

Between Whispers and Screams: Loudness Standard Deviation as a Proxy for Explicit Content Detection in US Romance Novels

Svenja Guhr

School of Information, University of California, Berkeley
South Hall, Berkeley, CA 94720-4600, USA
sguhr@berkeley.edu

Abstract

This study proposes and tests loudness standard deviation (SD) of fictional sound events as an acoustically grounded proxy for detecting explicit content in romance fiction. Working with a subcorpus of novels from the Harlequin *Men Made in America* series, scenes were annotated for character and ambient sound with loudness levels. Additionally, the scenes were annotated on a ternary severity scale with two content advisory categories drawn from the PG-story taxonomy, SEX & NUDITY and VIOLENCE & SCARINESS (Tsai et al., 2024), and tested whether within-scene loudness SD of character and ambient sound correlates with either category. Loudness standard deviation analyses of character and ambient sounds in scenes featuring explicit content reveal that erotic scenes are acoustically marked by significantly higher variability in character-produced sounds, reflecting the dynamic range from whispered dialogue to vocalized arousal, while no significant correlation was found between high ambient sound loudness SD and scenes of elevated VIOLENCE & SCARINESS.

1 Introduction

Near the end of Carla Cassidy’s *Midnight Wishes* (1997), two characters are tied back-to-back to chairs in a burning room. The scene unfolds across two passages of mounting acoustic intensity: first an eerie silence broken only by frantic breathing and muffled struggling, then screaming, coughing, the growl of the fire, the snap and crackle of burning fabric, and finally a moment of quiet resignation in which the two protagonists entwine their fingers and wait for death through being burned alive. In Guhr’s sound annotation framework (Guhr and Algee-Hewitt, 2024; Guhr, 2026; Guhr and Mahlberg, 2026), this scene moves from ambient silence through character sounds of escalating distress to the roaring din of the fire itself, with loudness values oscillating across the full scale.

A content advisory system would flag it without hesitation: sustained, explicit VIOLENCE & SCARINESS, with detailed depictions of smoke inhalation, physical restraint, and psychological terror. It is also the climax of a Harlequin romance novel. The Harlequin series *Men Made in America* (MMiA) delivers precisely this unsettling mix of violent, erotic, and tender romance elements that readers browsing for a cosy love story might not anticipate. Anne Stuart’s *Tangled Lies* (1984), for instance, opens its very first page with a bomb blast, its fictional soundscape spiking from Christmas carols to an explosion before subsiding into distant screams and approaching sirens. What unites these scenes is that both are encoded in the text’s acoustic structure: dynamic oscillation between loud events and silence produces a distributional pattern in the loudness annotations that I argue constitutes a candidate textual signal for the presence of content that rating systems, such as those for movies, would identify.

This paper proposes and tests a method for locating such scenes computationally, using sound and loudness annotations applied to the corpus of 50 MMiA romance novels, with a subset of 10 novels being manually encoded with scene boundaries in TEI XML (Guhr et al., 2025). Doing so I want to find out to what extent the standard deviation of loudness levels across sound events per scene (loudness SD) serves as a candidate proxy for identifying scenes flagged under two content advisory categories drawn from the PG-story taxonomy, SEX & NUDITY, and VIOLENCE & SCARINESS (Tsai et al., 2024), annotated on a ternary severity scale (*absent*, *weak*, *strong*) extended from Tsai et al.’s binary scheme to capture the gradations that are narratologically significant in adult literary fiction.

I hypothesize, first, that scenes annotated as *weak* or *strong* under VIOLENCE & SCARINESS will show significantly higher loudness SD in ambient sounds than *absent* scenes, reflecting the dy-

dynamic oscillation between loud events and silence characteristic of action and threat (H1) (see also Guhr and Algee-Hewitt (2024)); and second, that scenes annotated under SEX & NUDITY will show a distinctive loudness profile with a significantly higher loudness SD in character sounds, a high diversity of loudness ranging from soft whispers and sighs to vocalized peaks, distinguishing them acoustically from both neutral and violent scenes through the predominance of character sounds over ambient sounds (H2).

The paper makes three contributions: (1) it extends Guhr’s sound annotation framework to the US-English romance genre and to scene-level content analysis; (2) it proposes a scene-level explicit content assessment annotation schema for literary fiction; and (3) it provides preliminary empirical evidence that loudness standard deviation is a candidate acoustic signal for identifying erotic scenes in romance fiction, while the parallel hypothesis for violent and scary scenes remains inconclusive given the limited data.

2 Related Work

The computational study of sound in literary prose builds on the auditory turn in the humanities (Bull and Back, 2003) and Schafer (1994)’s categorization of sounds. While traditional literary studies approaches rely on single-text analysis and close reading (Ackermann, 2003; Picker, 2003; Schweighauser, 2006; Glotova, 2021; Molyneux, 2022; Foley, 2023), early computational approaches identify sound words and references to auditory experiences via dictionary matching algorithms (Bernhart, 2008; Häußler et al., 2024). More recent work moves from transformer-based ambient sound word detection (Guhr and Algee-Hewitt, 2024) toward the classification of sound events for German-language fiction, defined as sound-representing narrative events with asserted *realis* (Vauth et al., 2021; Sims et al., 2019), distinguishing ambient from character sound events and assigning loudness levels on a 0; 1–5 scale (Guhr, 2024). Guhr and Mahlberg (2026) extend parts of this framework to English-language fiction via cross-lingual BERT fine-tuning, confirming its transferability to 19th-century British prose.

To analyze how sound event distributions relate to narrative content, a spatially and temporally coherent unit of analysis is required, such as scenes. Scene segments, defined by consistency

in time, place, and characters and understood to center around a particular action (Gius et al., 2019, 2021), constitute self-contained units within which loudness SD can be meaningfully computed and compared (Zehe et al., 2021, 2025).

For explicit content assessment, I adopt two categories from the PG-story taxonomy (Tsai et al., 2024) for content safety in narrative text. Automated content rating has a longer history in film, where audio-based approaches detect violent and sexual content from acoustic features (Gianakopoulos et al., 2006), and text-based approaches classify risk-relevant content directly from movie scripts (Martinez et al., 2019). Zhang et al. (2021) predict MPAA ratings from film dialogue using a graded severity scheme (*none*, *mild*, *moderate*, and *severe*). To my knowledge, however, no existing work applies content advisory annotation at the scene level to adult literary fiction, nor proposes an acoustically grounded narratological feature as a proxy for content advisory relevance. I address both gaps, retaining the SEX & NUDITY and VIOLENCE & SCARINESS categories from Tsai et al. (2024) and extending their binary scheme to a ternary scale (*absent*, *weak*, *strong*). While PG-story was originally designed for child-safe story generation, its content categories are sufficiently general to transfer to adult literary fiction: the categories name phenomena (sexual content, nudity, violence, scariness) rather than presuppose a particular audience or register. What changes in adult romance fiction is not whether they are applicable, but how they are calibrated: the ternary extension is specifically designed to capture the gradations between implicit and explicit representation that are narratologically meaningful in this genre, and that a binary *present/absent* scheme would collapse.

3 Method

3.1 Data Set

The corpus consists of 50 Harlequin romance novels from the *Men Made in America* (MMiA) series (1982–2002), each set in a different US-State and written exclusively by female authors. Despite their shared romance premise, the novels frequently blend detective fiction, Gothic suspense, and high-stakes drama, making the series a generically rich testing ground. All volumes have been digitized, OCR-processed, manually cleaned, and encoded in TEI XML. 20% of the corpus has been manually annotated for scene segments.

entity	all	train	devel	test (internal)	test (external)
ambient sound	222	175	24	23	21
character sound	3,073	2,441	318	314	77
unlabeled words	212,578	170,297	21,212	21,069	9,823
total words	215,873	172,913	21,554	21,406	9,921

Table 1: Distribution of training, development, internal and external test set for fine-tuning the US-English-language SoundBERT.

3.2 Sound Classification and Loudness Labeling

Fictional sound events are operationalized following Guhr (2026), who defines a fictional sound as a sound event with asserted *realis* in the fiction¹, explicitly represented by sound words in the narrative. Sound events are classified into ambient and character sound events and assigned loudness levels on a 0; 1–5 scale via dictionary matching. The loudness dictionary was originally validated for 19th-century German literary prose (Guhr, 2026) and subsequently extended to 19th-century British English (Guhr and Mahlberg, 2026); its application to 20th-century US romance fiction is a further transfer step.²

The loudness scale (0.0–5.0) is an *ordinal* scale: its values encode a ranked ordering of perceptual loudness categories, but intervals between adjacent levels are not empirically validated as equal. Computing arithmetic means and standard deviations from ordinal values is strictly speaking statistically impermissible, since it presupposes equal-interval structure that an ordinal scale does not guarantee (Mari et al., 2023; Gius, 2026); human loudness perception follows an approximately logarithmic rather than linear progression (Stevens, 1946), further complicating this assumption. Following Guhr (2026) and the principle that measurement results are valid within their model context (Gius, 2026), loudness mean and SD are retained as a *corpus-comparative heuristic* for relative comparison across scenes within a consistently applied operationalization, interpreted as *categorical spread* rather than a claim about absolute acoustic quantities.

While average loudness levels enable compara-

¹The definition builds on the understanding of narratological event by Vauth et al. (2021) and asserted *realis* by Sims et al. (2019).

²Ambiguous sound words (i.e. items not present in the dictionary) are assigned no loudness value and excluded from the SD computation; only scenes with at least two labeled sound events of the relevant type are included in the statistical analysis (see section 4).

tive analysis of sonic profiles across texts, authors, or literary movements (Guhr, 2026; Häußler et al., 2024), they suffer from a balancing effect: a scene containing equal parts soft whispers at level 2 and loud peaks at level 4 produces a misleading mean of level 3, obscuring the dynamic range that characterizes specific scene types. The standard deviation (SD) of loudness levels addresses this by measuring the degree of variation within a scene: a high SD indicates substantial sonic variation, with some sounds significantly louder or softer than others. By computing SD separately for ambient and character sound events, it becomes possible to investigate whether specific scene types carry distinguishable acoustic signatures (Guhr, 2026). This study is the first to apply this SD-based approach to US-English romance fiction, foregrounding it as the primary analytical feature and hypothesizing that high ambient SD signals violent and threatening scenes through dynamic oscillation between loud events and silence (H1), while high character SD signals erotic scenes through the whisper-to-scream-of-pleasure profile characteristic of intimate encounters (H2).

For adaptation to 20th-century US-English romance fiction, three MMiA novels were manually annotated for character and ambient sound events and enriched with loudness levels (see Table 1), following the annotation guidelines of Guhr (2024) and using an adapted version of the translated English loudness dictionary Guhr and Mahlberg (2026). A state-of-the-art English BERT model (Devlin et al., 2019) was fine-tuned in the NEISS NTEE software environment (Zöllner et al., 2021). Evaluated against the manually annotated gold standard, the model achieves a span-level classification F1 of 0.79 (any overlap) and 0.66 before post-processing, rising to 0.81 and 0.67 respectively after post-processing, with a Gamma coefficient (Malthet et al., 2015) of 0.72. These results compare favorably to the 0.66 entity-level F1 reported for 19th-century British English (Guhr and Mahlberg, 2026,

277), though direct comparison is complicated due to the different test sets and English-language variations: the model reliably identifies sound-bearing passages and assigns the correct label, but often does not reproduce precise annotator boundaries, a known challenge in span-level literary annotation.³

3.3 Adapting the Explicit Content Taxonomy

For the explicit content assessment of scenes, I adopt two of the five categories from the PG-story content safety taxonomy (Tsai et al., 2024, 79): SEX & NUDITY, defined as text containing “sexual activities, implied sex reference, or nudity”; and VIOLENCE & SCARINESS, defined as text containing “violent, risky or unhealthy dangerous activities; reference to weapons, anti-social, or self-harming behaviors; intense frightening or scary situations.”⁴ Where Tsai et al. (2024) apply a binary *present/absent* annotation, I extend both retained categories to a ternary severity scale (*absent, weak, strong*) to capture gradations in adult literary fiction (see operationalization in the Tables 8 and 7). The extension from Tsai et al. (2024)’s binary scheme is motivated by the romance genre specifically: implicit references to sexual content via creative metaphors and an explicitly sexual scene are both *present* under a binary annotation, same counts for a scene containing violence or a scene being about violence, yet, they differ fundamentally in narrative function, editorial content policy, and, I hypothesize, in their loudness variance profiles. The *absent* category additionally serves as the reference group in the statistical analysis (see section 4).

A total of 100 scenes from the MMiA corpus were manually annotated for explicit content to serve as training data (109,227 words). An additional 30 scenes were independently annotated by two annotators (15,632 words), yielding an inter-annotator agreement of $\kappa = 0.82$ (Cohen’s kappa) for the SEX & NUDITY class and $\kappa = 0.79$ for

³LLMs were not employed for the classification task: although they have been applied to annotation in computational literary studies (Bamman et al., 2024), their tendency toward hallucination in structured annotation tasks (Semin et al., 2026) makes them less suitable than fine-tuned discriminative models for reproducible span-level TEI XML annotation, particularly given that the copyright status of the MMiA corpus restricts processing to smaller local models.

⁴The remaining three categories (Profanity & Slurs, Substance Consumption, and Discrimination & Bias) are excluded on acoustic grounds: none maps cleanly onto either the ambient or character sound channel in ways that would produce theoretically interpretable loudness variance signals at this corpus scale.

the VIOLENCE & SCARINESS class. This doubly-annotated subset serves as the held-out test set for evaluating the automated prediction of both content categories.

Since the MMiA series is structurally sparse with respect to scenes of the two categories, class imbalance would render any classifier trained solely on the annotated corpus unreliable.⁵ To address this, synthetic training passages were generated using the Anthropic *claude-sonnet-4-20250514* model, producing 150 romance novel passages distributed across the three severity levels (*absent, weak, strong*) for each content advisory category⁶, following the operationalization definitions in Table 7 and Table 8, and matching the prose register, genre conventions, and US-English setting of the MMiA corpus. The combined training set comprises 250 passages (100 human-annotated + 150 synthetic), with a stratified 80/20 split yielding the 200/50 train/validation distribution reported in Table 2. The 30 doubly-annotated scenes are held out entirely as an external test set and do not contribute to training.

To automate the assessment of SEX & NUDITY and VIOLENCE & SCARINESS in scenes, I fine-tuned XLM-RoBERTa large⁷ (Conneau et al., 2019), adapting it for the three-class ordinal classification task. To address class imbalance in the training data, I applied balanced class weighting and used a stratified 80/20 train/validation split with a fixed random seed ($s = 42$) for reproducibility.

Model performance is evaluated on a held-out internal validation set (20% of the training data) as

⁵LLM-generated synthetic data for training augmentation in low-resource classification tasks has been shown to be an effective strategy for addressing class imbalance and data scarcity (Long et al., 2024), and is particularly appropriate here given that the copyrighted MMiA texts cannot be redistributed as open training data. To assess stylistic consistency, five randomly sampled synthetic passages per severity level were compared against human-written MMiA passages: synthetic passages reproduced the genre conventions (e.g. first-person interiority, dialogue-driven intimacy with low to moderate level of sensuality, “[w]armhearted stories offer[ing] a range of tones, from light humor to drama,” “fast-paced and plot-driven,” (see Harlequin Submission Guidelines for the *American Romance* series: <https://harlequin.submittable.com/submit> via the Internet Archive Way Back Machine, version from October 2015) but tended toward a slightly more contemporary register, which stylistic gap is a known limitation.

⁶Used prompt: “Please generate [n] passages of max. 350 words each for the class [level] [category] following the operationalization definitions [xy]. Output a csv.”

⁷A state-of-the-art multilingual transformer model pre-trained on 2.5 TB of text across 100 languages using a masked language modeling objective (Conneau et al., 2019).

task	class	train	val	total (%)	weight
VIOLENCE/SCARINESS	<i>absent</i>	68	17	85 (34.0%)	0.980
	<i>weak</i>	50	13	63 (25.2%)	1.323
	<i>strong</i>	82	20	102 (40.8%)	0.817
SEX/NUDITY	<i>absent</i>	66	16	82 (32.8%)	1.016
	<i>weak</i>	66	17	83 (33.2%)	1.004
	<i>strong</i>	68	17	85 (34.0%)	0.980
total scenes		200	50	250	

Table 2: Training and validation set statistics for the two classification tasks. Class weights are computed inversely proportional to class frequency to address label imbalance.

Class	P	R	F1	Sup
<i>absent</i>	0.81	0.81	0.81	16
<i>weak</i>	0.81	0.76	0.79	17
<i>strong</i>	0.89	0.94	0.91	17
macro avg	0.84	0.84	0.84	50
weighted avg	0.84	0.84	0.84	50

Table 3: SEX & NUDITY: validation ($n = 50$).

Class	P	R	F1	Sup
<i>absent</i>	0.85	1.00	0.92	17
<i>weak</i>	1.00	0.77	0.87	13
<i>strong</i>	1.00	1.00	1.00	20
macro avg	0.95	0.92	0.93	50
weighted avg	0.95	0.94	0.94	50

Table 5: VIOLENCE & SCARINESS: validation ($n = 50$).

well as on the doubly-annotated external test set. The SEX & NUDITY classifier achieves an overall accuracy of 0.70 and a macro-averaged F1 of 0.63 on the external test set of 30 passages. Performance is strongest on the majority class *absent* (F1 = 0.80), while the *weak* class proves most difficult to identify (F1 = 0.43), likely reflecting its position as a transitional category between the two poles. Linear and quadratic weighted κ of 0.48 and 0.56, respectively, indicate moderate ordinal agreement, and a MAE of 0.33 confirms that misclassifications tend to occur between adjacent classes rather than across the full scale.

The VIOLENCE & SCARINESS classifier performs substantially higher on the internal validation set (accuracy = 0.94, macro F1 = 0.93) than on the external test set (accuracy = 0.47, macro F1 = 0.30), indicating a pronounced generalization gap compared to the SEX & NUDITY model. On the external test set, the classifier defaults heavily towards the *absent* class (F1 = 0.62, recall = 0.93), fails entirely on the *weak* category (F1 = 0.00), and identifies *strong* scenes with high pre-

Class	P	R	F1	Sup
<i>absent</i>	0.70	0.94	0.80	17
<i>weak</i>	0.75	0.30	0.43	10
<i>strong</i>	0.67	0.67	0.67	3
macro avg	0.70	0.64	0.63	30
weighted avg	0.71	0.70	0.66	30

Table 4: SEX & NUDITY: external test ($n = 30$).

Class	P	R	F1	Sup
<i>absent</i>	0.46	0.93	0.62	14
<i>weak</i>	0.00	0.00	0.00	10
<i>strong</i>	1.00	0.17	0.29	6
macro avg	0.49	0.37	0.30	30
weighted avg	0.42	0.47	0.35	30

Table 6: VIOLENCE & SCARINESS: external test ($n = 30$).

cision but very low recall (precision = 1.00, recall = 0.17). Cohen’s κ of 0.024 and a MAE of 0.70 confirm that errors frequently span non-adjacent ordinal categories, in stark contrast to the SEX & NUDITY model whose misclassifications tend to remain within neighboring classes.

These results establish an asymmetry that shapes the analysis to follow: the SEX & NUDITY classifier reaches a level of reliability sufficient for exploratory corpus analysis, while the VIOLENCE & SCARINESS classifier’s near-chance external performance renders its predictions unreliable as a standalone instrument. The latter’s outputs are nonetheless retained in the analysis, treated not as reliable labels but as a weak signal to be interpreted against the statistical correlations and with the caveat that classification errors are both frequent and severe.

4 Analysis Results

“Are you going to be a noisy lover, Amy?”
<character_sound loudness="2.5"><speaker gender="man">he</speaker> whispered, laughing softly</character_sound>. “Let’s see.”

In the erotic scene of Diana Palmer’s *Love by Proxy* (1985), the two protagonists find themselves alone together, the tension of their slow-burning attraction finally breaking into something more. The scene unfolds not through explicit description but through sound representation: “with rough breathing and wild heartbeats,” the narrator depicts the interplay of whispering and moaning until

<character_sound loudness="4.0"><speaker gender="woman">She</speaker> began to cry from the spiraling tension</character_sound>.

These are not incidental details but a deliberate sequence of alternating low- and high-intensity *character sound* events that together construct the acoustic atmosphere of the erotic scene. Averaging these loudness levels would produce a misleading picture of moderate, undifferentiated intensity: the peaks and troughs canceling each other out into an unremarkable mean. Standard deviation, by contrast, captures precisely this oscillation: it is sensitive to the dynamic range between a barely audible whisper and a vocalized peak, making the scene stand out in comparison to the others.

To investigate the correlation between a high SD of *character sound* loudness and represented explicit content, I conducted two parallel analyses: (1) I first examined whether *character sound* loudness within-scene standard deviation (SD) correlates with erotic content (here scenes annotated for explicit SEX & NUDITY, following the operationalization in the Table 7), using *ambient sound* as a within-corpus control. (2) I second reversed the logic: I examined whether *ambient sound* loudness mean and SD correlate with scenes annotated for explicit VIOLENCE & SCARINESS, using *character sound* as the control. The underlying intuition is that character-produced sounds should be the primary acoustic carrier of erotic content, reflecting the trajectory from whispered dialogue to vocalized arousal, whereas *ambient sound*, which captures environmental and atmospheric noise, should be more diagnostic of violent or threatening scenes.

I worked with a MMiA subcorpus of ten TEI XML-annotated romance novels comprising 815

scenes in total. Each scene was assigned a SEX & NUDITY and a VIOLENCE & SCARINESS score on the ternary severity scale (0 = *absent*, 1 = *weak*, 2 = *strong*), and contained <character_sound> and <ambient_sound> elements with loudness ratings on a 0; 1–5 scale.

4.1 Character Sound Loudness SD and SEX & NUDITY

Restricted to the 589 scenes with at least two *character sound* events⁸, I found a small but statistically reliable positive relationship between *character sound* SD and SEX & NUDITY score: Spearman $\rho = 0.164$ ($p < 0.001$), point-biserial $r = 0.116$ ($p < 0.01$), and AUC = 0.578, indicating that *character sound* SD discriminates erotic from non-erotic scenes at a rate approximately 16% above chance. A sensitivity analysis comparing only strongly erotic scenes (score 2, $n = 82$) against non-erotic scenes (score 0, $n = 218$) even yielded a larger effect (rank-biserial $r = 0.310$, $p < 0.001$), with mean SD rising from 0.470 to 0.644, a 37% increase. By contrast, no equivalent effect was found for the mean loudness of *character sound*, nor for *ambient sound* SD (Spearman $\rho = 0.060$, $p = 0.48$; AUC = 0.534), the latter result based on the 74 scenes with at least two *ambient sound* events. The null result for *ambient sound* is theoretically informative: it rules out the possibility that erotic scenes are simply acoustically busier overall. The significant increase in loudness variability is localized specifically to character-produced sounds, supporting the interpretation that it reflects the dynamic vocal and physical expressiveness of the characters rather than a general inflation of acoustic activity in the scene.

4.2 Ambient Sound Loudness SD and VIOLENCE & SCARINESS

In my second analysis, I tested whether *ambient sound* loudness SD is associated with *violent/scary* scenes (here scenes annotated for explicit VIOLENCE & SCARINESS, following the operationalization in the Table 8), comparing within-scene loudness SD for both *character sound* and *ambient sound* against the VIOLENCE & SCARINESS annotation.⁹ *Violent/scary* scenes showed markedly higher *character sound* SD than *non-violent* scenes

⁸Two sound events of the same category are the minimum required to compute SD.

⁹Annotation of 815 scenes: 588 contained at least two *character sound* events (556 *non-violent*, 32 *violent/scary*).

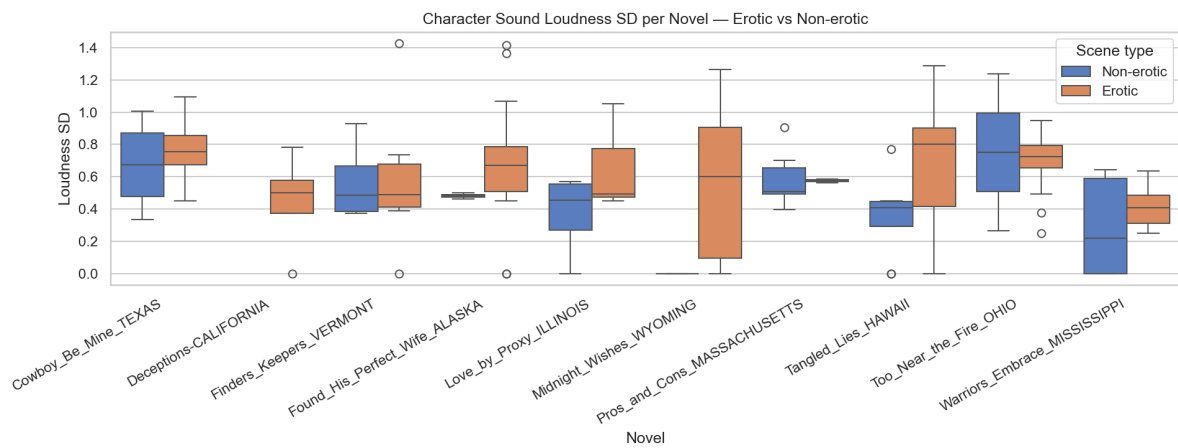


Figure 1: Character sound loudness SD per scene, classified in *erotic* vs. *non-erotic* scenes of ten MMiA novels.

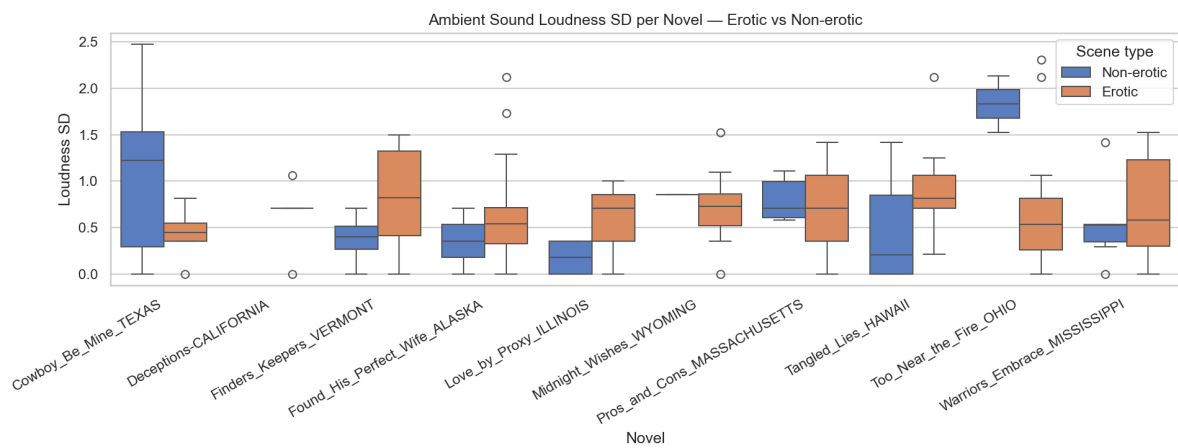


Figure 2: Ambient sound loudness SD per scene, classified in *erotic* vs. *non-erotic* scenes of ten MMiA novels.

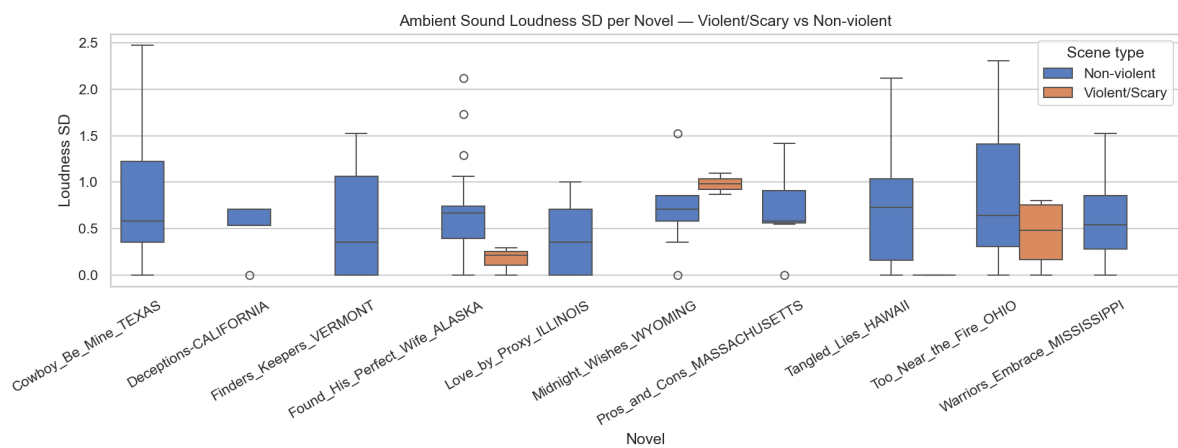
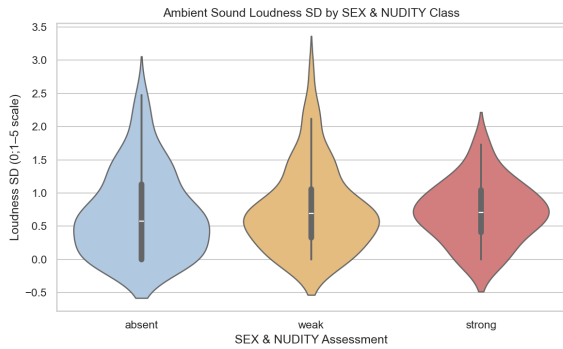
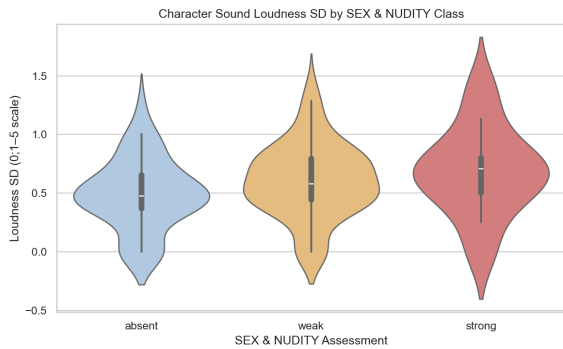


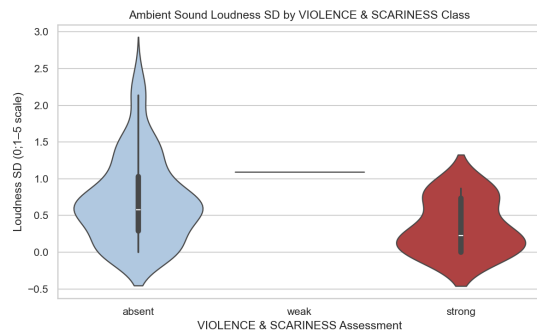
Figure 3: Ambient sound loudness SD per scene, classified in *violent/scary* vs. *non-violent* scenes of ten MMiA novels.



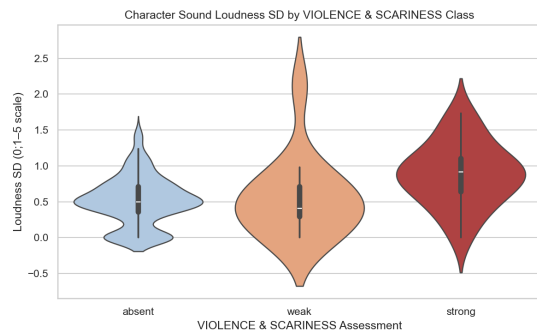
(a) Distribution of *ambient sound* loudness SD per scene across SEX & NUDITY assessment classes (*absent*, *weak*, *strong*).



(b) Distribution of *character sound* loudness SD per scene across SEX & NUDITY assessment classes (*absent*, *weak*, *strong*).



(c) Distribution of *ambient sound* loudness SD per scene across VIOLENCE & SCARINESS assessment classes (*absent*, *weak*, *strong*). All *weak* scenes contain fewer than 2 ambient sound events and therefore yield no loudness SD values.



(d) Distribution of *character sound* loudness SD per scene across VIOLENCE & SCARINESS assessment classes (*absent*, *weak*, *strong*).

(*violent*: $M = 0.749$; *non-violent*: $M = 0.511$), with the effect concentrated in scenes of *strong* VIOLENCE & SCARINESS (score 2: $M = 0.883$). A one-tailed Mann-Whitney test confirmed a statistically significant difference ($U = 11,333.5$, $p = .004$, rank-bi-serial $r = -0.274$), supported by Spearman $\rho = 0.111$ ($p = .007$) and AUC = 0.637; a sensitivity check restricted to scenes with *strong* versus *absent* VIOLENCE & SCARINESS produced the largest effect ($r = -0.508$, $p < .001$). By contrast, only 138 scenes contained at least two *ambient sound* events (124 *non-violent*, 14 *violent/scary*), and *ambient sound* SD showed no meaningful difference between scene types (*non-violent*: $M = 0.714$; *violent*: $M = 0.682$), with all tests returning null results (Mann-Whitney $p = .661$; Spearman $\rho = -0.043$; AUC = 0.467). This dissociation suggests that the romance novel writers of the corpus texts encode VIOLENCE & SCARINESS primarily through vocal behavior rather than through shifts in the surrounding soundscape, though both analyses are limited by class imbalance, and the *ambient sound* null result should be interpreted with caution given the small sample size for annotated *ambient sounds*. One possible explanation is that *violent/scary* scenes foreground interpersonal conflict and emotional arousal, which are inherently tied to the voice, while the ambient soundscape may be held constant or described independently of scene intensity or may even be not represented at all.

It should be noted, however, that the *ambient sound* analysis is substantially underpowered (only 14 *violent/scary* scenes), so the null result must be interpreted with caution; a larger corpus with denser *ambient sound* annotation would be required to rule out a genuine effect. The class imbalance in both analyses, 32 violent scenes out of 588 for *character sound*, 14 out of 138 for *ambient sound*, also warrants caution in generalizing the effect sizes reported here. In general, providing a large-scale scene-annotated novel corpus would be helpful for content moderation and editorial studies in future work.

5 Conclusion

This paper proposed and tested loudness SD as an acoustically grounded proxy for content advisory relevance in romance fiction. H2 is confirmed: *character sound* loudness SD could be detected as a correlate of explicit content featuring SEX &

NUDITY. H1, by contrast, given the small number of detected relevant scenes containing sufficient ambient sound events ($n = 138$), could neither be confirmed nor reasonably falsified: the ambient sound analysis returned null results across all tests, but the limited data points preclude a definitive conclusion. More scenes providing ambient sound events would be needed. However, in general romance fiction tend to provide fewer ambient sound representations, while focusing on soundful character interaction.

The SEX & NUDITY classifier achieved surprisingly good results (macro F1 = 0.63 on the external test set) given the heuristic nature of the approach and its reliance on synthetically generated training data. This suggests that the category is sufficiently well operationalized to support automated detection, even under low-resource conditions, and has the potential to be further refined with more data. The VIOLENCE & SCARINESS classifier proved considerably harder to finetune and generalize (macro F1 = 0.30), which can be attributed to the operationalization of the category itself. Following Tsai et al. (2024), the class bundles violence and scariness into a single label, yet the two phenomena are narratologically and acoustically distinct: scariness in particular is a broad and diffuse category whose textual and acoustic markers remain underspecified. The detection of scary scenes, therefore, remains an open desideratum, and a more fine-grained operationalization of VIOLENCE & SCARINESS will be part of future work.

Beyond these specific findings, the study demonstrates the analytical value of scene segments as a structural unit for corpus-level content analysis. Applied to different editions or adaptations of the same text, loudness SD and content advisory annotation could be used to detect omissions, alterations, or tonal shifts with respect to explicit content across versions, a use case with practical relevance for editorial studies, content moderation, and computational literary studies more broadly.

Limitations

Some limitations should be noted: (1) The statistical analyses rest on a subcorpus of 10 novels, yielding small sample sizes, particularly for *violent/scary* scenes (32 scenes with *character sound* events, 14 with *ambient sound* events), which means that the null result for H1 cannot be inter-

preted as a definitive falsification. (2) The VIOLENCE & SCARINESS category additionally bundles two narratologically distinct phenomena: scariness, encompassing psychological tension, suspense, and atmosphere, does not map cleanly onto the acoustic features examined here, and its conflation with physical violence likely contributes to the poor classifier generalization (macro F1 = 0.30). A more fine-grained operationalization separating the two is needed before definitive conclusions can be drawn. (3) Finally, the study is restricted to a single series of romance novels, and the generalizability of the identified acoustic signals to other romance series or literary traditions remains open for future research.

Code & Data

Code is available in the GitHub repository on https://github.com/SvenjaGuhr/Explicit_Content_Detection. Due to copyright restrictions, the 20th-century romance novels used in this study cannot be redistributed. However, the LLM-generated scenes are included in the repository, and the codebase is designed to be reusable for application to other datasets. The code and data is released under the GNU GPL v3.0 license. The author does not permit the use of the code or accompanying data in support of censorship, surveillance, or book-banning initiatives.

Acknowledgments

Many thanks go out to Mackenzie Cramer for research support. I gratefully acknowledge Hannah Walser and the Stanford Literary Lab for purchasing and providing access to the book corpus used in this study.

References

- Max Ackermann. 2003. *Die Kultur des Hörens: Wahrnehmung und Fiktion; Texte vom Beginn des 20. Jahrhunderts*. Falkenberg.
- David Bamman, Kent K. Chang, Lucy Li, and Naitian Zhou. 2024. [On Classification with Large Language Models in Cultural Analytics](#). In *Proceedings of the Computational Humanities Research Conference 2024*, volume 3834 of *CEUR Workshop Proceedings*, pages 494–527. CEUR.
- Toni Bernhart. 2008. [Stadt hören: Auditive Wahrnehmung in Berlin Alexanderplatz von Alfred Döblin](#). *Zeitschrift für Literaturwissenschaft und Linguistik*, 38(1):51–67.

- Michael Bull and Les Back. 2003. *The Auditory Culture Reader*. Sensory Formations Series. Berg.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Unsupervised cross-lingual representation learning at scale](#). *arXiv*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. ACL.
- Matt Foley. 2023. *Gothic Voices: The Vococentric Soundworld of Gothic Writing*. Cambridge University Press.
- Theodoros Giannakopoulos, Dimitrios Kosmopoulos, Andreas Aristidou, and Sergios Theodoridis. 2006. [Violence content classification using audio features](#). In Grigoris Antoniou, George Potamias, Costas Spyropoulos, and Dimitris Plexousakis, editors, *Advances in Artificial Intelligence*, volume 3955, pages 502–507. Springer.
- Evelyn Gius. 2026. [Warum wir in den Digital Humanities messen \(sollten\)](#). In *Book of Abstracts of DHd2026*, Vienna. Zenodo.
- Evelyn Gius, Fotis Jannidis, Markus Krug, Albin Zehe, Andreas Hotho, Frank Puppe, Jonathan Krebs, Nils Reiter, Natalie Wiedmer, and Leonard Konle. 2019. [Detection of Scenes in Fiction](#). In *Book of Abstracts of DH2019*. ADHO.
- Evelyn Gius, Carla Sökefeld, Lea Dümpelmann, Lucas Kaufmann, Anneke Schreiber, Svenja Guhr, Nathalie Wiedmer, and Fotis Jannidis. 2021. [Guidelines for Detection of Scenes](#). Zenodo.
- Elena Glotova. 2021. *Soundscapes in Nineteenth-Century Gothic Short Stories*. Umeå University.
- Svenja Guhr. 2024. *Raise Your Voice. Character Sound in German Literary Prose*. Technical University of Darmstadt. Unpublished.
- Svenja Guhr. 2026. *Raise Your Voice. Character Sound in German-language Fiction*. Digitale Literaturwissenschaft. Metzler, Springer Nature. In print.
- Svenja Guhr and Mark Algee-Hewitt. 2024. [What’s that scary sound? Ambient sound in gothic fiction](#). *Journal of Computational Literary Studies*, 2(1).
- Svenja Guhr and Michaela Mahlberg. 2026. [Soundful Dickens. the narrative functions of sounds in Dickens’s fictional worlds](#). In *Book of Abstracts of DHd2026*. Zenodo.
- Svenja Guhr, Huijun Mao, and Fengyi Lin. 2025. [Rethinking scene segmentation. advancing automated detection of scene changes in literary texts](#). In *Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2025)*, pages 79–86. ACL.
- Julian Häußler, Svenja Guhr, and Evelyn Gius. 2024. [Lautstärke und Konflikt in Realismus und Naturalismus](#). In *Book of Abstracts of DHd2024*. Zenodo.
- Lin Long, Rui Wang, Ruixuan Xiao, Junbo Zhao, Xiao Ding, Gang Chen, and Haobo Wang. 2024. [On LLMs-driven synthetic data generation, curation, and evaluation: A survey](#). In *Findings of the Association for Computational Linguistics ACL 2024*, pages 11065–11082. ACL.
- Luca Mari, Mark Wilson, and Andrew Maul. 2023. *Measurement Across the Sciences: Developing a Shared Concept System for Measurement*. Springer Series in Measurement Science and Technology. Springer Nature.
- Victor R. Martinez, Krishna Somandepalli, Karan Singla, Anil Ramakrishna, Yalda T. Uhls, and Shrikanth Narayanan. 2019. [Violence rating prediction from movie scripts](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 671–678.
- Yann Mathet, Antoine Widlöcher, and Jean-Philippe Métivier. 2015. [The unified and holistic method \$\gamma\$ for inter-annotator agreement measure and alignment](#). *Computational Linguistics*, 41(3):437–479.
- Frances Molyneux. 2022. *Sound’s Effects: Sonic Epistemology in the Victorian Novel*. Stanford University.
- John M. Picker. 2003. *Victorian Soundscapes*. Oxford University Press.
- R. Murray Schafer. 1994. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books.
- Philipp Schweighauser. 2006. *The Noises of American Literature, 1890-1985: toward a History of Literary Acoustics*. University Press of Florida.
- Danil Semin, Ondřej Dušek, and Zdeněk Kasner. 2026. [Strategies for span labeling with large language models](#). *arXiv*.
- Matthew Sims, Jong Ho Park, and David Bamman. 2019. [Literary event detection](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3623–3634. ACL.
- Stanley Smith Stevens. 1946. [On the theory of scales of measurement](#). *Science*, 103(2684):677–680.
- Alicia Y. Tsai, Shereen Oraby, Anjali Narayan-Chen, Alessandra Cervone, Spandana Gella, Apurv Verma, Tagyoung Chung, Jing Huang, and Nanyun Peng.

2024. [PG-story: Taxonomy, dataset, and evaluation for ensuring child-safe content for story generation](#). In *Proceedings of the 3rd Workshop on NLP for Positive Impact*, pages 78–97. ACL.
- Michael Vauth, Hans Ole Hatzel, Evelyn Gius, and Chris Biemann. 2021. [Automated event annotation in literary texts](#). In *Proceedings of Computational Humanities Research Conference 2021*, volume 2989 of *CEUR Workshop Proceedings*, pages 333–345. CEUR.
- Albin Zehe, Elisabeth Fischer, and Andreas Hotho. 2025. [Assessing the state of the art in scene segmentation](#). In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 9922–9941. ACL.
- Albin Zehe, Leonard Konle, Svenja Guhr, Lea Dümpelmann, Evelyn Gius, Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, and Annekea Schreiber. 2021. [Shared task on scene segmentation @ KONVENS 2021](#). In *Proceedings of the Shared Task on Scene Segmentation*, volume 3001 of *CEUR Workshop Proceedings*, pages 1–21. CEUR.
- Yigeng Zhang, Mahsa Shafaei, Fabio Gonzalez, and Tamar Solorio. 2021. [From none to severe: Predicting severity in movie scripts](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3951–3956. ACL.
- Jochen Zöllner, Konrad Sperfeld, Christoph Wick, and Roger Labahn. 2021. [Optimizing Small BERTs Trained for German NER](#). *Information*, 12(11):443.

Content Rating	Explanation	Sample Passage
SEX & NUDITY <i>absent</i>	No sexual or intimate content is present: the scene contains no physical intimacy, no expression of desire or arousal, and no implied or explicit sexual activity. Longing or emotional attachment may be present but is not rendered in physical or sensory terms.	<i>Tangled Lies</i> : Rachel, however, wasn't quite so fortunate. For a long time she lay in her bed, staring out her window, her mouth still tingling with a forbidden longing. And when she fell asleep this time, she was crying.
SEX & NUDITY <i>weak</i>	Sexual or intimate content is present but encoded implicitly through literary devices: metaphor, euphemism, physical suggestion, or narrative indirection. Desire or arousal is described, physical intimacy is initiated or implied, but the sexual encounter is not rendered explicitly on the page. The scene gestures toward sexuality through creative or literary means rather than depicting it directly.	<i>Love by Proxy</i> : Apparently there was a small conference going on. [...] "Yes?" the big man asked coldly, looking up with eyes that were every bit as dark as the straight black hair that fell forward onto a broad forehead. Amelia smiled wickedly. "Message for you, sir," she said. And she let the coat drop. The two men grouped around the desk stared, gaping, with appreciative smiles and big eyes. The bigger man stood erect and looked angry. Amelia had a passable voice no threat to the Met, of course, put passable. She began to gyrate in her outlandish belly dancer's costume to the tune of the birthday song, inching slowly closer to the big, dark man. He didn't look very receptive. In fact, he looked as if he'd like to pitch her out the window. That was even better. She laughed huskily as she went closer, her hips twitching, her skirts flying, her arms uplifted with the small cymbals on her fingers to show the high, soft curve of her breasts in their metallic casings. "Happy birthday, honey," she added at the end, and just for pure spite, she went on tip-toe to kiss him full on his hard, chiseled mouth with as much enthusiasm as she could muster.
SEX & NUDITY <i>strong</i>	Sexual activity is rendered explicitly as the primary event of the scene: a consummated sexual encounter is described in physical and sensory detail, with direct depiction of bodily response and intimacy. The scene leaves nothing to literary imagination; explicitness rather than suggestion is the dominant representational mode.	<i>Midnight Wishes</i> : By the time his fingers slid down to encounter the barrier of her panties, she wanted him more than she'd ever wanted anything in her life. With his gaze locked with hers, he slowly, sensually, pulled her panties off. When he touched her at the center of her being, tears of pleasure blurred her vision. He shifted positions and hovered over her for a moment, then slowly entered her. As their bodies joined, she continued to look into his eyes. In their dark depths, she saw not only passion, but caring, tenderness... emotions that stirred her more deeply than any physical act could accomplish alone. He moved, stroking urgency and need within her. As each thrust went deeper, became more powerful, she closed her eyes and gave herself completely to him.

Table 7: Operationalization of the retained PG-story content advisory category SEX & NUDITY (Tsai et al., 2024) on a ternary scale (*absent*, *weak*, *strong*), with sample passages drawn from the *Men Made in America* corpus. The scale distinguishes between the absence of category-relevant content, its implicit representation through literary devices such as metaphor, euphemism, or narrative suggestion (*weak*), and its explicit rendering as the primary event of the scene (*strong*).

Content Rating	Explanation	Sample Passage
VIOLENCE & SCARINESS <i>absent</i>	No category-relevant content is present: the scene contains no weapons, no physical threat, no fear-inducing situation, and no dangerous activity. Interpersonal tension or conflict may be present but is not rendered in terms of physical danger or threat to bodily integrity.	<i>Deceptions</i> : Adele rose gracefully from her chair and quietly approached Lisa. "The producer wants both of you, Lisa. That's what I've been trying to tell you. These period movies are chancy at best, and they need bankable stars to ensure the public will at least turn out — even if it's only to see the two of you on screen together again." She waited for the diminutive actress to absorb the information she was hearing. Lisa's black, almond-shaped eyes stared at her, the pain within them causing Adele to suffer a twinge of guilt until she recalled what she was trying to accomplish. "There is no way I can work with Drew again. Please don't ask that of me." Adele studied her for a moment in silence. Lisa's midnight black hair tumbled around her face and over her shoulders, almost touching her waist. The woman was beautiful - and fiery — and a damned good actress.
VIOLENCE & SCARINESS <i>weak</i>	Violence or threat is present but encoded implicitly through atmosphere, anticipation, or narrative suggestion: a weapon appears as a prop, danger is sensed or anticipated, or fear is the dominant emotional register of the scene. Physical harm is not depicted on the page; the scene gestures toward violence or scariness through literary and atmospheric means rather than rendering it directly.	<i>Finders Keepers</i> : Holly Wingate Paynter leaned her crowbar against the paint-chipped doorjamb and warily eyed the floor of the cold, dark master bedroom. A leaky ceiling and years of abandonment and neglect had left their mark. It didn't look good. Holly considered skipping the bedroom, but she'd prowled through every other room in the woefully dilapidated Danvers House, soaking up its creepy atmosphere. She scanned the room with her powerful flashlight and frowned as she considered possible routes. She wasn't particularly worried, just didn't want to accidentally leave her footprints in the rotten wood. Being a Wingate, she wasn't exactly at the Danvers House by invitation. Best to avoid any careless mistakes.
VIOLENCE & SCARINESS <i>strong</i>	Violence, captivity, or intense terror is rendered explicitly as the primary event of the scene: physical harm, serious injury, or sustained life-threatening threat is described in direct sensory and physical detail. The scene leaves nothing to literary imagination; explicitness rather than suggestion or atmosphere is the dominant representational mode, and a reader would identify its defining content as an act of violence or a life-threatening situation.	<i>Midnight Wishes</i> : An involuntary scream erupted from her raw throat as the curtains caught fire, the flames licking up toward the ceiling. Terror screamed inside her as she felt the hot breath of death surrounding her.

Table 8: Operationalization of the retained PG-story content advisory category VIOLENCE & SCARINESS (Tsai et al., 2024) on a ternary scale (*absent*, *weak*, *strong*), with sample passages drawn from the *Men Made in America* corpus. The scale distinguishes between the absence of category-relevant content, its implicit representation through literary atmosphere, anticipation, or suggestion (*weak*), and its explicit rendering as the primary event of the scene (*strong*).