Listening for Ideology: Automatic Analysis of Character Speech in Historical Nazi Propaganda Films

Nicolas Ruth, Manuel Burghardt, Andreas Niekler
Computational Humanities Group, Institute for Computer Science, Leipzig University
{nicolas.ruth, burghardt, aniekler}@informatik.uni-leipzig.de

Abstract

While the visual dimension of film has been widely explored in the digital humanities through methods such as "distant viewing," the audio layer has received less attention despite its crucial role in meaning-making. We address this gap with a four-step pipeline that combines speaker diarization, audio gender classification, automatic speech recognition (ASR), and LLM-based psycholinguistic analysis to infer character traits from film dialogues. Applying this method to a set of Nazi propaganda films, we find that despite the challenges of speaker diarization due to noisy historical film audio, modern ASR and GPT-based analyses produce character profiles consistent with existing film research. Our proposed pipeline advances distant reading of film dialogue, complements visual analyses and enables a scalable study of ideology in historical cinema. A case study of female characters in NS films identifies three recurring types, centered on the ideological figure of the mother in National Socialism.

Keywords: computational film analysis, NS ideology, propaganda film, audio analysis, character speech analysis

1. Introduction

The National Socialist (NS) regime's extensive use of film as a propaganda tool demonstrates the central role cinema played in shaping public opinion and disseminating ideology. Following Joseph Goebbels' call for a "revolution of the mind" (Leiser, as cited in Hardinghaus, 2008), the Nazi state sought to control all aspects of cultural production, establishing a centralized film industry dedicated to ideological messaging and even advocating for a new cinematic style (Giesen and Hobsch, 2005).

Scholarly research has long sought to understand the mechanisms by which film encodes ideology (Kellner, 1991; Ryan and Kellner, 1988), typically by focusing on the critical analysis of individual works, emphasizing visual style, narrative structure, and socio-political context. In recent years, however, computational methods have begun to reshape the field. Inspired by the metaphor of "distant reading" (Moretti, 2000) and its extension to the visual domain in approaches such as "distant viewing" (Arnold and Tilton, 2019) and "deep watching" (Bermeitinger et al., 2019), digital film studies increasingly deploy computer vision and machine learning to analyze large-scale visual corpora. These approaches represent a significant methodological advance, enabling the systematic study of visual patterns across hundreds of films. Whereas the visual layer of film has become a focal point of computational analysis, the audio layer, comprising spoken language, music, and sound effects, has received significantly less attention. Several projects have begun to address this gap by integrating multiple modalities into film analysis (Burghardt et al., 2024). *VIAN* (Halter et al., 2019), *TIB AV-Analytics* (Springstein et al., 2023), and *Zoetrope* (Liebl and Burghardt, 2023) are just some recent examples for tools that provide frameworks for the annotation, visualization, and exploration of multimodal film data. However, even within these projects, the audio layer tends to be treated as secondary to the visual layer, and the specific task of automatically analyzing character speech remains largely unexplored. This gap is particularly striking given the centrality of speech and dialogue in the construction of filmic meaning (Kozloff, 2000; Bednarek, 2018). As for the case of NS propaganda films, the way characters speak, the words they use, and the interpersonal dynamics encoded in their dialogue all contribute to an overarching ideological agenda.

Although significant progress has been made in automatic speech processing and analysis using deep learning models, these approaches have not yet been applied to historical film material – particularly to the study of film characters and their role in conveying ideologies. We address this gap by proposing a four-step audio analysis pipeline designed to process historical film audio and generate character-level insights. The pipeline begins with (1) speaker diarization, which segments the audio by speaker and identifies distinct voices. It then (2) applies automatic voice gender detection followed by (3) automatic speech recognition (ASR), to convert the detected segments into text. Finally, (4) it uses a GPT-based language analysis to infer character traits from the aggregated speech of individual characters. Together, these components enable the automated construction of character profiles and provide a basis for large-scale analysis of

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 5828–5838
11–16 May 2026.

ideological patterns.

We evaluate this pipeline on a selection of three so-called *Vorbehaltsfilme* (restricted Nazi films), kindly provided by the Friedrich Wilhelm Murnau foundation for research purposes. The investigated films are "Hitlerjunge Quex" (1933), "Jud Süß" (1940), and "Kopf hoch, Johannes!" (1941), which span a range of genres and ideological functions. From a digital humanities perspective, we are particularly interested in assessing how well our pipeline works for historical films, which present a wide range of problems in terms of audio quality, but also with regard to the language being used, which itself has its ideological and linguistic intricacies (Klemperer, 1947).

2. Related Work

Steps 1–3 of our proposed pipeline build on established technologies for speech analysis and transcription, with a particular focus on evaluating their performance on historical language data from films. Step 4, by contrast, introduces a novel approach that leverages state-of-the-art large language models (LLMs) to derive psychoanalytic insights into character profiles. As there is no directly comparable work for this method, the following section reviews related studies that have analyzed the language of fictional characters more generally using computational techniques.

The study of characters has long been central to film and media scholarship, particularly in analyses of ideology (Eder, 2025). Characters convey ideological messages, embody values, enact power relations, and generally model behaviors for audiences. In the context of Nazi cinema, character construction was a key mechanism of propaganda, used to normalize antisemitism, promote racial hierarchies, and depict obedience to the state as virtuous (Giesen and Hobsch, 2005; Niven, 2022). Traditional film scholarship has examined these dynamics through critical analysis of key films, revealing how character traits and relationships encode ideological narratives (Hardinghaus, 2008). Computational approaches to automatic character analysis in films remain relatively rare, particularly in historical film contexts. Existing work in computational narratology and character modeling so far has focused primarily on textual corpora, using techniques such as sentiment analysis (Nalisnick and Baird, 2013; Schmidt and Burghardt, 2018) and network analysis (Moretti, 2011; Agarwal et al., 2012) to identify character functions and relationships.

Our work draws on recent technological advances that have successfully used LLMs to extract character attributes from movie scripts (Baruah and Narayanan, 2024). We demonstrate that LLMs can also be used to infer character traits directly from transcribed dialogue, even when transcription quality is imperfect. The task at hand is closely connected to the field of psycholinguistics, where computer-based NLP methods are increasingly used to measure psychologically grounded personality traits (Herderich et al., 2024). They include techniques such as topic modeling, which can be applied to psychological interview data, and dictionary-based tools like *LIWC* (Boyd et al., 2022) or *Empath* (Fast et al., 2016). These approaches try to identify personality traits by analyzing linguistic patterns in written language. Some researchers have argued that language use can be a stronger indicator of personality than self-report questionnaires (Boyd and Pennebaker, 2017; Boyd et al., 2020). By integrating character analysis into a broader audio processing pipeline, we provide a methodological foundation for scaling up such analyses to large corpora of historical films.

3. Audio Analysis pipeline

This section provides a detailed description of how each of the four stages of the analysis pipeline was implemented.

3.1. Speaker Diarization

The goal of the pipeline is the automated analysis of film character profiles. This analysis is based on the speech segments of the characters. To automatically assign spoken text elements to characters, speaker diarization is required as a first step. Speaker diarization refers to the process of identifying and segmenting speakers within audio or video data. Its goal is to automatically extract speech segments and assign them to individual speakers throughout the entire medium (Plaquet and Bredin, 2023). Current state-of-the-art implementations include Nvidia's NeMo and Pyannote's speaker-diarization-3.1 pipeline (La Javaness R&D, 2023). This technological field is characterized by a high level of adaptability and complexity, originating from the combination of different sub-tasks. Pyannote and NeMo are largely based on a cascaded system that combines several sub-processes, including Voice Activity Detection (VAD), Speaker Change Detection, Overlapped Speech Detection (OSD), and Speaker Identification using embeddings and clustering methods. Each of these components requires individual configuration and fine-tuning, which makes the evaluation process extensive and demanding. Moreover, newer neural end-to-end approaches such as NeMo's Sortformer[1] model aim to simplify these processes but are still limited

[1] https://huggingface.co/nvidia/diar_sortformer_4spk-v1

in scope, for instance by supporting only up to four speakers. In particular, Nvidia NeMo offers a wide range of parameters and options, which increases the flexibility of the system but also adds further layers of complexity, making exhaustive testing and comparison challenging. The presented pipeline uses Pyannote and NeMo and compares them to each other.

3.2. Automatic Voice Gender Detection

The next step of the audio pipeline aims to determine the gender of the speakers based on the audio segments. This allows for an analysis of gender-specific differences in film character profiles. For this task, the pipeline implements an audio analysis model developed by Audeering[2]: wav2vec2-large-robust-24-ft-age-gender[3]. The model is the result of fine-tuning of all layers of Facebook's Wav2Vec model[4]. The datasets used for fine-tuning inlcude aGender[5] or Mozilla Common Voice[6].

Reported accuracy values for the gender classifcation range from 0.90 to 0.98 for female speakers and from 0.91 to 1.00 for male speakers, depending on the test dataset (Burkhardt et al., 2023). The model output includes an age estimation and confidence scores for the classes male, female, and child. It is important to note that this model's binary gender identification can only align with diverse gender identities if the dataset it is used on reflects such a binary division. In this context, the binary gender framework represented in the historical material of National Socialism corresponds to the assumptions of the model.

3.3. Automatic Speech Recognition

The following step in the pipeline enables a shift in analysis from the signal level to the text analysis level by automatically transcribing the textual content of the speech segments. Two derivatives of OpenAI's Whisper were tested. Whisper is characterized by a high degree of robustness to background noise in audio signals, as demonstrated in the original paper, where it was tested with bar noise. Given the wide range of ambient sounds that characterise film audio, this feature is of particular significance. In this study, the model sizes large-v2 and medium were evaluated, as model size does not necessarily correlate with transcription quality. For datasets such as TED-LIUM3[7], WSJ[8], CallHome[9] or AMI-IHM[10], Whisper large-v2 produced higher word error rates (WER) than Whisper medium (Radford et al., 2022). The models were used with the default parameterization of the OpenAI's Whisper library.[11]

3.4. GPT-Based Psycholinguistic Analysis

The psycholinguistic analysis of character traits builds on the preceding pipeline steps to filtera list of speech segments for each individual character. The extracted character dialogue elements are submitted via the OpenAI-API to the model gpt-3.5-turbo-0125[12] using the following prompting strategy:

System Prompt

You are an expert in analyzing fictional characters. Analyze this character speech and answer these questions if possible. Keep it short and give your answer in the format of a Python dictionary with the numbering as keys!

1. Describe the character's personality in five words.
2. What are the character's main goals?
3. What are the character's main motivations?
4. How does the character interact with other characters?
5. *Name the title of the film if you recognize it!*

Text of the character:

Input

```
<INSERT TRANSCRIBED SPEECH SEG-
MENTS AS LIST>
```

The prompt addresses various character dimensions of film, which are based on existing theoretical frameworks (Chatman, 1978; Eder, 2025). In

[2] https://www.audeering.com/de/

[3] https://huggingface.co/audeering/wav2vec2-large-robust-24-ft-age-gender

[4] https://huggingface.co/facebook/wav2vec2-large-robust

[5] https://catalogue.elra.info/en-us/repository/browse/ELRA-S0365/

[6] https://commonvoice.mozilla.org/de

[7] https://www.openslr.org/51/

[8] https://catalog.ldc.upenn.edu/LDC93S6A

[9] https://catalog.ldc.upenn.edu/LDC97S42

[10] https://groups.inf.ed.ac.uk/ami/corpus

[11] https://pypi.org/project/openai-whisper/

[12] Note: This model is now deprecated, and the experiments were conducted when it was still current. As no improvement in the procedure is anticipated with a newer model, these results are reported without conducting new experiments.

question (5), we test whether the model was able to identify the film solely based on the speech passages of a given character, which was never the case. The question was included to evaluate if the analysis is based on the character's speech or on prior knowledge of the films.

4. Evaluation Results

The four-step audio analysis pipeline described above was systematically evaluated with respect to its performance on historical film audio and its ability to generate meaningful character-level analyses. Because each stage of the pipeline presents distinct challenges and contributes differently to the overall goal of character trait extraction, we evaluate them individually by means of quantitative and qualitative methods.

4.1. Evaluation of Speaker Diarization

Speaker diarization is a particularly complex problem in historical film contexts due to degraded audio quality, background noise, overlapping speech, and a lack of labeled training data. Traditional diarization systems assume controlled recording conditions, an assumption that does not hold in archival film audio. The research data consists of historical German film audio, which includes strong background noise, uncertain and high speaker numbers, and very long audio segments. Moreover, the complex processing diarization framework introduces extra methodological difficulties and potential sources of error. The evaluation is based on a manually created ground truth and uses the Diarization Error Rate (DER) as a metric. The DER measures the percentage of wrongly classified speech segments. It includes the error classes: false detection, missed speech, and wrong speaker identification (Fiscus et al., 2006). A value of zero would indicate perfect results.

The speaker diarization task was tested using different configurations of the Pyannote pipeline as well as Nvidia's NeMo model. These included (1) the standard Pyannote model, (2) audio preprocessing using spectral gating noise reduction[13] and (3) neural voice isolation technique[14], (4) a German segmentation model[15], (5) a custom segmentation model fine-tuned on "Jud Süß", and (6) a custom correction method that reduced speaker confusion using face embedding clusters. This approach is based on face extraction using Insightfaces[16] *buffalo_l* model and using HDBScan clustering algorithm. Afterwards it corrects speaker attributions by similarity of the appearance of similar faces. In (7), approach (6) is combined with the custom trained segmentation model of approach (5). In addition (8), the NeMo model was evaluated. NeMo was applied using the cascaded system with Clustering Diarizer, the *diar_infer_general.yaml* configuration and 30 fixed speakers

ID	Speaker Diarization Approach	Avg DER
1	Pyannote Standard	73.3
2	Pyannote with Spectral Gating Voice Isolation	94.1
3	Pyannote with Neural Voice Isolation	94.1
4	Pyannote with German Segmentation	84.5
5	Pyannote with Custom Segmentation Model	82.6
6	Pyannote with Face Enhancement	67.3
7	Pyannote with Face Enhancement and Custom Segmentation	77.3
8	**Nvidia NeMo**	**58.84**

Table 1: Average Diarization Error Rate (DER) in % results for different speaker diarization approaches. The Nvidia NeMo system achieved the best overall performance.

The results are shown in Table 1. Among all tested systems, NeMo achieved the lowest average diarization error rate (DER) of 58.84%, outperforming all Pyannote configurations. The approaches initially tested on Pyannote were not explored for Nvidia NeMo, because it became evident that the technological solutions are not able to solve the problem at this time. Although the experiments present an evaluation as a step on the way to full automation, speaker diarization remains the most challenging component in the pipeline. Therefore, the following parts in our pipeline still rely on manually verified ground truth for speaker identity. In the future, cross-modal methods, which combine audio and visual information, should be tested for this type of complex historical film data (see (Sharma and Narayanan, 2022); (Cheng et al., 2025)).

4.2. Evaluation of Audio-based Gender Detection

For the evaluation of the automated gender recognition, a ground truth test set was created by assigning a gender label to each individual character, including a separate category for children. This information was then transferred to the character's

[13] https://pypi.org/project/noisereduce/

[14] https://huggingface.co/speechbrain/sepformer-dns4-16k- enhancement

[15] https://huggingface.co/diarizers-community/speaker-segmentation-fine-tuned-callhome-deu

[16] https://github.com/deepinsight/insightface

speech segments, for which the gender was subsequently classified.

Class	Precision	Recall	F1-Score	Support
Female	0.42	0.96	0.58	164
Male	0.99	0.88	0.93	1,147
Child	1.00	0.06	0.11	86
Accuracy				0.83

Table 2: Classification performance of the gender classification model.

Class	Precision	Recall	F1-Score	Support
Female	0.68	0.95	0.79	20
Male	0.98	0.98	0.93	141
Child	0	0	0	8
Accuracy				0.93

Table 3: Classification results of gender classification. Segments > 10s.

As shown in Table 2, the inclusion of all speech segments results in an accuracy of 0.83. When only speech segments with a minimum length of ten seconds are classified, accuracy increases to 0.93 (see table 3). The results for the recognition of male voices are promising already. For the female class, however, there is an imbalance between precision and recall. The classifier reliably identifies positive cases but also wrongly classifies many speech segments as female voices even though they are not. Notably, these false positives are male voices; thus, no confusion with children's voices occurs, which might have been expected due to pitch similarities.

In the recognition of children's voices, which are generally underrepresented in the test dataset, clear weaknesses of the model become apparent. As shown in Table 2, for 86 child voice segments the model achieves a precision of 1 and a recall of 0.06. When the classifier identifies a child, the classification is correct, but this occurs only rarely. We assume several influencing factors. One is the audio quality, which is often affected by background noise. Another factor is that previous research has shown that models of this kind can be influenced by linguistic features. This phenomenon may manifest in the use of specific words or phrases that correlate with certain classes (Wagner et al., 2023). In the context of a narrative historical German film, the language of a child character is likely to be constructed differently and to address other topics than in real recordings of children.

The definition of what defines a child is also not clear-cut. For data annotation, a general rule was applied: a character was labeled as a child if they appeared visibly younger than an adolescent. Since the classification currently applies only to speech segments and does not directly determine the gender of a film character, an average voting approach was used across all speech segments of a character with a minimum duration. In the resulting gender assignments, errors occurred only in the misclassification of the child characters *Johannes von Redel* and *Wilhelm Panse* from the film "Kopf hoch, Johannes!" as a female. These errors were manually corrected for the following evaluation. Other errors were automatically compensated for through the average voting process and the character's gender correctly classified.

4.3. Evaluation of Automatic Speech Recognition

For the following analysis, we conduct a detailed qualitative examination of observed ASR errors. While standard evaluation metrics such as Word Error Rate (WER) and Character Error Rate (CER) provide a single aggregate score, they do not reflect the semantic severity of individual errors. This limitation becomes particularly relevant when ASR output serves as input for downstream language models, where minor orthographic errors are less important, but semantic deviations can significantly affect model behavior. To capture such nuances, several semantic evaluation metrics have been proposed, including BERTScore (Zhang et al., 2020), SemDist (Kim et al., 2021), Semantic-WER (Roy, 2021), SeMaScore (Sasindran et al., 2024) and Aligned Semantic Distance (ASD) (Rugayan et al., 2023). While these metrics offer improved correlation with meaning preservation, they still produce single-value summaries. Therefore, they fail to reveal the specific nature or distribution of error types. They are thus better suited for large-scale model comparison rather than for detailed case analyses. Because this study investigates the application of existing ASR systems within a highly specialized historical context, a more fine-grained qualitative assessment was decided for. A sample of 100 randomly selected speech segments from "Jud Süß" was analyzed, and transcriptions from different ASR models were compared manually. Despite a high level of accuracy, this evaluation revealed four predominant categories of errors:

A) *Specialized vocabulary*: Historical and context-specific terms such as *Durchlaucht*, *Staatsstreich*, or *Rabbuni* were often misrecognized. These rare tokens frequently degraded the accuracy of entire sentences, suggesting that models attempt to enforce semantic coherence when uncertain.

B) *Acoustic degradation*: Extreme low-volume passages or heavy background noise led to frequent hallucinations posing particular risks for semantic downstream tasks.

C) *Token repetition*: Some Whisper models exhibited looping behavior, producing repeated tokens. This was mitigated post hoc through the automatic removal of redundant n-grams.

D) *Language and dialectal variation*: The presence of Yiddish language and Berlin dialect resulted in systematically higher error rates, underscoring the models' limited adaptability to non-standard speech varieties.

The evaluation showed similar error patterns of the models. In general there was a slightly higher performance of large-v2, which was subsequently chosen for the following analysis.

4.4. Evaluation of GPT-Based Character Analysis

Since the methodology follows an exploratory approach, no strict metrics can be calculated to assess the quality of the character analyses. Therefore, the evaluation of the procedure will discuss its performance qualitatively based on the character descriptions of prominent film figures as presented in film studies and historical literature. Given the space limitations, the discussion here focuses on a representative excerpt. The difficulties arising from the complexity and ambiguity of film characters will be discussed in section 6.

Dimension	Model's answer
Personality traits	cunning, manipulative, business-minded, loyal, strategic
Goals	wealth, power, and influence
Motivations	preserving wealth and power
Interactions	tactical and manipulative

Table 4: GPT-based analysis of the character *Joseph Süß Oppenheimer* in the film "Jud Süß".

In the notoriously antisemitic film "Jud Süß", National Socialist antisemitism culminates in the antagonist *Joseph Süß Oppenheimer*. *Oppenheimer* becomes the advisor to the eccentric and decadently living *Duke of Württemberg*. He manipulates the *Duke* into acting against his own people, schemes to gain power, and commits assaults on women. *Oppenheimer's* character embodies antisemitic resentments in a way reminiscent of caricatures, making him a direct personification of NS ideology. He is manipulative and seeks power for "his people," power over Germans and Christians. In his portrayal, he appears "lustful, deceitful, destructive, and power-hungry" (Niven, 2022). These descriptions can also be observed in the character traits generated solely from the speech segments, as shown in Table 4. The same applies to the *Duke* (see table 5), who symbolizes a corrupted establishment.

Dimension	Model's answer
Personality traits	authoritative, impulsive, autocratic, emotional and overbearing
Goals	governing and securing loyalty
Motivations	power, security and stability
Interactions	governing and securing loyalty

Table 5: GPT-based analysis of the character *Duke of Württemberg* in the film "Jud Süß (1940).

In addition to these two characters, whose essence was well captured by the approach, other notable patterns can also be observed. Notably, the analyses in "Jud Süß" reveal a bias shaped by contemporary moral standards. The model does not adopt the film's antisemitic worldview but instead frames antisemitic characters as antagonists. This becomes evident for the case of *Aktuar Karl Faber*, who was originally portrayed as a heroic figure, who fights against *Oppenheimer* and makes a lot of anti-Semitic statements. He was automatically characterized as *intelligent, manipulative, and ruthless*, motivated by *power and financial gain*, and thus hinting negative attributes. This indicates that the model evaluates characters through an ethical lens derived from modern training data rather than the historical context of the film. Similarly, the portrayal of *Oppenheimer*'s companion Levy was less aligned with the film's intended depiction. GPT described him as *concerned, loyal, and skilled*, motivated by *self-preservation and justice*, emphasizing interpersonal dynamics rather than his supposed villainy. This may be due to dialectal transcription errors or the model's focus on linguistic rather than visual cues, but also highlights limitations of the approach.

The film "Kopf hoch, Johannes!" focuses on the character triangle of *Johannes*, a boy with adjustment difficulties, his caring aunt *Julieta*, and his strict father, who gradually softens over the course of the film (Giesen and Hobsch, 2005). This relational dynamic is reflected in the automated analysis. *Johannes* is characterized by the model as *searching, vulnerable, impulsive, lonely and rebellious* with the goals of *feeling accepted*, *finding love and finding his place* and his interactions are described as *defensive, distant and vulnerable.* His father is described as *hard, strict, determined, suspicious, vulnerable and self-sacrificing* with the goal to *protect his family* and interacts in a *repellent*

and controlling manner. *Julieta* is characterized as *caring, patient, loyal, understanding and self-sacrificing* with the goals of *ensuring the well-being of her son* and *fostering understanding between her son and his father.* She interacts in a *concerned, loving and understanding* way. These identified personality traits accurately reflect a core dynamic of the film.

Overall, the model performed robustly in identifying personality traits, key attributes, goals, motivations, and interaction types, providing valuable insights into character personality and narrative function.

5. Case Study: Gender Representation in Nazi Propaganda Films

To demonstrate the analytical potential of our audio-based pipeline beyond technical evaluation, we present a case study that applies the proposed pipeline to an important question in film and cultural studies: how are female characters constructed and ideologically instrumentalized in Nazi propaganda cinema (Vaupel, 2005). This analysis is of particular social relevance, as the audience in Nazi Germany became increasingly female while men were engaged in the war. Film propaganda functioned as a means to expand and sustain the National Socialist sphere of power by deliberately targeting women (Vaupel, 2005). This case study illustrates how computational approaches can support interpretive research questions in the humanities and highlights the capacity of language-based character analysis to reveal recurring ideological patterns. For all female characters in the three investigated films, results were generated using the developed pipeline for the categories "Personality traits", "Goals," "Motivation," and "Interaction with other Characters." These results were examined through an exploratory qualitative analysis, from which character categories were derived.

5.1. Female Character Categories

Cat. 1: Traditional mother figures – The first category of female characters represents the mother figure as a central element of National socialist ideology. The personality traits of the female characters in the three films, which were identified through the pipeline, point to this main category. It represents the *loyal, helpful, caring, patient, loving, devoted*, but also *concerned and struggeling* woman. Included in this category are *Mother Heini*, *HJ supporter grandma* and the *Nurse* from "Hitlerjunge Quex", as well as *Julieta Merck*, *Mother Panse* and the *Nurse* from "Kopf hoch, Johannes!", as well as the *Duchess of Württemberg* from "Jud Süß". The female characters aim to *protect their families, maintain peace, offer help, show care, assist others, or bring happiness*. They are motivated by intrinsic factors such as the *protection of the family, benevolence, love, hope, or loyalty*. They interact in *concerned, friendly, polite, compassionate, loving, and supportive ways*.

Cat. 2: Younger women as mothers – In this category we find younger women who are strongly related to the concept of the mother role, but with elements of youth. It is highlighted in slight deviations that are observed in *Ulla Dörries*, a young National Socialist, who fights together with other young Nazis. She is also *impulsive*. Similarly, *Dorothea Sturm* is characterized slightly differently, she also appears *confused and concerned*. Nevertheless, both characters resemble the described category one in their personality traits, though they seem to be influenced by their youth.

Cat. 3: Non-conforming representations of femininity – This category includes two characters that act as counter-types to maternalist and domestic NS norms. One prime example here would be the case of *Gerda* from "Hitlerjunge Quex", a female character who deviates strongly from the first category. The model describes her as *combative, confrontational, direct, controlling, and commanding*. In the film, *Gerda* functions as an ally of the antagonists. She is a communist woman who, on behalf of her leaders, seduces and persuades the Hitler Youth member *Grundler* to turn away from National Socialism. Her goals are described as *survival, power, and dominance* and her motivations as *fear, pride, and a sense of superiority*. Finally, according to the model, she interacts *dominant, overbearing, and confrontational*. The automated analyses also reveals a certain ambiguity in her character, as her goal is also *survival*, driven by her motivation *fear*. This aligns to the films content, as *Gerda* is pressured into her actions by her communist leaders.

Another case, were the automated analysis hints to deviations is the *Daughter of a Landstand* in "Jud Süß". The model describes her as *confident, strong, and rebellious*. In the film she resists her controlling father because she wants to attend a ball hosted by the *Duke*. However, in the film's plot, as she attends the ball she is subjected to sexual violence by the *Duke* and *Joseph Süß Oppenheimer*. This final example vividly demonstrates how women who deviate from the National Socialist ideal of femininity are punished within the narrative – their suffering could be interpreted as an implicit warning to the female audience.

5.2. Discussion: Ideological elements in the depictions of femininity

The detected main category one of women in the films is strongly oriented toward the mother figure, a central ideological element of National Socialist thought. Hitler describes the role of the woman as a mother as a counterpart to the male heroism on the battlefield (Wagenaar, 2023). In the idealized primary role as a mother, women were expected to contribute to the preservation of the people by sustaining the family. This balanced duality was shaped by the National Socialist ideal of the ethnic-family model, a key ideological element. This element becomes quite clear by the portrayed characteristics of femininity in the films. The three films already reveal a negative or cautionary depiction of autonomy, dominance, and self-confidence in the propagation of the female ideal. Even within this limited sample of National Socialist propaganda films, it becomes evident that an exploratory analysis of divergent portrayals of female roles can provide valuable insights and help outline the image of womanhood expected within the German nation. In the context of a scaled-up analysis, these factors could be examined in relation to the specific period of the Nazi regime, film genre, character importance, age, and other variables. In this way, a more differentiated and complex picture can be developed of which target groups were addressed, what kind of image was propagated, and how women were expected to identify with it.

6. Reflection

The presented evaluation and the case study demonstrate both the methodological potential and the interpretive value of computational audio analysis for historical film research. In this section, we reflect on the findings and discuss methodological challenges and limitations.

One key insight from our experiments concerns the limitations of closed-source models, which pose serious challenges to reproducibility. The filters and settings applied within such systems are typically opaque, a particularly critical issue for historical analysis, as it introduces potential biases that can significantly affect results. Consequently, any exploratory analysis must be conducted with careful methodological reflection, following the standards of qualitative research. Both model-related biases and researcher biases—especially confirmation bias stemming from prior hypotheses—need to be explicitly acknowledged and critically addressed.

Another important point is that the insights into character traits obtained with this approach are based on an exploratory analysis. This makes the results interpretative and more difficult to compare, which can be challenging when addressing questions across large datasets. Consequently, the evaluation of the method is primarily qualitative. A future approach could be to give the models less freedom in analyzing character traits and instead move toward zero-shot classification using pre-defined character categories. This would improve comparability and open-up testing with created baseline datasets. However, an exploratory analysis of the complex medium of film aligns well with the principles of Distant Viewing. Therefore, in line with its intended purpose, the exploratory analysis provides the opportunity to identify and analyze unexpected or divergent categories and to generate research questions from them. The presented approach offers considerable potential for investigating character profiles in relation to variables such as film genre, historical context within the National Socialist period, and the films' reception or popularity. Future research could address questions such as: In which films do women get portrayed as mother figures? In which roles do they diverge and at what time during the Nazi regime?

Nevertheless, it is important to note that a thorough character analysis must also account for ambiguities and subtleties of characters, which can be lost in the presented approach. This is particularly evident in the absence of a temporal dimension in the character analysis. If a character develops and changes over the course of a film, this is currently not captured. While increased scalability will inevitably affect the complexity of the analysis, temporal aspects could potentially be incorporated in the future, adjusting and extending the existing the pipeline.

Additionally, it is important to approach the films with source-critical scrutiny. In particular, for historical material, it is necessary to examine which edited version of the film is being analyzed to avoid biases by changed versions of the films. Finally, future work should focus on improving the speaker diarization in this pipeline.

7. Conclusion and Outlook

Our results show that while speaker diarization remains a methodological bottleneck — particularly given the challenges posed by historical audio recordings — state-of-the-art ASR models such as Whisper (Radford et al., 2022) perform robustly even under noisy conditions. Moreover, GPT-based analyses of transcribed speech yield character trait profiles that align closely with established interpretations in film scholarship (Niven, 2022; Giesen and Hobsch, 2005), demonstrating the potential of large language models to support interpretive work in film studies. The proposed pipeline represents a first step toward this goal, offering a scalable, ex-

tensible framework for audio-based film analysis. It also points the way toward future multimodal approaches that integrate visual and auditory data to achieve a more holistic understanding of how ideology operates across media layers.

8. Ethics Statement

This study involves the analysis of Vorbehaltsfilme (restricted Nazi propaganda films), which contain antisemitic, racist, and otherwise harmful content. These films are preserved by the Friedrich Wilhelm Murnau foundation and made available solely for academic research under controlled conditions. All analyses in this paper were conducted within a critical scholarly framework aimed at understanding the mechanisms of ideological construction in historical cinema. No material derived from the films is publicly distributed, and no attempt has been made to reproduce or disseminate harmful content outside of this research context. The study adheres to ethical principles of responsible data handling and historical sensitivity, acknowledging the continued impact of these materials and their potential for misuse.

9. Author Contributions

Nicolas Ruth: Conceptualization; Methodology; Software; Validation; Formal Analysis; Investigation; Data Curation; Writing – Original Draft; Visualization. *Manuel Burghardt*: Conceptualization; Writing – Original Draft; Writing – Review Editing; Supervision. *Andreas Niekler*: Conceptualization; Methodology.

10. Acknowledgements

We thank the Friedrich Wilhelm Murnau foundation for granting access to the digitized Vorbehaltsfilme used in this research. Their support was essential in enabling the development and evaluation of the proposed audio analysis pipeline.

11. Bibliographical References

Apoorv Agarwal, Augusto Corvalan, Jacob Jensen, and Owen Rambow. 2012. Social network analysis of alice in wonderland. In *Proceedings of the NAACL-HLT 2012 Workshop on computational linguistics for literature*, pages 88–96.

Taylor Arnold and Lauren Tilton. 2019. Distant viewing: analyzing large visual corpora. *Digital Scholarship in the Humanities*, 34(Supplement_1):i3–i16.

Sabyasachee Baruah and Shrikanth Narayanan. 2024. Character attribute extraction from movie scripts using llms. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8270–8275. IEEE.

Monika Bednarek. 2018. *Language and television series: A linguistic approach to TV dialogue*. Cambridge University Press.

Bernhard Bermeitinger, Sebastian Gassner, Siegfried Handschuh, Gernot Howanitz, Erik Radisch, and Malte Rehbein. 2019. Deep watching: Towards new methods of analyzing visual media in cultural studies. In *Books of Abstracts – ADHO DH Conference, Utrecht*.

Ryan L. Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W. Pennebaker. 2022. The development and psychometric properties of liwc-22. Technical report, University of Texas at Austin. Accessed October 17, 2025.

Ryan L. Boyd, Paola Pasca, and Kevin Lanning. 2020. The personality panorama: Conceptualizing personality through big behavioural data. *European Journal of Personality*, 34(5):599–612.

Ryan L Boyd and James W Pennebaker. 2017. Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18:63–68. Big data in the behavioural sciences.

Manuel Burghardt, John A Bateman, Eric Müller-Budack, and Ralph Ewerth. 2024. Computational tools and methods for film and video analysis. *Compendium computational theology*, 1:147–173.

Felix Burkhardt, Johannes Wagner, Hagen Wierstorf, Florian Eyben, and Björn Schuller. 2023. Speech-based age and gender prediction with transformers.

Seymour Chatman. 1978. *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell University Press, Ithaca, NY.

Luyao Cheng, Hui Wang, Chong Deng, Siqi Zheng, Yafeng Chen, Rongjie Huang, Qinglin Zhang, Qian Chen, Xihao Li, and Wen Wang. 2025. Integrating audio, visual, and semantic information for enhanced multimodal speaker diarization on multi-party conversation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 19914–19928, Vienna, Austria. Association for Computational Linguistics.

Jens Eder. 2025. *Characters in Film and Other Media: Theory, Analysis, Interpretation*, 1 edition. Open Book Publishers, Cambridge, UK.

Ethan Fast, Binbin Chen, and Michael S. Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 4647–4657, New York, NY, USA. Association for Computing Machinery.

Jonathan G. Fiscus, Nicolas Radde, John S. Garofolo, Audrey Le, Jerome Ajot, and Christophe Laprun. 2006. The rich transcription 2005 spring meeting recognition evaluation. In *Machine Learning for Multimodal Interaction*, pages 369–389, Berlin, Heidelberg. Springer Berlin Heidelberg.

Rolf Giesen and Manfred Hobsch. 2005. *Hitlerjunge Quex, Jud Süß und Kolberg. Die Propagafilme des Dritten Reiches Dokumente und Materialiern zum NS-Film*. Schwarzkopf & Schwarzkopf Verlag, Berlin.

Gaudenz Halter, Rafael Ballester-Ripoll, Barbara Flueckiger, and Renato Pajarola. 2019. Vian: A visual annotation tool for film analysis. In *Computer Graphics Forum*, volume 38(3), pages 119–129. Wiley Online Library.

Christian Hardinghaus. 2008. *Filmpropaganda für den Holocaust?: Eine Studie anhand der Hetzfilme "Der ewige Jude" und "Jud Süß"*. Tectum-Verl., Marburg.

Alina Herderich, Heribert H. Freudenthaler, and David Garcia. 2024. A computational method to reveal psychological constructs from text data. *Psychological Methods*.

Douglas Kellner. 1991. Film, politics, and ideology: Reflections on hollywood film in the age of reagan. *Velvet Light Trap*, 27(2):9–24.

Suyoun Kim, Abhinav Arora, Duc Le, Ching-Feng Yeh, Christian Fuegen, Ozlem Kalinli, and Michael L. Seltzer. 2021. Semantic distance: A new metric for asr performance analysis towards spoken language understanding.

Victor Klemperer. 1947. *LTI – Notizbuch eines Philologen*. Reclam.

Sarah Kozloff. 2000. *Overhearing film dialogue*. Univ. of California Press.

La Javaness R&D. 2023. Speaker diarization: An introductory overview. Accessed October 17, 2025.

Bernhard Liebl and Manuel Burghardt. 2023. *Designing a Prototype for Visual Exploration of Narrative Patterns in NewsVideos*. Gesellschaft für Informatik eV.

Franco Moretti. 2000. Conjectures on world literature. *New left review*, 2(1):54–68.

Franco Moretti. 2011. Network theory, plot analysis. In *Stanford Literary Lab Pamphlet 2*.

Eric T Nalisnick and Henry S Baird. 2013. Character-to-character sentiment analysis in shakespeare's plays. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 479–483.

Bill Niven. 2022. *Jud Süß. Das Lange Leben eines Propagandafilms*. Mitteldeutscher Verlag, Halle.

Alexis Plaquet and Hervé Bredin. 2023. Powerset multi-class cross entropy loss for neural speaker diarization. In *INTERSPEECH 2023*. ISCA.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision.

Somnath Roy. 2021. Semantic-wer: A unified metric for the evaluation of asr transcript for end usability.

Janine Rugayan, Giampiero Salvi, and Torbjørn Svendsen. 2023. Perceptual and task-oriented assessment of a semantic metric for asr evaluation. In *Interspeech 2023*, pages 2158–2162.

Michael Ryan and Douglas Kellner. 1988. Camera politica. *The politics and ideology of contemporary Hollywood film. Bloomington*.

Zitha Sasindran, Harsha Yelchuri, and T. V. Prabhakar. 2024. Semascore: A new evaluation metric for automatic speech recognition tasks. In *Interspeech 2024*, page 4558–4562. ISCA.

Thomas Schmidt and Manuel Burghardt. 2018. An evaluation of lexicon-based sentiment analysis techniques for the plays of gotthold ephraim lessing. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, Santa Fe, New Mexico, USA*. Association for Computational Linguistics.

Rahul Sharma and Shrikanth Narayanan. 2022. Using active speaker faces for diarization in tv shows.

Matthias Springstein, Markos Stamatakis, Margret Plank, Julian Sittel, Roman Mauer, Oksana Bulgakowa, Ralph Ewerth, and Eric Müller-Budack. 2023. Tib av-analytics: A web-based platform for scholarly video analysis and film studies. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3195–3199.

Angela Vaupel. 2005. *Frauen im NS-Film*. Verlag Dr. Kovac, Hamburg.

Maike Wagenaar. 2023. *Das Frauen- und Mutterbild im Nationalsozialismus und seine Auswirkungen bis heute. Eine sozialpsychologische Untersuchung zu unbewussten Übernahmen*. Budrich Academic Press, Berlin.

Johannes Wagner, Andreas Triantafyllopoulos, Hagen Wierstorf, Maximilian Schmitt, Felix Burkhardt, Florian Eyben, and Björn W. Schuller. 2023. Dawn of the transformer era in speech emotion recognition: Closing the valence gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10745–10759.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert.