Nineteen Years of ASMR on YouTube: A Multilingual, Theme-Level Analysis of 42,268 Videos

Md Shadab Alam[a,*], Pavlo Bazilinskyy[a]

[a]*Eindhoven University of Technology, Department of Industrial Design, Eindhoven, 5612AZ, North Brabant, The Netherlands*

Abstract

ASMR videos have become a major genre on online platforms, yet their large scale characteristics remain underexplored. Using the YouTube Data API and a pytube-fix workflow, we assemble a dataset of 42,268 ASMR videos from 8,587 channels (2008 to 2026, 34 languages) enriched with duration, views, likes, inferred language, theme flags, and lemmatised title and description text. English dominates (76.94% of videos), followed by Korean, Japanese, Spanish, Dutch, and Portuguese. Across the corpus, the mean growth is 1,786.69 views per day and the duration analysis shows that short videos (under 10 minutes) average 3,435.52 views per day versus 1,005.91 for 10 to 30 minute content, while very long (over 180 minutes) videos reach 2,481.93 views per day. Theme detection indicates that drive themed content (17.14%), sleep related content (17.13%), and visual trigger content (15.54%) are particularly prevalent, with whisper (10.83%) and binaural videos (8.55%) also common. K means clustering on multimodal text, language, and engagement features, visualised with tSNE, yields 11 content clusters (1 to 29,321 videos) and a small set of extremely high growth outliers.

Keywords: ASMR, YouTube, Corpus analysis, t-SNE, Content themes

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 [1, 2].

*Corresponding author.

Email addresses: `m.s.alam@tue.nl` (Md Shadab Alam), `p.bazilinskyy@tue.nl` (Pavlo Bazilinskyy)

Viewers often describe ASMR as a form of self-care, reporting reductions in stress and anxiety, improved mood, and better sleep [3, 4]. Experimental work has increasingly established ASMR as a reproducible psychophysiological phenomenon rather than a purely anecdotal curiosity. Studies such as Poerio et al. [1] and Engelbregt et al. [5] 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. [6], emphasise heterogeneity in methods, stimuli, and results and situate ASMR in relation to neighbouring constructs such as frisson, synaesthesia, and misophonia.

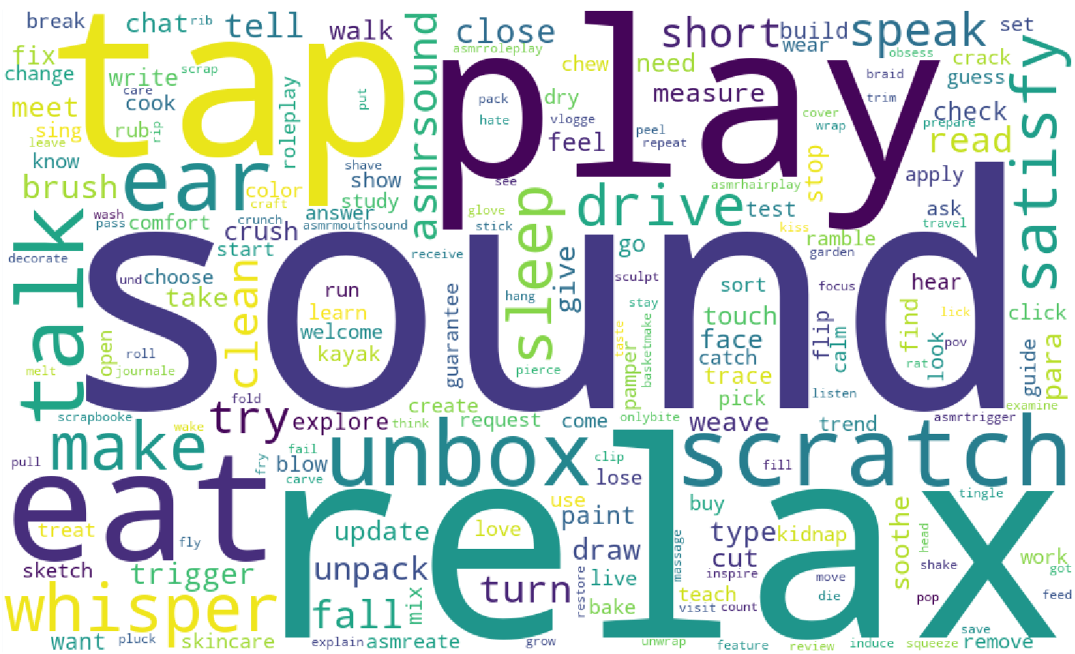

Figure 1: Word cloud of the 200 most frequent verb lemmas extracted from the titles of all 42,268 ASMR videos in the dataset. After lemmatisation, removal of a custom stopword list, and filtering to tokens tagged as verbs, each lemma is shown with font size proportional to its corpus frequency.

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 [7] and Smith and Snider [8] analyse how ASMR creators and viewers negotiate pleasure, stigma, and shifting boundaries between public and private life. Later contributions such as Klausen [9] 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 suggest that ASMR or ASMR-like content may be especially important for younger audiences, with trend studies reporting high self-reported use among 16–24-year-olds and Generation Z for relaxation, studying and sleep support [11, 12]. In the peer-reviewed literature, ASMR research has similarly concentrated on young adults–experimental studies often recruit participants in their early 20s–so the strongest empirical evidence currently available mainly reflects younger, online-active cohorts rather than population-representative age gradients [2, 5, 13].

In parallel, a growing body of work asks how video length relates to attention and engagement on digital platforms. Large-scale analyses of YouTube viewing behaviour show that audience retention generally declines rapidly early in playback and that view-duration trajectories vary substantially across videos, channels, and contexts, implying that there is no single universally optimal length [14]. 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 [15]. 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 [16]. 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" [17]. However, these studies focus on general, promotional or instructional content rather than ASMR, leaving open whether a relaxation-orientated genre follows similar length–engagement patterns or displays its own session-length preferences.

Historically, ASMR on YouTube has developed through overlapping phases. Scholarly accounts frequently point to WhisperingLife's early whisper-only upload *Whisper 1 — hello!* (2009) as a landmark in the emergence of intentionally produced "trigger" content and in consolidating an identifiable whisper/ASMR community around dedicated channels rather than incidental triggers in other genres (`https://www.youtube.com/watch?v=IHtgPbfTgKc`) [18, 19]. 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 [18, 19]. A subsequent phase of mainstreaming and platform growth (approximately 2012–2018) saw ASMR become increasingly legible as a distinct YouTube genre, as creators and audiences learnt to work with platform features (search, tags, recommendation and monetisation systems) to produce, find, and sustain ASMR content on scale [20, 19]. 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, and eating-trigger content) alongside increasingly sophisticated production practices, including higher-fidelity and spatial audio recording [19, 20]. Most recently, a phase of commercialisation and platform changes (2020–present) has been marked by sponsored and branded ASMR content and creator strategy shifts in response to algorithmic governance and monetisation pressures [21, 22, 20].

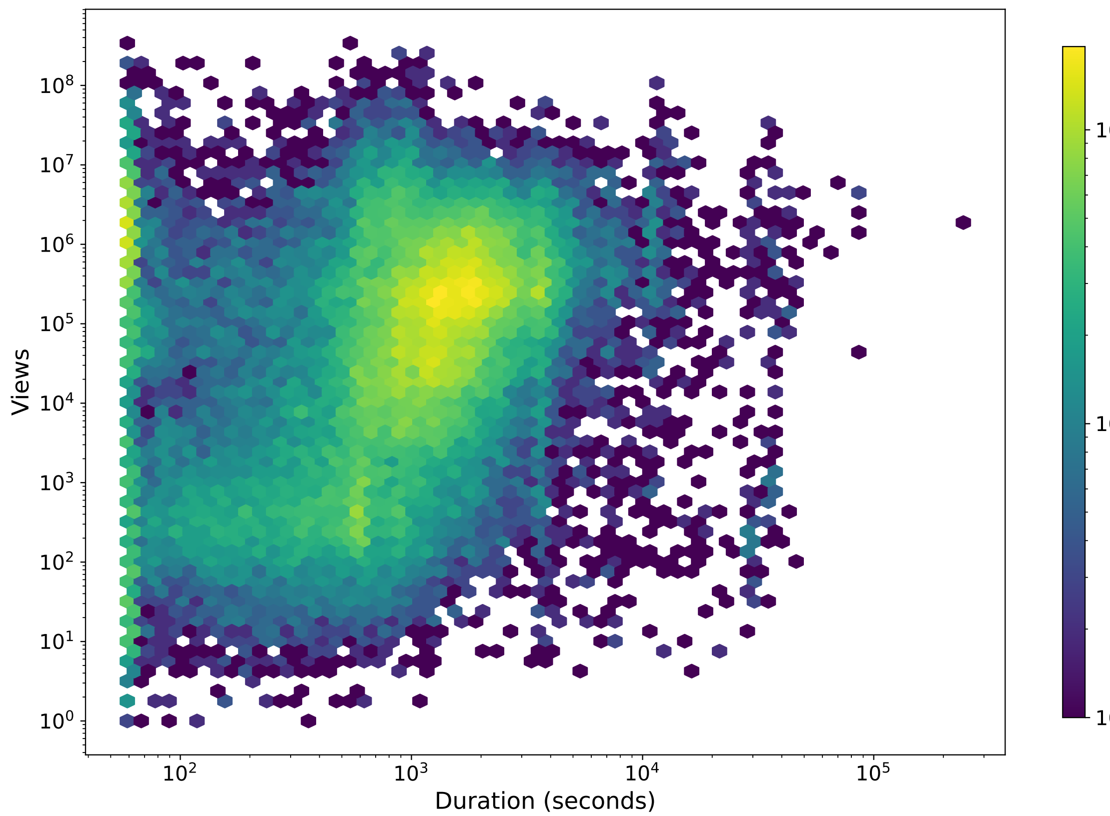

Figure 2: Log–log hexagonal bin plot showing each video's duration in seconds (x-axis) against its cumulative view count at the time of data collection in January 2026 (y-axis). The plot includes 42,259 ASMR videos with positive duration and view counts; each hexagon aggregates multiple videos, with colour intensity indicating the number of videos in that bin.

The 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. [23], 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 [24] and Łapińska [25], 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 sub-genres. Studies such as Wang [26], Klausen [19], and Abdallah [27] combine multimodal discourse analysis with concepts including haptic audio-visuality, ambient co-presence, and digital intimacy 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. [28] and Song et al. [29] 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 [20, 30] investigate how ASMR creators navigate YouTube's affordances, monetisation regimes, and community norms, while Portas Ruiz [21] and Feiz et al. [22] examine the integration of ASMR into influencer marketing and advertising. More general work on YouTube engagement, such as Liikkanen and Salovaara [31], 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 42,268 videos uploaded between 2008 and 2026 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 an additional retrieval workflow implemented in the *pytubefix* library (`https://pytubefix.readthedocs.io`). The analysis is conducted on publicly available, video-level information (e.g. titles, descriptions, and engagement statistics) and does not reproduce or redistribute audiovisual content; where examples are discussed, we cite the original YouTube URLs. According to YouTube's guidance on copyright and fair use, such uses may be permissible for research purposes, noting that fair use determinations are context-specific and that applicable copyright exceptions vary between jurisdictions [1]. 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 6 January 2026. 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

[1] `https://support.google.com/youtube/answer/9783148?hl=en&sjid=15816069292466875886-EU`

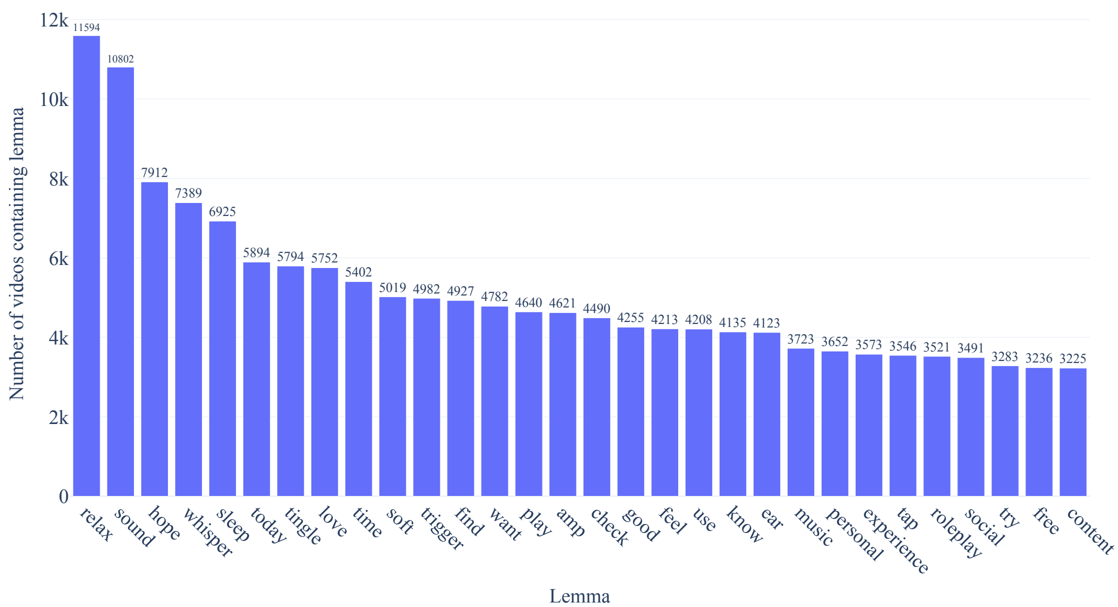

Figure 3: Visualisation of the 30 most frequent lemmatised content words in the combined titles and description of all 42,268 ASMR videos. The x-axis lists lemmas, and the y-axis shows the number of distinct videos in which each lemma appears at least once (document frequency).

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 seconds. 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 111,915.09 at k = 4 to 68,339.70 at k = 11, but the marginal gain drops substantially beyond this point (for example, only a 2.89% reduction to 66,367.53 at k = 12, and then a further 15.80% reduction spread over eight additional clusters, i.e. on average 1.98% per additional cluster up to 55,878.46 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 42,268 ASMR videos collected from 8,587 distinct channels, uploaded between 1 January 2008 and 6 January 2026. All likes and views for each video were updated on 17 January 2026. In all videos, 34 different languages were identified. English accounted for 32,521 videos (76.94%), followed by Korean (n=1,483, 3.51%), Japanese (n=1,328, 3.14%), Spanish (n=1,110, 2.63%), Dutch (n=923, 2.18%), Portuguese (n=838, 1.98%), French (n=671, 1.59%), Russian (n=667, 1.58%) and German (n=407, 0.96%). The average video duration was 1,547.26 s (SD = 3,237.04). The mean number of views per video was 1,196,672.40 (SD = 7,104,220.79), and the mean number of likes was 21,079.16 (SD = 127,953.86). The derived metric of views per day had a mean of 1,786.69 (SD = 18,102.02).

For text-based analyses, a word cloud was generated from concatenated titles and descriptions of all 42,268 videos. A spaCy-based 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=11594), *sound* (n=10802), *hope* (n=7912), *whisper* (n=7389), *sleep* (n=6925), *today* (n=5894), *tingle* (n=5794), *love* (n=5752), *time* (n=5402) and *soft* (n=5019) (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=4579; 10.83%), no-talking (n=2336; 5.53%), sleep-related (n=7241; 17.13%), binaural or 3D-audio (n=3615; 8.55%), role-play (n=5694; 13.47%), ear-cleaning or ear-focused (n=1382; 3.27%), mukbang or eating (n=3013; 7.13%), keyboard or typing (n=1752; 4.14%), visual or hand-movement triggers (n=6572; 15.55%), and drive-themed content (n=7246; 17.14%). 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 25,421 slow-growth videos (60.14%), 14,532 medium-growth videos (34.38%), 2,283 fast-growth videos (5.40%), and 32 unknown (0.08%).

Videos were also classified into duration buckets. Short form videos under 10 min accounted for 13,047 videos (30.87%), medium length videos between 10 and 30 min for 19,146 videos (45.30%), upper medium videos between 30 and 60 min for 7,280 videos (17.22%), long form videos between 60 and 180 min for 2,237 videos (5.29%), and very long videos exceeding 180 min for 558 videos (1.32%); no videos 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 2,448,759 views with an engagement rate of 0.03, whereas videos under 10 min averaged 1,338,694 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 42,259 videos with positive duration and non zero views (Figure 2). Duration ranged from 59 s to 244,211 s (median = 993 s, IQR = 478 to 1,752 s), with a mean of 1,547.56 s (SD = 3,237.31). Views ranged from 1 to 3.40×10^8 (median = 77,924; IQR = 4,146 to 461,280), with a mean of 1,196,672.40 (SD = 7,104,220.79). For the 42,259 videos with positive views, $\log_{10}$(views) had a mean of 4.62 and a standard deviation

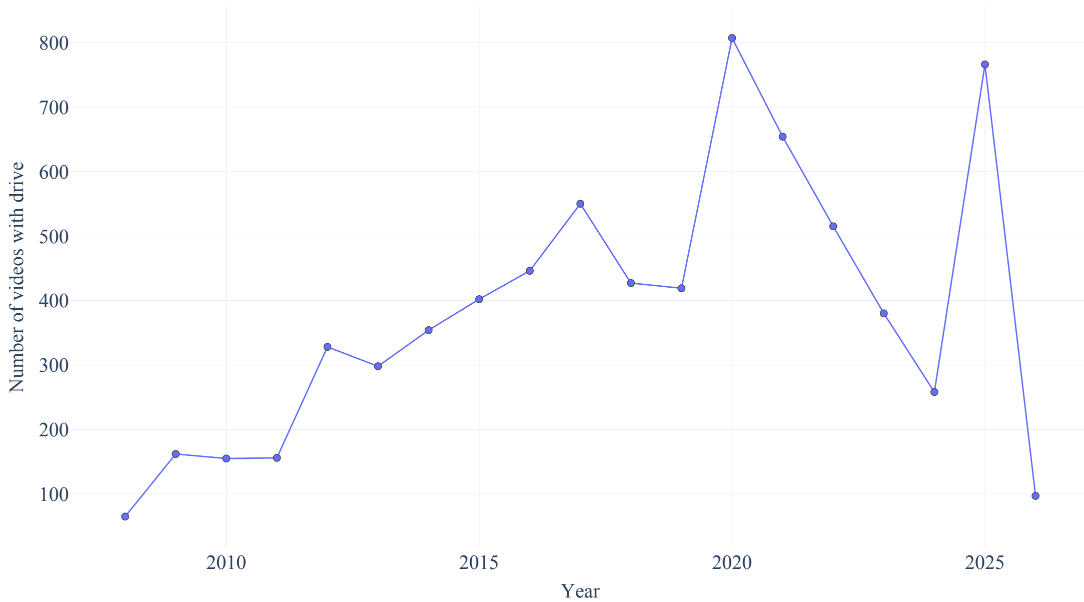

Figure 4: Yearly count of "driving"-themed ASMR videos, where the driving theme is defined via titles or descriptions containing the lemma *drive* or related phrases (e.g. "driving", "drive with me", "car", "road trip").

of 1.44. The D'Agostino Pearson normality test yielded $k^2 = 1971.77$ and $p = 0$. The Shapiro Wilk test (on a subsample) yielded $W = 0.974$ and $p = 2.04 \times 10^{-29}$.

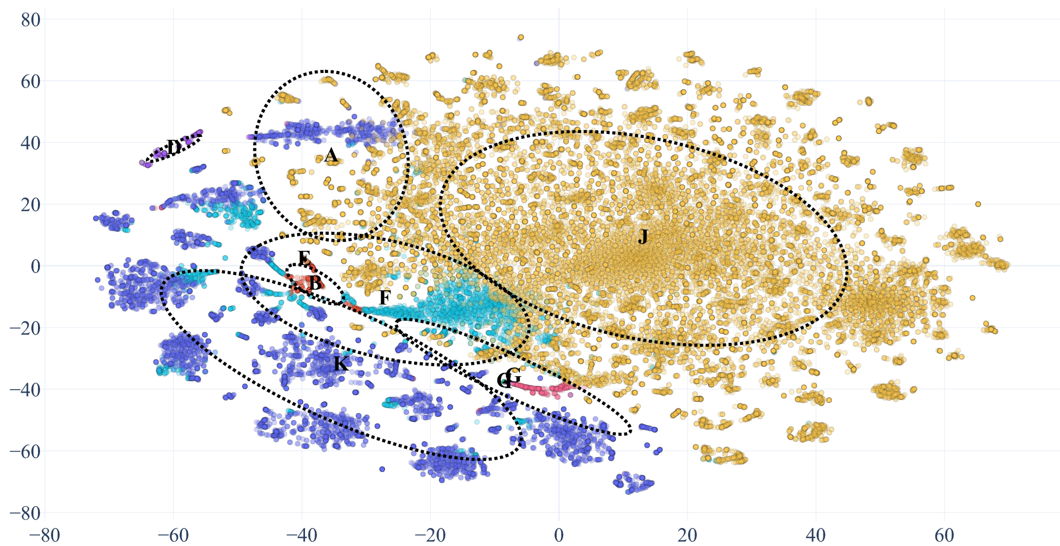

Figure 5: Two-dimensional t-SNE projection of all videos in a joint feature space combining text (TF–IDF over titles and descriptions), duration, engagement rate, views per day, and language. Each point is a video, coloured by its k-means cluster assignment, with faint ellipses indicating the approximate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

The descriptive statistics of views per day for the focal themes are summarised in Table 4. Whisper videos (n=4,572) had a mean of 768.55 views/day, drive themed videos (n=7,239) had 1,538.14 views/day, no talking videos (n=2,334) had 1,796.56 views/day, sleep related videos (n=7,231) had 1,374.01 views/day, and binaural videos (n=3,612) had 458.78 views/day.

Theme trends were computed for all years with valid upload dates (2008 to 2026). For no talking content, the overall trend comprised 19 yearly observations (2,334 videos), with 443 language year combinations included in the by language breakdown. For binaural content, the overall trend also included 19 yearly observations (3,612 videos). A lemma based temporal trend was also computed for the lemma "drive" (Figure 4).

Finally, we use a tSNE embedding to visualise similarities among videos in the learnt feature space. The two dimensional projection (shown in Figure 5) places all 42,268 videos in 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 [32], the clusters labelled A to K in the figure range in size from 1 to 29,321 videos. The largest groups, clusters A and G (29,321 and 3,223 videos, respectively), are predominantly 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 817.44 and 983.94 views per day. A sizeable non English cluster B (8,109 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 1,545.03 views per day. Additional medium sized clusters, such as F and J (842 and 211 videos, respectively), capture multilingual roleplay and visually rich formats with a mean growth of 2,481.62 and 841.44 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 (36 videos) grows more modestly at 4.49 views per day, while a closely related cluster D (261 videos) also exhibits low growth at 6.39 views per day. At the extreme, clusters E (238 videos) and K (21 videos) consist largely of hyper viral content: E is characterised by eating and mukbang themes and attains mean growth of 82,047.71 views per day, whereas K comprises ultra short, clip like videos with mean growth of 359,013.56 views per day. Cluster I (5 videos) forms another hyper viral outlier group with mean growth of 1,291,381.28 views per day, while a small outlier cluster H (1 video) grows at 2,047.25 views per day. Taken together, the tSNE 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 provides an ecosystem-level account of ASMR video production on YouTube by combining large-scale metadata analysis with lexical modelling, rule-based theme detection, and unsupervised clustering. Whereas much prior research has focused on individual creators, specific triggers, or comment-based ethnographies, our dataset captures 42,268 videos uploaded between 2008 and 2026 from 8,587 channels across 34 languages. This scale and temporal range allow us to situate familiar ASMR formats within a broader and more heterogeneous media ecology and to examine how subgenres emerge, expand, and stabilise over time.

Language	n	Views	Views/day	Likes	Likes/day	Engagement ($\times10^{-2}$)
English	**32521**	1,093,007 (6,650,935)	1,514.42 (15,323.28)	19,544 (131,833)	36.00 (304.84)	2.40
Korean	1483	3,200,851 (9,398,295)	3,271.30 (9,675.31)	36,695 (81,360)	40.36 (99.68)	2.19
Japanese	1328	1,662,393 (6,032,126)	3,477.25 (12,475.65)	26,151 (90,323)	68.83 (221.27)	1.77
Spanish	1110	945,360 (2,758,214)	1,952.60 (6,422.83)	27,676 (105,703)	64.34 (174.12)	3.99
Dutch	923	59,847 (213,997)	36.82 (320.46)	1,274 (4,250)	0.71 (3.95)	2.21
Portuguese	838	1,504,314 (5,477,248)	5,029.58 (23,024.96)	48,636 (154,765)	161.82 (542.13)	6.43
French	671	560,139 (6,681,862)	2,516.78 (50,840.02)	11,286 (94,919)	48.27 (712.50)	3.53
Russian	667	347,718 (1,302,000)	314.71 (1,352.15)	8,397 (34,517)	7.95 (35.60)	2.80
German	407	651,808 (7,187,105)	894.69 (9,091.17)	15,748 (177,503)	23.82 (224.35)	3.35
Vietnamese	224	**12,594,350** (35,345,130)	7,663.87 (19,955.64)	117,411 (287,026)	68.15 (154.13)	1.80
Italian	230	905,868 (4,638,399)	2,324.48 (11,248.58)	25,619 (136,175)	60.85 (265.26)	3.09
Estonian	204	2,170,180 (13,624,480)	1,508.19 (7,816.07)	26,155 (91,124)	25.88 (96.53)	5.43
Indonesian	268	1,926,826 (8,974,991)	6,451.85 (30,336.05)	37,496 (183,369)	130.86 (606.51)	8.32
Filipino	242	240,724 (1,737,224)	7,870.41 (107,371.62)	4,810 (27,226)	118.27 (1,480.39)	4.23
Polish	71	282,768 (633,214)	259.94 (565.74)	7,162 (20,310)	10.64 (34.90)	3.03
Turkish	236	427,376 (1,570,326)	578.44 (1,514.66)	7,025 (33,035)	14.57 (42.67)	3.66
Unknown	161	397,732 (1,719,617)	969.25 (3,789.31)	11,629 (45,205)	29.02 (79.85)	6.39
Swahili	65	104,910 (631,043)	515.78 (3,131.02)	14,615 (35,282)	72.22 (175.37)	**9.08**
Norwegian	76	273,546 (893,755)	1,618.75 (7,116.13)	12,370 (45,796)	58.89 (193.99)	3.48
Afrikaans	49	214,105 (692,368)	363.93 (1,049.11)	6,734 (24,947)	14.47 (43.22)	4.98
Catalan	48	820,419 (2,297,143)	2,627.57 (10,002.39)	24,510 (54,665)	103.90 (382.30)	4.42
Danish	84	1,258,277 (5,939,378)	4,153.41 (13,450.83)	33,671 (164,776)	109.25 (304.77)	3.90
Bulgarian	57	267,408 (603,370)	527.27 (1,509.87)	5,302 (10,931)	10.64 (27.96)	2.65
Hungarian	40	159,746 (812,496)	60.69 (258.63)	7,913 (24,860)	3.44 (8.97)	4.08
Thai	62	1,103,466 (2,245,385)	951.25 (2,920.04)	21,125 (38,634)	21.09 (71.91)	2.72
Arabic	114	456,208 (2,173,089)	1,003.03 (6,942.28)	12,038 (61,941)	31.22 (250.97)	5.05
Romanian	24	3,414,458 (14,516,580)	5,072.05 (21,643.63)	88,763 (397,175)	147.90 (595.46)	4.98
Finnish	18	143,001 (156,841)	237.90 (453.08)	2,708 (2,893)	4.05 (8.66)	7.22
Ukrainian	9	199,311 (345,170)	428.68 (657.34)	9,367 (16,025)	18.67 (27.55)	3.57
Swedish	21	9,961,138 (42,935,880)	**29,874.52** (131,788.12)	183,480 (737,339)	542.97 (2,263.02)	4.96
Greek	4	66,948 (76,478)	60.88 (83.65)	1,523 (1,737)	1.59 (2.46)	3.58
Chinese (Simplified)	4	111,607 (216,619)	251.50 (501.06)	1,414 (2,778)	3.21 (6.41)	2.99
Chinese (Traditional)	2	178 (33)	0.08 (0.03)	3 (NA)	0.00 (NA)	1.94
Czech	7	122,594 (144,773)	178.02 (172.55)	4,257 (5,401)	6.40 (4.84)	6.14

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

Lexical analysis (Figure 3) indicates that ASMR titles and descriptions are organised around a stable vocabulary of relaxation, sensory experience, and affective orientation. The most frequent lemmas include *relax*, *sound*, *sleep*, *tingle*, *whisper*, *soft*, *love*, *hope*, *today*, and *time*. This lexicon aligns with experimental and ethnographic accounts that frame ASMR as a practice used for relaxation and sleep support [19, 1, 9]. At the same time, the prominence of affectively directed language (e.g., *hope*, *love*) and high-frequency interpersonal verbs in descriptions (e.g., *feel*, *want*, *find*, *know*, *check*) supports interpretations of ASMR as a technologically mediated form of comfort, care, and interpersonal address [7, 22, 8, 24]. Crucially, these patterns are visible at corpus scale, suggesting that such framing is not limited to a small set of highly visible channels.

The distribution of verb lemmas further reflects diversification in ASMR production. Title verbs such as *tap*, *scratch*, *talk*, *eat*, and *unbox* map onto established

Title length bucket	n	Views (SD)	Views/day (SD)	Likes (SD)	Likes/day (SD)	Engagement ($\times 10^{-2}$)
$\leq$ 5 words	6854	761,443 (3,972,762)	1,543.29 (21,967.83)	16,386 (99,829)	37.94 (339.43)	3.106
6 to 10 words	19775	878,334 (5,625,352)	1,509.42 (19,542.32)	18,766 (127,705)	41.31 (379.67)	2.512
11 to 20 words	15502	1,733,052 (9,252,941)	2,213.77 (13,811.34)	25,345 (138,041)	42.21 (243.91)	2.607
$>$ 20 words	137	8,210,510 (20,562,410)	5,667.08 (14,133.86)	83,899 (174,999)	52.85 (103.21)	3.618

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

Duration bucket	n	Views (SD)	Views/day (SD)	Likes (SD)	Likes/day (SD)	Engagement ($\times 10^{-2}$)
Under 10 min	13047	1,338,694 (8,851,847)	3,435.52 (31,746.01)	36,814 (230,803)	103.05 (633.98)	3.48 (5.57)
10 to 30 min	19146	1,190,007 (7,215,989)	1,005.91 (5,431.35)	16,484 (68,992)	17.89 (63.38)	2.51 (4.58)
30 to 60 min	7280	843,556 (2,478,901)	888.15 (2,251.89)	12,574 (33,471)	20.53 (67.93)	2.00 (2.34)
60 to 180 min	2237	1,262,812 (3,593,064)	1,605.63 (3,890.41)	16,330 (35,821)	30.44 (76.41)	2.18 (2.52)
Over 180 min	558	2,448,759 (8,335,487)	2,481.93 (5,261.99)	27,631 (89,223)	37.33 (73.68)	3.06 (5.17)

Table 3: Summary statistics by duration bucket.

trigger families and hybrid formats that merge ASMR with adjacent genres such as product interaction and eating content [23, 9]. In parallel, the prominence of the lemma *drive* in the temporal analysis (Figure 4) and the high prevalence of drive-themed content (n=7,246; 17.14%) indicate that car- and travel-adjacent soundscapes are no longer marginal within the explicitly labelled ASMR corpus. One plausible interpretation is that creators increasingly treat everyday infrastructures and ambient mobility as reliable sources of continuous sound, enabled by accessible recording hardware and reinforced by platform incentives that may reward longer sessions in some contexts. While this interpretation remains speculative, the observed prevalence suggests that driving-oriented ASMR should be treated as a major format alongside sleep framing and whisper-centred production.

Language-level results (Table 1) underscore the globalisation of ASMR production. English remains dominant (76.94%), but substantial activity is evident in Korean, Japanese, Spanish, Portuguese, French, Russian, and other language communities. Swedish exhibits the highest mean views per day among languages with at least 20 videos, although this estimate is based on a relatively small sample and should be interpreted cautiously. Engagement rates show a different pattern, with Swahili exhibiting the highest mean engagement and comparatively elevated engagement also observed for Portuguese and Arabic. Together, these differences suggest that ASMR is not merely replicated across languages but is actively reshaped within local creative cultures, consistent with arguments that ASMR practices are culturally situated rather than uniform across linguistic contexts [25, 24]. The present findings therefore support a shift from English-centric descriptions towards comparative work

Theme	n_{theme}	n_{vpd}	Views (SD)	Views/day (SD)	Likes (SD)	Likes/day (SD)	Engagement ($\times 10^{-2}$)
Binaural / 3D audio	3615	3612	779,406 (2,784,397)	458.78 (1,481.40)	10,033 (33,334)	9.23 (32.35)	1.66 (1.41)
Drive	7246	7239	1,228,129 (8,229,579)	1,538.14 (17,972.99)	19,680 (140,276)	35.58 (389.00)	2.40 (3.60)
No talking	2336	2334	2,389,971 (10,997,020)	1,796.56 (6,740.47)	29,893 (94,270)	26.41 (98.23)	2.73 (7.82)
Sleep related	7241	7231	1,076,671 (4,628,780)	1,374.01 (5,791.20)	17,177 (98,234)	31.78 (152.72)	2.23 (3.20)
Whisper	4579	4572	643,671 (2,806,356)	768.55 (4,028.99)	10,217 (43,346)	19.39 (102.61)	1.95 (1.53)

Table 4: Summary statistics for videos containing each thematic category (theme present only). n_{theme} is the number of videos flagged with the theme. n_{vpd} is the subset with non-missing views/day used for views/day and likes/day.

that attends to platform-native performance metrics across linguistic communities.

Duration-based patterns point to a mixed ecology of ASMR formats. Short videos under 10 minutes account for a substantial share of production (30.87%) and show the highest mean views per day, consistent with quick-access trigger content embedded in everyday routines. Very long videos exceeding 180 minutes remain rare (1.32%) but accumulate comparatively high mean total views and relatively high mean views per day, consistent with uses as background ambience and sleep support. Across the corpus, the heavy-tailed distribution of views and strong deviations from normality in $\log_{10}$(views) confirm that ASMR, like other YouTube genres [31], exhibits pronounced inequality: a small number of highly successful videos capture disproportionate attention, while the majority attract modest audiences.

Theme-based indicators further nuance common assumptions about which formats "perform best." When views/day is computed over videos with valid growth values, no-talking videos (n=2,334; 1,796.56 views/day) and drive-themed videos (n=7,239; 1,538.14 views/day) show higher mean growth than whisper videos (n=4,572; 768.55 views/day), and sleep-related videos also outperform whisper on average (n=7,231; 1,374.01 views/day). Binaural or 3D-audio content remains the lowest-growth category among the focal themes (n=3,612; 458.78 views/day). These patterns suggest that contemporary visibility is not driven solely by canonical whisper-based presentation or by technically sophisticated audio setups. Instead, growth appears to reflect an interplay of creative choices, audience use cases, and recommender dynamics, with particular advantage for formats that reduce linguistic dependence (no talking) or supply continuous ambience (drive, sleep).

Clustering analysis provides an additional lens on ASMR's internal heterogeneity. Using an eleven-cluster solution in a joint feature space combining text, behavioural metrics, and language, we identify groups that differ markedly in typical duration, language composition, views per day, and mean engagement. A very large, predominantly English cluster (29,321 videos) shows moderate mean growth (817.44 views/day), consistent with a broad, general-purpose ASMR core. A second large cluster (8,109 videos) is multilingual and dominated by Korean, Japanese, Dutch, Spanish, and Russian content, with higher mean growth (1,545.03 views/day), indicating a substantial non-English production regime that is not simply peripheral to the English core. Additional medium-sized clusters capture distinct format logics, including a long-duration, sleep-heavy cluster (842 videos; mean duration 160.77 minutes; mean growth 2,481.62 views/day) and another long-form sleep-oriented cluster with lower growth (211 videos; mean duration 594.40 minutes; mean growth 841.44 views/day). The clustering results also reveal very low-growth clusters with minimal mean views/day, alongside several hyper-viral clusters with exceptionally high mean growth, including a short-clip cluster (21 videos; 359,013.56 views/day) and a small outlier cluster (5 videos; 1,291,381.28 views/day). The coexistence of extreme high- and low-growth clusters reinforces that engagement metrics can be unstable in sparse or low-view regimes and that growth and engagement should be interpreted jointly rather than in isolation.

Finally, the t-SNE visualisation of the clustering solution suggests multiple dense regions separated by low-density transition zones, consistent with a genre organised around several semi-independent creative strategies rather than a single continuum. Dense regions plausibly correspond to widely shared production formulae (e.g., general-purpose whisper/talk mixes or sleep-support conventions), whereas smaller clusters and isolated points reflect unusual combinations of themes, languages, and presentation styles, including niche high-growth configurations. Overall, the results portray YouTube ASMR as a multipolar field structured by overlapping stylistic and functional logics, shaped both by long-standing genre conventions and by the growing prominence of non-verbal, ambience-oriented, and clip-like formats that can scale rapidly under platform distribution dynamics.

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 ASMR as it is explicitly labelled by creators and surfaced through keyword search, rather than on the full universe of ASMR adjacent or implicitly soothing content. This constraint is especially salient given the scale of the contemporary corpus and the increased prominence of shorter, clip like, and hybrid formats that may not use the label consistently. Short form variants of soothing or ASMR like content have become more visible on TikTok, Instagram Reels, Facebook, and YouTube Shorts, but such materials fall largely outside our sampling frame. The present findings therefore characterise patterns in explicitly labelled YouTube ASMR rather than the full cross platform ecosystem. Future work could extend the corpus by combining richer query sets (for example, sleep, no talking, whisper, binaural, roleplay, tapping, 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 (for example, microphone technique, pacing, or camera work) remain opaque at the metadata level. This limitation is amplified by the expanded multilingual coverage, because theme keywords and idiomatic trigger descriptions vary across languages and communities. Subsequent studies should incorporate multimodal features, including audio and video based descriptors (for example, spectrogram features, spatial audio cues, visual trigger detection, gesture and camera movement analysis), to obtain a more faithful representation of the sensory content of ASMR videos and to reduce reliance on language specific surface patterns.

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. Moreover, growth metrics such as views per day can reflect both audience demand and platform distribution effects, which may vary substantially across platforms and between long form and short form regimes. 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 and theme patterns observed here extend beyond YouTube.

Methodologically, the use of tSNE 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. This caution is particularly important given the extreme cluster size imbalance in the updated dataset, where a small number of very large clusters coexist with tiny outlier clusters that can be visually emphasised. 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 sampling strategy spans the period from January 2008 to January 2026 and partitions 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, using hierarchical or time series models that account for varying data density, constructing more balanced sampling schemes across time, language, and platform, or 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 [33], the authors openly provide these research artefacts to support reproducibility, collaboration and further advancements in the field. The full dataset collected during the study, together with all analysis and visualisation code, is available at `https://www.dropbox.com/scl/fo/7fyi4syvs059bctagj8ix/AOyIQjQrfQBCGgTpqONHQC4?rlkey=ro13uemf57ownq2gmparfoxw8&st=tj5jlorh&dl=0`. A mentioned verison of code is available at `https://github.com/Shaadalam9/ASMR-analysis`.

References

[1] G. L. Poerio, E. Blakey, T. J. Hostler, T. Veltri, More than a feeling: Autonomous sensory meridian response (asmr) is characterized by reliable changes in affect and physiology, PloS one 13 (6) (2018) e0196645. doi:https://doi.org/10.1371/journal.pone.0196645.

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

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

[4] N. Woods, J. M. Turner-Cobb, '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 (3) (2023) 2337. doi:10.3390/ijerph20032337.

[5] H. Engelbregt, K. Brinkman, C. Van Geest, M. Irrmischer, J. B. Deijen, 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 (6) (2022) 1727–1742. doi:10.1007/s00221-022-06377-9.

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

[7] J. Andersen, Now you've got the shiveries: Affect, intimacy, and the asmr whisper community, Television & New Media 16 (8) (2015) 683–700. doi:https://doi.org/10.1177/1527476414556184.

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

[9] . H. The case, 'safe and sound': What technologically-mediated asmr is capable of through sound, SoundEffects 8 (1) (2019).
URL https://www.soundeffects.dk/article/view/115035

[10] R. Gallagher, Eliciting euphoria online: The aesthetics of "asmr" video culture, Film Criticism 40 (2) (2016). doi:https://doi.org/10.3998/fc.13761232.0040.202.

[11] ASMR University, Global asmr survey by deezer uncovers helpful data about asmr, accessed: 2025-12-09 (3 2022).
URL https://asmruniversity.com/2022/03/14/deezer-asmr-survey/

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

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

[14] M. Park, M. Naaman, J. Berger, A data-driven study of view duration on youtube, in: Proceedings of the international AAAI conference on web and social media, Vol. 10, 2016, pp. 651–654. doi:https://doi.org/10.1609/icwsm.v10i1.14781.

[15] A. C. Munaro, R. Hübner Barcelos, E. C. Francisco Maffezzolli, J. P. Santos Rodrigues, E. Cabrera Paraiso, To engage or not engage? the features of video content on youtube affecting digital consumer engagement, Journal of consumer behaviour 20 (5) (2021) 1336–1352. doi:https://doi.org/10.1002/cb.1939.

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

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

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

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

[20] J. Maddox, On the limits of platform-centric research: Youtube, asmr, and affordance bilingualism, International Journal of Communication 15 (2021) 21.

[21] E. Portas Ruiz, Youtube, the attention economy and digital audience interest in sponsored asmr videos, The Anáhuac journal 22 (2) (2022) 12–41. doi:https://doi.org/10.36105/theanahuacjour.2022v22n2.01.

[22] D. Feiz, S. Alipour, M. Asgharinajib, Investigating the effect of advertisements based on autonomous sensory meridian response (asmr) on brand persona with netnographic approach, Quarterly Journal of Brand Management 9 (2) (2022) 13–42. doi:10.22051/bmr.2022.39386.2305.

[23] S. Niu, H. S. Manon, A. Bartolome, N. B. Ha, K. Veazey, 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, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3491102.3517563.

[24] Y. Triani, A. I. Kamil, A. Arifin, I. Chalid, D. Fitri, F. A. Aulia, A netnographic analysis of asmr consumption on youtube among individuals with sleep and anxiety disorders, Jurnal Ilmiah Guru Madrasah 4 (1) (2025) 155–166. doi:10.69548/jigm.v4i1.53.

[25] J. Łapińska, 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 (19/3) (2022) 35–52. doi:10.26881/bp.2022.3.02.

[26] Q. Wang, Digital intimacy: a multimodal discourse analysis on asmr videos, Open Access Library Journal 10 (8) (2023) 1–9. doi:10.4236/oalib.1110538.

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

[28] P. P. Zarazaga, G. E. Henter, Z. Malisz, 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, 2023, pp. 1–5. doi:10.1109/ICASSP49357.2023.10095965.

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

[30] J. Maddox, What do creators and viewers owe each other? microcelebrity, reciprocity, and transactional tingles in the asmr youtube community, First Monday (2021). doi:https://doi.org/10.5210/fm.v26i1.10804.

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

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

[33] P. Ebel, P. Bazilinskyy, M. Colley, C. M. Goodridge, P. Hock, C. P. Janssen, H. Sandhaus, A. R. Srinivasan, P. Wintersberger, 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, AutomotiveUI '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 94–105. doi:10.1145/3640792.3675730.