Voices Across Decades: A Multimodal Diachronic Corpus of German Bundestag Debates (GerParlDia-MM)

Ingo Siegert[1,2]
[1] Mobile Dialog Systems, Otto von Guericke University Magdeburg, Germany
[2] Intelligent Assistive Systems for Psychotherapy, University Hospital Magdeburg, Germany

Abstract

This paper presents a multimodal diachronic corpus of German parliamentary debates spanning 1949 – 2025. The dataset focuses on speakers with exceptionally long political careers in the Bundestag, covering at least six parliamentary terms for female and eight for male members, comprising 75 individuals (43 men/32 female) and 2,136 speeches. The corpus integrates audio, video (when available), and official transcripts, enriched with metadata on date, party affiliation, and legislative term. Transcripts were temporally aligned with parliamentary media recordings, and non-speech segments were automatically removed. The corpus enables research on voice aging, intra-speaker variability, and longitudinal political language, and supports benchmarking of ASR and speaker recognition across decades. Thus, this corpus bridges the gap between short-term speech corpora and single-speaker longitudinal datasets, offering a unique foundation for studying change in voice, style, and rhetoric over more than seventy years of German parliamentary history.

Keywords: Longitudinal Speech Corpus, Multimodal Resources, Diachronic Linguistics, Parliamentary Speeches

1. Introduction

Longitudinal speech data are essential for understanding how voices and language evolve over time and for assessing the robustness of current speech technologies under such variation (Richter et al., 2020; Siegert et al., 2020). Research in speech technology, forensic phonetics, and diachronic linguistics has shown that age-related changes in voice quality, prosody, and articulation can substantially affect automatic speech and speaker recognition performance (Singh et al., 2023; Vipperla et al., 2010). However, most available corpora capture only short-term variability or focus on individual speakers rather than systematic long-term observations (Siegert et al., 2019; Tornow et al., 2016).

Parliamentary debates provide a unique setting for longitudinal speech research: they combine consistent recording conditions, well-documented contexts, and rich metadata about speakers and time periods. Despite these advantages, no existing dataset has yet combined audio, video, and text from German parliamentary sessions in a way that enabling an unprecedented foundation for studying diachronic changes in voice, style, and rhetoric.

To address this gap, the **German Parliamentary Diachronic Multimodal Corpus (GerParlDia-MM)**, a longitudinal resource covering more than seventy years of German parliamentary debates (1949–2025) is introduced. It unites audio, video, and official transcripts for 75 long-serving members of parliament, each represented by at least one speech per year. The corpus provides a foundation for research on voice aging, stylistic and rhetorical change, and the longitudinal performance of automatic speech processing systems. By integrating multimodal data, institutional context, and detailed metadata, it extends short-term speech corpora and single-speaker longitudinal datasets. The resource is fully reproducible as an open benchmark for multimodal and diachronic speech research. To illustrate potential research applications, the corpus supports analyses of both linguistic and acoustic change over time. For example, longitudinal measurements of lexical diversity and syntactic structure can reveal shifts in rhetorical style across decades, while acoustic analyses enable the study of age-related changes in prosodic features such as pitch range, speech rate, or voice quality. Such analyses provide insights into both institutional discourse evolution and the robustness of automatic speech processing systems under long-term speaker variation.

2. Related Work

Longitudinal corpora have been explored in various domains, yet few provide consistent recordings of the same speakers over extended periods. Short-term multi-speaker datasets such as CSLT-Chronos (Wang et al., 2016) or VoxCeleb (Nagrani et al., 2020) offer wide coverage but span only a few years, limiting analyses of diachronic change. Longitudinal resources like TCDSA (Kelly et al., 2012) or the multilingual VoxAging corpus (Ai et al., 2025) include recordings over several decades but remain restricted in sample density and institutional consistency.

Single-speaker resources demonstrate the potential of institutional contexts for longitudinal analysis. For instance, the Merkel Podcast Corpus (Saha et al., 2022) and studies of Queen Eliza-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 6289–6297
11–16 May 2026.

beth II's broadcasts (Harrington et al., 2000). While these corpora allow for detailed prosodic and stylistic analyses, they lack broader speaker diversity.

For the German language, several parliamentary resources exist, notably GermaParl (Blaette and Leonhardt, 2023) and OpenDiscourse (Richter et al., 2020), which provide richly annotated plenary texts since 1949. However, these resources are primarily text-centered and do not include systematically linked and temporally aligned audio or video recordings. As a result, they do not support multimodal longitudinal analyses of within-speaker change across decades.

The German Parliamentary Diachronic Multimodal Corpus (GerParlDia-MM) introduced in this paper extends these efforts by linking official transcripts with corresponding audiovisual materials and detailed speaker metadata. It thereby provides the first multimodal and diachronically continuous resource enabling empirical analyses of political speech, voice aging, and stylistic change over more than seven decades.

3. Data Collection & Processing

The GerParlDia-MM was constructed through a multi-stage workflow involving speaker selection, media retrieval, transcription alignment, and metadata integration.

3.1. Data Sources

The foundation of the corpus is the official parliamentary metadata provided by the German Bundestag as Open Data[1] in XML format. These files contain biographical records of all members of parliament since 1949, including persistent `PoliticianID` identifiers that enable cross-referencing between datasets. From these data, all members serving across multiple legislative terms were identified. To reflect historical differences in gender representation, asymmetric thresholds were applied: at least eight terms for male and six terms for female members. This adjustment ensures comparable temporal coverage despite unequal participation in earlier decades.

3.2. Speaker Selection

From the complete list of 5,223 members of the German Bundestag (as of August 2025), individuals whose parliamentary careers span at least eight legislative terms for men and six for women were identified, reflecting gender-specific representation differences in earlier decades. This yielded 75 long-serving members, 43 men and 32 women, whose activity extends from the 1st legislative period (1949–1953) to the ongoing 21st period, which began in March 2025.

The longitudinal coverage per person ranges from about 18 to 53 years, averaging roughly three decades of active parliamentary service (see Table 1). *Rita Schwarzelühr-Sutter* (2006–2024; 18.3 years) continues to serve in the current Bundestag, while *Helmut Schmidt* (1954–2006; 52.6 years) delivered a special plenary address in 2006, almost twenty years after his regular departure in 1987. The longest female trajectory belongs to *Herta Däubler-Gmelin* (1973–2009; 35.4 years).

Measure	Male	Female
Min duration	26.8	21.9
Max duration	52.6	35.4
Avg duration	32.6	26.3
Min entry age	22.0	26.0
Avg entry age	34.0	36.2
Avg exit age	68.3	62.2
Max exit age	81.0	74.0

Table 1: Duration of parliamentary careers in years and average entry/exit age among long-term Bundestag members.

On average, speakers entered parliament at 34.9 years of age (range 22–47) and left at 65.7 years (range 45–81), indicating a typical political career of roughly 30 years. The corpus covers all major parties represented since 1949: SPD (33), CDU (20), CSU (9), BÜNDNIS 90/DIE GRÜNEN (6), FDP (4), DIE LINKE (3), and one non-affiliated member. This diversity ensures balanced ideological and generational representation across post-war German politics.

3.3. Retrieval of Speeches

For each selected member, the corresponding speeches were collected from the Bundestag Mediathek whenever available and, for earlier sessions, from the Parliamentary Archives. To guarantee linguistic substance, only speeches with a duration of at least 120 s ($\approx$ two minutes) were included. Across all candidates, a theoretical maximum of 2,459 speeches was expected. In total, 2,136 items (87 %) were successfully retrieved. Missing years primarily result from sessions without recordings or from members who did not deliver a speech in that term. Eight speakers are still active in the current legislative period, so additional speeches will be integrated in future updates (see Figure 1).

3.4. Alignment of Transcripts

Speech transcripts were temporally aligned with the corresponding audio using WhisperX (Bain et al.,

[1] https://www.bundestag.de/services/opendata

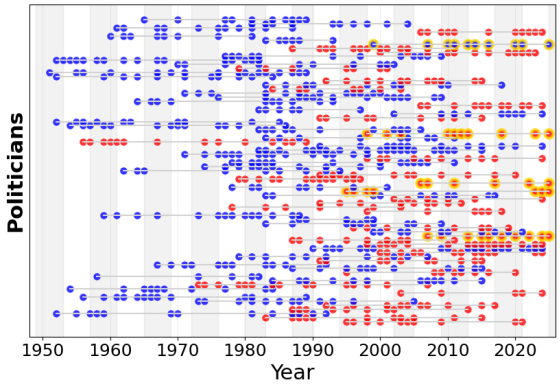

Figure 1: Overview of available speeches per speaker per year (male = blue, female = red). Highlighted speakers are current members of the Bundestag, indicating that further speeches may be added in the future.

2023), which provides sentence-level segmentation and alignment. This enabled time-accurate mapping between text and audio for prosodic and acoustic analyses.

To identify the actual start and end of each speech, excluding the president's announcements or subsequent agenda items, a hybrid approach combined regular expressions with a large language model (`meta-llama-3.1-8b-instruct`). Typical greeting and closing patterns were automatically recognized, and the resulting onset/offset timestamps were stored as metadata, ensuring consistent temporal boundaries across all items.

3.5. Integration with OpenDiscourse and Plenary Protocols

Each speech was linked to its corresponding `PoliticianID` from the Bundestag Open Data records and, where available, to the `speeches.id` entry from the OpenDiscourse database (Richter et al., 2020). As OpenDiscourse currently covers debates until May 2021, all later speeches were connected to their official plenary protocol, whose persistent URL was added to the metadata. This dual linkage ensures longitudinal traceability and facilitates interoperability with other parliamentary corpora such as GermaParl (Blaette and Leonhardt, 2023).

3.6. Metadata Structure

Metadata were compiled at both person and speech levels. On the person level, available information includes name, gender, date of birth and death (if applicable), education, profession, and party affiliation. Speech-level metadata include date, legislative term, media type (audio/video), technical parameters (codec, bitrate, resolution), duration and start time, cross-references to OpenDiscourse and plenary protocols, Mediathek or archive identifiers, and automatically computed audio/video quality scores (see Section 5.1). These structured metadata ensure reproducibility and interoperability across linguistic and political domains.

3.7. Quality Control

All metadata were validated through semi-automated consistency checks across speaker-year combinations, ensuring that each entry contained one verified speech per person and year. This guarantees data completeness and transparent provenance for downstream analyses.

4. Dataset Description

The harmonized corpus covers more than seven decades of German parliamentary debates. It contains 2,136 speeches from 75 long-term members of the Bundestag (1949 – 2025), each combining audio, video (when available), and aligned transcripts with rich metadata.

Table 2: Sources, modalities, and quality levels represented in the corpus.

Category	Count	Share (%)
Total speeches	2,136	100
From archive	1,443	67.4
From media library	698	32.6
With video	1,141	53.3
Linked to OpenDiscourse	1,951	91.2

About two thirds of all recordings originate from the Parliamentary Archive and one third from the Mediathek. Over half include synchronized video, reflecting the transition from audio-only documentation in the early decades to digital full-HD recordings in recent years. Audio formats evolve from early low-bitrate AAC to high-quality PCM, mirroring the technological progression of the Bundestag's media infrastructure and providing valuable variability for testing robustness in automatic speech processing.

All speeches are linked to their official plenary protocol and, where available, to the corresponding OpenDiscourse entry, ensuring longitudinal traceability and reproducibility across data sources.

The corpus totals 1.26 TB of audiovisual data and approximately 27.7 hours of recorded speech, offering substantial temporal depth and multimodal coverage. Together with its standardized metadata, the resource provides a reliable foundation for diachronic analyses of voice, style, and rhetoric.

Table 3: Aggregate corpus statistics (N=2,136 speeches).

Measure	Mean	Total
Speech duration (min)	18.30	–
Audio size (GB)	0.51	504.50
Video size (GB)	0.69	790.30
Token count	1,285	2.5 M

Over multiple legislative periods, the dataset reflects major technological transitions in parliamentary media production. Early analog archive recordings (1950s – 1970s) exhibit limited frequency range and higher noise, while digital PCM formats introduced in the mid-1980s markedly improved clarity and stability. From the 2000s onward, the transition to AAC compression and later to full-HD video (after 2017) further enhanced technical quality, though increased spontaneity and background activity in modern plenary sessions add natural acoustic variability. This temporal heterogeneity provides both challenges and opportunities for studying the robustness of speech and speaker recognition systems over time.

5. Dataset Analysis

To ensure that the GerParlDia-MM provides reliable and consistent material for longitudinal research, a series of quantitative assessments were conducted. These analyses cover technical audio and video quality, paralinguistic and interactional activity, and linguistic stability over the entire observation period". Together, they illustrate the multimodal depth and both the longitudinal and diachronic breadth of the dataset.

5.1. Media Quality Assessment

The perceptual quality of all audio and video recordings was evaluated using two reference-free, state-of-the-art models: NISQA for speech and DOVER for video.

Audio quality. For each file, the NISQA model (Mittag et al., 2021) estimated mean opinion scores (MOS-like) on a five-point scale. Results were aggregated by decade to visualize historical trends in recording fidelity. Across all 2,136 speeches, the mean NISQA score was 2.95 (SD = 0.51), corresponding to moderate-to-good intelligibility. As expected, early analog recordings from the 1950s – 1970s scored lowest (mean $\approx$ 2.6), followed by steady improvement during the transition to digital PCM formats in the mid-1980s (peak $\approx$ 3.1). Interestingly, more recent recordings (post-2010) show a mild decline (mean $\approx$ 2.9), likely caused not by technical degradation but by increased spontaneity, audience noise, and interactional overlap in modern sessions. Figure 2 illustrates this non-linear trend, reflecting both technological progress and the evolving acoustic environment of parliamentary debates.

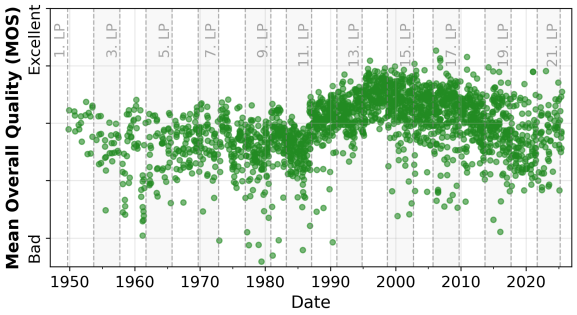

Figure 2: Estimated MOS audio quality over time using NISQA (mean per decade).

Video quality. To estimate visual quality, the DOVER model (Wu et al., 2023) was applied to all available videos. Unlike traditional metrics, DOVER disentangles *technical* and *aesthetic* quality dimensions without requiring reference data, making it suitable for historical material. Both indicators improved sharply after the shift from analog PAL/NTSC recordings to digital formats around 2010 and again after the introduction of full-HD video in 2017. The technical score increased continuously due to higher resolution and encoding stability, while the aesthetic score plateaued once camera framing and color balance reached a standardized broadcast style. A slight decline after 2022 likely relates to the introduction of automated camera control and adaptive compression in live streaming setups.

Overall, these objective metrics confirm a clear trajectory of media improvement over the corpus' 70-year span while retaining natural variability essential for benchmarking speech- and video-processing systems.

5.2. Scene and Acoustic Event Analysis

To capture interactional dynamics, visual scene detection and acoustic event recognition were applied to all video data. Scene transitions were extracted using PySceneDetect[2] (content-based mode), while acoustic events such as applause, laughter, or interjections were identified with YAMNet[3]. Across all sessions, an average of 1.8 scene changes per minute (SD = 0.7) was observed, typically corresponding to alternating camera perspec-

[2] https://www.scenedetect.com/

[3] https://github.com/tensorflow/models/tree/master/research/audioset/yamnet

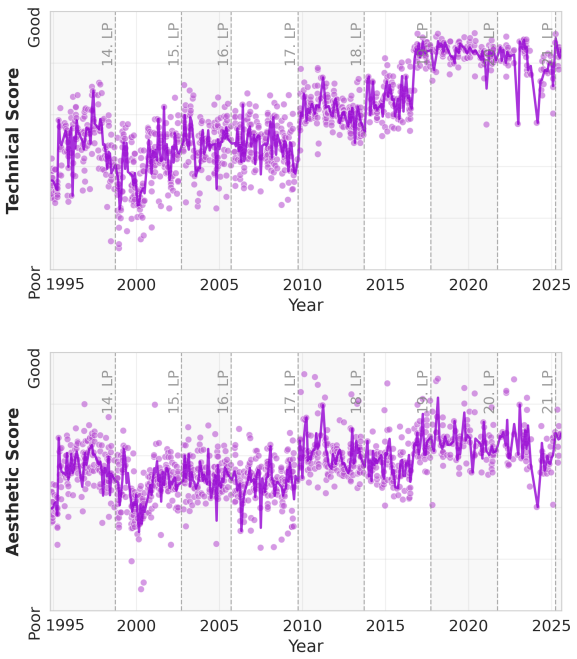

Figure 3: Estimated video quality using DOVER: (top) technical, (bottom) aesthetic scores.

tives between the speaker and the audience. Applause occurred in roughly 9 % of speeches and laughter or interruptions in another 4 %, serving as valuable anchors for studying audience engagement, speaker charisma, or emotional resonance, see Figure 4.

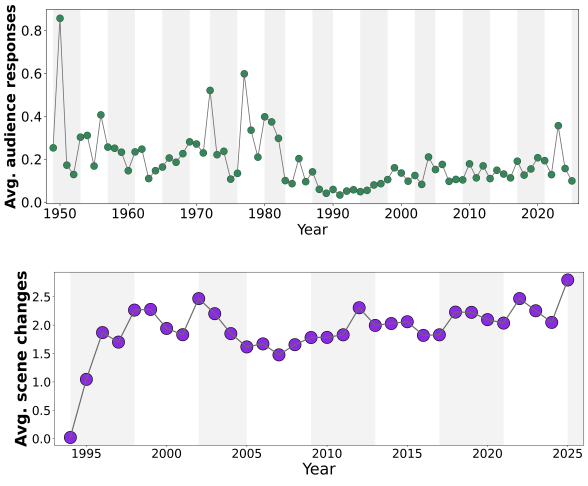

Figure 4: Automatically identified acoustic events (applause, laughter, and other reactions) using YAMNet (above) and visual scene changes using PySceneDetect (below). Both are normalized per media clip length and measured per year.

5.3. Textual Analysis

The textual dimension was examined using official plenary transcripts (see Section 3.5). Tokenization and POS tagging were performed with spaCy[4] (German model v3.7). The corpus contains approximately 2.5 million tokens, averaging 1,285 words per speech. Table 5 summarizes the evolution of key linguistic metrics across seven decades of parliamentary speech.

Lexical diversity was initially assessed using the type–token ratio (TTR). Since TTR is known to be sensitive to text length, we additionally computed three widely used length-robust measures: the moving-average type-token ratio (MATTR) (McCarthy and Jarvis, 2010) as well as MTLD and HD-D (McCarthy and Jarvis, 2010), see Table 4. The results indicate only modest changes in lexical diversity across decades. MATTR decreases slightly from 0.780 in the 1950s to 0.769 in the 2020s. HD-D shows a comparable small decline (0.8938 to 0.8790). MTLD exhibits a somewhat stronger decrease (129.31 to 110.35), suggesting an increase in lexical repetition rather than a substantial reduction in vocabulary range, potentially reflecting changes in rhetorical style or message framing in parliamentary speech over time.

Table 4: Lexical diversity by decade using length-robust metrics.

Decade	MATTR_{100}	MTLD	HD-D
1950s	0.780	129.31	0.894
1960s	0.777	130.74	0.890
1970s	0.789	143.62	0.897
1980s	0.786	139.50	0.894
1990s	0.789	135.81	0.895
2000s	0.781	124.23	0.880
2010s	0.770	120.62	0.878
2020s	0.769	110.35	0.879

To assess whether the corpus-level trends might be driven by speaker turnover, two long-serving members of parliament were examined: Herta Däubler-Gmelin (1972–2009) and Helmut Schmidt (1953–1987). Their lexical diversity values are comparable to the corpus-level averages (MTLD: 131.09 and 128.46; HD-D: 0.8989 and 0.9036), indicating that these speakers fall well within the overall distribution of lexical diversity values in the corpus. This suggests that the observed aggregate patterns are unlikely to be caused by a small number of highly distinctive speakers.

5.4. Discussion of Variability

These analyses highlight substantial temporal and technical variability intrinsic to a corpus spanning

[4] https://spacy.io/

Table 5: Average tokens per speech and mean sentence length per decade.

Decade	Tokens/Speech	Mean Sent. Len.
1950s	1,658	19.3
1960s	1,504	19.2
1970s	2,138	19.3
1980s	1,300	17.3
1990s	1,258	16.6
2000s	1,263	16.1
2010s	1,042	16.1
2020s	803	16.4

more than seven decades. Despite evolving recording conditions, the dataset remains nearly balanced across gender with 43 male and 32 female speakers. The combination of stable linguistic patterns, measurable changes in media quality, and multimodal coverage makes it an exceptional foundation for cross-disciplinary research in speech technology, linguistics, and political communication. It also provides a realistic testbed for benchmarking the diachronic robustness of speech- and speaker-recognition systems.

6. Legal and Ethical Considerations

Following the quantitative analyses, this section addresses the legal and ethical framework governing the data sources and their use in scientific contexts.

The *GerParlDia-MM* integrates multimodal data from the German Bundestag across several official sources, each governed by distinct legal frameworks. The dataset comprises metadata and textual transcriptions of plenary speeches and persistent identifiers referencing audiovisual material from the Bundestag Mediathek and the Parliamentary Archives. While the released corpus itself includes no audiovisual media, it enables the alignment of transcripts and metadata with the corresponding official recordings via unique IDs or Mediathek links.

Text data: Plenary protocols and printed proceedings of the German Bundestag are, according to § 4(1) of the *Archivordnung des Deutschen Bundestages* (Deutscher Bundestag, 2021a), "documents that were intended for publication at the time of their creation and therefore are not subject to any protection period." Consequently, these textual materials are publicly accessible and may be used, processed, and analyzed for scientific and educational purposes without additional authorization. Their use in this corpus fully complies with the legal provisions on archival and information freedom.

Audiovisual data: The referenced audiovisual material originates from the official *Bundestag Mediathek* and is covered by the *Nutzungsbedingungen für das Audio- und Videomaterial des Parlamentsfernsehens* (Deutscher Bundestag, 2021b). These terms explicitly permit free use for educational, cultural, and parliamentary purposes, provided that no commercial or advertising use takes place and that the material is not altered in a misleading or defamatory manner. All materials must retain their source attribution ("Deutscher Bundestag") or the original broadcaster logo. The corpus therefore only provides identifiers and links to the Mediathek, ensuring that users can access the material within the framework of the Bundestag's licensing terms.

Archival material: Older or non-digitized sessions are preserved under the *Archivordnung* and the *Nutzungsordnung für das Parlamentsarchiv* (Deutscher Bundestag, 2021c). These documents may be used for scientific research upon application and approval by the Parliamentary Archives (§ 3 *Archivordnung*). Any publication or reproduction of such material requires explicit written permission and appropriate citation as specified by the archive. No such archival reproductions will thus be distributed within the corpus release.

All components of the *GerParlDia-MM* are thus sourced from publicly accessible and legally compliant materials. The project follows the ethical principles of transparency, traceability, and respect for personal rights. Only speech data that was intended for public dissemination in parliamentary contexts is included, and no personal or non-public data has been processed.

Researchers may request collaborative use of the audiovisual materials under the Bundestag's Mediathek and archival terms, ensuring consistent ethical safeguards. This ensures that any extended multimodal analyses are conducted under the same legal and ethical safeguards as the core corpus.

7. Limitations

While the GerParlDia-MM provides unprecedented diachronic and multimodal coverage, several limitations remain.

First, the dataset is constrained by the availability and quality of official recordings. Early decades rely on analog archive material with limited frequency range and higher background noise, which restricts certain acoustic analyses. Similarly, video coverage only begins in the mid-1990s, limiting multimodal investigations for earlier periods. In addition, changes in recording and broadcasting technology over the decades—including microphone systems, audio processing pipelines, and video production practices—may influence acoustic and visual feature extraction independently of speaker behavior. However, the textual analyses presented in this

study rely on official plenary transcripts, which remain comparatively stable across legislative periods.

Second, the corpus design emphasizes continuity and speaker longevity rather than population representativeness. Although the selection criteria ensure temporal balance, they inevitably favor political elites with long parliamentary careers. As a result, linguistic and prosodic patterns should primarily be interpreted as reflections of institutional parliamentary discourse rather than general spoken German.

Third, parliamentary speeches vary substantially in length, ranging from short interventions to extended policy statements. Although lexical diversity was measured using length-robust metrics such as MATTR, MTLD, and HD-D, unequal speech lengths may still influence aggregated statistics.

Fourth, thematic shifts in parliamentary debates across decades may influence lexical and rhetorical patterns independently of broader linguistic change.

Finally, some metadata gaps remain, particularly regarding missing speeches, incomplete transcripts, or evolving transcription standards across legislative terms, which may introduce minor inconsistencies in linguistic annotation. Future releases will continue to address these issues through automated validation routines and incremental improvements based on corpus maintenance and user feedback.

8. Conclusion & Outlook

The GerParlDia-MM constitutes the first multimodal diachronic resource of German parliamentary speech. By integrating seventy-five long-serving members of parliament and more than two thousand speeches from 1949 to 2025, it enables longitudinal analyses that reach beyond short-term variation and isolated case studies. The combination of aligned transcripts, audio, and video data allows researchers to investigate diachronic changes in linguistic style and rhetoric, and longitudinal developments in prosody and presentation associated with aging and institutional change.

The corpus opens new opportunities for applications across multiple research domains. In *speech technology*, it provides a benchmark for evaluating the long-term robustness of automatic speech recognition, speaker verification, and diarization systems under historically diverse recording conditions. In *linguistics and phonetics*, it enables the exploration of voice aging, charisma, and prosodic adaptation within a consistent communicative setting. In *computational linguistics and multimodal studies*, it supports analyses of diachronic lexical trends, gesture-prosody relations, and rhetorical framing. Beyond language technology, it also offers a longitudinal foundation for social and political sciences to examine representation, participation, and communication styles in the Bundestag over time.

Looking ahead, future work will focus on extending the corpus to upcoming legislative terms and refining its multimodal alignment. Additional annotations for prosody, gesture, emotion, and sentiment will be integrated, and automated pipelines for segmentation and quality assessment will be further optimized. By maintaining open access and transparent documentation, the corpus aims to serve as a living resource that evolves with the German parliamentary record and thus supporting reproducibility, interdisciplinary collaboration, and sustained research on the dynamics of spoken political discourse.

9. Availability

The *GerParlDia-MM* dataset and its documentation are publicly described at https://mobile-dialog-systeme.github.io/GerParlDia-MM/.

The release includes structured metadata, aligned transcripts, and persistent identifiers linking each speech to the corresponding audiovisual materials in the Bundestag Mediathek and the Parliamentary Archives.

Due to legal restrictions, audiovisual media files are not redistributed directly. Instead, stable identifiers and access URLs are provided to ensure reproducibility and verifiability of all referenced materials.

Versioned archives of the dataset metadata and documentation are deposited in a long-term repository (Zenodo). Persistent DOI references to these releases are listed on the project website to ensure sustainability and citability of the resource.

10. Acknowledgements

The staff of Referat ID 2 (Parlamentsarchiv des Deutschen Bundestages) is thanked for their kind support and helpful information regarding archival access and metadata structure. The efforts of student assistants and annotators involved in data preparation and validation are also gratefully acknowledged, in particular A.B., C.L., and R.B.

This research was funded by the BMFTR within the project Medinym (KI-basierte Anonymisierung personenbezogener Patientendaten in klinischen Text- und Sprachdatenbeständen), funded by the European Union – NextGenerationEU.

11. Bibliographical References

Zhiqi Ai, Meixuan Bao, Zhiyong Chen, Zhi Yang, Xinnuo Li, and Shugong Xu. 2025. VoxAging: Continuously Tracking Speaker Aging with a Large-Scale Longitudinal Dataset in English and Mandarin. In *Interspeech 2025*, pages 3648–3652.

Max Bain, Jaesung Huh, Tengda Han, Andrea Vedaldi, and Andrew Zisserman. 2023. Whisperx: Time-accurate speech transcription of long-form audio. *arXiv preprint arXiv:2303.00747*.

Andreas Blaette and Christoph Leonhardt. 2023. Germaparl corpus of plenary protocols (v2.0.1). Zenodo.

Deutscher Bundestag. 2021a. Archivordnung für den deutschen bundestag. https://www.bundestag.de/resource/blob/190272/archivordnung.pdf. Beschluss des Ältestenrates vom 21 March 2019, veröffentlicht 18 October 2021.

Deutscher Bundestag. 2021b. Nutzungsbedingungen für das audio- und videomaterial des parlamentsfernsehens. https://www.bundestag.de/resource/blob/296016/b2b8e3ed04b91bbfb235cfed975f1a69/nutzungsbedingungen_de-data.pdf. Parlamentsfernsehen, 25 March 2021.

Deutscher Bundestag. 2021c. Nutzungsordnung für das parlamentsarchiv. https://www.bundestag.de/resource/blob/190274/Nutzungsordnung.pdf. Anlage gemäß §3(10) der Archivordnung für den Deutschen Bundestag, veröffentlicht 18 October 2021.

Jonathan Harrington, Sallyanne Palethorpe, and Catherine Watson. 2000. Monophthongal vowel changes in received pronunciation: an acoustic analysis of the queen's christmas broadcasts. *Journal of the International Phonetic Association*, 30(1–2):63–78.

Finnian Kelly, Andrzej Drygajlo, and Naomi Harte. 2012. Speaker verification with long-term ageing data. In *5th IAPR International Conference on Biometrics (ICB)*, pages 478–483. IEEE.

Philip M. McCarthy and Scott Jarvis. 2010. Mtld, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2):381–392.

Gabriel Mittag, Babak Naderi, and Sebastian Möller. 2021. Nisqa: A deep cnn-self-attention model for multidimensional speech quality prediction with crowdsourced datasets. In *Proceedings of Interspeech*, pages 2127–2131.

Arsha Nagrani, Joon Son Chung, Weidi Xie, and Andrew Zisserman. 2020. Voxceleb: Large-scale speaker verification in the wild. *Computer Speech & Language*, 60:101027.

Florian Richter, Philipp Koch, Oliver Franke, Jakob Kraus, Fabrizio Kuruc, Anja Thiem, Judith Högerl, Stella Heine, and Konstantin Schöps. 2020. Open discourse. Harvard Dataverse.

Debjoy Saha, Shravan Nayak, and Timo Baumann. 2022. Merkel podcast corpus: A multimodal dataset compiled from 16 years of angela merkel's weekly video podcasts. *arXiv preprint arXiv:2205.12194*.

Ingo Siegert, Jannik Nietzold, Ralph Heinemann, and Andreas Wendemuth. 2019. The restaurant booking corpus – content-identical comparative human-human and human-computer simulated telephone conversations. In *Elektronische Sprachsignalverarbeitung 2019. Tagungsband der 30. Konferenz*, volume 93 of *Studientexte zur Sprachkommunikation*, pages 126–133, Dresden, Germany. TUDpress.

Ingo Siegert, Yamini Sinha, Oliver Jokisch, and Andreas Wendemuth. 2020. Recognition performance of selected speech recognition apis – a longitudinal study. In *Speech and Computer*, pages 520–529, Cham. Springer International Publishing.

Vishwanath Pratap Singh, Md Sahidullah, and Tomi Kinnunen. 2023. Speaker verification across ages: Investigating deep speaker embedding sensitivity to age mismatch in enrollment and test speech. In *Interspeech 2023*, pages 1948–1952.

Michael Tornow, Martin Krippl, Svea Bade, Angelina Thiers, Ingo Siegert, Sebastian Handrich, Julia Krüger, and Andreas Wendemuth Lutz Schega. 2016. Integrated health and fitness (igf)-corpus - ten-modal highly synchronized subject-dispositional and emotional human machine interactions. In *Multimodal Corpora: Computer vision and language processing (MMC 2016). Workshop Programme*, Portorož, Slovenia.

Ravi Chander Vipperla, Steve Renals, and Joe Frankel. 2010. Ageing voices: The effect of changes in voice parameters on asr performance. *EURASIP Journal on Audio, Speech, and Music Processing*.

Linlin Wang, Jun Wang, Lantian Li, Thomas Fang Zheng, and Frank K. Soong. 2016. Improving speaker verification performance against long-term speaker variability. *Speech Communication*, 79:14–29.

Haoning Wu, Erli Zhang, Liang Liao, Chaofeng Chen, Jingwen Hou Hou, Annan Wang, Wenxiu Sun Sun, Qiong Yan, and Weisi Lin. 2023. Exploring video quality assessment on user generated contents from aesthetic and technical perspectives. In *International Conference on Computer Vision (ICCV)*.