ASSESSING SPEAKING MODES IN RADIO NEWS USING TOPIC CLASSIFICATION AND ACOUSTIC PARAMETERS

Sven Grawunder[1,2]*, Ute Gradmann*[1]

[1]*Martin Luther University Halle-Wittenberg,* [2]*Max Planck Institute for Evolutionary Anthropology*

{ sven.grawunder|ute.gradmann } @sprechwiss.uni-halle.de

Abstract: This study investigates an approach to the automatic classification of speaking modes in German radio news. Based on a large, manually annotated corpus of broadcast news, we combine acoustic analysis, text-based methods, and positional metadata to characterize and distinguish news modes. Acoustic parameters, lexical distributions, and normalized segment positions are analyzed separately and jointly. A supervised machine learning model integrating these features achieves nearly 90% classification accuracy across most modes. An ablation study highlights the relative contribution of textual, acoustic, and positional features. The results demonstrate the viability of transparent, multimodal approaches for speaking-mode classification in broadcast news.

1 Introduction – Speaking Modes in News Speech

Research on broadcast news speech (radio and television) has revealed a high degree of differentiation in speaking modes, which affect speaking patterns in various ways depending on text type, topic, and underlying communicative tasks (e.g., announcing vs. reporting) [1]. These speaking modes also function as distinct ways of addressing the audience. They originate in concepts of journalism and media production, reflecting the perspectives of editors and presenters [2], as well as in the program format of the station (e.g. news radio, youth radio, adult contemporary, Top-40)[1] [3].

Consequently, the performance of texts within a news program ("show") is deliberately differentiated [4, cf. chaps. 1, 7, and 9]. Previous studies have demonstrated that parameters such as speaking rate and fundamental frequency (F0) vary systematically across speaking modes, particularly between news bulletins, transitional moderation, and reporting [1, 5, 6]. Thus, differences between modes associated with so-called *core news* (news bulletins) and *soft news* (e.g., sports, stock market, celebrities) can be expected. However, existing research on news speech has primarily focused on *core news*, a tendency that has also been observed for other languages [7] and media [6].

These findings underline the need for explicit labelling of speaking modes in order to facilitate analyses of their interaction with other media-production-related factors as well as speaker-specific characteristics [8]. To address this need, the present paper combines automatic topic classification with the assessment of phonetic parameters to establish a basis for further investigation of a transparent speaking-mode labelling approach. In addition, a structural component—namely the position of individual mode elements within the recorded news program—is taken into account. A central aim of this study is to preserve methodological transparency, particularly with respect to the algorithms employed, in order to enable a clear understanding of the decision pathways involved.

[1] https://radiodatacenter.net/en/glossary/formatradio/

2 Data – The News Arch

In the present study, only radio broadcast news are used for analysis. This type of audio data provides a controlled framework with consistent parameters, including pragmatic setting, communicative task, duration, content, professional speakers, and technical conditions [9, 10].

The systematic collection of German radio news from public broadcasting stations was initiated in 2003 [11] and is referred to as the *NachrichtenArche* (News Arch). Since then, annual recordings of public German radio stations have been conducted on November 11 at 1 p.m., ensuring continuity in the construction of a consistently annotated corpus. The subcorpus used in the present study is part of the *NachrichtenArche* and comprises 543 broadcast news recordings from the period 2003 to 2019.

Each recording in the News Arch has been manually segmented and annotated using Praat TextGrids [12]. The annotations cover linguistic categories, media-related information, and metadata such as time stamps and broadcast station. The annotation layers relevant for this study include speaker ID (containing information such as speaker gender), speaking mode, topic, orthographic transcription with segmentation into report units, and sentence segmentation. The *topic* category was assigned by annotators based on the content of each report (e.g., "Koalitionsverhandlungen").

The *mode* category is of central importance for the present study. Annotators selected one of the following categories: 1 = news bulletin (core news), 2 = report, 3 = weather, 4 = traffic news, 5 = news overview, 6 = stock market, 7 = jingle, 8 = original sound bite (verbatim report), 9 = sports, 10 = station identification, 31 = transition (intro), and 32 = transition (outro) [cf. 1, 8, 10].

For the purposes of this study, the material was segmented into report units (bulletins), resulting in a total of 6,454 items with durations ranging from 0.5 seconds to 4.5 minutes. Of these units, 3,442 were produced by male speakers, 2,487 by female speakers, and 525 by speakers of unspecified gender.

3 Methods – Three Approaches

The processing of the manually segmented units was carried out using the programming languages *Python* and *R* [13], the acoustic analysis software *Praat* [12], and its Python interface *Parselmouth* [14]. Text-based analyses further employed a pre-trained tool based on a large language model.

Acoustic analysis : The acoustic analysis comprised pause detection, fundamental frequency (F0) extraction, speaking rate, and articulation rate. Speaking rate and articulation rate were calculated using an adapted version of a widely used custom script [15]. As reliable speaker gender information was not consistently available at this stage, the median F0 and F0 range were measured in semitones.

The resulting acoustic measures were subjected to multivariate analysis using principal component analysis (PCA), implemented with the base `R` function `prcomp()`. The analysis was conducted under the assumption of seven meaningful clusters corresponding to speaking modes that reflect radio production factors as well as news topics [3].

Text-based analysis : The text-based analysis of different news modes consisted of two components: a lemmatization-based analysis of mode-specific lexical distributions and an approach employing a pre-trained language model. First, the orthographic transcripts were grouped according to speaking mode. Tokens were then lemmatized using the *spaCy* lemmatizer [16] and

counted, resulting in ranked type-frequency lists for each mode.

The second text-based approach aimed at automatically assigning segments to speaking modes and expanded the scope of the News Arch by incorporating data from the German pre-trained language model *gbert-large* [17].

Position-based analysis : The third, position-based approach builds on the findings of Simon and Altmeppen [18], who demonstrated in a comparative study of four radio stations that *core news* and *soft news* tend to occur at statistically different positions within broadcast news programs. Following this approach, each item in the present, more fine-grained mode framework was mapped onto a normalized position scale ranging from 0 to 10.

For each news recording, the sequence of segments was counted and normalized to this scale, and the position of each segment was defined relative to its occurrence within the program and its associated speaking mode.

4 Results

4.1 Acoustic Signal based Approach

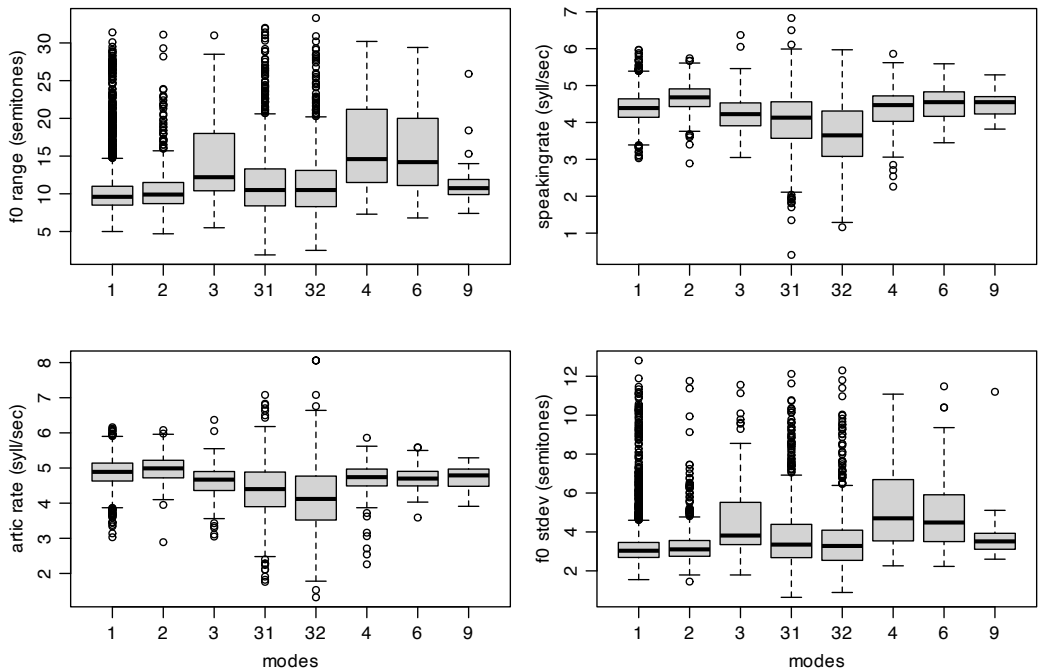

Figure 1 – Boxplots of acoustic signal parameters by speaking mode (1 = news bulletin, 2 = report, 3 = weather, 4 = traffic, 5 = overview, 6 = stock market, 7 = jingle, 8 = original sound bite, 9 = sports, 10 = station ID, 31 = transition (intro), and 32 = transition (outro))

Core news with modes 1 and 2 exhibit the lowest values for F0 standard deviation and F0 range (see Figure 1). *Soft news*, on the other hand, show a higher F0 range Accordingly, principal component 1 (PC1) is primarily loaded with F0-related parameters, along with higher probability scores for the topics of politics and weather. Principal component 2 (PC2) is mainly associated with speaking rate and articulation rate, as well as higher topic scores for *sports* and *economy*. In the PCA plot (Fig. 2 right), the relatively compact clustering of *core news* bulletins (mode 1) and reports (mode 2), corresponding to *politics* and *economy* (Fig. 2 left), stands in contrast to the more dispersed grouping of soft news items such as weather reports (mode 3) and stock market updates (mode 6).

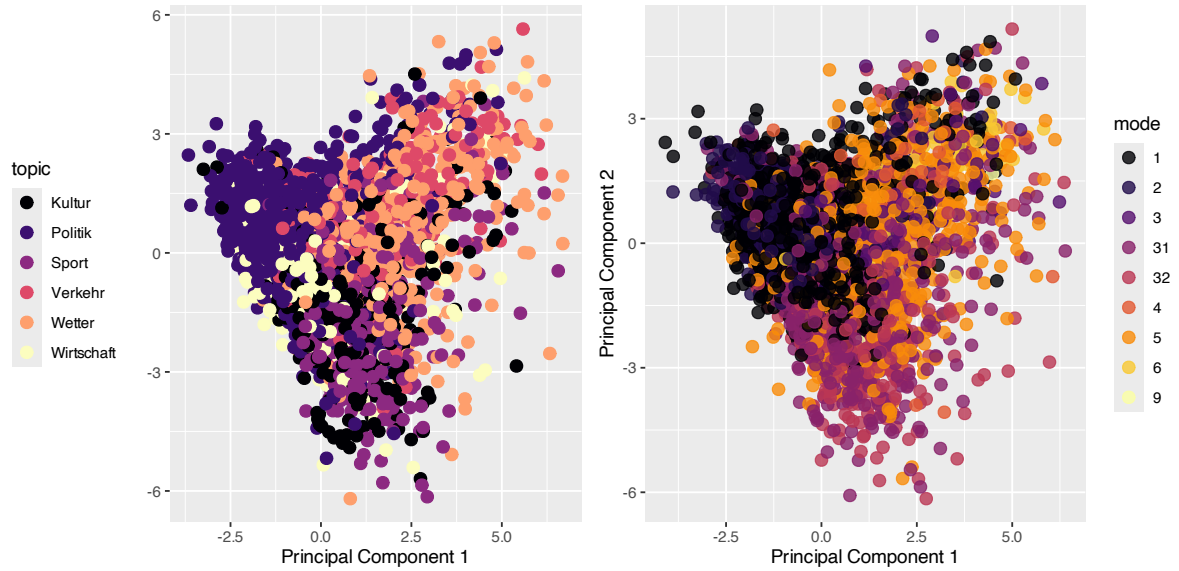

Figure 2 – PCA with max positioning topic class [culture, politics, sports, traffic, weather, economy] (left) ; PCA with all speaking modes (right)

4.2 Text-Based Approach

With regard to the first text-based analysis, we find that mode 1 (news / bulletins) contains 14,436 types, reports 10,326 types, weather 1,293 types, traffic news 1,286 types, news overview 1,766 types, interviews 1,469 types, sports 587 types, station indications 122 types, transition (intro) 2,264 types and transition (outro) 954. The type token ratio of the modes differs as well:

	Number of Types	Number of Tokens	Type Token Ratio
mode 1	14,436	158,937	0.09
mode 2	10,326	88,507	0.12
mode 3	1,293	11,049	0.12
mode 4	1,293	6,528	0,20
mode 5	1,766	7,424	0.24
mode 8	1,469	5,816	0.25
mode 9	587	1,337	0.43
mode 10	122	336	0.36
mode 31	2,264	9,275	0.24
mode 32	945	3,525	0.27

Table 1 – Type count, token count and type token ratio of modes

The ranking of some topic-specific types in table 2 illustrates the lexical differences of the news modes.

Table 2 illustrates the lexical differences marking the news modes. Exemplary lemmata were chosen and listed here, e.g. *wetter* and *grad* occur much more frequently in mode 3 (and 5) than in others, *stau* occurs more often in traffic news. First person pronouns are ranked much higher in mode 8 than in the others, which is in line with Grawunder and Kettel [5] and Burger [19, 103] who state that in reporting news the speaker's "I" is set aside while in interviews and original sound clips reporters do report from a subjective perspective.

For the second text-based approach, the pre-trained zero-shot classification tool *SVALabs - gbert large zeroshot nli* [20] was implemented assigning text-snippets to chosen categories. In this case, the 6,454 segments were assigned by the algorithm to the categories *politics*, *econ-*

mode:	1	2	3	4	5	8	9	10	31	32
type:	rank	rank	rank	rank	rank	rank	rank	rank	rank	rank
wetter	2003	1311	36	1166	**5**	430	-	12	29	41
grad	2761	1166	**4**	1172	**4**	394	-	-	-	-
fußball	3736	-	-	-	1022	-	463	-	658	-
stau	7001	-	947	**10**	-	-	-	70	416	162
ich	5696	249	186	94	913	**13**	-	-	112	77
wir	487	81	85	31	130	**9**	-	-	129	93
nachricht	1328	1191	270	1058	220	-	-	**2**	44	**3**

Table 2 – Examples of lemmata, illustrating the difference in frequency of occurrence per mode

omy, *culture*, *sports*, *weather* and *traffic* following the customary categories of news defined by Schwenke [3]. With the sum of probabilities constrained to $P = 1.0$, each item received a distribution across all six topics. The topic with the highest probability was taken as the assigned category, resulting in 6,111 classified items. By this method, text snippets were automatically classified using external data and the results will be related to the acoustic analysis later on.

4.3 Position-Based Approach

Figure 3 illustrates the distribution of speaking modes across normalized positions within the news recordings. The observed patterns corroborate previous findings by Simon and Altmeppen [18], showing that, following the news overview, new bulletins typically occur immediately after the station identification. These bulletins are commonly followed by reports, which frequently incorporate original sound bites. Subsequently, a transitional moderation often leads to segments such as stock market reports, traffic news, and, finally, weather forecasts.

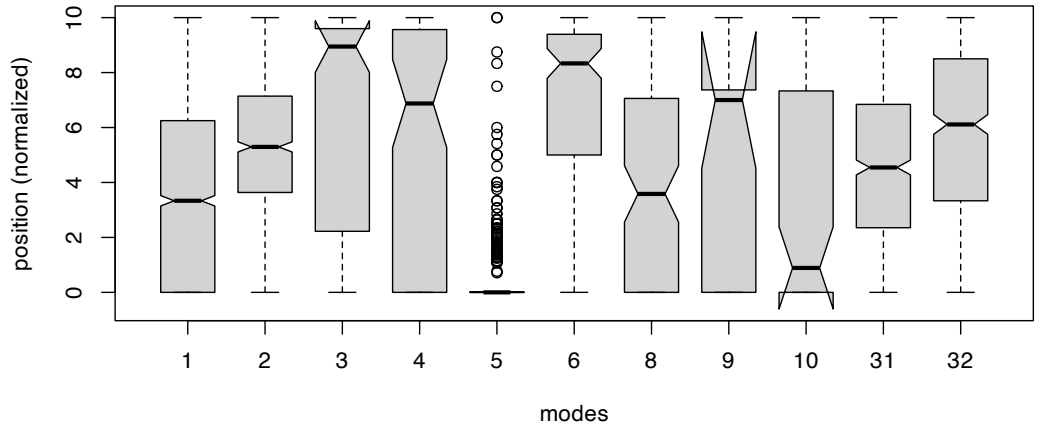

Figure 3 – Boxplots showing the (normalized) position of modes within a news recording

5 Discussion – Combining Approaches in a Machine Learning Classification Task

The results of the acoustic analysis corroborate previous findings based on a smaller subset of the News Arch [5, 21], showing lower speaking and articulation rates, as well as smaller F0 ranges for *core news* [6]. Speaking rate patterns were largely consistent with prior research [7, 22, 23], confirming that bulletin-style (*core*) news exhibits a more controlled and homogeneous

delivery, whereas the other text modes, such as traffic news and stock market, can range more freely.

Our analyses further demonstrate that news modes can be characterized and distinguished not only by acoustic parameters but also by textual features and metadata, such as segment position within a broadcast. This multimodal differentiation suggests that machine learning approaches can effectively leverage these complementary sources of information to predict news modes.

5.1 Combined Machine Learning Classification of Speaking Modes

To evaluate the predictive potential of these parameters, i.e. features, we implemented a supervised machine learning classification experiment. A test set of 10% of the items in each mode (603 items total) was held out, while the remaining 90% served as training data. Classification was performed using the Python packages *torch*, *scikit-learn*, and *sentenceTransformers*.

In three runs, 540, 541 and 537 items out of the 603 were predicted correctly reaching a score of 89.44%. Examinating the results in every mode category separately reveals a more nuanced picture, as - obviously - larger categories or very distinct categories like *weather* can be predicted more easily. Category 6 was so sparse that it was not included in the training experiment.

Mode	Number in training set	Masked in test set	Precision	Recall	Accuracy	F1
1	2457	272	0,94	0,96	0,95	0,95
2	810	90	0,94	0,88	0,97	0,91
3	258	28	0,96	0,93	1,00	0,95
4	66	7	1,00	0,57	1,00	0,73
5	605	67	0,86	0,96	0,98	0,91
8	107	11	0,89	0,73	0,99	0,80
9	20	2	0,00	0,00	0,99	-
10	54	6	0,83	0,83	1,00	0,83
31	622	69	0,74	0,71	0,94	0,73
32	462	501	0,78	0,78	0,96	0,78
Total:		**603**	**0,80**	**0,74**	**0,98**	**0,84**

Table 3 – Classification performance metrics per mode

5.2 Ablation Study

To assess the contribution of individual data types, we conducted an ablation study in which training omitted one feature set at a time: textual information, segment position, voice parameters (F0), or speaking and articulation rates. In all cases, predictive performance declined, with the largest drop observed when textual data were omitted (Table 4). These findings highlight the dominant role of lexical information in mode classification, while also confirming the complementary value of acoustic and structural features.

Omitted data:	Text	Position	Voice (F0)	Speakingrate
Predicted correctly (%):	75,12	86,73	88,56	87,40

Table 4 – Ablation study results: prediction of test set omitting training data

6 Conclusion & Outlook

The present study demonstrates that radio news speaking modes can be classified with high accuracy using a combination of acoustic, textual, and positional features. Our results indicate that a multimodal approach is viable for automatic speaking-mode detection and can complement traditional manual annotation methods.

For future work, several directions are promising. First, automatic segmentation that incorporates prosodic cues may further enhance classification performance [24]. Second, integrating additional contextual factors, such as broadcast profile (e.g., news radio, youth radio) or speaker-specific information including gender and number of speaker turns, could improve the granularity and interpretability of predictions. Finally, embedding this approach into broader workflows for the analysis of radio news development and evolution [10] could facilitate longitudinal studies and support applications in linguistic, phonetic or media research.

References

[1] GRAWUNDER, S.: *Die Erforschung des Sprechens mittels Nachrichtenkorpora: Die Nachrichtenarche der ARD*. In I. BOSE and D. SCHWIESAU (eds.), *Nachrichten schreiben, sprechen, hören: Forschungen zur Hörverständlichkeit und Radionachrichten*, pp. 147–155. Frank & Timme, Berlin, 2011.

[2] VON LA ROCHE, W. and A. BUCHHOLZ (eds.): *Radio-Journalismus: Ein Handbuch für Ausbildung und Praxis im Hörfunk*. Springer Fachmedien Wiesbaden, 2017. doi:10.1007/978-3-658-10796-3.

[3] SCHWENKE, A.: *Nachrichten: mehr als informativ: Nachrichten-Sprechstile im Radio zwischen Formatierung und Hörverständlichkeit*. Schriften zur Sprechwissenschaft und Phonetik, 20. Frank & Timme, Verlag für wissenschaftliche Literatur, 2020.

[4] BOSE, I. and D. SCHWIESAU (eds.): *Nachrichten schreiben, sprechen, hören: Forschungen zur Hörverständlichkeit von Radionachrichten*. Frank & Timme, Berlin, 2011.

[5] GRAWUNDER, S. and S. KETTEL: *Anmoderieren/Überleiten/Antexten. Was passiert zwischen Meldung und Bericht in Hörfunknachrichten? SPIEL : eine Zeitschrift zur Medienkultur. Neue Folge*, 1(1/2), pp. 151–170, 2015. doi:10.3726/80123_151.

[6] DIAS, T. E. C., P. C. MARTINS, L. C. TEIXEIRA, and A. C. C. GAMA: *Análise da variação prosódica em diferentes estilos de reportagens telejornalísticas*. *Audiology - Communication Research*, 20(3), pp. 210–214, 2015. doi:10.1590/2317-6431-2014-1528.

[7] RODERO ANTÓN, E.: *A comparative analysis of speech rate and perception in radio bulletins*. *Text & Talk. 2012 June 12; 32 (3): 391-411*, 2012. 10230/27802.

[8] DRAXLER, C., F. KLEBER, S. GRAWUNDER, and J. TROUVAIN: *Teilautomatisierter Workflow zur Aufbereitung großer Audiodatenmengen für signalbasierte Analysen*. In S. GRAWUNDER (ed.), *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2025*, pp. 138–145. TUDpress, Dresden, 2025. URL https://www.essv.de/pdf/2025_138_145.pdf.

[9] COTO-SOLANO, R.: *Computational sociophonetics using automatic speech recognition*. *Language and Linguistics Compass*, 16(9), p. e12474, 2022. doi:10.1111/lnc3.12474.

[10] KLEBER, F., C. DRAXLER, J. TROUVAIN, and S. GRAWUNDER: *Rundfunknachrichten als linguistisch-phonetische Ressource für die Sprachwandelforschung: Ein halbautomatisiertes Verfahren und eine Demosprachdatenbank*. In A. BRUNNER, S. HANSEN, C. LANG, N. D. T.

TU, and S. WOLFER (eds.), *Deutsch Im Wandel – Tools, Methoden, Ressourcen. Beiträge Zur Methodenmesse Der IDS-Jahrestagung 1.*, vol. 16 of *IDSopen*. IDS-Verlag, Mannheim, 2026. doi:10.21248/idsopen.16.2026.67.

[11] SCHWIESAU, D., S. GRAWUNDER, and I. BOSE: *Die Nachrichtenarche der ARD*. In I. BOSE and D. SCHWIESAU (eds.), *Nachrichten schreiben, sprechen, hören: Forschungen zur Hörverständlichkeit von Radionachrichten*, pp. 147–155. Frank & Timme, Berlin, 2011.

[12] BOERSMA, P. and D. WEENINK: *Praat: doing phonetics by computer [computer program]*. 2024. Retrieved 27 October 2024 from `http://www.praat.org/`.

[13] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL `https://www.R-project.org/`.

[14] JADOUL, Y., B. THOMPSON, and B. DE BOER: *Introducing parselmouth: A python interface to praat*. *Journal of Phonetics*, 71, pp. 1–15, 2018. doi:10.1016/j.wocn.2018.07.001.

[15] DE JONG, N. H. and T. WEMPE: *Praat script to detect syllable nuclei and measure speech rate automatically*. 41(2), pp. 385–390, 2009. doi:10.3758/BRM.41.2.385.

[16] HONNIBAL, M. and I. MONTANI: *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*, 2017.

[17] CHAN, B., S. SCHWETER, and T. MÖLLER: *German's next language model*. In D. SCOTT, N. BEL, and C. ZONG (eds.), *Proc. of the 28th Intern. Conf. on Computational Linguistics*, pp. 6788–6796. Inter. Committee on Comp. Ling., Barcelona, Spain (Online), 2020. doi:10.18653/v1/2020.coling-main.598.

[18] SIMON, B., YVES and K.-D. ALTMEPPEN: *Hörfunknachrichten in Deutschland*. Nomos, Baden-Baden, 1 edn., 2016.

[19] BURGER, H.: *Sprache der Massenmedien*. De Gruyter, Berlin, Boston, 1984. doi:10.1515/9783110855913.

[20] WOCHATZ, N.: *Svalabs - gbert large zeroshot nli*. 2021. URL `https://huggingface.co/svalabs/gbert-large-zeroshot-nli`.

[21] GASSER, E., B. AHN, D. J. NAPOLI, and Z. ZHOU: *Production, perception, and communicative goals of American newscaster speech*. *Language in Society*, 48(2), pp. 233–259, 2019. doi:10.1017/S0047404518001392.

[22] RODERO, E.: *The Principle of Distinctive and Contrastive Coherence of Prosody in Radio News: An Analysis of Perception and Recognition*. *Journal of Nonverbal Behavior*, 39(1), pp. 79–92, 2015. doi:10.1007/s10919-014-0201-5.

[23] FINKELSTEIN, M. and O. AMIR: *Speaking Rate among Professional Radio Newscasters: Hebrew Speakers*. *Studies in Media and Communication*, 1(1), pp. p131–139, 2013. doi:10.11114/smc.v1i1.119.

[24] LEVOW, G.-A.: *Prosody-based topic segmentation for Mandarin broadcast news*. In *Proceedings of HLT-NAACL 2004: Short Papers*, pp. 137–140. Association for Computational Linguistics, Boston, Massachusetts, USA, 2004. URL `https://aclanthology.org/N04-4035/`.