Eighteen Years of ASMR on YouTube: A Multilingual, Theme-Level Analysis of 20,087 Videos

Md Shadab Alam*
m.s.alam@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Pavlo Bazilinskyy
p.bazilinskyy@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

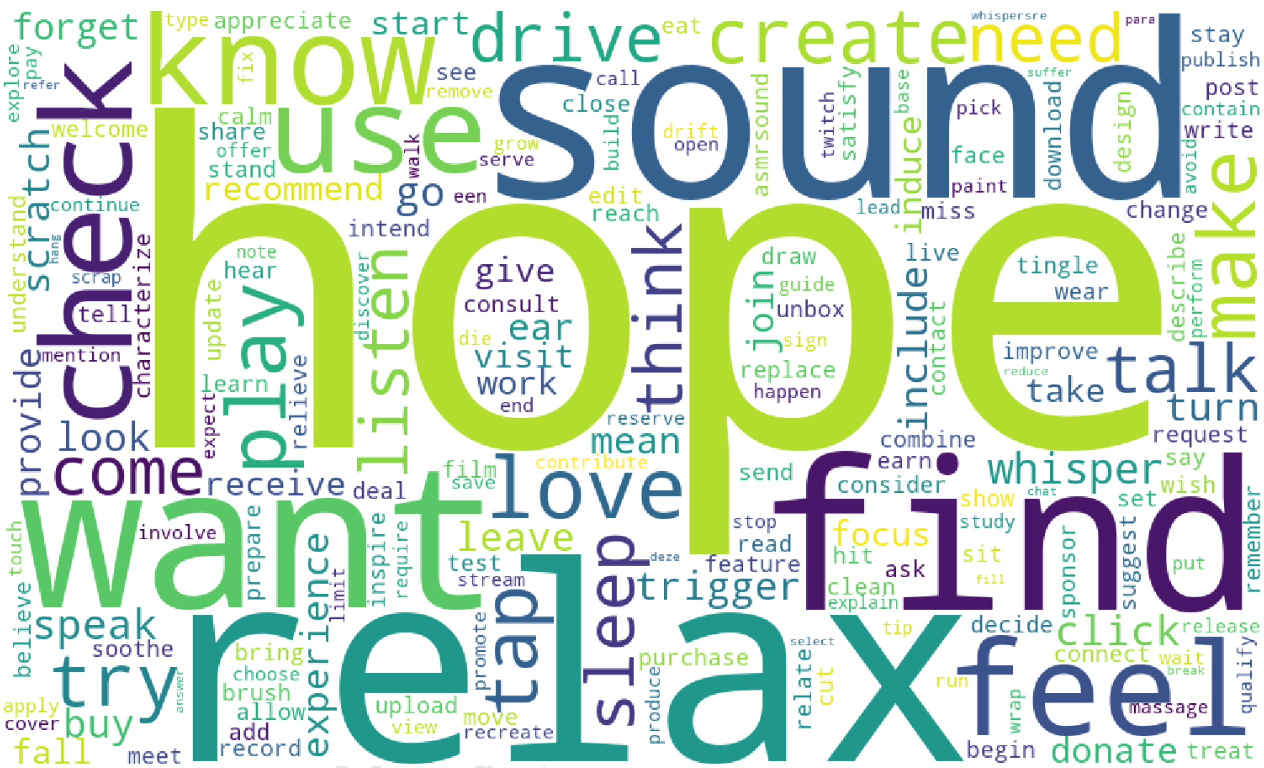

Figure 1: Word cloud of the 200 most frequent verb lemmas extracted from the titles and descriptions of all 20,087 ASMR videos in the dataset. After lemmatisation, removal of a custom stopword list, and filtering to tokens tagged as verbs, each lemma is shown with font size proportional to its corpus frequency, highlighting actions such as "relax", "sleep", "whisper", and "tingle".

Abstract

ASMR videos have become a major genre on online platforms, yet their large-scale characteristics remain underexplored. Using YouTube Data API and a pytubefix workflow, we assemble a dat0aset of 20,087 ASMR videos from 4,076 channels (2008–2025, 40 languages) enriched with duration, views, likes, inferred language, theme flags, and lemmatised title description text. English dominates (82.19% of videos), followed by Korean, Japanese, Spanish, Dutch, and Portuguese. Across the corpus, the mean growth is 2,146.25 views per day and the duration analysis shows that short videos (<10 minutes) average 4,128.62 views per day versus 1,225.65 for 10- to 30 minute content, while very long (>180 minutes) videos reach 5,228.64 views per day. Theme detection indicates that sleep-related (17.79%) and visual-trigger content (16.30%) are particularly prevalent, with whisper (11.49%) and binaural videos (10.29%) also common, while driving-themed videos remain rare (9.84%). K-means clustering on multimodal text, language, and engagement features, visualised with t-SNE, yields 11 content clusters (9–7,300 videos) and a small set of extremely high-growth videos.

Keywords

ASMR, YouTube, corpus analysis, t-SNE, Content themes

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Md Shadab Alam and Pavlo Bazilinskyy. 2025. Eighteen Years of ASMR on YouTube: A Multilingual, Theme-Level Analysis of 20,087 Videos. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Autonomous sensory meridian response (ASMR) refers to a tingling, soothing sensation that some individuals experience in response to specific audio-visual stimuli such as whispering, gentle tapping, or simulated close personal attention [5, 25]. Viewers often describe ASMR as a form of self-care, reporting reductions in stress and anxiety, improved mood, and better sleep [29, 36]. Experimental work has increasingly established ASMR as a reproducible psychophysiological phenomenon rather than a purely anecdotal curiosity. Studies such as Poerio et al. [25] and Engelbregt et al. [7] document systematic changes in affect, cardiovascular responses, electrodermal activity, and neural dynamics during exposure to ASMR, while narrative and comparative reviews, particularly Mahady et al. [20], emphasise heterogeneity in methods, stimuli, and results and situate ASMR in relation to neighbouring constructs such as frisson, synaesthesia, and misophonia.

From a media and communication perspective, ASMR is also a socio-cultural practice in which intimacy, care, and affect are mediated through networked screens and headphones. Early qualitative work highlighted the grassroots "whisper community" that formed around online forums and pioneering YouTube channels, showing how whispered speech and low-fi recording practices foster mediated co-presence and "non-standard intimacy". Andersen [3] and Smith and Snider [30] analyse how ASMR creators and viewers negotiate pleasure, stigma, and shifting boundaries between public and private life. Later contributions such as Klausen [32] and Gallagher [10] further theorise ASMR as a mode of sonic and audiovisual intimacy linked to broader developments in online video culture.

Industry surveys and platform reports also suggest that this soothing video culture is especially salient for younger audiences: global listener and trend studies point to a disproportionately high uptake among 16–24-year-olds and Generation Z, who report using ASMR or ASMR-like content to relax, study, or fall asleep [4, 33]. Consistent with this demographic skew, experimental ASMR studies typically recruit young adult samples (mean ages around 20 years), indicating that heavy users are disproportionately drawn from younger cohorts [7, 28].

In parallel, a growing body of work asks how video length relates to attention and engagement on digital platforms. Large-scale industry analyses of online video report steep drops in average viewer retention after the first couple of minutes, with a secondary "sweet spot" for mid-length content around 6–12 minutes [9, 38]. Academic analyses of YouTube influencers similarly find that medium- and long-form videos tend to attract more views, likes, and comments than very short clips [23]. In short-form environments such as social-media feeds and in-feed advertising, experimental work instead points to inverted-U effects with optimal lengths on the order of a few tens of seconds [27]. Educational video research on Massive Open Online Courses (MOOCs) also documents sharp engagement declines beyond approximately six minutes, while cautioning against a universal "six minute rule" [12]. However, these studies focus on general, promotional or instructional content rather than ASMR, leaving open whether a relaxation-orientated genre aimed primarily at young viewers follows similar length–engagement patterns or displays its own session-length preferences.

Historically, ASMR on YouTube has developed through overlapping phases. A frequently cited early landmark is WhisperingLife's short video *Whisper 1 — hello!* (26 March 2009), a 106 s whisper-only clip in which the creator explains her longstanding affinity with whispering and introduces a channel devoted to whispered speech (https://www.youtube.com/watch?v=IHtgPbfTgKc). This video is often described as the first intentional ASMR "trigger" video on YouTube and helped crystallise the emerging whisper community around dedicated channels rather than incidental triggers in other genres [11]. In the early origins and niche-community phase (pre-2012), ASMR circulated mainly through small forums and dedicated whisper or soft-spoken role-play channels, following the coining of the term "autonomous sensory meridian response" in 2010. A subsequent phase of mainstreaming and platform growth (approximately 2012–2018) brought substantial journalistic attention and recognition of ASMR as a distinctive genre within YouTube's search and recommendation systems; industry accounts portrayed ASMR as one of the fastest-growing trends on the platform [21, 22]. As the genre matured, a phase of diversification and professionalisation (roughly 2016–2020) introduced recognisable sub-genres (e.g., medical role play, sound-focused "no talking" videos, mukbang and eating-trigger content) alongside increasingly sophisticated production practices, including binaural microphones and multi-camera setups. Most recently, a phase of commercialisation and platform changes (2020–present) has been marked by branded ASMR campaigns and influencer collaborations, the incorporation of ASMR signals into advertising and short-form video formats, and ongoing adjustments in creator strategies in response to algorithmic change, demonetisation pressures, and cross-platform publishing [2, 13].

A growing empirical literature describes different facets of this evolving ASMR ecosystem on YouTube, but it remains dispersed across disciplinary and methodological domains. One strand comprises content-analytic studies that systematically code the visual, auditory, and interactional features of ASMR videos and relate these to viewer engagement. Niu et al. [24], for example, examine interaction modalities and parasocial cues across a large sample of videos, while other studies document trigger types, performer characteristics, and formal conventions in YouTube ASMR content. A second strand focusses on comments and everyday uses of ASMR through netnography and qualitative content analysis. Examples include Triani [34] and Łapińska [15], which show how viewers frame ASMR as self-care, negotiate authenticity and intimacy, and articulate the meanings of "tingles" in comment threads and online discussions.

A third cluster of work approaches ASMR as a multimodal and discursive phenomenon, often concentrating on role-play subgenres. Studies such as Wang [35], Klausen [14], and Abdallah [1] combine multimodal discourse analysis with concepts including haptic audio-visuality, ambient co-presence, and digital intimacy

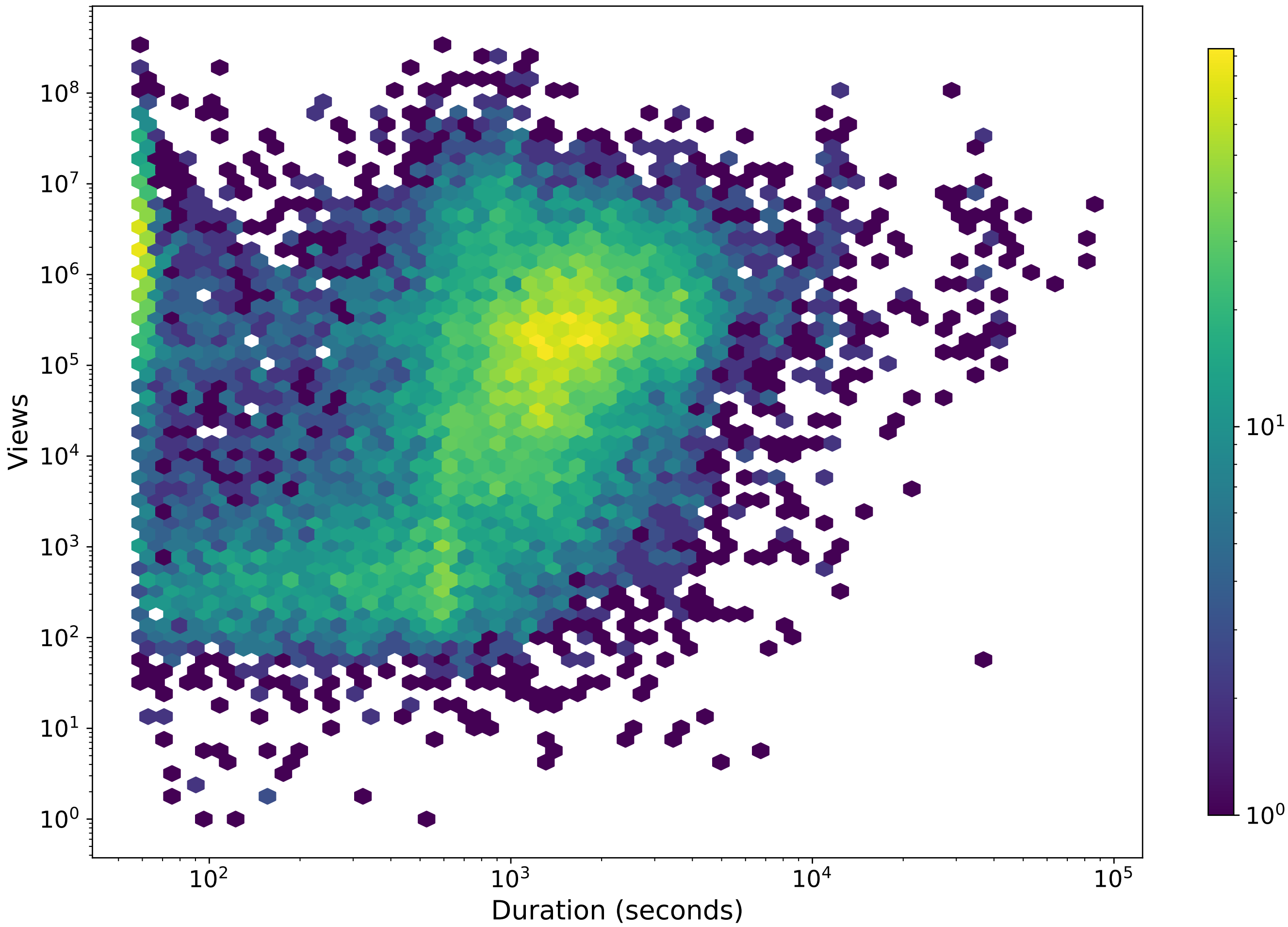

Figure 2: Log–log hexagonal bin plot showing each video's duration in seconds (x-axis) against its cumulative view count at the time of data collection in December 2025 (y-axis). The plot includes 20,080 ASMR videos with positive duration and view counts; each hexagon aggregates multiple videos, with colour intensity indicating the number of videos in that bin.

to show how camera positioning, gaze, voice, gesture, and sound design are orchestrated to simulate physical closeness and care. Related work on ASMR role play and whispered speech feeds into speech-technology and HCI research: Zarazaga et al. [37] and Song et al. [31] treat ASMR as a large-scale resource for whispered speech and unvoiced language identification. A fourth strand situates ASMR within broader audience and platform dynamics on YouTube and social media. Studies such as Maddox [18, 19] investigate how ASMR creators navigate YouTube's affordances, monetisation regimes, and community norms, while Portas Ruiz [26] and Feiz et al. [8] examine the integration of ASMR into influencer marketing and advertising. More general work on YouTube engagement, such as Liikkanen and Salovaara [16], offers methodological and conceptual resources to understand ASMR as an instance of a wider set of native media practices.

1.1 Aim of the study

The aim of this study is to provide a quantitative, multilingual characterisation of ASMR content on YouTube using 20,087 videos uploaded between 2008 and 2025 and retrieved through a large-scale keyword-based pipeline centred on the query "ASMR". We combine video-level metadata, language information, title and description text, rule-based theme annotations, and behavioural measures such as views per day and mean engagement. Our analysis asks: (i) how ASMR videos are distributed across languages, formats, and title styles; (ii) how prevalent major ASMR themes (e.g. whispering, no talking, sleep, binaural, role-play, mukbang, driving) are and how they differ in reach and mean engagement; (iii) how duration and other structural features relate to popularity; (iv) how ASMR content has evolved over time; and (v) how videos cluster into recurrent content types when represented in a joint feature space across the full observation period.

2 Method

All data collection and analysis code used in this study is available as supplementary material and in a public GitHub repository (see section 6). We collect ASMR-related video data from YouTube (https://www.youtube.com) using a keyword-based pipeline that combines the official YouTube Data API v3 (https://developers.google.com/youtube/v3) with a scraping-based workflow implemented in the *pytubefix* library (https://pytubefix.readthedocs.io). In all experiments, we use a single English query keyword, "ASMR", which we pass identically to both the YouTube Data API search endpoint and the *pytubefix* search interface; in both cases, we restrict the results to standard videos (excluding non-video items such as channels and playlists).

To increase coverage beyond what the API alone returns and to obtain a richer set of metadata, the pipeline comprises two discovery branches. In the API-based branch, we use the YouTube Data API to perform paginated searches for the query "ASMR", restricted to standard videos. Each search retrieves up to 50 results per page and up to 100 pages (up to 5,000 candidates per search). To mitigate this per-search cap while covering the full history of ASMR content, we partition the study period into consecutive three-month upload windows and run separate API searches for each window, starting on 1 January 2008 and ending on 8 December 2025. In the scraping-based branch, we use *pytubefix* to issue the same keyword query against the public YouTube search interface, again restricted to videos and using YouTube's default ranking. For each video identifier discovered by either branch, we request the corresponding watch page using *pytubefix* and parse it to extract extended metadata, including title, description, duration in seconds, view and like counts, channel identifier, author name, upload date, and an initial language estimate based on the combined title–description text. When both API- and scraping-based metadata are available for the same video, we merge them into a single record; where fields are missing, we fall back to whichever source provides the information. If neither source provides a language label, we apply automatic language detection over the concatenated title and description using an off-the-shelf language-identification package.

All discovered videos are subjected to a uniform set of inclusion criteria. First, to enforce topical relevance, we require that the lowercase query keyword "asmr" appear in the video title (case-insensitive substring match); videos whose titles do not contain this token are discarded, even if they were retrieved by the API or the scraper. Second, to exclude YouTube Shorts and extremely brief clips, we require an estimated duration of at least 60 s. Videos with missing or non-parsable durations are conservatively treated as short and removed. After filtering, videos discovered through the two branches are deduplicated by video identifier, resulting in a unique set of ASMR-related videos for subsequent analysis.

For each video that passed these inclusion criteria, we constructed a structured record with the following per-video fields:

(1) **Identifiers and channel metadata:** the unique video identifier, the associated channel identifier, and the channel's display name or author field.
(2) **Textual fields:** the video title and description, as returned by YouTube at the time of collection. In all subsequent text-based analyses, we treat the concatenation of title and description as a single document.
(3) **Temporal information:** the upload timestamp in UTC, the derived calendar date, and the upload year and month. For each video we also compute the number of days since upload relative to a fixed reference date, used in growth-related measures.
(4) **Duration:** the video duration in seconds and minutes, obtained by parsing ISO 8601 duration(https://www.iso.org/iso-8601-date-and-time-format.html) strings and/or watch-page metadata. We further discretise duration into coarse buckets with five levels: under 10 min, 10–30 min, 30–60 min, 60–180 min, and over 180 min, plus an *unknown* category for rare cases with missing values.
(5) **Engagement statistics:** the total number of views and likes at the time of collection. From these we derive (i) the number of views per day since upload,
$$\text{views per day}(v) \;=\; \frac{\text{views}_v}{\text{days since upload}_v}\,, \tag{1}$$
and (ii) a per-video engagement rate defined as the ratio of likes to views whenever view counts are strictly positive,
$$\text{engagement}(v) \;=\; \frac{\text{likes}_v}{\text{views}_v}\,. \tag{2}$$
Videos with zero or missing view counts are assigned a missing engagement value. When reporting aggregate results for a subset of videos S (e.g. videos in a given language, duration bucket, or cluster), we use the arithmetic mean of this per-video engagement rate,
$$\text{mean engagement}(S) \;=\; \frac{1}{|S|}\sum_{v\in S}\text{engagement}(v)\,, \tag{3}$$
and refer to this quantity as *mean engagement*. As auxiliary context, we also obtain channel-level statistics from the YouTube Data API (total view count and total number of uploaded videos per channel) and compute an average number of views per uploaded video; this channel-average statistic is used to form a relative-views measure for some descriptive analyses but is not a primary focus of the present study.
(6) **Language:** a normalised language code inferred from a combination of platform metadata (default audio or interface language) and automatic language detection on the concatenated title–description text, performed using the `langdetect` Python package (https://pypi.org/project/langdetect/). Where platform metadata and automatic detection disagree, we manually normalised obvious aliases (e.g. different codes for English) and treated the remainder as distinct categories.
(7) **Title style features:** automatically derived indicators that characterise title formatting. These include the number of words and characters in the title and binary flags for stylistic devices: presence of brackets or parentheses, all-caps words of length at least three, exclamation marks, question marks, hashtags, and explicit "no talking" tags (e.g. "no talking", "no-talk"). These features are used in analyses relating title style to views and engagement.

(8) **Content themes:** ten Boolean indicators capturing broad ASMR themes derived from the lemmatised title–description text: whisper-focused content, no-talking or speech-free content, sleep-related content, binaural or 3D audio, role-play scenarios, ear-focused treatments, eating and mukbang-style content, keyboard and typing sounds, visually emphasised triggers, and driving-related content. Each indicator is set to true if the video's text matches a rule-based pattern for that theme and false otherwise.

(9) **Growth category:** a categorical label that discretises the views-per-day measure into four levels: *slow*, *medium*, and *fast* for increasing ranges of views per day, and *unknown* for missing or non-positive values.

For text-based analyses, we operate on the concatenation of each video's title and description, treating this as a single document irrespective of how the video was discovered. Prior to further processing, we apply light normalisation: URL substrings are removed and line breaks are replaced by spaces, but emoji and most punctuation are preserved to retain potentially meaningful tokens. We then apply a stop-word filter that combines the built-in English stop-word list from the `wordcloud` package with a custom list tailored to ASMR YouTube content. The custom list removes (a) the token *ASMR* itself and closely related platform-specific tokens (e.g. *gmail*, *comment*, *channel*), (b) frequent English function words and pronouns (e.g. *the*, *and*, *you*, *this*), (c) common social-media filler such as *thanks*, *subscribe*, *follow*, *like*, *watch*, *video*, (d) common French function words (e.g. *le*, *la*, *des*, *et*, *pour*), (e) all single-letter tokens, (f) isolated punctuation marks, and (g) standalone digits.

To obtain linguistically informed lexical profiles, we use spaCy's English language model to lemmatise the cleaned text. Non-alphabetic tokens and tokens marked as stop words by spaCy are discarded, and some lemma families are normalised to a shared canonical form (e.g. *whispers* and *whispering* are mapped to the lemma *whisper*). For each video, we form a set of distinct content lemmas so that repeated occurrences of the same lemma within a video contribute at most one count for that video. Aggregating across the corpus, we count, for each lemma, the number of videos in which it appears at least once and rank lemmas by this document-frequency measure. These counts are used to construct summary tables and bar-chart visualisations of the most frequent lemmas in ASMR titles and descriptions.

From the same lemmatised title–description text, we derive the rule-based theme indicators described above. Whisper-focused content is flagged when lemmas related to *whisper* occur. No-talking or speech-free videos are identified either when the text contains explicit phrases such as "no talking", "no-talk", or "without talking", or when a lemma such as *talk* or *speak* is preceded by a negation. Sleep-related content is detected via lemmas such as *sleep* and *insomnia* or explicit phrases like "for sleep". Binaural and 3D audio are captured by mentions of *binaural*, "3D audio", "3D sound", "3Dio", and "8D audio/sound". Role-play scenarios are detected via explicit terms such as *roleplay*, abbreviations like "RP", and lemmas associated with examinations and services (e.g. *exam*, *checkup*, *haircut*, *barber*). Ear-focused treatments are flagged by phrases such as "ear cleaning", "ear massage", "ear exam", "ear attention", "ear brushing", or local co-occurrence of lemmas *ear* or *otoscope* with *clean*, *brush*, *massage* or *attention*. Eating and mukbang-style videos are detected via mentions of *mukbang*, "eating ASMR", "eating sounds", and related phrases. Keyboard and typing sounds are flagged by lemmas such as *keyboard* and *type*. Visually emphasised triggers are identified by phrases including "visual triggers", "hand movements", "visuals", "slow movements", "trigger assortment" or related lemmas. Driving-related content is detected when lemmas such as *drive* appear or when the text contains phrases like "driving", "drive with me", "car" or "road trip".

To characterise heterogeneity in ASMR video types, we perform an unsupervised clustering analysis over a joint feature space that combines textual, behavioural, and language information. For the textual component, we represent each video's concatenated title–description as a TF–IDF-weighted bag of words over unigrams and bigrams. To reduce sparsity and noise in this representation, we restrict the TF–IDF vocabulary to the 5,000 terms with the highest overall informativeness and require that a term appear in at least 5 videos to be included; this limits the dimensionality of the feature space, improves computational efficiency, and discards ultra-rare tokens (e.g., idiosyncratic names or typographical errors) that are unlikely to contribute to stable, interpretable clusters. For the behavioural component, we use three numeric variables: duration in minutes, engagement rate (Equation 2), and views per day since upload. The detected language is encoded as a categorical factor using one-hot encoding. These components are combined into a single feature matrix using a column-wise preprocessing pipeline implemented with `scikit-learn` (https://scikit-learn.org/).

In this representation, we fit the k-means models for $k \in [4, \ldots, 20]$ and use the elbow method on the sum of squared errors (inertia) within the cluster to select the number of clusters. Inertia decreases from 51,183.59 at k = 4 to 32,397.73 at k = 11, but the marginal gain drops substantially beyond this point (for example, only a 3.39% reduction to 31,300.59 at k = 12, and then a further 14.76% reduction spread over eight additional clusters, i.e., on average 1.85% per additional cluster up to 26,680.08 at k = 20). We therefore choose k = 11, which lies in the elbow region and balances parsimony with a sufficiently fine-grained separation of different types of ASMR content.

3 Results

The final dataset contained 20,087 ASMR videos collected from 4,076 distinct channels, uploaded between 1 January 2008 and 7 December 2025. All likes and views for each video were updated on 8 December 2025. In all videos, 40 different languages were identified. English accounted for 16,509 videos (82.19%), followed by Korean (n=516, 2.57%), Japanese (n=489, 2.43%), Spanish (n=449, 2.24%), Dutch (n=386, 1.92%), Portuguese (n=338, 1.68%), French (n=270, 1.34%), Russian (n=207, 1.03%) and German (n=192, 0.96%). The average video duration was 1,481.59 s (SD = 2,640.38). The mean number of views per video was 1,413,206.56 (SD = 8,312,664.60), and the mean number of likes was 26,530.70 (SD = 162,057.27). The derived metric of views per day had a mean of 2,163.34 (SD = 12,919.59) (see Figure 3).

For text-based analyses, a word cloud was generated from concatenated titles and descriptions of all 20,087 videos. A spaCy-based

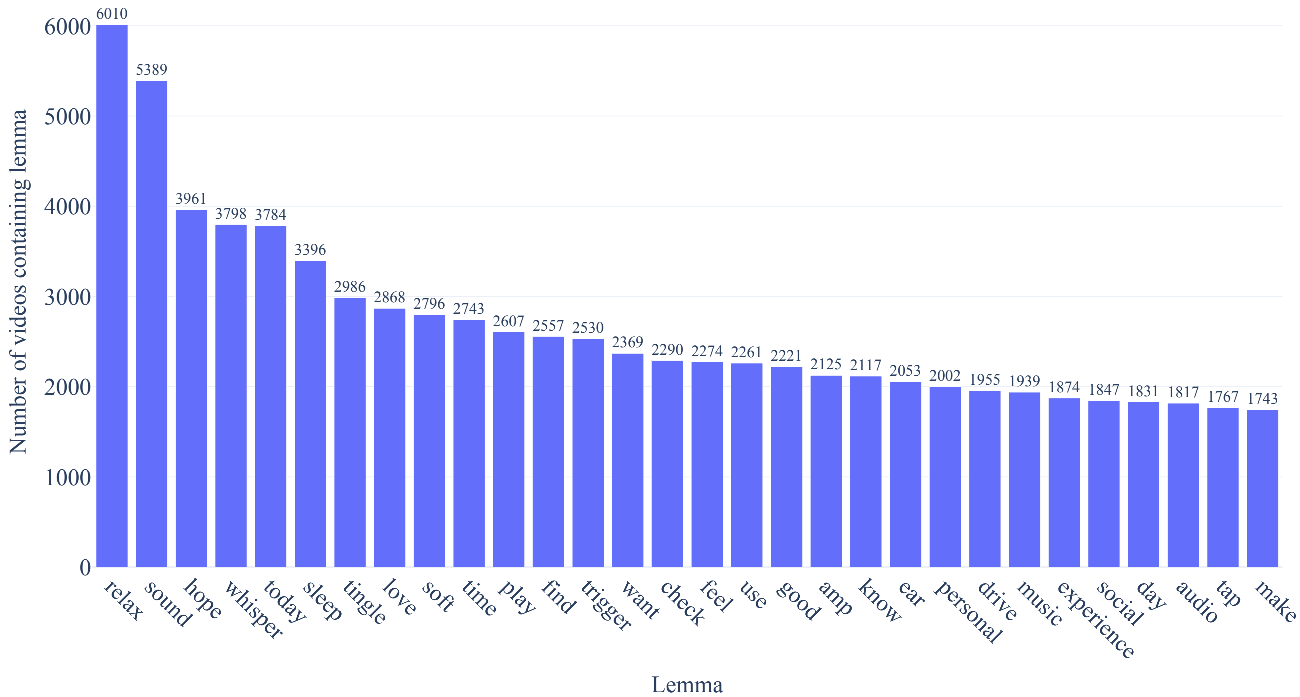

Figure 3: Visualisation of the 30 most frequent lemmatised content words in the combined titles and descriptions of all 20,087 ASMR videos. The x-axis lists lemmas, and the y-axis shows the number of distinct videos in which each lemma appears at least once (document frequency).

lemma analysis over all videos yielded a table of the 30 most frequent lemmas, with each lemma counted at most once per video. The ten most frequent lemmas were *relax* (n=6010), *sound* (n=5389), *hope* (n=3961), *whisper* (n=3798), *today* (n=3784), *sleep* (n=3396), *tingle* (n=2986), *love* (n=2868), *soft* (n=2796) and *time* (n=2743) (see Figure 3).

Automatic theme detection was performed using a rule-based pipeline on the lemmatised titles and descriptions. Theme-specific lemmas (e.g., *whisper*, *sleep*, *roleplay*, *mukbang*, *keyboard*, *drive*) and surface-pattern expressions (e.g., "no talking", "3D audio", "ear cleaning", "hand movements", "visual triggers") were matched. This procedure yielded ten theme labels. The number (and proportion) of videos with each theme were: whisper (n=2308; 11.50%), no-talking (n=1079; 5.37%), sleep-related (n=3574; 17.79%), binaural or 3D-audio (n=2067; 10.29%), role-play (n=2632; 13.11%), ear-cleaning or ear-focused (n=627; 3.12%), mukbang or eating (n=1026; 5.11%), keyboard or typing (n=895; 4.46%), visual or hand-movement triggers (n=3274; 16.30%), and drive-themed content (n=1976; 9.84%). Growth categories were derived using fixed thresholds on views per day: videos with fewer than 1,000 views/day were labelled slow-growth, those with 1,000–10,000 views/day medium-growth, and those exceeding 10,000 views/day fast-growth. Videos with zero or missing values were assigned an unknown category. This resulted in 11,663 slow-growth videos (58.06%), 6,822 medium-growth videos (33.96%), 1,595 fast-growth videos (7.94%), and 7 unknown (0.03%).

Videos were also classified into duration buckets. Short-form videos under 10 min accounted for 6,025 videos (30.00%), medium-length videos between 10 and 30 min for 9,228 videos (45.94%), upper-medium videos between 30 and 60 min for 3,590 videos (17.87%), long-form videos between 60 and 180 min for 1,061 videos (5.28%), and very long videos exceeding 180 min for 182 videos (0.91%); one video had unknown duration. Mean views and engagement rates for each duration bucket are summarised in Table 3. For example, videos longer than 180 min averaged 6,004,632 views with an engagement rate of 0.02, whereas videos under 10 min averaged 1,648,006 views with an engagement rate of 0.03. Language-level summaries appear in Table 1, and title-length statistics in Table 2.

The relationship between duration and popularity was examined for the subset of 20,080 videos with positive duration and non-zero views (Figure 2). Duration ranged from 59 s to 86,400 s (median = 1,031 s, IQR = 502–1,762 s), with a mean of 1,481.94 s (SD = 2,640.76). Views ranged from 1 to 3.40×10^8 (median = 97,498; IQR = 5,269–501,339), with a mean of 1,413,277.00 (SD = 8,312,866.00). For the 20,081 videos with positive views, $\log_{10}(\text{views})$ had a mean of 4.73 and a standard deviation of 1.36. The D'Agostino–Pearson normality test yielded $k^2 = 1009.80$ and $p = 5.30 \times 10^{-220}$. The Shapiro–Wilk test (on a subsample) yielded $W = 0.98$ and $p = 6.29 \times 10^{-28}$.

Descriptive statistics of views per day for the focal themes are summarised in Table 4. Whisper videos (n=2,308) had a mean of 1,010.73 views/day, drive-themed videos (n=4,596) had 1,285.45 views/day, no-talking videos (n=1,079) had 2,118.87 views/day, sleep-related videos (n=3,574) had 1,780.96 views/day, and binaural videos (n=2,067) had 488.68 views/day.

Theme trends were computed for all years with valid upload dates (2008–2025). For no-talking content, the overall trend comprised 18 yearly observations (1,079 videos), with 370 language–year combinations included in the by-language breakdown. For binaural content, the overall trend also comprised 18 yearly observations (2,067 videos). A lemma-based temporal trend was also computed for the lemma "drive" (Figure 4).

Finally, we use a t-SNE embedding to visualise similarities among videos in the learned feature space. The two-dimensional projection (shown in Figure 5) places all 20,087 videos into a small number of dense regions separated by sparser transition zones and a handful of clear outliers. With the $k = 11$ solution selected via the elbow method [17], the clusters labelled A–K in the figure range in size from 9 to 7,290 videos. The largest groups, clusters A and G (6,756 and 7,290 videos, respectively), are almost exclusively English-language and together account for the majority of the corpus; they exhibit mixed ASMR themes with moderate prevalence of whispering, sleep-related titles, binaural recordings, roleplay, visual triggers and driving-related content, and attain mean growth of approximately 269.64 and 837.42 views per day. A sizeable non-English cluster B (2,588 videos) is dominated by Japanese, Korean, Dutch, Spanish, and Russian content, shows elevated rates of mukbang/eating and roleplay descriptors, and achieves a mean of 1,822.06 views per day. Additional medium-sized clusters, such as F and J (2,380 and 301 videos, respectively), capture multilingual roleplay and visually rich formats with mean growth of 2,291.47 and 901.46 views per day.

Several smaller clusters are strongly enriched for specific themes and achieve markedly higher or lower growth than the main groups. Cluster C (367 videos) combines a high share of videos of no talk and sleep with frequent mentions of ear cleaning and binaural sound, and reaches a mean growth of 3, 404.79 views per day, while a closely related cluster D (56 videos) of videos focused on very sleep shows a similarly elevated growth of 3,419.98 views per day. At the extreme, clusters E (61 videos) and K (9 videos) consist largely of hyper-viral content: E is characterised by eating and mukbang themes and attains mean growth of 138,092.77 views per day, whereas K comprises ultra-short, clip-like videos with mean growth of 373,380.05 views per day. Cluster I (268 videos) forms another high-performing group with many mukbang-orientated and visually salient videos (mean growth 39,577.73 views per day), while a small outlier cluster H (11 videos) is dominated by driving-related titles and grows more modestly at 56.61 views per day. Taken together, the t-SNE visualisation and cluster-level statistics show that the ASMR ecosystem is structured around a few large, predominantly English clusters of general-purpose ASMR, complemented by non-English and theme-specialised clusters (sleep/no-talking, binaural ear cleaning, mukbang, and short clips) that differ systematically in language mix and audience growth.

4 Discussion

This study aimed to provide an ecosystem-level account of ASMR video production on YouTube by combining large-scale metadata analysis with lexical modelling, thematic detection, and unsupervised clustering. Although much previous work has concentrated on individual creators, specific trigger types, or comment-based ethnographies, our dataset spans 20,087 videos uploaded between 2008 and 2025 across 4,076 channels and 40 languages. This coverage makes it possible to situate well-known ASMR formats within a broader and more heterogeneous media ecology, and to trace how new sub-genres emerge over time.

Lexical analysis (Figure 3) shows that ASMR titles and descriptions revolve around a highly consistent vocabulary of care, calm, and sensory experience. Terms such as *relax*, *sound*, *sleep*, and *tingle* remain fundamental, echoing how participants describe ASMR as a practice that promotes sleep or self-soothing in controlled experiments and ethnographic accounts [14, 25, 32]. At the same time, creators frequently invoke *support*, *hope*, and *personal* attention, reinforcing the arguments that ASMR functions as a technologically mediated form of intimacy and comfort [3, 8, 30, 34]. Our analysis extends these claims across a much larger and more diverse corpus, showing that such language is not restricted to a handful of channels but forms a structural component of the genre.

The distribution of lemmas also shows diversification within the field. Frequent references to *ear*, *tap*, *check* and *roleplay* correspond to well-established trigger categories such as ear cleaning and medical roleplay [24, 32]. Meanwhile, terms such as *drive* and *amp* point to emerging micro-genres and cross-platform influences. The growth of drive-themed ASMR (Figure 4), which increases sharply around 2020 and rises again in 2025, likely reflects several converging factors: the widespread availability of inexpensive, high-quality recording hardware (from smartphones to dashcams) and free large-scale hosting on YouTube lower the barrier to recording hours of in-car ambience, while COVID-19 lockdowns may have created both more time for creators to experiment and more demand from viewers for "slow" vicarious travel experiences. Although this interpretation is necessarily speculative, it illustrates how ASMR readily absorbs everyday environments and infrastructure changes into new sensory formats.

Our language-level analysis (Table 1) highlights the ongoing globalisation of ASMR production. English remains dominant, but substantial activity is evident in Korean, Japanese, Spanish, Portuguese, French, and other language communities. In particular, Korean videos achieve the highest mean views per day among languages with sufficient sample sizes (> 100), whereas Portuguese-language videos achieve the highest mean engagement rates. These patterns indicate that ASMR is not simply replicated in other languages, but is actively reinterpreted and reshaped within local creative cultures, aligning with the arguments that ASMR practices are culturally situated rather than uniform across linguistic contexts [15, 34]. Our findings therefore support a shift away from English-centric analyses and towards comparative approaches that attend to platform-native performance metrics.

The duration analysis reveals a bifurcated ecology of ASMR formats. Shorter videos (under 10 minutes) attract the highest views per day and engagement rates, consistent with rapid-consumption,

Language	n	Views	Views/day	Likes	Likes/day	Engagement ($\times10^{-2}$)
English	**16509**	1,091,721 (6,203,731)	1,587.29 (11,407.73)	22,333 (164,306)	45.54 (258.30)	2.06
Korean	516	8,467,757 (28,957,360)	7,983.91 (19,698.55)	73,458 (166,193)	89.99 (199.08)	1.62
Japanese	489	2,193,631 (8,685,242)	5,146.20 (17,034.15)	30,104 (75,795)	95.86 (324.26)	1.59
Spanish	449	1,420,348 (3,904,452)	3,697.33 (12,054.50)	48,245 (163,663)	125.71 (369.47)	4.16
Dutch	386	128,189 (906,270)	86.16 (795.18)	2,027 (10,605)	1.45 (11.00)	2.10
Portuguese	338	1,855,969 (4,933,720)	8,410.54 (25,046.59)	59,425 (139,358)	287.34 (755.39)	5.62
French	270	385,701 (850,888)	992.51 (1,415.48)	10,220 (24,479)	40.91 (63.62)	3.77
Russian	207	607,243 (1,883,459)	698.27 (2,085.58)	15,093 (54,815)	19.86 (65.14)	2.57
German	192	3,123,096 (12,642,060)	3,077.25 (14,736.05)	60,369 (276,648)	64.33 (344.64)	2.87
Vietnamese	89	9,461,394 (27,300,900)	6,739.50 (16,829.52)	91,960 (209,525)	61.18 (125.72)	1.61
Italian	85	1,527,641 (7,406,688)	4,140.57 (18,148.61)	44,225 (218,961)	118.67 (436.37)	3.21
Estonian	61	7,105,667 (29,876,660)	5,389.77 (16,781.96)	72,543 (243,933)	84.91 (200.21)	1.66
Indonesian	54	7,156,125 (16,729,040)	**22,268.99** (50,331.24)	161,156 (395,447)	432.95 (953.43)	2.07
Filipino	54	1,248,766 (4,953,066)	1,016.80 (3,880.97)	15,003 (31,078)	15.27 (35.17)	2.99
Polish	49	231,341 (743,593)	288.94 (709.08)	6,039 (23,874)	11.15 (26.58)	3.52
Turkish	43	543,981 (1,103,585)	1,814.31 (3,333.17)	6,062 (17,250)	52.88 (109.07)	2.53
Unknown	42	1,377,681 (3,070,454)	2,704.70 (4,281.72)	43,014 (87,035)	88.99 (125.27)	3.44
Swahili	38	128,605 (756,727)	755.68 (4,644.02)	58,220 (72,896)	338.08 (474.37)	2.74
Norwegian	35	1,448,922 (5,192,202)	1,909.15 (4,906.03)	40,046 (96,008)	93.63 (227.35)	2.63
Afrikaans	27	2,466,746 (8,393,231)	2,383.12 (6,009.05)	58,213 (162,243)	85.89 (207.77)	2.74
Catalan	26	**17,385,450** (40,062,890)	14,373.16 (31,671.12)	121,239 (207,271)	141.14 (201.69)	3.21
Danish	22	4,782,763 (11,892,660)	12,183.09 (25,664.08)	112,973 (310,282)	290.62 (567.77)	2.00
Bulgarian	21	432,035 (884,986)	916.44 (2,406.58)	7,929 (15,748)	17.81 (42.22)	2.19
Hungarian	20	522,314 (1,352,275)	888.79 (2,788.99)	33,773 (40,218)	18.53 (17.96)	2.08
Thai	20	893,241 (1,510,951)	1,816.45 (6,278.55)	19,045 (28,531)	44.07 (156.69)	2.66
Arabic	15	1,007,054 (2,258,661)	1,110.97 (1,333.12)	22,874 (63,976)	23.05 (26.34)	4.23
Romanian	14	6,719,441 (18,729,090)	9,624.07 (29,843.93)	177,373 (522,437)	269.38 (831.70)	2.26
Finnish	6	230,087 (147,290)	243.16 (379.03)	4,173 (3,147)	1.83 (1.46)	1.78
Ukrainian	3	1,826 (1,867)	0.92 (1.38)	388 (—)	0.25 (—)	**9.76**
Swedish	2	2,301,014 (1,982,614)	3,490.38 (4,336.36)	85,512 (112,190)	147.40 (204.33)	2.57
Greek	2	75,354 (102,859)	33.14 (45.75)	1,315 (1,817)	0.58 (0.81)	1.45
Chinese	2	2,642 (3,451)	115.87 (163.72)	156 (—)	7.11 (—)	3.07
Czech	1	1,895 (—)	125.23 (—)	112 (—)	7.40 (—)	5.91

Table 1: Language-level statistics: means (with SD in brackets) for views, views/day, likes, likes/day, and engagement.

Title length bucket	n	Views (SD)	Views/day (SD)	Likes (SD)	Likes/day (SD)	Engagement ($\times10^{-2}$)
$\leq$ 5 words	2829	1,011,354 (5,149,352)	1,743.45 (9,883.83)	23,450 (134,771)	49.20 (213.11)	2.066
6–10 words	9966	969,262 (5,832,315)	1,645.86 (10,191.98)	23,891 (152,642)	55.69 (288.59)	2.247
11–20 words	7239	2,086,838 (11,276,160)	2,798.37 (15,769.20)	30,284 (182,808)	59.18 (300.04)	2.250
> 20 words	53	14,308,240 (27,385,360)	10,339.56 (20,469.27)	127,678 (157,159)	87.14 (116.86)	1.668

Table 2: Summary statistics for ASMR videos by title length bucket.

high-intensity trigger content that fits into everyday routines. In contrast, very long videos (more than 180 minutes) accumulate substantial total views despite lower interaction rates, suggesting that they serve as background or sleep-support material. This dual structure highlights that ASMR fulfils both active and passive modes of media use. The heavy-tailed distribution of views further confirms that ASMR, like other YouTube genres [16], is characterised by extreme inequality: a small number of highly successful videos dominate attention, while the majority attract modest audiences.

Theme-based indicators nuance assumptions about which ASMR formats "perform best". In our sample, no-talking videos (n=1,079; 2,118.87 views/day), sleep-related videos (n=3,574; 1,780.96 views/day), and drive-themed videos (n=4,596; 1,285.45 views/day) achieve equal or higher mean growth than whisper videos (n=2,308; 1,010.73 views/day), while binaural/"3D audio" content (n=2,067; 488.68 views/day) tends to underperform on average. Sleep-related and drive videos also show substantial variability in views/day and engagement, consistent with a wider upper tail and greater scope for high-performing outliers. Overall, these patterns suggest that ASMR success is not determined solely by adhering to canonical whisper/no-talking formats or by using technically sophisticated audio setups; instead, visibility appears to depend on a complex interplay of creative choices, audience expectations, and platform recommender dynamics.

Clustering analysis provides additional information on ASMR's internal heterogeneity. Using an eleven-cluster solution over the joint feature space that combines textual, behavioural, and language information (selected via the elbow method), we observe groups that differ markedly in typical duration, language composition,

Duration bucket	*n*	Views (SD)	Views/day (SD)	Likes (SD)	Likes/day (SD)	Engagement ($\times 10^{-2}$)
Under 10 min	6025	1,648,006 (10,262,270)	4,017.20 (20,880.82)	59,939 (323,016)	168.48 (575.35)	2.59 (3.00)
10–30 min	9228	1,366,389 (8,472,354)	1,208.09 (6,689.44)	18,141 (75,487)	22.12 (76.31)	2.17 (2.24)
30–60 min	3590	898,566 (2,382,665)	1,062.02 (2,861.40)	13,148 (37,002)	26.96 (88.90)	1.98 (1.67)
60–180 min	1061	1,443,199 (4,257,191)	1,889.08 (4,270.75)	17,721 (40,285)	37.21 (89.59)	2.18 (3.41)
Over 180 min	182	6,004,632 (16,057,790)	5,183.55 (9,148.62)	62,400 (152,433)	73.27 (112.04)	1.71 (1.39)
Unknown	1	4,078 (—)	376.35 (—)	1,238 (—)	114.25 (—)	30.36 (—)

Table 3: Summary statistics by duration bucket.

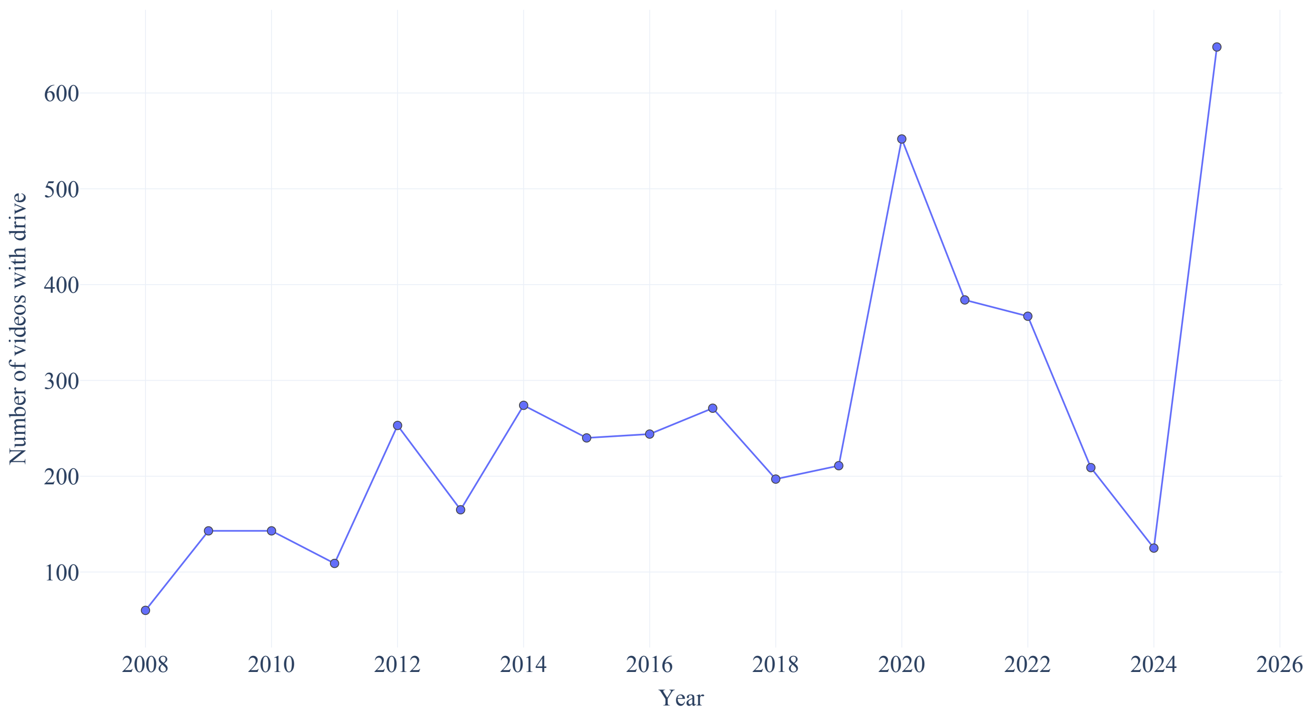

Figure 4: Yearly count of "driving"-themed ASMR videos, where the driving theme is defined via titles or descriptions containing the lemma *drive* or related phrases (e.g., "driving", "drive with me", "car", "road trip").

Theme	*n*	Views (SD)	Views/day (SD)	Likes (SD)	Likes/day (SD)	Engagement ($\times 10^{-2}$)
Binaural / 3D audio	2067	862,100 (3,339,958)	488.68 (1,778.72)	10,050 (34,167)	9.70 (39.89)	1.59 (1.58)
Drive	4596	1,050,389 (7,664,015)	1,285.45 (7,962.17)	17,328 (131,604)	35.05 (208.34)	2.25 (2.87)
No talking	1079	2,855,311 (12,859,000)	2,118.87 (7,665.70)	33,906 (108,391)	30.47 (84.07)	2.18 (2.02)
Sleep-related	3574	1,412,007 (6,368,745)	1,780.96 (7,277.76)	21,430 (152,177)	43.48 (203.31)	2.05 (1.92)
Whisper	2308	662,140 (2,276,008)	1,010.73 (5,027.76)	11,076 (42,800)	27.94 (133.14)	1.92 (1.69)

Table 4: Summary statistics for videos containing each thematic category (theme-present only).

views per day, and mean engagement. Several large, predominantly English clusters correspond to general-purpose ASMR formats that mix whispering, sleep-related framing, binaural sound, roleplay, visual triggers, and driving-related content, and together account for the bulk of production. Other clusters capture more specialised niches: non-English videos with a strong presence of Japanese, Korean, Dutch, Spanish, and Russian; long-form sleep and no-talking formats enriched for ear-cleaning and binaural descriptors; and multiple mukbang- and visually oriented groups with markedly higher growth. At the extreme, one very small cluster is dominated by ultra-short, clip-like videos that achieve exceptionally high rates of views per day, indicating a Shorts-like pattern of consumption.

The t-SNE-based visualisations of this clustering solution reveal several dense clusters separated by low-density regions, suggesting that ASMR production is organised around multiple semi-independent creative strategies rather than a single continuum. Dense regions correspond to widely shared "formulae", such as conventional whisper-talk role plays or standard sleep-support

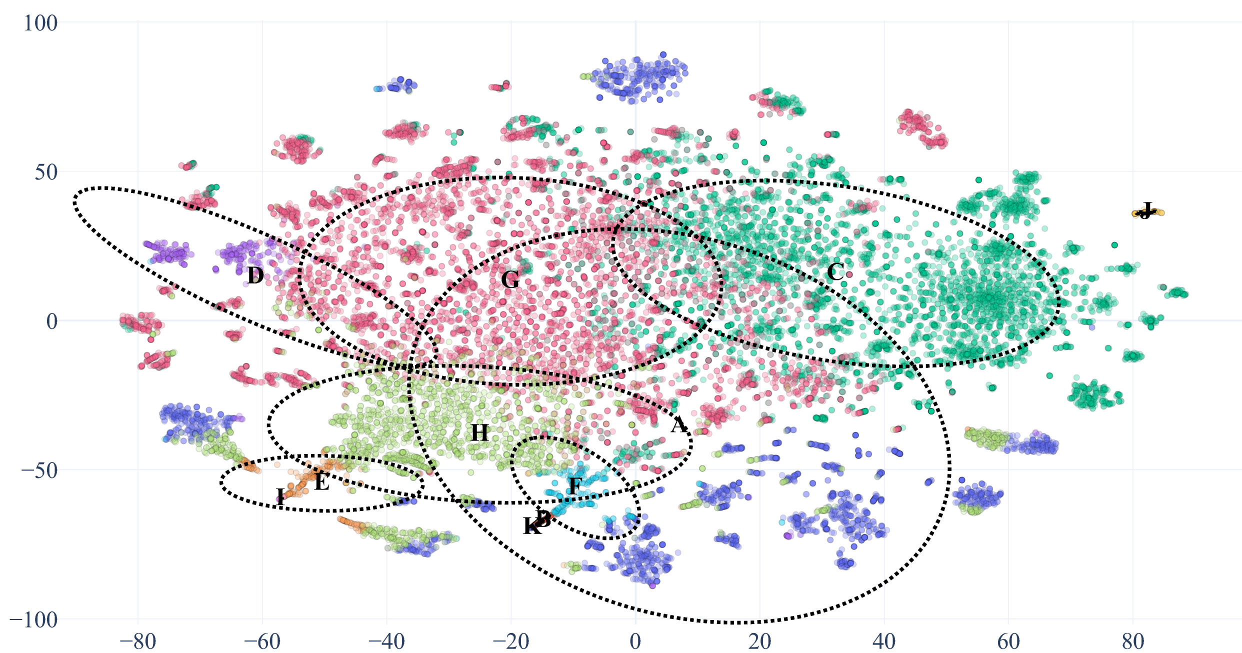

Figure 5: Two-dimensional t-SNE projection of all videos in a joint feature space combining text (TF–IDF over titles and descriptions), duration, engagement rate, views per day, and language. Each point is a video, coloured by its k-means cluster assignment, with faint ellipses indicating the approximate spatial extent of clusters. The plot shows how different ASMR video types group together in terms of content, style, and performance.

formats, while more isolated points and small clusters reflect unusual combinations of triggers, languages and presentation styles, including niche high-growth configurations. Taken together, these results portray ASMR on YouTube as a multipolar field defined by overlapping stylistic and functional logics, structured both by long-standing genre conventions (e.g., sleep and relaxation scripts) and by newer infrastructural and social developments (e.g., the prominence of mukbang and short-form, high-velocity content).

5 Limitations and future work

This study has several important limitations that also suggest concrete directions for future research. Our corpus is restricted to YouTube videos longer than 60 s and retrieved via a keyword-based pipeline centred on the query "ASMR". In practice, this means that we focus on the long-form ASMR tradition that has developed around session-length content lasting many minutes or hours, and on videos whose creators explicitly label them as ASMR in the title. Platforms that historically imposed strict limits of only a few seconds are structurally less hospitable to such formats and therefore host relatively little canonical ASMR, although short-form variants of "soothing" or ASMR-like content have become more visible on TikTok, Instagram Reels, Facebook, and YouTube Shorts. Because such materials fall largely outside our sampling frame, the present findings characterise patterns in long-form, explicitly labelled YouTube ASMR rather than the full cross-platform ecosystem. Future work could extend the corpus by combining richer query sets (e.g., sleep-, whisper-, or trigger-related terms) with short-form data from multiple platforms and explicitly compare stylistic patterns, trigger types, and engagement dynamics across duration regimes and ecosystems.

A further limitation is that our analysis is entirely based on platform metadata and textual information (titles, descriptions, and related fields). Such metadata is creator-dependent, often incomplete, and not standardised, which can lead to under-detection or misclassification of ASMR themes. Many core ASMR triggers—especially auditory triggers such as tapping, scratching, mouth sounds, or brushing—cannot be reliably inferred from text alone, and nuanced differences in performance style (e.g., microphone technique, pacing, or camera work) remain opaque at the metadata level. Subsequent studies should incorporate multimodal features, including audio- and video-based descriptors (e.g., spectrogram features, visual trigger detection, gesture and camera-movement analysis), to obtain a more faithful representation of the sensory content of ASMR videos.

The focus on a single platform also introduces platform-specific biases. YouTube has its own recommendation algorithms, audience composition, and production norms, which shape how ASMR content is produced, surfaced, and engaged with. As a result, the engagement metrics, language distributions, and temporal trends observed here may not generalise to other platforms where ASMR culture, audience behaviour, and content curation differ. A cross-platform perspective that combines data from ecosystems such as TikTok, Instagram, and Facebook would allow researchers to examine how platform design, affordances, and recommendation logic influence ASMR production and consumption, and to test whether the length–engagement patterns observed here are specific to long-form YouTube videos or extend to short-form formats.

Methodologically, the use of t-SNE as a non-linear embedding for visualising similarities among videos comes with well-known constraints: it is sensitive to hyperparameters, initialisation, and random seeds, and it does not provide a straightforward global notion of distance. The resulting two-dimensional maps should therefore be interpreted as exploratory visualisations rather than as a definitive taxonomy of ASMR subgenres. Future work could explore more interpretable and reproducible embedding strategies, such as transformer-based text embeddings, multimodal (audio–visual) representations, or supervised embeddings aligned with manually annotated ASMR categories, and then revisit clustering or community detection using these representations.

Finally, even though our sliding-window sampling strategy spans the period from January 2008 to December 2025 and partition searches into three-month upload windows, the temporal and linguistic coverage of the dataset remains uneven: some years and some languages are sparsely represented, and keyword-plus date-bounded retrieval cannot guarantee complete recall for high-volume periods. This imbalance can make longitudinal interpretations less stable for early years or low-sample languages and can inflate the apparent importance of heavily represented recent periods. Future studies should explicitly model such imbalances, for example, by using hierarchical or time-series models that account for varying data density, by constructing more balanced sampling schemes across time, language, and platform, or by combining keyword-based retrieval with channel-centric sampling. Together, these extensions—short-form and cross-platform data, multimodal content analysis, more robust embeddings, and temporally aware modelling—would provide a more comprehensive picture of how ASMR is produced, experienced, and transformed across the contemporary media landscape.

6 Supplementary Material

In line with current open science practices and recommendations for transparency in user research [6], the authors openly provide these research artefacts to support reproducibility, collaboration and further advancements in the field. The analysis code and the responses of the participants are available at https://doi.org/10.4121/c4c88e1e-543e-4b37-9c31-af9431a8aa01. The code is available at https://github.com/Shaadalam9/ASMR-analysis.

CRediT authorship contribution statement

6.1 Authors' contributions

Conceptualisation: Md Shadab Alam, Pavlo Bazilinskyy; Methodology: Md Shadab Alam; Data Curation: Md Shadab Alam; Formal analysis and investigation: Md Shadab Alam, Pavlo Bazilinskyy; Writing - original draft preparation: Md Shadab Alam; Writing - review and editing: Md Shadab Alam, Pavlo Bazilinskyy; Resources: Pavlo Bazilinskyy; Software: Md Shadab Alam; Investigation: Md Shadab Alam; Supervision: Pavlo Bazilinskyy;

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Abdallah, D., Engström, M., 2020. So far away, yet so close: A study on how intimacy is attempted to be produced in girlfriend roleplay asmr videos on youtube. URL: http://www.diva-portal.org/smash/get/diva2:1392245/FULLTEXT01.pdf.

[2] Admetrics, 2024. Asmr marketing: Transforming brand engagement. URL: https://www.admetrics.io/en/post/asmr-marketing. accessed: 2025-12-10.

[3] Andersen, J., 2015. Now you've got the shiveries: Affect, intimacy, and the asmr whisper community. Television & New Media 16, 683–700. doi:https://doi.org/10.1177/1527476414556184.

[4] ASMR University, 2022. Global asmr survey by deezer uncovers helpful data about asmr. URL: https://asmruniversity.com/2022/03/14/deezer-asmr-survey/. accessed: 2025-12-09.

[5] Barratt, E.L., Davis, N.J., 2015. Autonomous sensory meridian response (asmr): a flow-like mental state. PeerJ 3, e851. doi:10.7717/peerj.851.

[6] Ebel, P., Bazilinskyy, P., Colley, M., Goodridge, C.M., Hock, P., Janssen, C.P., Sandhaus, H., Srinivasan, A.R., Wintersberger, P., 2024. Changing lanes toward open science: Openness and transparency in automotive user research, in: Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 94–105. URL: https://doi.org/10.1145/3640792.3675730, doi:10.1145/3640792.3675730.

[7] Engelbregt, H., Brinkman, K., Van Geest, C., Irrmischer, M., Deijen, J.B., 2022. The effects of autonomous sensory meridian response (asmr) on mood, attention, heart rate, skin conductance and eeg in healthy young adults. Experimental Brain Research 240, 1727–1742. doi:10.1007/s00221-022-06377-9.

[8] Feiz, D., Alipour, S., Asgharinajib, M., 2022. Investigating the effect of advertisements based on autonomous sensory meridian response (asmr) on brand persona with netnographic approach. Quarterly Journal of Brand Management 9, 13–42. doi:10.22051/bmr.2022.39386.2305.

[9] Fishman, E., 2016. How long should your next video be? URL: https://eddl.tru.ca/wp-content/uploads/2019/08/EDDL5101_W5_Fisherman_2016.pdf. wistia Learning Center; accessed via EDDL5101 course PDF.

[10] Gallagher, R., 2016. Eliciting euphoria online: The aesthetics of "asmr" video culture. Film Criticism 40. doi:https://doi.org/10.3998/fc.13761232.0040.202.

[11] Garro, D., 2017. Autonomous meridian sensory response–from internet subculture to audiovisual therapy, in: Electronic Visualisation and the Arts (EVA 2017), BCS Learning & Development. doi:10.14236/ewic/EVA2017.79.

[12] Guo, P.J., Kim, J., Rubin, R., 2014. How video production affects student engagement: An empirical study of mooc videos, in: Proceedings of the first ACM conference on Learning@ scale conference, pp. 41–50. doi:https://doi.org/10.1145/2556325.2566239.

[13] IZEA, 2023. How influencers and brands can use the asmr trend. URL: https://izea.com/resources/how-influencers-and-brands-can-use-the-asmr-trend/. accessed: 2025-12-10.

[14] Klausen, H.B., 2021. Asmr explained: Role play videos as a form of touching with the eyes and the ears. First Monday doi:https://doi.org/10.5210/fm.v26i9.11691.

[15] Łapińska, J., 2022. Tingles, sparkles, shivers: Language of affect in online discussions on autonomous sensory meridian response. Beyond Philology An International Journal of Linguistics, Literary Studies and English Language Teaching , 35–52doi:10.26881/bp.2022.3.02.

[16] Liikkanen, L.A., Salovaara, A., 2015. Music on youtube: User engagement with traditional, user-appropriated and derivative videos. Computers in human behavior 50, 108–124. doi:https://doi.org/10.1016/j.chb.2015.01.067.

[17] MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations, in: Le Cam, L.M., Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, CA. pp. 281–297. URL: http://projecteuclid.org/euclid.bsmsp/1200512992.

[18] Maddox, J., 2021a. On the limits of platform-centric research: Youtube, asmr, and affordance bilingualism. International Journal of Communication 15, 21. URL: https://ijoc.org/index.php/ijoc/article/view/16305.

[19] Maddox, J., 2021b. What do creators and viewers owe each other? microcelebrity, reciprocity, and transactional tingles in the asmr youtube community. First

Monday doi:https://doi.org/10.5210/fm.v26i1.10804.

[20] Mahady, A., Takac, M., De Foe, A., 2023. What is autonomous sensory meridian response (asmr)? a narrative review and comparative analysis of related phenomena. Consciousness and cognition 109, 103477. doi:https://doi.org/10.1016/j.concog.2023.103477.

[21] Miller, M., 2017. Asmr, the internet subculture of "sounds that feel good," is going mainstream. Fast Company URL: https://www.fastcompany.com/90135533/asmr-the-internet-subculture-of-sounds-that-feels-good-is-going-mainstream. accessed: 2025-12-10.

[22] Mooney, A., Klein, J., 2016. Asmr videos are the biggest youtube trend you've never heard of. Think with Google URL: https://www.thinkwithgoogle.com/intl/en-emea/consumer-insights/consumer-trends/asmr-videos-are-biggest-youtube-trend-youve-never-heard/. accessed: 2025-12-10.

[23] Munaro, A.C., Hübner Barcelos, R., Francisco Maffezzolli, E.C., Santos Rodrigues, J.P., Cabrera Paraiso, E., 2021. To engage or not engage? the features of video content on youtube affecting digital consumer engagement. Journal of consumer behaviour 20, 1336–1352. doi:https://doi.org/10.1002/cb.1939.

[24] Niu, S., Manon, H.S., Bartolome, A., Ha, N.B., Veazey, K., 2022. Close-up and whispering: An understanding of multimodal and parasocial interactions in youtube asmr videos, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. doi:10.1145/3491102.3517563.

[25] Poerio, G.L., Blakey, E., Hostler, T.J., Veltri, T., 2018. More than a feeling: Autonomous sensory meridian response (asmr) is characterized by reliable changes in affect and physiology. PloS one 13, e0196645. doi:https://doi.org/10.1371/journal.pone.0196645.

[26] Portas Ruiz, E., 2022. Youtube, the attention economy and digital audience interest in sponsored asmr videos. The Anáhuac journal 22, 12–41. doi:https://doi.org/10.36105/theanahuacjour.2022v22n2.01.

[27] Reichstein, T., Dost, F., 2025. How video length shapes engagement, in: Proceedings of the 2025 International Marketing Trends Conference, Venice, Italy. pp. 1–9. URL: https://archives.marketing-trends-congress.com/2025/pages/PDF/060.pdf.

[28] Sakurai, N., Nagasaka, K., Sasaki, K., Yuguchi, Y., Takahashi, S., Kasai, S., Onishi, H., Kodama, N., 2023a. The relaxation effect of autonomous sensory meridian response depends on personal preference. Frontiers in Human Neuroscience 17, 1249176. doi:https://doi.org/10.3389/fnhum.2023.1249176.

[29] Sakurai, N., Nagasaka, K., Takahashi, S., Kasai, S., Onishi, H., Kodama, N., 2023b. Brain function effects of autonomous sensory meridian response (asmr) video viewing. Frontiers in Neuroscience 17, 1025745. doi:10.3389/fnins.2023.1025745.

[30] Smith, N., Snider, A.M., 2019. Asmr, affect and digitally-mediated intimacy. Emotion, Space and Society 30, 41–48. doi:https://doi.org/10.1016/j.emospa.2018.11.002.

[31] Song, M., Yang, Z., Parada-Cabaleiro, E., Jing, X., Yamamoto, Y., Schuller, B., 2023. Identifying languages in a novel dataset: Asmr-whispered speech. Frontiers in Neuroscience 17, 1120311. doi:https://doi.org/10.3389/fnins.2023.1120311.

[32] The case, .H., 2019. 'safe and sound': What technologically-mediated asmr is capable of through sound. SoundEffects 8. URL: https://www.soundeffects.dk/article/view/115035.

[33] Think with Google, 2016. Asmr videos are the biggest youtube trend you've never heard. URL: https://www.thinkwithgoogle.com/intl/en-emea/consumer-insights/consumer-trends/asmr-videos-are-biggest-youtube-trend-youve-never-heard/. accessed: 2025-12-09.

[34] Triani, Y., Kamil, A.I., Arifin, A., Chalid, I., Fitri, D., Aulia, F.A., 2025. A netnographic analysis of asmr consumption on youtube among individuals with sleep and anxiety disorders. Jurnal Ilmiah Guru Madrasah 4, 155–166. doi:10.69548/jigm.v4i1.53.

[35] Wang, Q., 2023. Digital intimacy: a multimodal discourse analysis on asmr videos. Open Access Library Journal 10, 1–9. doi:10.4236/oalib.1110538.

[36] Woods, N., Turner-Cobb, J.M., 2023. 'it's like taking a sleeping pill': Student experience of autonomous sensory meridian response (asmr) to promote health and mental wellbeing. International journal of environmental research and public health 20, 2337. doi:10.3390/ijerph20032337.

[37] Zarazaga, P.P., Henter, G.E., Malisz, Z., 2023. A processing framework to access large quantities of whispered speech found in asmr, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 1–5. doi:10.1109/ICASSP49357.2023.10095965.

[38] Zelios Agency, 2025. How long should your video be in 2025? marketing video length best practices. URL: https://zelios.agency/best-video-length/. accessed: 2025-12-09.

Received 20 February 2025; revised 12 March 2025; accepted 5 June 2025