiobinder.com/blog/types-of-camera-shots-sizes-in-film/>



Figure 2: Examples of visual, sound, touch, and scenario subcategories

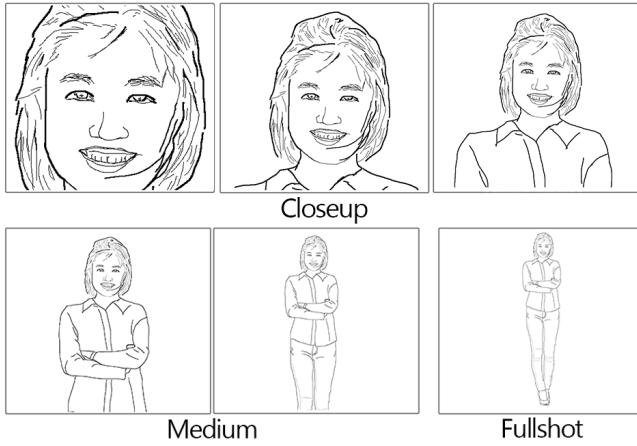


Figure 3: Closeup, medium, and full shot examples in spatial proximity

4.2 Video Data Annotation

This work uses Amazon Mechanical Turk (MTurk) to annotate the ASMR videos. Each task consists of two steps. Each participant was asked to watch each video for three minutes in the beginning, one minute in the middle, and one minute in the end. Then the participant was asked to annotate the five multimodal interactions and three parasocial attractions by answering multi-choice questions. Example pictures were provided to explain each visual and proximity subcategory. Example video clips were provided to explain each sound and touch subcategory. A qualification test was performed to pre-screen qualified participants. To be qualified, the MTurkers must indicate that they have watched at least 10 ASMR videos before, do not feel ASMR videos disturbing or unsatisfying, and experience a tingling sensation and relaxation after watching ASMR videos. To test participants' ASMR knowledge, a pre-screen question asked them to pick two typical ASMR videos from the other four non-ASMR videos. To ensure annotation quality, we only invited MTurkers who have completed more than 5000 tasks on MTurk with an approval rate greater than 97%. A simple math question and a question asking participants to choose two ASMR

videos from four video descriptions were deployed in each task as attention tests. MTurkers must answer both attention tests correctly to get the work accepted. Otherwise, the task is rejected and re-released to other participants.

Before annotating all the data, we first test the agreement between MTurk workers and the expert annotations. MTurkers also completed the 88 videos that the experts have annotated. The annotations of all subcategories between experts and MTurkers reached a substantial agreement, with kappa scores ranging from 0.67 to 0.80 (Table 3). At the end, the annotation was completed by 47 MTurk participants with an average completion time of 8.9 minutes. Each accepted task was paid at the rate of USD \$1.50. MTurkers report 167 videos with problems (99 unavailable, 28 age-restricted, and 40 non-English). After removing the 167 videos, 2663 videos are used for final data analysis. These videos have an average of 225,143.69 views ($SD = 654717.5$) and 5180.89 likes ($SD = 9554.74$).

4.3 Comment Data Processing

483,583 comments of the 2663 videos were collected for analysis, with each video having 181.59 comments on average ($SD = 87.81$). Among 2663 videos, 687 videos have 300 comments. The comment analysis uses two different text processing methods, LIWC [51] and PMI [8], to obtain words related to social connection and intimacy, sensory perception, and relaxation and sleepiness. For each video, we merge all the collected comments into one corpus and use LIWC to calculate the percentages of words related to social processes and sensory perception. Social processes include words like "mate," "talk," and "they" and all non-first-person-singular personal pronouns, as well as verbs that suggest human interaction (talking, sharing) [51]. Sensory perception includes bodily and perceptual words. The body category under biological processes contains words such as "cheek," "hands," and "spit." Perceptual processes recognize words related to perception, including "look," "heard," and "feeling." We also generate a text document with all comments combined and obtained the emotional tone and social-, body-, and perception-word percentages in the entire comment corpus. This step allows us to examine the overall sentiment in ASMR video comments and compare ASMR comments with the base rates of expressive writing, natural speech, and Twitter data [51].

LIWC does not provide words related to intimacy, relaxation, and sleep processes. Therefore, we use the pointwise mutual information (PMI) technique [8] to recognize words and phrases associated with those processes. Similar approaches have been widely applied in social media analysis (e.g., [14, 55]). The PMI measures the likelihood that two terms occur together in a corpus. The PMIs are calculated based on the 8,914,289 comments from the filtered 85,734 videos. Comments are pre-processed to remove stop words and punctuation to retain only meaningful words (including regular words and emojis). Words are then processed to find bigrams of common phrases. The keywords to generate associated word lists are “intimate,” “relax,” and “sleep.” To filter out too-rare and too-common words, identified associated words must appear in at least 1000 comments and no more than 1/10 of all comments. We choose the top 3% of the qualified terms with the highest PMIs as the word lists for each keyword. Table 4 lists example words associated with each keyword. Then we apply a similar approach as in LIWC to count the percentages of words from each list in each video’s comments.

Table 4: Example words associated with intimacy, relaxation, and sleep identified by PMI. The word list is used to count the percentages of words associated with each feeling.

Keyword	Number of words	Example words with top PMIs
intimate	85	intimate, intimacy, connection, sexual, relationships, sensual, romantic, personal, desire, events, atmosphere, relationship, emotionally, casual, decision, approach, scenario, witness, creation, client, private, audience, sacrifice, strangely, destiny, partner, remain, alternate, detail, interact
relax	157	relax, informative, unwind, tense, stressful day, wonderfully, educational, incredibly, stressful, entertain, stevie, strangely, lavender, atmosphere, 😊, extremely, meds, super, amazingly, movements, peaceful, manner, drift, stress, visually, surprisingly, serum, ambient, traditional
sleep	155	sleep, have trouble, peacefully, drift, schedule, clinic, nights, meds, pills, autoplay, 4am, insomnia, alarm, pill, tonight, trouble, brush teeth, aid, auto play, medication, 1am, night, induce, nightmares, wake, 2am, help, put, pm, 3am

4.4 Statistical Method

The visual, taste, scenario annotations, and the three parasocial subcategories are stored as multi-categorical nominal variables. Sound and touch, the two subcategories with multiple possible choices, are saved as dummy variables (1 is containing the interaction, otherwise 0). For RQ2, Pearson’s Chi-squared test (contingency table) is used to identify significant associations between multimodal interaction and parasocial attraction subcategories. For comment analysis in RQ3, we first perform regression analysis to identify multimodal and parasocial subcategories that significantly predict each feeling word percentage. For each feeling word, two least-squared regression (LSR) models are built, with one using multimodal interactions as independent variables and the other using parasocial attraction variables. The two models are multivariate regressions with all modality factors (or all attraction factors) serving as independent variables simultaneously. We perform posthoc analysis with the Steel-Dwass method for each significant factor to identify differences between subcategory pairs. Nonparametric comparisons are performed due to the word frequencies being not normally distributed. In all statistical testings, the significant threshold ($alpha = 0.05$) is adjusted with the Bonferroni method.

5 RESULTS

5.1 RQ1: Multimodal Interactions in ASMR Videos

RQ1 seeks to overview the interaction modalities used in ASMR videos and suggest how prevalent triggers are (see Figure 4 left). **Visual.** Face-to-face is the most common visual setting. Around 2/3 of videos present ASMRtists themselves in front of the camera. The ASMRtists also perform triggers in other modalities like making whispering and object sounds, manipulating objects, or reaching to the camera at the same time (see Figure 4 right). Mukbang, object only, and images are also common visual settings. Only a small proportion of ASMR videos use video games, animals, or other visual interactions. **Sound.** We notice ASMRtists tend to mix multiple sound effects in ASMR videos. Only 896 (33.65%) videos contain only one type of sound. 1081 (40.59%) of videos use two audio triggers. 687 (25.80%) videos use three or more different types of sound effects. The most common sound in ASMR videos is whispering and soft speaking. Other common sound effects include object sounds, mouth effects, body and cloth sounds, mic effects, and ambient sounds. **Touch.** More than half ($N = 1573, 59.07\%$) of ASMR videos use at least one touch trigger in the video. The most common touch interaction is touching objects in the video to generate tingling sounds. 29.1% of videos have ASMRtists reaching toward the camera and pretended to touch the viewers’ face or body. Less than 10% of videos contain touching ASMRtists’ own body parts or a person in the video. **Taste.** Tasting is not a commonly used interaction modality in ASMR videos. Only 12.35% of videos use tasting triggers, mostly in Mukbang videos (Figure 4 right). **Scenario.** 71.65% of videos do not use any roleplays in the videos. The most common roleplay is in service scenarios in which ASMRtists perform services or treatment processes. Less than 10% of videos are fantasy or romance roleplays.

The analysis of the distribution of multimodal interactions in ASMR videos suggests most ASMRtists choose to look at the camera to mimic face-to-face interactions with the viewer. ASMR videos are sound-diverse and rich. The most common sound interactions are whispering to the viewers and sounds made by manipulating trigger objects. Around 1/3 of ASMR videos use touch interaction by touching objects and/or touching the viewers. Taste interaction is used to perform Mukbang videos. The majority of the ASMR videos do not have roleplays and plotted scenarios. The most common ASMR scenario is service or treatment roleplays.

5.2 RQ2: Parasocial Attractions and ASMR Experience Patterns

RQ2 examines parasocial attractions in ASMR videos and their associated interaction modalities to evoke ASMR experiences. Figure 5 left illustrates their distribution, and Figure 5 right shows all the positive and negative associations.

5.2.1 ASMR for Social Experiences. The analysis of social attraction shows that 1843 (69.21%) of the videos have the ASMRtists talking to or talking with the viewer (Figure 5 left). Pearson’s chi-squared test suggests the face-to-face presentation (*visual.face_to_face*), whispering sounds (*sound.whispering*), and virtually touching the viewer through camera-reaching (*touch.viewer*) are significantly

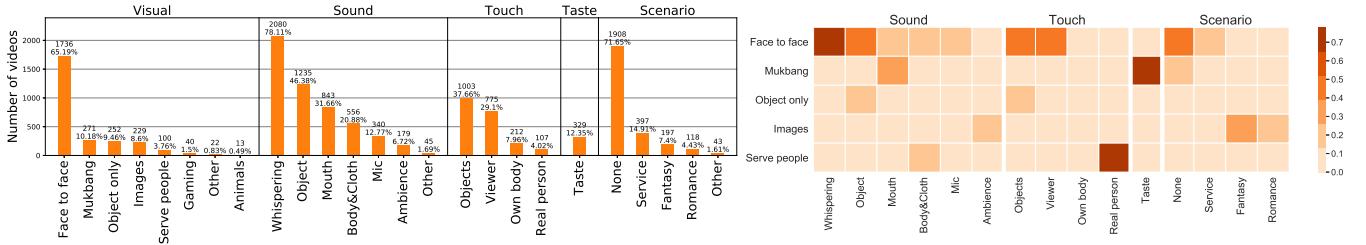


Figure 4: Left: The distribution of videos in the subcategories of the five interaction modalities. Right: The co-appearance rates of visual triggers and triggers of other modalities, calculated using the Jaccard index.

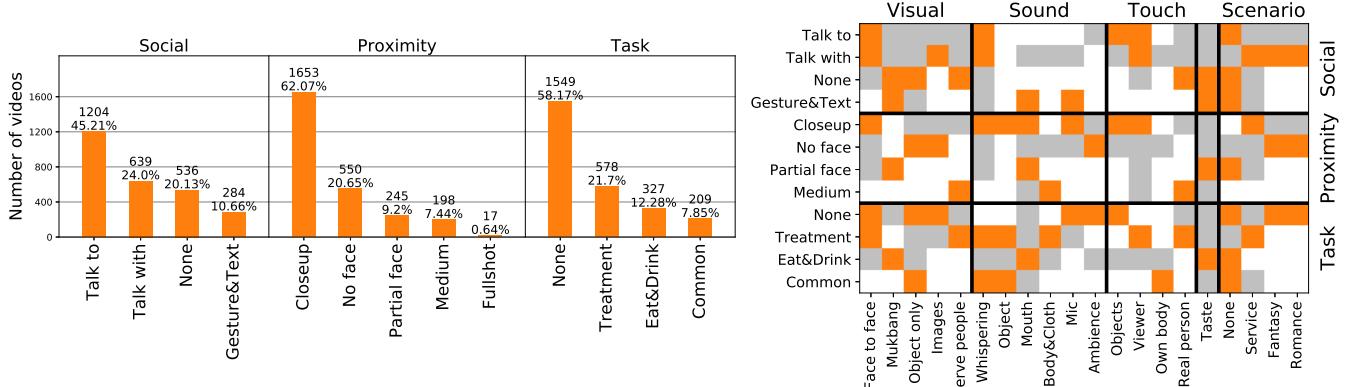


Figure 5: Left: the distribution of videos in subcategories of parasocial attractions. Right: The significant associations between subcategories of multimodal interactions and parasocial attractions. Orange squares are significant positive associations. Grey squares are significant negative associations.

associated with talk-to or and talk-with videos (*social.talk_to* and *social.talk_with*). Videos that talk to the viewers are also significantly associated with touching objects (*touch.objects*). Talking with the viewers is also associated with all three types of role-plays (Figure 5 right). Among all 1843 talk-to and talk-with videos, 1366 have the ASMRtists interacting with the viewers in “face-to-face” settings and making whispering sounds (e.g., Figure 6-a). 645 videos also pretend to touch the viewers through camera reaching (e.g., Figure 6-b). 504 videos talk to the viewers while the ASMRtists touch objects to make tingling sounds (e.g., Figure 6-c). Talk-with videos use scenarios in which the ASMRtist role-plays a service provider ($N = 273$, Figure 6-d), a fantasy character ($N = 108$, Figure 6-e), or an intimate partner ($N = 81$, Figure 6-f). Although most talk-with videos show the performer looking at the viewer face-to-face, 116 out of 639 talk-with videos are videos with static or no images (*visual.images*, e.g., Figure 6-f). Non-socializing videos are associated with Mukbang (*visual.mukbang*), object-only (*visual.object_only*), serving other people (*visual.serve_people*), touching another person in the video (*touch.real_person*), tasting (*taste*), and non-roleplays (*scenario.none*).

The high percentage of videos with conversational content suggests that social connection is a common experience incorporated by ASMRtists. ASMR performers deliver social connection experiences with multimodal interactions such as face-to-face whispering,

hand-reaching, and one-sided or back-and-forth conversation. The ASMRtists tend to touch objects to generate tingling sounds while talking to the viewers. ASMRtists also perform service, fantasy, and romantic roleplays to engage the viewer in an emulated conversation, in which many ASMRtists pretend that they can hear the viewer’s responses. Videos without socialization are videos of food eating, serving another person, or merely manipulating objects.

5.2.2 ASMR for Intimate Interaction. The result of spatial proximity suggests that the majority (1868, 70.15%) of the ASMR videos have the ASMRtists presenting their full faces in closeup, medium, and full shot camera distances (Figure 5 left). Only 550 (20.65%) videos do not have human appearances, and 245 (9.2%) videos show partial faces. Closeup is the most used shot scale used in ASMR videos, suggesting most ASMRtists seek to simulate nearness with the viewer by positioning themselves in close camera proximity. The association analysis suggests that face-to-face (*visual.face_to_face*), whispering sounds (*sound.whispering*), object sounds (*sound.object*), mouth sounds (*sound.mouth*), mic effects (*sound.mic*), and the touching of objects (*touch.objects*) and viewers (*touch.viewer*) significantly associate with closeup camera proximity (*proximity.closeup*, Figure 5 right). In all 1653 closeup videos, 1490 videos whispers to the viewer near the camera (e.g., Figure 7-a). 809 videos have ASMRtists making various trigger sounds (e.g., Figure 7-b). 591 and 245 closeup videos contain mouth effects (e.g., Figure 7-c) and mic



(a) The ASMRtist softly whispers to the viewer. (b) The ASMRtist talks with viewer and pretends to reach viewer's face in a treatment.



(c) The ASMRtist manipulates a bottle near the mic while intro-venting in a nerve exam on the viewer in a treatment roleplay.



(e) The ASMRtist pretends to recruit the viewer into their punk band in a fantasy roleplay. (f) The ASMRtist talks with the viewer in a romantic, image-only roleplay.

Figure 6: Example ASMR videos showing ASMR interactions for social attraction.

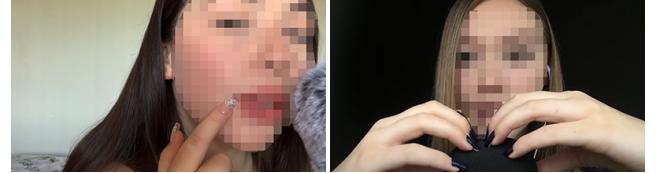
effects (e.g., Figure 7-d). For touch interactions, closeup videos also consist of manipulating objects (e.g., Figure 7-b) and pretending to touch the viewers (e.g., Figure 7-e). In these videos, ASMRtists make mouth sounds near the mic or interact with the mic to engender the sound of physical closeness. Although some videos do not show the ASMRtist on screen, 59 of them simulate intimate and romantic roleplays in their conversations with the viewers (Figure 7-f).

These results suggest that ASMRtists seek to use cameras and microphones to emulate intimate interactions with the viewers. Common approaches include positioning near the camera, near-mic whispering, manipulating objects, reaching hands to the camera, and making mouth and mic sounds. Even in videos without the performer's physical presence, ASMRtists can emulate intimate roleplays and conversations to express intimacy.

5.2.3 ASMR for Activity Observation. The analysis of task attraction shows most of the videos do not contain a clear task. More than half of the videos ($N = 1549, 58.17\%$) are *task.none*. The associations suggest three main types of videos that do not have specific tasks. The first type is ASMRtists using the face-to-face camera setting ($N = 1085$) and chatting with the viewer and/or making random sounds (e.g., Figure 8-a). The second main type is chatting in fantasy or romance roleplays ($N = 210$, e.g., Figure 6-f and 7-f). The third type is object-only videos ($N = 191$), in which ASMRtists manipulate trigger objects without meaningful purposes



(a) An ASMRtist whispers near the camera/mic to simulate close-ear speaking. (b) The ASMRtist manipulates pliers near the camera to generate tingling sound.



(c) The ASMRtist smacks lips near the mic to emulate close-ear mic and whispers trigger words feelings. (d) The ASMRtist scratches the ear near the mic while ASMRtist whispers trigger words (coconut, pickle, etc.).



(e) The ASMRtist pretends to do makeup on viewer's face while ASMRtist emulates cleaning viewer's ears as a sister. (f) An audio-only ASMR, where viewer is asleep.

Figure 7: Example ASMR videos showing ASMR interactions for intimacy.

(e.g., Figure 8-b and c). More than 40% of videos contain treatment, eating and drinking, and common daily tasks. Among 578 treatment videos, ASMRtists whisper to the viewer ($N = 507$), make object ($N = 317$) and body/cloth sounds ($N = 164$), and emulate service scenarios ($N = 381$). Treatment videos (*task.treatment*) are significantly associated with face-to-face (*visual.face_to_face*, $N = 456$) and touching the viewer (*touch.viewer*, $N = 304$), as those treatment videos pretend to perform the service on the viewer (e.g., Figure 6-b and d and Figure 7-e). Treatment videos (*task.treatment*) are also significantly associated with performing service on another person (*visual.serve_people*, $N = 85$) and touching them (*touch.real_person*, $N = 83$, e.g., Figure 8-d). 337 videos consist of tasks of eating and drinking (e.g., Figure 8-e and f), most of which present a large quantity of food in the video (*visual.mukbang*, $N = 259$) and make mouth sounds (*sound.mouth*, $N = 301$). Mukbang task is significantly associated with non-social (*social.none*) and partial face presentations (*proximity.partial_face*). The third major task is common daily activities (e.g., Figure 8-g and h). 216 (8.11%) videos show daily tasks in which ASMRtists perform everyday activities. Common daily tasks (*Task.common*) are significantly associated with object-only presentations (*visual.object_only*, $N = 43$), whispering sounds (*sound.whispering*, $N = 184$), and object sounds (*sound.object*, $N = 122$). Daily tasks (*task.common*) are also significantly associated with touching own body (*touch.own_body*,

$N = 38$) because these videos contain activities such as applying makeup (e.g., Figure 8-h).

These results indicate that ASMRtists tend to present activities without a clear, purposeful task. The taskless videos include mundane, repetitive, and unintentional actions, facing the viewers or only showing the trigger objects. Videos with particular tasks involve treating the viewer or another person, eating a large quantity of food, or other common everyday activities. Those activities are also considered soft, clicking, slow, or repetitive, which are likely to induce ASMR experiences [5].

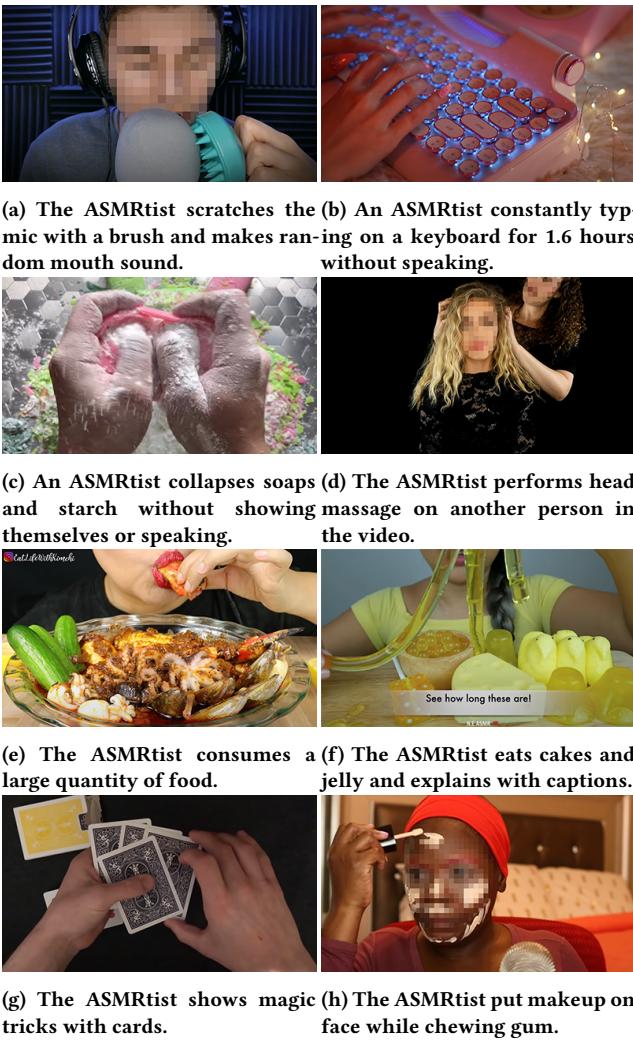


Figure 8: Example ASMR videos showing no-task and with-task ASMR interactions.

5.3 RQ3: Viewers' Comments about the Feelings of ASMR Experience

RQ3 explores viewers' feelings about different ASMR interactions by calculating the percentages of the LIWC and PMI identified keywords. We first compare the linguistic attributes of all ASMR video

comments with the base rates of expressive writing, natural speech, and Twitter data [51]. Figure 9 shows the results. The emotion tone score of all ASMR comments is 99 (50 is neutral), higher than the other three types of textual data, indicating viewers' overall positive reaction to the ASMR videos. ASMR comments have comparable social word frequencies, suggesting viewers have similar social expression as in other texts. The body words (7.68%) and perception words (7.63%) in ASMR comments are higher than the other three texts. This shows that viewers write more in the comments about things associated with body and perception processes. The overall positive emotion and high frequency of body and perception words imply that viewers obtained sensational pleasure from watching the ASMR videos (see example comments in Appendix A).

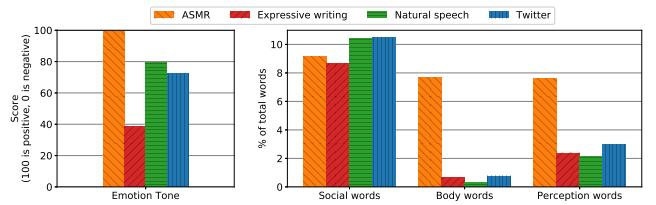


Figure 9: Comparison of emotion tone and percentages of social, body, and perception words between ASMR videos and expressive writing, natural speech, and Twitter [51]

5.3.1 Social and Intimacy. The LSR model that predicts social word frequencies by the parasocial attraction subcategories suggests that parasocial attraction factors are significant predictors (Table 5). The model that predicts social word frequencies by interaction modalities shows that visual and sound are significant predictors. The posthoc analysis shows that ASMR videos that leverage social attraction techniques lead to higher use of social words in the comments (Figure 10 top). The social word frequencies in talk-to and talk-with (*social.talk_to* and *social.talk_with*) videos are significantly higher than gesture/text videos (*social.gesture&text*) and non-social videos (*social.none*). Similarly, videos with ASMRtists whispering sounds have more social comments than videos without communication. For spatial proximity, videos with the ASMRtists being closeup (*proximity.closeup*), and medium distance (*proximity.medium*) have higher social word frequencies than videos without ASMRtists' appearances. With regard to visual modalities, videos that use visual settings of static images (mostly audio-only roleplays), face-to-face interaction, and serving people have higher social word frequencies than Mukbang and object manipulation videos. These results indicate that the social attraction techniques used in ASMR videos, such as talking to/with the viewers, showing themselves face-to-face, and whispering led viewers to express more social processes in the comments than non-social ASMR videos.

The LSR models which predict the intimate word frequency show that social, proximity, and task, are significant parasocial predictors (Table 5). Visual, sound, and scenario are the significant multimodal interaction predictors. The posthoc analysis shows that talking with or to viewers leads to more comments related to intimacy. Figure 10 bottom shows the results. Videos with whispering sounds also have significantly more intimate words in the

Table 5: Results of multivariate LSR models that predict the percentages of feeling words in comments. Only the p-values of significant predictors are presented. Adjusted $\alpha = 0.0083$. *Multi-choice factors, p is from the dummy variable with the smallest p-value.

Dependent variable	LSR of Parasocial Attractions						LSR of Multimodal Interactions							
	F	p	r^2	p_{social}	$p_{physical}$	p_{task}	F	p	r^2	p_{visual}	p_{sound}^*	p_{touch}^*	p_{taste}	$p_{scenario}$
Social words	32.35	<0.0001	0.11	<0.0001	<0.0001	-	18.44	<0.0001	0.13	<0.0001	<0.0001	-	-	-
Intimacy words	24.03	<0.0001	0.08	<0.0001	0.0012	<0.0001	16.92	<0.0001	0.12	0.0016	<0.0001	-	-	0.0001
Body words	49.36	<0.0001	0.16	<0.0001	<0.0001	<0.0001	39.03	<0.0001	0.25	<0.0001	<0.0001	0.0002	0.0060	0.0003
Perception words	29.55	<0.0001	0.10	<0.0001	<0.0001	0.0006	27.21	<0.0001	0.18	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Relax words	43.28	<0.0001	0.14	0.0003	<0.0001	<0.0001	30.42	<0.0001	0.20	<0.0001	<0.0001	<0.0001	0.00035	<0.0001
Sleep words	36.73	<0.0001	0.12	0.0065	0.0004	<0.0001	23.03	<0.0001	0.16	-	<0.0001	<0.0001	<0.0001	<0.0001

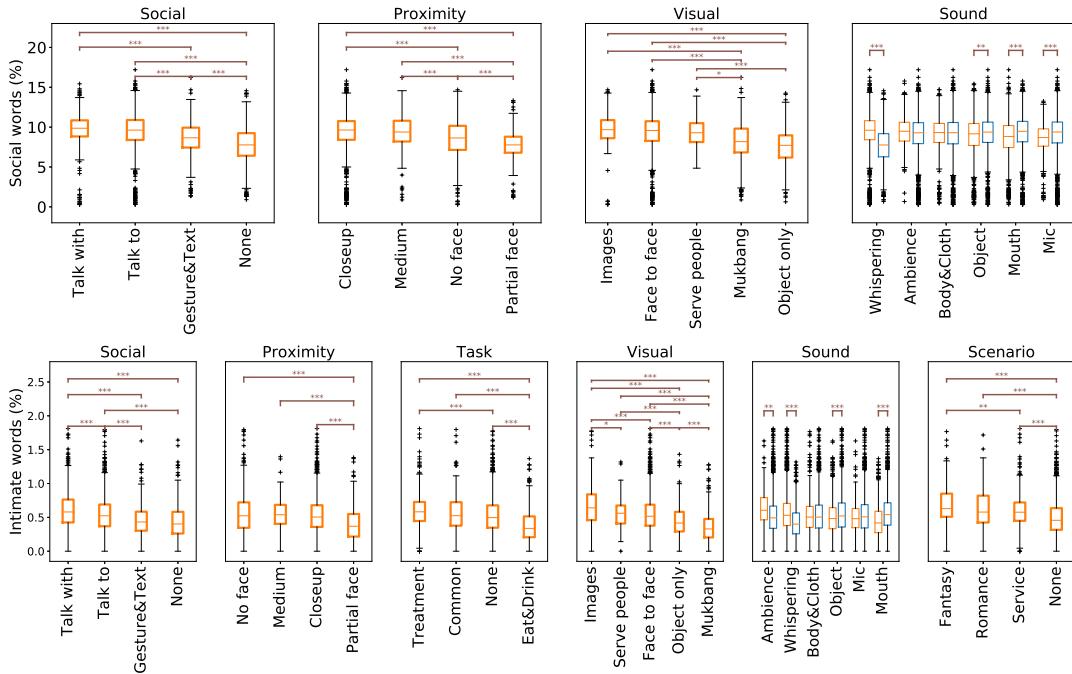


Figure 10: Social and intimate word frequencies between subcategories of significant multimodal and parasocial predictors. Ordered by the average percentage in descending order. Horizontal bars show significant differences ($p^* < 0.05$, $p^{} < 0.01$, $p^{***} < 0.001$).**

comments than videos without whispering. Mukbang videos have significantly lower intimate words in comments. The comparison of intimate words shows that comments to roleplay videos have more viewers' intimate expressions than videos without roleplays. Since many fantasy and romance roleplays are voice-only, videos without performers' faces (*proximity.no_face*), with static images (*visual.images*), and with ambient sounds (*sound.ambience*) have significantly higher numbers of intimate words in the comments.

The comparison of social and intimate words suggests that the ASMR videos with social interactions – such as presenting the ASMRtist in the videos and whispering to the viewers – are more likely to receive viewers' social responses than ASMR videos without socialization. Roleplays lead to more intimate reactions in the comments than non-roleplay videos. In contrast, Mukbang and object-only videos have lower social and intimate expressions.

5.3.2 Body and Perception. Viewers share their body and perceptual feelings or comment on ASMRtists' body or actions (see Appendix A). The LSR model suggests social, spatial, and task attractions can significantly predict the use of body and perception words (Table 5). All multimodal interaction subcategories are significant predictors of body words. Visual, sound, touch, and scenario are significant predictors of perception words. Posthoc shows that Mukbang videos lead to the highest sensory words in comments (Figure 11). Eating or drinking videos (*task.eat&drink*) have significantly more sensory responses than other task-oriented and taskless videos. Comments to videos that contain mouth sounds have significantly more body and perception words. Taste interactions lead to significantly more body words. Since Mukbang videos tend to show partial faces and only use gestures and text to communicate (e.g., Figure 8-e and f), the gesture/text (*social.gesture&text*)

and the partial face (*proximity.partial_face*) videos have the highest sensory word use in the comments. We also noticed that videos without tasks (*task.none*) have higher body words. Non-roleplay videos (*scenario.none*) have the highest usages of body and perception words. Object-only videos (*visual.object_only*, e.g., Figure 8-b and c) also have a higher mentioning of the body and perception words. Videos with object sounds (*sound.object*, e.g., Figure 6-c and Figure 7-b) and mic sounds (*sound.mic*, e.g., Figure 7-d and 8-a) have significantly more comments with body and perception words. These results imply that presenting and consuming a large quantity of food in ASMR videos, as well as videos without tasks or roleplays are more likely to induce viewers' feelings related to sensory perception.

5.3.3 Relaxation and Sleepiness. The LSR model that predicts relaxation words shows that all parasocial and multimodal subcategories are significant predictors (Table 5). The model that predicts sleep words suggests that all parasocial and multimodal subcategories except for visual interaction are significant predictors. Posthoc analysis between different categories shows that videos related to treatment and intimate interactions have the highest percentage of relax and sleep words in comments (Figure 12). Videos with treatment performance (*task.treatment* and *scenario.service*) have the highest percentages for both measurements. Videos that involve performing services on another person (*visual.serve_people* and *touch.real_people*) also lead to more relaxation expression. Videos that pretend to touch the viewer by reaching toward the camera (*touch.viewer*) have significantly more sleep words than videos without this interaction. ASMR videos with mic sounds (*sound.mic*), in which ASMRtists get close up to the camera and make near-ear mic sound effects (e.g., Figure 7-d and Figure 8-a), lead viewers to comment more about relaxation and sleepiness.

These results imply that videos showing treatment processes and physical intimacy induce feelings of relaxation and sleepiness for viewers more often. Videos with near-ear microphone effects also incite feelings of relaxation and sleepiness. It should be noted that although the visual and touch settings of treatment ASMRs seek to simulate physical intimacy with the viewers, viewers do not express more intimacy in the comments. Instead, the emulation of close proximity interactions in treatments lets viewers express more relaxation and sleepiness.

6 DISCUSSION

The analysis of multimodal interactions and parasocial attractions describes the common patterns used to induce ASMR experiences through audio-visual media. Our work depicts common ASMR interactions but does not contrast their effects with other online content. This section summarizes that the ASMRtists deliver ASMR effects through three experience patterns: multimodal social connection, relaxing physical intimacy, and sensory-rich activity observation.

6.1 ASMR as an Experience of Multimodal Social Connection

Prior research primarily examine ASMR triggers' characteristics, and their different physiological effects on the viewers [6, 54, 56]. Although ASMR videos are considered as a new pathway to connect creators and their viewers [40], there is little knowledge regarding

what specific interactions ASMRtists perform to best establish social connections. We find the social experience in YouTube ASMR videos is commonly offered and multi-modeled. Around 65% of videos in our data contain the performer looking at the camera, which gives the illusion of a "face-to-face" interaction between the actor and the viewer, and 78% of videos involve whispering. 70% of ASMR videos have ASMRtists communicating verbally to the viewer, with 24% of videos pretending that the performer can hear viewers' reply (talk-with videos). 59.07% of videos used at least one type of touch interaction. The pervasive use of conversational content reveals that sound effects are not the only drivers of ASMR; ASMRtists engage viewers and induce ASMR through experiences of one-sided social connection. These results are consistent with the significance of face-to-face interactions noted in prior research [65]. In a multimodal conversation, the performer faces the viewer, communicates in whispers, touches the viewers through camera reaching, introduces and manipulates triggers, and emulates imagined scenarios. Since ASMR needs to be triggered with appropriate stimuli [5], and because not all triggers "work" for all viewers, the diverse modalities allow viewers to try out and encounter triggers that can bring ASMR sensations. The social interactions could also foster the feeling of co-presence with the ASMR performer [71]. Interaction modalities such as whispering with/to the viewers and being spatially close up to the camera lead viewers to write more about the social processes in the comments. Even in audio-only videos without visual presentations, ASMRtists play fantasy and romantic roles and chat with the viewers in stories. The analysis of viewer comments suggests that viewers tend to leave more intimate comments to videos with those ASMR components.

These findings imply new pathways to design parasocial experiences with ASMR effects. ASMR interaction techniques can provide social exposure that increases closeness in asynchronous video communication. Video-based technologies incorporating ASMR effects and multimodal ASMR interactions may augment parasocial connection experiences. Since ASMR may offer positive affect, as well as intimate and relaxing experiences for viewers [3, 53], face-to-face video communications can leverage ASMR interactions to transfer the process of speaking-listening to a richer experience with tingling sensations. The social experience pattern captured from ASMRtists' videos implies that technologies can incorporate ASMR interactions in multiple modalities such as whispering, camera-reaching, emulated back-and-forth conversation, and trigger manipulation in order to induce viewers' ASMR sensations. Users need both time and variety in order to see if tingles develop, and the multi-modalities allow for that temporal unfolding and variety. For example, applications such as video conferencing tools, podcasts, and social audio apps can potentially introduce multimodal ASMR to reduce the exhaustion and fatigue from long-time use [44]. Voice-based virtual assistants [50] may also include ASMR effects to reduce the robotic sound.

6.2 ASMR as an Experience of Relaxing Physical Intimacy

Leveraging attraction and interaction techniques to demonstrate intimacy is also a typical pattern in ASMRtists' videos. Prior research has explored ASMR as an experience of digital intimacy

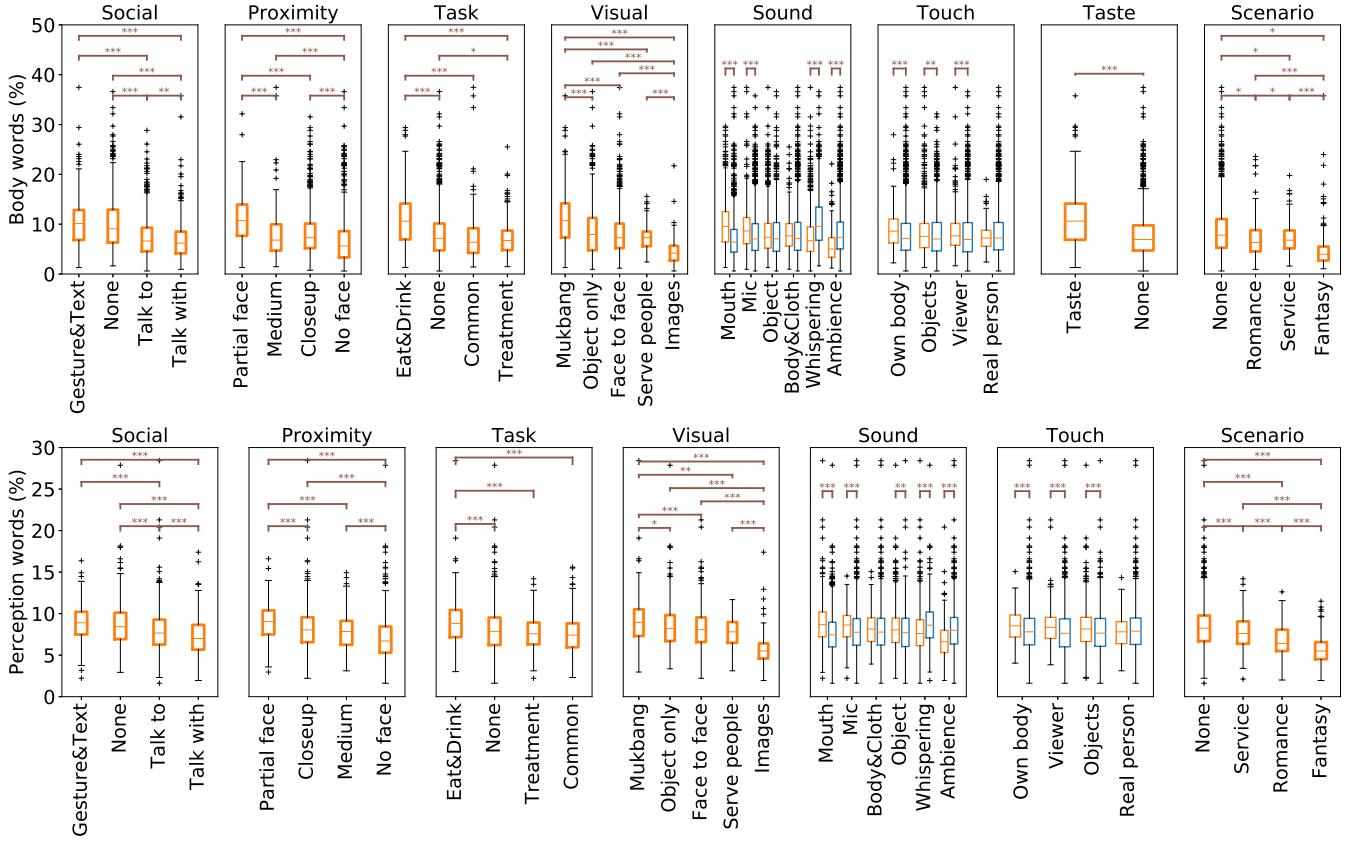


Figure 11: Body and perception word frequencies between subcategories of significant multimodal and parasocial predictors. Ordered by the average percentage in descending order. Horizontal bars show significant differences ($p^* < 0.05$, $p^{} < 0.01$, $p^{***} < 0.001$).**

[3, 31] as well as the ways ASMRtists create roleplay videos to foster intimate feelings with the viewers [71]. This paper overviews ASMRtists' techniques to design intimate experiences and how these techniques relax viewers and help with sleeping. We find that the most common camera shot type in ASMR videos is closeup – framing the performer's face at a near distance while excluding most of their body. Around 30% of videos have the ASMRtists pretending to touch the viewers through camera reaching. About 30% of videos also make close-mic mouth sounds, and 12% manipulate the microphone itself to simulate physical intimacy through sound interactions. These interactions are commonly performed in service-oriented videos such those involving massage, haircuts, makeup applications, etc. However, our comment analysis suggests that viewers do not express more intimacy to videos with intimate interactions than other videos. On the other hand, videos with close interactions have more comments regarding relaxation and sleepiness-related words. Our results imply that although ASMRtists virtually approach the viewers, viewers expressed relaxing and calming experiences more than intimacy to such videos.

Our findings suggest new opportunities to design ASMR-based applications to present intimacy and deliver soothing experiences. For example, ASMR interactions allow service providers such as

masseurs and Reiki masters to offer virtual treatments through ASMR videos. This virtual therapy could provide a possible solution when in-person service is unavailable, or for users who cannot afford in-person treatment. People separated from loving relationships [21] or patients living in stressful hospital settings [67] need intimate interactions. ASMR effects with close-mic whispers and near-camera touching could potentially engender a feeling of intimacy to induce relaxing experiences. Virtual social encounters with ASMR performers could also provide alternatives for people with social difficulties (e.g., due to autism or social anxiety) to enjoy safe, calm, regularized social experiences on demand [29]. To augment such experiences, designers can create new ASMR video interactions. For example, ASMRtists use the talk-with and camera-reaching techniques to mimic physical proximity. Novel interaction techniques such as VR, AR, and other telepresence technologies can be integrated to augment the social and virtual presence during ASMR videos. However, we want to remind the HCI community of the existence of sexual ASMR videos (S-ASMR) that are intentionally made for sexuality and sexual arousal [65]. The design for intimacy needs to differentiate ASMR videos from S-ASMR videos, especially to avoid young kids accessing an S-ASMR video without parental guidance.

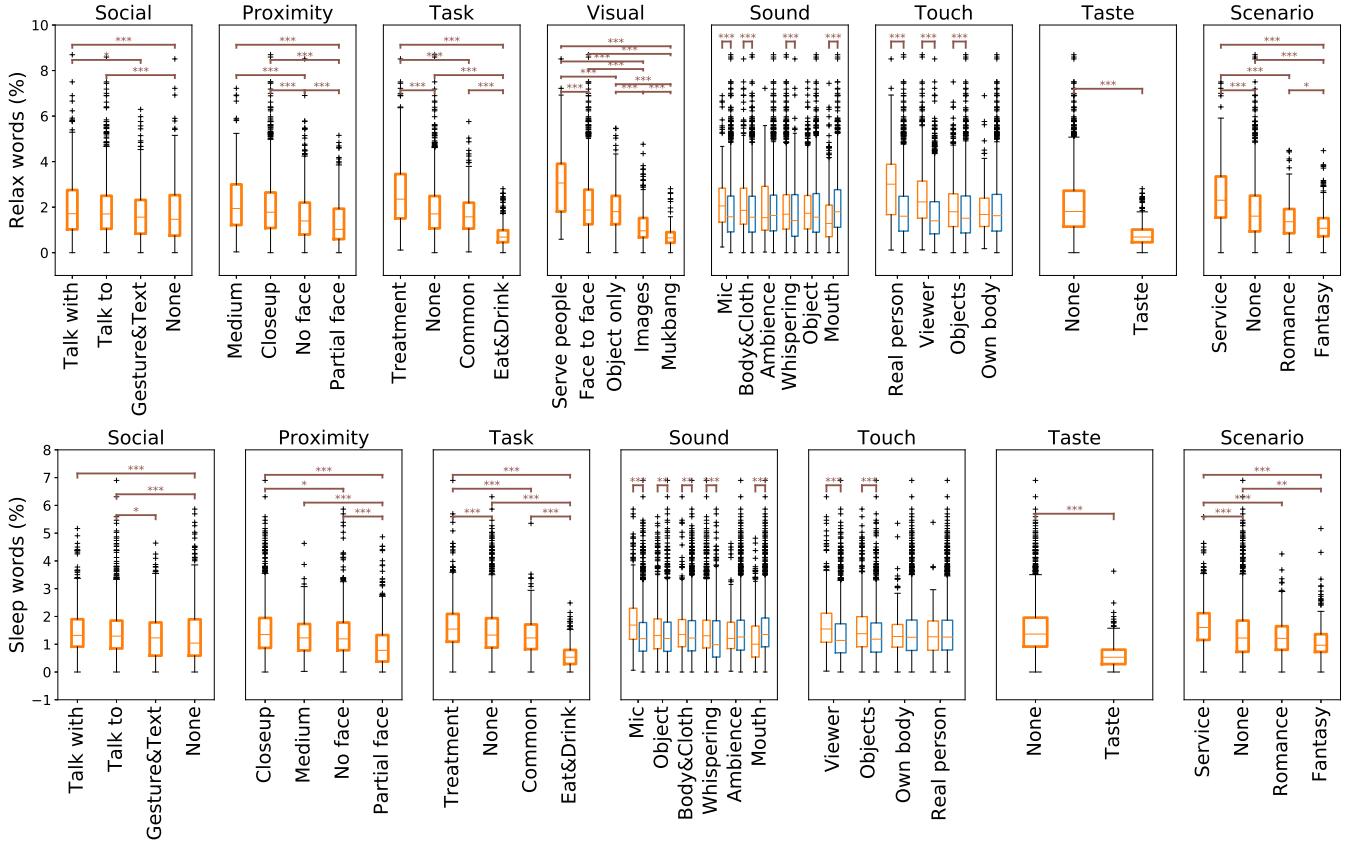


Figure 12: Relax and sleep word frequencies between subcategories of significant multimodal and parasocial predictors. Ordered by the average percentage in descending order. Horizontal bars show significant differences ($p^* < 0.05$, $p^{} < 0.01$, $p^{***} < 0.001$).**

6.3 ASMR as an Experience of Sensory-rich Activity Observation

Prior research have studied roleplay ASMRs as a primary type [35, 41, 65, 68, 71]. However, our findings suggest that more than 70% of videos in our data do not have roleplay scenarios. Also, in contrast to the wide use of social attraction and spatial proximity, most ASMRtists' videos do not use task attractions to elicit ASMR experiences. Only around 40% of videos in our dataset have identifiable tasks and goals. These numbers indicate that intentional ASMR videos are not limited to roleplays; future work should include the diverse non-roleplays and taskless videos when examining ASMR performance and effects. Videos with tasks include the performance of various physical treatments, eating a large quantities of food, and displaying mundane activities such as playing cards or putting on makeup. Taskless videos can involve casual chatting or object or mic manipulation without showing the performer. The infrequent appearance in videos of purposeful tasks implies that ASMR effects do not require attention or real acts of care to take place. Therefore, many ASMRtists choose not to demonstrate abilities by completing tasks or making clear storylines, but instead remain focused exclusively on the production of triggering effects. The analysis of viewer comments further reveals that eating/drinking videos

and videos without tasks or roleplays are associated with viewers' comments about the body and perceptual processes, indicating that these videos are prone to trigger bodily and perceptual experiences.

Although personal attention, care, and intimacy are common elicitors of ASMR [3], our findings suggest many ASMRtists also adopt the "taskless" activities in ASMR videos. Those videos don't particularly care about addressing viewers, adopting a stance of calculated indifference, and this disinvested attitude is designed to induce ASMR feelings. Therefore, videos that does not require close attention except for observing peaceful and repetitive activities – videos such as crafting process demonstrations, instructions for applying makeup or skincare, and tutorials on organizing everyday objects – may consider employing ASMR effects. Prior research suggested that ASMR is an ambient sensory effect in YouTube study-with-me videos [37]. Videos like these can potentially reduce the human presence and intentionally make tingling sounds in the background to trigger ASMR feelings. However, videos that include slow and dull tasks may evoke ASMR feelings unintentionally and could make viewers lose focus and feel sleepy. In those cases, ASMR may need to be avoided if the video is geared toward learning and requires attention. Designers may also consider conveying sensory-rich experiences through Mukbang ASMR or sound-focused ASMR.

Watching food-eating videos has shown benefits to mitigate homesickness [30]. People watch Mukbang videos to gain multi-sensory immersion and “commensality.” [4] ASMR can be a sensory experience incorporated in human-food interaction [15, 68]. Interaction designers can induce ASMR experiences by mouth and mic sounds to augment sensory pleasure. Technologies for sensory reality and relaxation (e.g., virtual reality for anxiety-related problems [48]) can incorporate ASMR techniques such as eating/drinking sounds or sound-focused scenes to induce sense of presence and relaxing experiences.

7 CONCLUSION AND FUTURE WORK

This work analyzed the multimodal interactions and parasocial experiences in 2663 YouTube ASMR videos. We annotated how ASMRtists use visual, sound, touch, taste, and roleplay triggers to deliver social, physical, and task attractiveness. We obtained the distribution of ASMR interaction modalities and parasocial attractions. The associations between interaction modalities and parasocial attractions reveal patterns of ASMR experiences. Feeling-oriented words were recognized from viewer comments in order to probe whether different ASMR interactions lead to different viewer feelings. Face-to-face orientation, whispering sounds, and touching objects are the most interaction modalities. Social interactions are common and multi-modeled. ASMRtists implement social attractions and virtually proximate the viewers, but most ASMR videos do not involve roleplays or contain purposeful tasks. Our results summarize that YouTube ASMR videos provide three experiences: multimodal social connection, relaxing physical intimacy, and sensory-rich activity observation. These experiential descriptions seek to foster future media productions on a wide array of platforms that include ASMR interactions and effects.

Moving forward, we hope this work serves as a seminal study to inspire more ASMR-augmented designs. There are also many open-ended questions to be addressed by HCI researchers and practitioners. First, one limitation of this work is that we only consider intentional ASMR created and shared by YouTube ASMRtists to induce ASMR experiences specifically. Prior studies noticed viewers also experience ASMR with videos such as Bob Ross’ *The Joy of Painting* and a recording of *Lectures on Quantum Field Theory* [20, 41], which are not made for ASMR but contains ASMR properties. We did not include unintentional videos without “ASMR” labels due to difficulties recognizing and collecting them from YouTube. We also consider intentional ASMR interactions to be purposefully designed and performed; therefore, easier to be adopted in design. Future research may compare and contrast the effects of the two ASMR video types. Second, this work does not interview actual viewers to obtain their in-situ feelings of ASMR interactions; viewers’ reactions to different ASMR interactions were obtained from video comments. It is possible that viewers do not externalize all of their feelings of intimacy or relaxation in comments. However, we believe this work provides an overview of ASMR interaction techniques that can guide future studies to examine ASMR-based intimacy and well-being in various use cases [39]. Future research needs to assess the actual effects of ASMR interactions of different people and in different contexts, especially when ASMR interactions are designed for people with social anxiety or disabilities.

Third, YouTube creators contribute vernacular creativity [9] to build parasocial relationships. HCI researchers should consider interviewing ASMRtists or involving them in participatory design to understand their preferences and difficulties in managing parasocial interactions. Last, the growing ASMR communities across different cultures [4, 37] encourages HCI studies to examine how ASMR videos affect the creator-viewer communications and relationships. It is valuable to expand ASMR research to non-English videos to have a cross-cultural understanding of ASMR.

REFERENCES

- [1] Nitin K Ahuja. 2013. "It feels good to be measured": clinical role-play, Walker Percy, and the tingles. *Perspectives in biology and medicine* 56, 3 (2013), 442–451. <https://doi.org/10.1353/pbm.2013.0022>
- [2] Muhammad Abubakar Alhassan and Diane Pennington. 2020. Detecting Critical Responses from Deliberate Self-Harm Videos on YouTube. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*. Association for Computing Machinery, New York, NY, USA, 383–386. <https://doi.org/10.1145/3343413.3378002>
- [3] Joceline Andersen. 2014. Now You've Got the Shivers: Affect, Intimacy, and the ASMR Whisper Community. *Television & New Media* 16, 8 (11 2014), 683–700. <https://doi.org/10.1177/1527476414556184>
- [4] Laurens Anjani, Terrance Mok, Anthony Tang, Lora Oehlberg, and Wooi Boon Goh. 2020. Why Do People Watch Others Eat Food? An Empirical Study on the Motivations and Practices of Mukbang Viewers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376567>
- [5] Emma L Barratt and Nick J Davis. 2015. Autonomous Sensory Meridian Response (ASMR): a flow-like mental state. *PeerJ* 3 (2015), e851.
- [6] Emma L Barratt, Charles Spence, and Nick J Davis. 2017. Sensory determinants of the autonomous sensory meridian response (ASMR): understanding the triggers. *PeerJ* 5 (2017), e3846. <https://doi.org/10.7717/peerj.3846>
- [7] Ava Bartolome, Nguyen Binh Ha, and Shuo Niu. 2021. Investigating Multimodal Interactions and Parasocial Attractiveness in YouTube ASMR Videos. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21)*. Association for Computing Machinery, New York, NY, USA, 14–18. <https://doi.org/10.1145/3462204.3481763>
- [8] Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL* 30 (2009), 31–40.
- [9] Jean Burgess. 2012. YouTube and the formalisation of amateur media. In *Amateur Media*. Routledge, 69–74.
- [10] Jean Burgess and Joshua Green. 2018. *YouTube: Online video and participatory culture*. John Wiley & Sons, Cambridge, UK.
- [11] Sean. Cannell and Benji. Travis. 2018. *YouTube secrets : the ultimate guide to growing your following and making money as a video influencer*. Lioncrest Publishing, 192 pages. <https://www.amazon.com/YouTube-Secrets-Ultimate-Following-Influencer-ebook/dp/B07GL48G7L>
- [12] Kathy Charmaz and Linda Liska Belgrave. 2015. Grounded Theory. <https://doi.org/10.1002/9781405165518.wbeosg070.pub2>
- [13] Harry Cheadle. 2012. ASMR, the Good Feeling no one can Explain. *VICE*. <http://www.vice.com/read/asmr-the-good-feeling-no-one-can-explain> (accessed October 15, 2014) (2012).
- [14] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. *Icwsm* 13 (2013), 1–10.
- [15] Jialin Deng, Yan Wang, Carlos Velasco, Ferran Altarriba Altarriba Bertran, Rob Comber, Marianna Obrist, Katherine Isbister, Charles Spence, and Florian 'Floyd' Mueller. 2021. The Future of Human-Food Interaction. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411763.3441312>
- [16] Yvonne D Eaves. 2001. A synthesis technique for grounded theory data analysis. *Journal of Advanced Nursing* 35, 5 (9 2001), 654–663. <https://doi.org/10.1046/j.1365-2648.2001.01897.x>
- [17] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding Topic Signals in Large-Scale Text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4647–4657. <https://doi.org/10.1145/2858036.2858535>
- [18] Beverley Fredborg, Jim Clark, and Stephen D Smith. 2017. An Examination of Personality Traits Associated with Autonomous Sensory Meridian Response (ASMR). *Frontiers in Psychology* 8 (2017), 247. <https://www.frontiersin.org/article/10.3389/fpsyg.2017.00247>

[19] Beverley K Fredborg, James M Clark, and Stephen D Smith. 2018. Mindfulness and autonomous sensory meridian response (ASMR). *PeerJ* 6 (8 2018), e5414. <https://doi.org/10.7717/peerj.5414>

[20] Rob Gallagher. 2016. Eliciting Euphoria Online: The Aesthetics of “ASMR” Video Culture. *FILM CRITICISM* 40, 2 (6 2016). <https://doi.org/10.3998/fc.13761232.0040.202>

[21] Yumei Gan. 2021. Capturing love at a distance: Multisensoriality in intimate video calls between migrant parents and their left-behind children. *Social Interaction. Video-Based Studies of Human Sociality* 4, 3 (2021). <https://doi.org/10.7146/si.v4i3.128148>

[22] David C Giles. 2002. Parasocial Interaction: A Review of the Literature and a Model for Future Research. *Media Psychology* 4, 3 (8 2002), 279–305. https://doi.org/10.1207/S1523785XMEP0403_04

[23] Paula Clare Harper. 2020. ASMR: bodily pleasure, online performance, digital modality. *Sound Studies* 6, 1 (1 2020), 95–98. <https://doi.org/10.1080/20551940.2019.1681574>

[24] Tilo Hartmann. 2017. Parasocial interaction, parasocial relationships, and well-being. In *The Routledge handbook of media use and well-being: International perspectives on theory and research on positive media effects*. Routledge/Taylor & Francis Group, New York, NY, US, 131–144. <https://doi.org/10.4324/9781315714752>

[25] Marc Hassenzahl. 2010. Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics* 3, 1 (1 2010), 1–95. <https://doi.org/10.2200/S00261ED1V01Y201003HCl008>

[26] Donald Horton and R Richard Wohl. 1956. Mass Communication and Para-Social Interaction. *Psychiatry* 19, 3 (8 1956), 215–229. <https://doi.org/10.1080/00332747.1956.11023049>

[27] Mu Hu, Mingli Zhang, and Nuan Luo. 2016. Understanding participation on video sharing communities: The role of self-construal and community interactivity. *Computers in Human Behavior* 62 (2016), 105–115. <https://doi.org/10.1016/j.chb.2016.03.077>

[28] Agnieszka B Janik McErlean and Michael J Banissi. 2017. Assessing Individual Variation in Personality and Empathy Traits in Self-Reported Autonomous Sensory Meridian Response. *Multisensory Research* 30, 6 (2017), 601–613. <https://doi.org/10.1163/22134808-00002571>

[29] Isabel L Kampmann, Paul M G Emmelkamp, and Nekhmedin Morina. 2016. Meta-analysis of technology-assisted interventions for social anxiety disorder. *Journal of Anxiety Disorders* 42 (2016), 71–84. <https://doi.org/10.1016/j.janxdis.2016.06.007>

[30] Ryan M Kelly, Yueyang Cheng, Dana McKay, Greg Wadley, and George Buchanan. 2021. “It’s About Missing Much More Than the People”: How Students Use Digital Technologies to Alleviate Homesickness. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI ’21)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445362>

[31] Helle Breth Klausen. 2019. ‘Safe and sound’: What technologically-mediated ASMR is capable of through sound. *SoundEffects - An Interdisciplinary Journal of Sound and Sound Experience* 8, 1 (2019), 87–103. <https://www.soundeffects.dk/article/view/115035>

[32] Josephine Klefeker, Libi Striegel, and Laura Devendorf. 2020. What HCI Can Learn from ASMR: Becoming Enchanted with the Mundane. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376741>

[33] Aleksandra Kovacevich and David Huron. 2019. Two studies of autonomous sensory meridian response (ASMR): The relationship between ASMR and music-induced frisson. *Empirical musicology review* 13, 1-2 (2019), 39–63.

[34] Kate Szer Kurtin, Nina O’Brien, Deya Roy, and Linda Dam. 2018. The Development of Parasocial Interaction Relationships on YouTube. *The Journal of Social Media in Society; Vol 7, No 1 (2018): The Journal of Social Media in Society* 7 (5 2018), 233–252. <https://www.thejsms.org/index.php/TSMRI/article/view/304>

[35] Joanna Łapińska. [n.d.]. Posthuman and Post-Cinematic Affect in ASMR “Fixing You” Videos. *LIVING AND THINKING IN THE POSTDIGITAL WORLD* ([n. d.]), 153.

[36] Joanna Łapińska. 2020. CREATIVITY OF HUMAN AND NON-HUMAN MATTER INTERWOVEN: AUTONOMOUS SENSORY MERIDIAN RESPONSE VIDEOS IN A POSTHUMAN PERSPECTIVE. *Creativity Studies* 13, 2 (5 2020), 336–350. <https://doi.org/10.3846/cs.2020.11703>

[37] Yoonjoo Lee, John Joon Young Chung, Jean Y Song, Minsuk Chang, and Juho Kim. 2021. Personalizing Ambience and Illusionary Presence: How People Use “Study with Me” Videos to Create Effective Studying Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI ’21)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445222>

[38] Mengjie Liu and Qiang Zhou. 2019. A Preliminary Compilation of a Digital Video Library on Triggering Autonomous Sensory Meridian Response (ASMR): A Trial Among 807 Chinese College Students. *Frontiers in Psychology* 10 (2019), 2274. <https://www.frontiersin.org/article/10.3389/fpsyg.2019.02274>

[39] Anna M Lomanowska and Matthieu J Guitton. 2016. Online intimacy and well-being in the digital age. *Internet Interventions* 4 (2016), 138–144. <https://doi.org/10.1016/j.invent.2016.06.005>

[40] Jessica Maddox. 2020. What do creators and viewers owe each other? Micro-celebrity, reciprocity, and transactional tinges in the ASMR YouTube community. *First Monday* 26, 1 SE - Articles (12 2020). <https://doi.org/10.5210/fm.v26i1.10804>

[41] Hugh S Manon. 2018. ASMR Mania, Trigger-Chasing, and the Anxiety of Digital Repletion. Springer International Publishing, Cham, 227–248. https://doi.org/10.1007/978-3-319-63817-1_12

[42] Allison Mooney, Jason Klein, Consumer Goods, and Consumer Trends. 2016. ASMR Videos Are the Biggest YouTube Trend You’ve Never Heard Of. *Google, Sept* (2016).

[43] Ricardo Morales, Daniela Ramirez-Benavides, and Mario Villena-Gonzalez. 2021. Autonomous Sensory Meridian Response self-reporters showed higher scores for cognitive reappraisal as an emotion regulation strategy. *PeerJ* 9 (2021), e11474. <https://doi.org/10.7717/peerj.11474>

[44] Robby Nadler. 2020. Understanding “Zoom fatigue”: Theorizing spatial dynamics as third skins in computer-mediated communication. *Computers and Composition* 58 (2020), 102613. <https://doi.org/10.1016/j.compcom.2020.102613>

[45] Iida Nissinen. 2018. Tingles, tea boxes and sticky sounds—ASMR read diffractively with agential realism. (2018).

[46] Shuo Niu, Ava Bartolome, Cat Mai, and Ha B Nguyen. 2021. #Stayhome #withme: How do youtubers help with covid-19 loneliness? In *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411764.3445397>

[47] Shuo Niu, Cat Mai, Katherine G McKim, and D Scott McCrickard. 2021. #TeamTrees: Investigating How YouTubers Participate in a Social Media Campaign. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (10 2021). <https://doi.org/10.1145/3479593>

[48] Theodore Oing and Julie Prescott. 2018. Implementations of Virtual Reality for Anxiety-Related Disorders: Systematic Review. *JMIR Serious Games* 6, 4 (2018), e10965. <https://doi.org/10.2196/10965>

[49] Sharon Oviatt. 1999. Ten Myths of Multimodal Interaction. *Commun. ACM* 42, 11 (11 1999), 74–81. <https://doi.org/10.1145/319382.319398>

[50] Emmi Parviainen and Marie Louise Juul Søndergaard. 2020. Experiential Qualities of Whispering with Voice Assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376187>

[51] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.

[52] Giulia Poerio. 2016. Could Insomnia Be Relieved with a YouTube Video? The Relaxation and Calm of ASMR. Springer International Publishing, Cham, 119–128. https://doi.org/10.1007/978-3-319-45264-7_15

[53] Giulia Poerio. 2016. Could Insomnia Be Relieved with a YouTube Video? The Relaxation and Calm of ASMR. In *The Restless Compendium : Interdisciplinary Investigations of Rest and Its Opposites*, Felicity Callard, Kimberley Staines, and James Wilkes (Eds.). Springer International Publishing, Cham, 119–128. https://doi.org/10.1007/978-3-319-45264-7_15

[54] Giulia Lara Poerio, Emma Blakey, Thomas J Hostler, and Theresa Veltri. 2018. More than a feeling: Autonomous sensory meridian response (ASMR) is characterized by reliable changes in affect and physiology. *PLOS ONE* 13, 6 (6 2018), e0196645. <https://doi.org/10.1371/journal.pone.0196645>

[55] Eugenia Ha Rim Rho and Melissa Mazmanian. 2020. Political Hashtags & the Lost Art of Democratic Discourse. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376542>

[56] Craig Richard. 2018. *Brain tingles: The secret to triggering autonomous sensory meridian response for improved sleep, stress relief, and head-to-toe euphoria*. Simon and Schuster.

[57] Dana Rotman, Jennifer Golbeck, and Jennifer Preece. 2009. The Community is Where the Rapport is – on Sense and Structure in the YouTube Community. In *Proceedings of the Fourth International Conference on Communities and Technologies (C&T ’09)*. Association for Computing Machinery, New York, NY, USA, 41–50. <https://doi.org/10.1145/1556460.1556467>

[58] Rebecca B Rubin and Michael P McHugh. 1987. Development of parasocial interaction relationships. *Journal of Broadcasting & Electronic Media* 31, 3 (6 1987), 279–292. <https://doi.org/10.1080/08838158709386664>

[59] Kristen M Scott, Simone Ashby, and Julian Hanna. 2020. “Human, All Too Human”: NOAA Weather Radio and the Emotional Impact of Synthetic Voices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3313831.3376338>

[60] RAJEEV SHARMA, VLADIMIR I PAVLOVIĆ, and THOMAS S HUANG. 2002. Toward Multimodal Human-Computer Interface. In *Advances in Image Processing and Understanding*. Series in Machine Perception and Artificial Intelligence, Vol. Volume 52. WORLD SCIENTIFIC, 349–365. https://doi.org/doi:10.1142/9789812776952_0014

[61] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. 2010. How Useful Are Your Comments? Analyzing and Predicting YouTube Comments and Comment Ratings. In *Proceedings of the 19th International Conference on*

World Wide Web (WWW '10). Association for Computing Machinery, New York, NY, USA, 891–900. <https://doi.org/10.1145/1772690.1772781>

[62] Naomi Smith and Anne-Marie Snider. 2019. ASMR, affect and digitally-mediated intimacy. *Emotion, Space and Society* 30 (2019), 41–48. <https://doi.org/10.1016/j.emospa.2018.11.002>

[63] Stephen D Smith, Beverley Katherine Fredborg, and Jennifer Kornelsen. 2020. Functional connectivity associated with five different categories of Autonomous Sensory Meridian Response (ASMR) triggers. *Consciousness and Cognition* 85 (2020), 103021. <https://doi.org/10.1016/j.concog.2020.103021>

[64] Stephen D Smith, Beverley Katherine Fredborg, and Jennifer Kornelsen. 2017. An examination of the default mode network in individuals with autonomous sensory meridian response (ASMR). *Social Neuroscience* 12, 4 (7 2017), 361–365. <https://doi.org/10.1080/17470919.2016.1188851>

[65] Rebecca Lurie Starr, Tianxiao Wang, and Christian Go. 2020. Sexuality vs. sensuality: The multimodal construction of affective stance in Chinese ASMR performances*. *Journal of Sociolinguistics* 24, 4 (9 2020), 492–513. <https://doi.org/10.1111/josl.12410>

[66] Amanda Stears. 2019. Three Things You Need to Know About ASMR Videos. <https://midiaresearch.com/blog/three-things-you-need-to-know-about-asmr-videos>

[67] Jayne Wallace, Anja Thieme, Gavin Wood, Guy Schofield, and Patrick Olivier. 2012. Enabling Self, Intimacy and a Sense of Home in Dementia: An Enquiry into Design in a Hospital Setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2629–2638. <https://doi.org/10.1145/2207676.2208654>

[68] Philip Weber, Thomas Ludwig, Sabrina Brodesser, and Laura Grönawald. 2021. “It’s a Kind of Art!”: Understanding Food Influencers as Influential Content Creators. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445607>

[69] Donghee Yvette Wohin, Guo Freeman, and Caitlin McLaughlin. 2018. Explaining Viewers’ Emotional, Instrumental, and Financial Support Provision for Live Streamers. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174048>

[70] Zhenhui Yuan, Gheorghita Ghinea, and Gabriel-Miro Muntean. 2015. Beyond Multimedia Adaptation: Quality of Experience-Aware Multi-Sensorial Media Delivery. *IEEE Transactions on Multimedia* 17, 1 (2015), 104–117. <https://doi.org/10.1109/TMM.2014.2371240>

[71] Michele Zappavigna. 2020. Digital intimacy and ambient embodied copresence in YouTube videos: construing visual and aural perspective in ASMR role play videos. *Visual Communication* (7 2020), 1470357220928102. <https://doi.org/10.1177/1470357220928102>

A EXAMPLE COMMENTS

Following are example comments with the highest counts of social, intimate, body, perception, relaxation, and sleepiness words. The percentages of word use in statistical analysis are calculated based on a text corpus that merges all comments of one video.

Social

- Hi 
- love you !!!!!!!!!!!!!!!
- Love you 
- Hi 
- Love you love you love you love you love you
- Fool ... [all the rest "fool"]
- Love you love 
- Hi 
- hi!!!!!!!
- Hi Hi Hi Hi Hi Hi Hi Hi Hi Hi

Intimacy

- Love love the set up, make up and atmosphere!
- Social distance bro social distance back up bro I was not ready
- Literally nobody does casual conversation and personal attention like ASMR Power of Sound

- This puts me to sleep every night for past 3 weeks
- Can you do a asmr dim light shh shh go to sleep now shh shush i tuck you in bed with lots shh shh
- Or a asmr dim light shh shush I tuck in bed its just a nightmare aww shh shh for crying and sleep
- Or best friend helps you fall asleep with lots of shh shh sweetheart i tuck you in bed with lots shh shh
- You helped me sleep better the past few nights.
- Your videos are helpful with insomnia and other sleep issues
- i was so stressed tonight and this helped ease my panic thankyou 
- I kept falling asleep and then waking up then falling asleep back-
TvT

B DATA PROTECTION IMPACT ASSESSMENT (DPIA)

All videos are publicly shared when gathered, and the researchers did not directly interact with any ASMRtists to collect any information about the video creators. We acknowledge this work uses automated data processing tools (LIWC and PMI) to recognize behavioral information from YouTube comments. To protect YouTube users' privacy, we de-linked the YouTube account information from

the collected comments after the data was collected. To our best knowledge, YouTube does not support searching comments. LIWC and PMI are lexicon-based techniques that count the appearance of words related to the measured dimensions. Only words related to common feelings such as social process, intimacy, perception, and relaxation are quantified. We did not intentionally recognize or process any identity information.

The data annotation on Amazon Mechanical Turk only involves identifying objective information from the video (e.g., how the ASMRtist interacts with trigger objects or the camera). All questions are multi-choice questions with pre-defined categories. We do not ask participants to provide personal information or subjectively describe any video content. To protect the participants who are uncomfortable with ASMR videos, all participants must pass a qualification test before the study. The participant must indicate that they had watched ASMR videos before and did not feel them disturbing. At the beginning of the annotation, we also informed that if the video content makes the participant uncomfortable, they should close the survey immediately. With the careful research steps, the IRB office at the authors' institute granted an exemption to the protocol of this study.