Who Speaks How? Automatic Dialect Intensity Trajectories from Audio Data

Anonymous ACL submission

Abstract

Dialect intensity analyses are commonly based on aggregated recording-level representations, thereby ignoring temporal variation within speaking situations. In this work, we investigate whether dialect intensity can be represented reliably as a trajectory within running speech using DIME, an embedding-based dialect intensity measure. We analyze the influence of window length and overlap on trajectory stability and show a clear trade-off between temporal resolution and structural reliability. While shorter segments increase temporal granularity, they also introduce higher local variability and reduced trajectory coherence. At the same time, the trajectories remain globally stable across segmentation settings, demonstrating that temporally resolved dialect intensity modeling is feasible. Building on these findings, we compare classical aggregated speaker type classification with a trajectory-based approach using recordings from the REDE corpus. The temporal analyses substantially alter the resulting speaker classifications and show that aggregated analyses systematically overestimate categorical switching behavior.

1 Introduction

Since the groundbreaking studies of Labov (1972), situational variation and style shifting have long been central topics in sociolinguistics. Speakers continuously adapt their speech behavior depending on communicative context, audience, and interactional setting (Bell, 1984). In parallel, quantitative approaches in dialectometry have increasingly modeled dialect variation through continuous distance-based measures (Nerbonne and Heeringa, 1997). While such variation has traditionally been investigated in larger units of discourse (sentences, texts) using aggregated linguistic metrics and average values, recent computational approaches increasingly allow continuous quantitative modeling of dialect intensity and language variation as speech unfolds.

Previous work in German variational linguistics distinguishes speakers according to their situational use of dialectal and standard-oriented varieties (Macha, 1991; Lausberg, 1993; Lenz, 2003). Building on this perspective, Lenz (2003) distinguishes between speakers exhibiting categorical switching behavior, gradual shifting, or little situational variation. These speaker types were later operationalized quantitatively by Lameli (2025).

Since these speaker types are primarily defined through observable situational variation, they are particularly compatible with continuous dialect intensity measures such as DIME (Anonym, 2026a). However, traditional speaker type assignment typically relies on a single aggregated value per speaking situation and therefore ignores temporal variation within situations (Lameli, 2026). As a result, aggregated recording-level analyses may obscure dynamic realization patterns and distort speaker type interpretation. The temporally resolved DIME trajectories make it possible to revisit these speaker types—for the first time—from a dynamic perspective.

Previous work demonstrated that DIME correlates with established phonetic reference measures and can quantify dialect intensity relative to Standard German using speech embeddings extracted from a dialect classification model (Anonym, 2026a,b). However, the original approach relied on aggregated recording-level statistics and therefore did not investigate temporal variation within recordings.

In this work, we investigate whether dialect intensity can be modeled reliably as a temporally evolving trajectory within running speech using DIME. More specifically, we address three research questions: (RQ1) How do window length and overlap affect the stability and validity of DIME trajectories? (RQ2) To what extent does temporal

resolution change speaker type assignment compared to classical aggregated approaches? (RQ3) Do temporally resolved trajectories reveal speaker profiles not captured by classical categorization? To answer these questions, we first analyze the influence of window length and overlap on trajectory stability and interpretability before comparing classical aggregated speaker type assignment with a temporally resolved trajectory-based approach.

2 DIME

Building on prior work, we use the Dialect Distance Measure from Classification Embeddings (DIME) introduced in Anonym (2026a) as a continuous measure of dialect intensity. The previous study demonstrated that DIME correlates with an established independent phonetic reference and can quantify dialectal distance relative to an external standard reference corpus. In the original setup, audio recordings were segmented into non-overlapping 10-second windows, from which segment-level speech embeddings were extracted using a trained dialect classification model. Instead of using the final classification output, activations of the last hidden layer are used as continuous speech representations. DIME combines the absolute distance to a standard reference centroid with projection onto a learned dialect–standard direction. By computing the score independently for short audio segments, DIME enables temporally resolved analyses of dialect intensity within recordings rather than restricting the analysis to a single aggregated value per recording.

3 Data

3.1 Speech Data

The speech data used in this study originates from the REDE project (Schmidt et al., 2020–; Fischer and Lameli, 2026), which contains recordings from multiple speaking situations eliciting different degrees of dialectal and standard-oriented speech production (Lameli, 2025). We focus on three recording conditions differing systematically in dialect intensity.

The first condition consists of Wenker sentences in dialect (WSD). In this setting, speakers are presented with the so-called Wenker sentences (Wenker, 2013) in Standard German and are instructed to translate them into their local dialect. This condition therefore represents explicitly dialect-intended speech. The second condition comprises Wenker sentences in Standard German (WSS). Here, speakers hear dialectal versions of the Wenker sentences in their local dialect and are asked to translate them into their most formal Standard German. Consequently, this condition represents standard-intended speech. The third condition consists of free conversations among friends (FG), in which speakers engage in spontaneous dialogue with a familiar conversation partner without direct interviewer interaction. Unlike WSD and WSS, FG is not explicitly guided towards either dialectal or standard-oriented production and is therefore expected to occupy an intermediate position between both extremes.

3.2 Standard Reference

Since DIME quantifies dialect intensity relative to a chosen reference corpus, selecting an appropriate standard reference is central. In order to obtain a reference that approximates Standard German as closely as possible, we use recordings from the German television news program *Tagesschau*. The corpus consists of recordings from three speakers covering broadcasts between 2018 and 2022. Tagesschau speakers are professionally trained and operate within a highly standardized broadcast setting. As such, they are expected to adhere closely to codified pronunciation norms (Krech et al., 2009). Their speech is therefore commonly used as an empirical approximation of spoken Standard German (Torres and Müller, 2021). The resulting embeddings are used to construct the standard reference centroid and the dialect—standard direction underlying the DIME score.

3.3 Dialect Intensity Reference Measure

As an external reference measure of dialect intensity, we use the *D-value* provided within the REDE project (see Herrgen and Schmidt (1989), Herrgen et al. (2001)), following Anonym (2026a). The D-value is a linguistically motivated measure that quantifies the phonetic distance of a speech sample to codified Standard German based on manually annotated phonetic transcriptions, with higher values indicating higher dialect intensity.

However, an important limitation for the present study is that D-values are only available as a single global value per speaker and recording condition. Consequently, no temporally resolved reference measure exists that would allow direct evaluation of local dialectal variation within recordings. While correlations with aggregated DIME scores

are therefore still considered, the temporal analysis in this work additionally relies on other characteristics of the extracted trajectories, such as their stability and variation across and within speaking situations (see subsection 4.2).

4 Methods & Segmentation Strategy

4.1 Window Lengths & Overlap

In the original DIME setup, recordings were segmented into non-overlapping windows of 10 seconds. While this window size provides comparatively stable segment-level representations, it also limits the temporal resolution of the resulting dialect intensity trajectories. Since the present work focuses on the temporal development of dialect intensity within recordings, we additionally investigate shorter window lengths in order to capture more localized variation in speech behavior.

At the same time, reducing the window size may lead to less stable embeddings and therefore noisier DIME trajectories due to the smaller amount of acoustic information available per segment. This introduces a trade-off between temporal resolution and robustness of the resulting dialect intensity estimates.

In addition to varying the window length, we also investigate the effect of overlapping windows. Using overlap increases the temporal density of the extracted trajectories and may act as a smoothing mechanism by reducing abrupt changes between adjacent segments.

4.2 Measures

To evaluate the influence of window length and overlap on temporally resolved DIME trajectories, we consider several complementary measures capturing (i) agreement with previous DIME results, (ii) local trajectory stability, (iii) similarity to the original 10-second setup, and (iv) variation across speakers and recording conditions.

Correlation with D-value. To ensure comparability with previous work, we continue to evaluate the relationship between DIME and the D-value. Segment-level DIME scores are aggregated per speaker and recording condition using the median, resulting in one score per speaker $\times$ situation cell. Pearson correlation with the corresponding D-values is then computed.

WSS–WSD Separation. In addition, we report the area under the ROC curve (AUC) for distinguishing WSS and WSD segments. This measure evaluates whether segment-level DIME scores continue to separate standard-intended from dialect-intended speech under different segmentation strategies.

Local Trajectory Variability. To quantify local variation within recordings, we compute the standard deviation of segment-level DIME scores separately for each speaker $\times$ situation cell. We then report the weighted median of these standard deviations, where cells containing more segments receive larger weights. Increased values indicate increased local variation, potentially caused by either genuine temporal dynamics or reduced embedding stability for shorter segments.

Standard Error of Aggregated Scores. While local trajectory variability captures the spread of segment-level scores, the standard error of the mean (SEM) estimates the uncertainty of the aggregated speaker-condition score. For each speaker $\times$ situation cell, SEM is computed as $SEM = \frac{SD}{\sqrt{n}}$, where SD denotes the within-cell standard deviation and n the number of segments.

Temporal Smoothness. To assess the continuity of DIME trajectories, we compute the Pearson correlation between adjacent segment scores within recordings.

Agreement with the 10-Second Baseline. We additionally compare all segmentation strategies against the original non-overlapping 10-second setup using Pearson correlation between scores and mean absolute score differences.

Mixed-Effects Analysis. Finally, we fit linear mixed-effects models of the form

$$\text{DIME} \sim C(\text{situation}) + (1 \mid \text{speaker})$$

where recording situation (C) is treated as a fixed effect and speaker as a random intercept. From these models, we report the estimated contrast between dialect-intended speech and FG as well as between FG and standard-intended speech. In addition, we report the estimated speaker-level variance and residual variance, capturing between-speaker variability and remaining unexplained variation, respectively.

4.3 Speaker Type Definition

Following Lenz (2003) and Lameli (2025), we distinguish between the speaker types *Switcher*, *Shifter*, and *Moveless*. Figure 1 illustrates the conceptual distinction between these speaker profiles.

Only speakers with at least 40 seconds of available speech material in both WSS and WSD, as

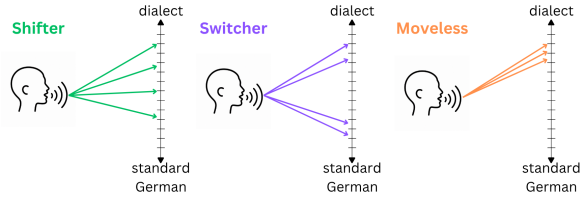

Figure 1: Conceptual illustration of the speaker types *Shifter*, *Switcher*, and *Moveless*. Shifters realize intermediate dialect intensity levels between dialectal and standard-oriented speech, whereas Switchers primarily alternate between distinct poles. Moveless speakers exhibit little variation.

well as at least 80 seconds in FG, were considered for speaker type assignment. This constraint ensures sufficient temporal coverage for stable trajectory estimation across all speaking situations.

Classical Speaker Type Assignment. Traditionally, dialect intensity is represented by a single aggregated value per speaker and recording situation (e.g. Lameli (2025)). Following this setup, speaker types are initially assigned based on the median DIME score of each speaker in the three recording situations WSD, WSS, and FG.

Speakers are classified as *Moveless* if all pairwise distances between the three situation-level medians remain smaller than the global standard deviation of all speaker-condition medians. These speakers therefore exhibit little overall situational variation.

For the remaining speakers, the two situations with the largest distance are identified. Speakers are classified as *Switcher* if the remaining distances are smaller than the global standard deviation. In this case, speakers exhibit two distinct poles corresponding to dialectal and standard-oriented speech without substantial intermediate realizations.

All remaining speakers are classified as *Shifter*. Unlike Switchers, these speakers exhibit more gradual transitions between dialectal and standard-oriented realizations and therefore occupy intermediate positions between both poles.

Temporal Speaker Type Assignment. While the classical approach relies on one aggregated value per recording situation, DIME provides temporally resolved dialect intensity trajectories based on segment-level scores and thereby enables speaker behavior to be analyzed within recording situations rather than only across them.

To estimate speaker-specific dialectal and standard-oriented poles while reducing the influence of extreme outliers, reference ranges are derived from the elicited WSD and WSS conditions. For WSS, the 10% and 50% quantiles are computed, representing highly standard-oriented and typical standard-oriented realizations for the speaker at hand. Analogously, for WSD, the 90% and 50% quantiles are computed, representing highly dialectal and typical dialectal realizations.

Based on these speaker-specific poles, each FG segment is assigned to one of three regions: a standard-oriented pole if the score falls below the WSS median, a dialect-oriented pole if the score exceeds the WSD median, and an intermediate region otherwise. Intermediate segments occurring between two neighboring pole realizations are treated as transitional segments rather than stable intermediate realizations.

To quantify the degree of polarization within FG, we compute the proportion of polar realizations relative to all non-transitional realizations:

$$P_{\mathrm{FG}} = \frac{p_{\mathrm{polar}}}{p_{\mathrm{polar}} + p_{\mathrm{middle}}}$$

where p_{polar} denotes the proportion of segments assigned to either pole and p_{middle} the proportion of stable intermediate realizations. High values therefore indicate trajectories dominated by realizations close to the dialectal and standard-oriented poles, whereas lower values indicate more continuous intermediate behavior. Figure 2 illustrates the assignment of FG segments to pole-oriented, stable intermediate, and transitional realizations.

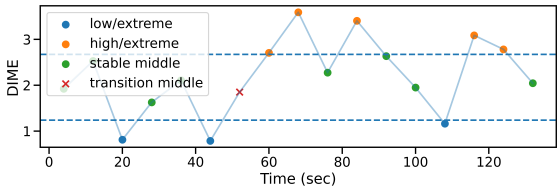

Figure 2: Illustration of the FG polarization measure. Segments are assigned to dialect-oriented and standard-oriented poles, stable intermediate realizations, or transitional intermediate realizations occurring between opposite poles.

Speakers are classified as *Moveless* if the speaker-specific span between the 90% quantile of WSD and the 10% quantile of WSS remains smaller than the lower quartile of all observed speaker spans. These speakers therefore exhibit only minimal overall dialectal range.

Among the remaining speakers, those with

strong FG polarization ($P_{\mathrm{FG}} \geq 0.8$) and realizations near both poles are classified as *Switcher*. Speakers with weaker FG polarization ($P_{\mathrm{FG}} < 0.8$) are classified as *Shifter*, indicating more gradual movement along the dialect intensity continuum. Finally, speakers with strong polarization but realizations concentrated around only one pole are assigned to a hybrid *Shifter/Switcher* category, since their trajectories do not allow a clear distinction between categorical switching and continuous shifting behavior.

Since no established temporal speaker type criteria exist, the selected thresholds are intended as heuristic operationalizations for exploring temporal realization patterns rather than as theoretically fixed category boundaries. Additional sensitivity analyses with alternative FG polarization thresholds ($P_{\mathrm{FG}} = 0.7$ and 0.9; Appendix A) yielded highly similar overall transition patterns, particularly the strong shift from classical *Switcher* to temporal *Shifter* behavior.

5 Experiments

5.1 Window Size and Overlap

The results in Table 1 show a clear influence of segment length on the stability and interpretability of the extracted dialect intensity trajectories. Overall correlations with the D-value remain remarkably stable for segment lengths between 10 and 3 seconds ($r \approx .65$–$.67$), with only the 1-second condition showing a noticeable decline ($r = .582$). This suggests that globally aggregated DIME scores remain largely robust even under substantially shorter segmentation strategies. A likely explanation is that aggregation at the speaker-condition level compensates for part of the increased local variability introduced by shorter segments.

In contrast, measures targeting local trajectory structure reveal a substantially different picture. The separation between WSS and WSD segments continuously decreases with shorter segment lengths, dropping from an AUC of $.804$ at 10 seconds to $.588$ at 1 second. Similarly, local trajectory variability increases strongly as segment length decreases, rising from $.731$ to values above 1.4. This increase may reflect both genuine short-term dialectal variation and reduced embedding stability caused by the smaller amount of acoustic information available per segment. In contrast, SEM decreases for shorter segment lengths due to the larger number of segment-level observations. However, this reduction does not indicate improved trajectory quality. Instead, it primarily reflects the increasing number of segment-level observations produced under shorter segmentation strategies. The divergence between decreasing SEM and increasing local variability highlights that a larger number of observations does not necessarily imply more reliable dialect intensity trajectories.

The mixed-effects analysis further supports this interpretation. The estimated contrast between dialect-intended speech and FG decreases gradually with shorter segment lengths. This effect is particularly pronounced for the contrast between FG and standard-intended speech, which nearly disappears for 1-second segments. This near-complete collapse of the FG–Standard contrast suggests that 1-second segments no longer reliably capture supra-segmental phonetic information distinguishing standard-oriented from spontaneous speech. At this temporal granularity, the embedding space may primarily encode local articulatory variation while losing broader prosodic and rhythmic patterns associated with style shifting. The results therefore suggest a practical lower bound for reliable DIME trajectory extraction. At the same time, speaker-level variance decreases substantially, whereas residual variance increases strongly. Together, these findings suggest that speaker-specific dialectal structure becomes increasingly obscured by local variation and instability at shorter segment lengths.

A similar pattern is observed for temporal smoothness. Correlations between adjacent segment scores decrease steadily as segment length decreases, indicating that the resulting trajectories become temporally less coherent and exhibit increasingly abrupt local fluctuations. Importantly, the overlap condition (8-second windows with 4-second overlap) substantially improves temporal smoothness while preserving nearly identical global correlations and separation performance compared to the non-overlapping setup. This suggests that overlap acts as an effective temporal smoothing mechanism without substantially altering the overall dialect intensity structure.

Finally, comparison with the original 10-second setup shows that segment length down to approximately 3 seconds remains highly consistent with the baseline configuration ($r > .96$). Only the 1-second condition exhibits a substantial decline in baseline agreement together with markedly larger absolute score deviations, indicating a structural drift of the extracted dialect intensity measure.

	10s	9s	8s	8s (4s ov.)	5s	3s	1s
Overall Correlation	.657***	.665***	.653***	.661***	.658***	.666***	.582***
Separation (AUC)	.804	.800	.794	.794	.760	.721	.588
Local Trajectory Variability	.731	.763	.833	.834	1.107	1.478	1.256
SEM	.201	.201	.205	.148	.215	.223	.109
Dialect–FG Contrast	.480	.478	.475	.471	.439	.401	.372
FG–Standard Contrast	.973	.973	.976	.979	.979	1.010	.046
Speaker Variance	.666	.665	.657	.666	.629	.615	.160
Residual Variance	.788	.859	.954	.949	1.504	2.452	1.633
Temporal Smoothness	.683***	.656***	.617***	.755***	.476***	.345***	.129***
Baseline Agreement	–	.982***	.983***	.988***	.979***	.966***	.657***
Baseline Deviation	–	0.128	0.148	0.124	0.328	0.712	0.874

Table 1: Evaluation of different window lengths and overlap configurations for DIME extraction. Reported metrics assess correlation with the D-value, discrimination between speaking conditions, temporal stability, and agreement with the original non-overlapping 10-second setup. Correlations marked with *** indicate $p < .001$.

Overall, the results reveal a clear trade-off between temporal resolution and structural stability. Shorter segments increase temporal granularity but also introduce higher local variability and reduced trajectory coherence. The most stable balance between temporal resolution and dialectal structure is observed for segment lengths between approximately 5 and 10 seconds, with overlapping windows providing additional temporal smoothing.

Importantly, the results further show that temporally resolved dialect intensity modeling is feasible at all. Despite the increased temporal granularity, DIME trajectories remain structurally stable across a wide range of segmentation settings and preserve meaningful distinctions between speaking situations. This suggests that dialect intensity can be modeled not only as a global property of recordings, but also as a dynamically evolving characteristic within running speech.

Based on these results, we use non-overlapping 8-second segments for the subsequent analyses. This setting preserves the global validity of the original 10-second setup while providing slightly higher temporal resolution. In contrast to the overlapping configuration, it avoids introducing additional dependency between adjacent observations.

5.2 Speaker Statistics

After applying the minimum duration constraints of 40 seconds for WSS and WSD as well as 80 seconds for FG, 307 out of 556 speakers remained for the speaker type analyses. The thresholds were chosen to ensure sufficient segment-level coverage under the selected 8-second segmentation setup, including at least five segments for WSS and WSD as well as sufficiently long FG trajectories (10 segments) for temporal analysis.

We measured the Mean DIME Scores per recording situation. As expected based on the elicitation design, WSS exhibits the lowest dialect intensity scores (0.834 ± 1.141) and therefore lies closest to the Tagesschau reference centroid, whereas WSD shows the highest dialect intensity values (2.288 ± 1.403). FG occupies an intermediate position between both extremes (1.821 ± 1.247). This pattern is consistent with the intended recording conditions, where WSS targets standard-oriented speech, WSD dialect-oriented speech, and FG spontaneous unguided interaction.

FG additionally exhibits both the largest number of segments and the highest variance (27.759 ± 15.841) (WSD 20.508 ± 5.695; WSS 20.782 ± 6.525). This can largely be attributed to the conversational nature of the recordings, which vary substantially in duration and allow for greater situational flexibility compared to the fixed Wenker sentence tasks.

5.3 Classical Speaker Type Analysis

Using the classical aggregated speaker representation, the global standard deviation of all speaker-condition medians of DIME was $\sigma = 1.13$. Based on this threshold, 226 speakers were classified as *Switcher*, 68 as *Moveless*, and only 13 as *Shifter*.

The resulting distribution suggests that the majority of speakers appear to alternate between distinct dialectal and standard-oriented poles when only aggregated situation-level medians are considered. Intermediate realizations are comparatively rare under this perspective.

However, the classical approach only considers one single representative value per recording situation and therefore ignores potential temporal structure within recordings. In particular, speakers

whose FG realizations fluctuate around an intermediate range may exhibit similar aggregated medians as speakers alternating categorically between dialectal and standard-oriented poles. Consequently, aggregated recording-level statistics systematically obscure temporal realization patterns.

5.4 Temporal Speaker Type Analysis

For the temporally resolved speaker classification, the lower quartile of all observed speaker spans was 2.665 and used as the threshold for identifying *Moveless* speakers. Figure 3 visualizes the distribution of speakers with respect to speaker span and FG polarization together with the selected decision thresholds.

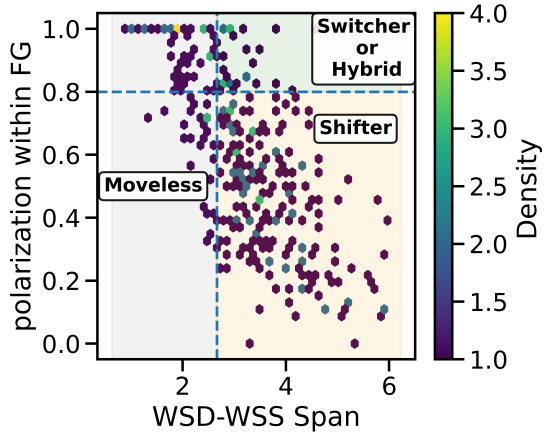

Figure 3: Distribution of speakers with respect to speaker span and FG polarization. Dashed lines indicate the thresholds used for temporal speaker type assignment.

In contrast to the classical aggregated analysis, the temporal analysis yields substantially different speaker type distributions. Most speakers are now classified as *Shifter* (200 speakers), followed by 77 *Moveless* speakers, 21 hybrid *Shifter/Switcher* speakers, and only 9 *Switcher* speakers.

Figure 4 illustrates a representative hybrid *Shifter/Switcher* trajectory. While the speaker clearly demonstrates both dialectal and standard-oriented competence in WSD and WSS, the FG trajectory remains concentrated near only one pole. Consequently, the temporal trajectory of FG alone does not allow determining whether intermediate realizations are part of the actively used repertoire or whether the speaker would exhibit categorical switching behavior under different interactional conditions.

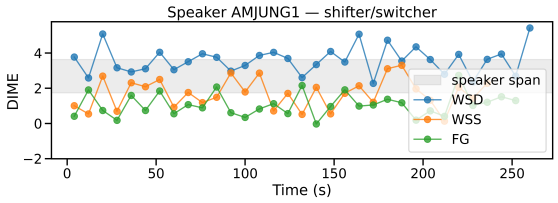

Figure 4: Example of a hybrid *Shifter/Switcher* speaker.

5.5 Speaker Type Transitions

Table 2 summarizes the transition between classical and temporal speaker type assignment.

		Temporal			
		Movel.	Shift.	Hybrid	Switch.
Classical	Moveless	43	17	1	7
	Shifter	0	10	3	0
	Switcher	34	173	17	2

Table 2: Transition matrix between classical and temporal speaker type assignment.

Only 55 of the 307 speakers retained the same speaker type under both approaches, whereas 252 speakers changed category. Overall agreement between both approaches was low ($\kappa = .083$), indicating that temporally resolved analysis alters the resulting speaker type interpretation. The most prominent transition occurs from classical *Switcher* to temporal *Shifter*, affecting 173 speakers.

This shift can largely be explained by the granularity of both approaches. The classical assignment is based solely on aggregated medians per recording situation and therefore only captures the global position of FG relative to WSS and WSD. Consequently, speakers exhibiting stable intermediate realizations within FG may obtain similar aggregated medians as speakers alternating categorically between dialectal and standard-oriented poles.

The temporally resolved trajectories reveal that many speakers previously classified as *Switcher* actually exhibit gradual and relatively stable intermediate realizations characteristic of *Shifter* behavior. The temporal analysis therefore suggests that aggregated recording-level statistics substantially overestimate categorical switching behavior by collapsing distinct temporal realization patterns into similar global representations.

6 Discussion

The results demonstrate that temporally resolved dialect intensity analysis provides more detailed

insights into speaker behavior than aggregated recording-level statistics. While no temporally resolved external ground truth is available, the trajectory structure is supported by several internal diagnostics. Global correlations with the D-value remain comparatively stable across most segment lengths, speaker-condition contrasts persist across segmentation settings, and overlapping and non-overlapping 8-second trajectories yield highly similar global structures. In addition, adjacent-segment correlations remain meaningful except for the 1-second condition. Together, these diagnostics suggest that the extracted trajectories are not merely segmentation artifacts, but reflect a stable latent dialect intensity structure.

At the same time, the experiments reveal a clear trade-off between temporal resolution and trajectory stability. Larger windows provide smoother and more stable trajectories but obscure short-term dialectal dynamics, whereas shorter windows increase temporal granularity at the cost of higher local variability and instability. Segment lengths between approximately 5 and 10 seconds provided the most robust balance between both aspects. In addition, overlapping windows improved temporal smoothness without altering the overall dialect intensity structure, indicating that overlap can function as a useful smoothing mechanism.

Beyond the methodological findings, the temporal analyses also change the resulting speaker typology. Under the classical aggregated approach, most speakers are classified as *Switcher*. However, temporally resolved trajectories reveal that many of these speakers actually exhibit relatively stable intermediate realizations characteristic of *Shifter* behavior. This discrepancy results largely from the fact that aggregated recording-level statistics collapse different temporal realization patterns into similar global summary values. Consequently, the distinction between *Shifter* and *Switcher* proposed in earlier variational linguistic work (Lenz, 2003) appears to depend strongly on the temporal granularity at which speaker behavior is analyzed. At the same time, the sensitivity of the classification to the selected thresholds and the emergence of hybrid speaker profiles suggest that these speaker types are better interpreted as regions along a continuum of situational variation rather than as strictly separated categories. The trajectories themselves may therefore be more informative than the resulting categorical assignments.

Finally, the introduction of the hybrid *Shifter/Switcher* category highlights that linguistic competence and realized speech behavior do not necessarily coincide. Some speakers clearly demonstrate the ability to realize both dialectal and standard-oriented poles in WSD and WSS, while nevertheless remaining concentrated near only one pole during spontaneous interaction in FG. In such cases, it remains unclear whether intermediate realizations are actively available but simply not used under the observed interactional conditions, or whether the speaker would exhibit more categorical switching behavior in different communicative settings. While the present work introduces this hybrid category primarily in the context of FG, similar mismatches between linguistic competence and realized behavior may also occur across other speaking situations and interactional contexts.

7 Conclusion

In this work, we investigated temporally resolved dialect intensity trajectories derived from the DIME measure. The results showed a clear trade-off between temporal resolution and trajectory stability. Shorter segments increase temporal granularity but also introduce more local variability and instability, whereas larger segments provide more stable representations at the cost of temporal detail. Segment lengths between approximately 5 and 10 seconds provided the best balance between both aspects, while overlapping windows improved temporal smoothness.

The experiments further showed that temporally resolved analyses change the interpretation of speaker behavior compared to classical recording-level approaches. The results further demonstrate that aggregated recording-level analyses systematically overestimate categorical switching behavior by collapsing distinct temporal realization patterns into similar global speaker representations.

Overall, the results demonstrate that temporally resolved dialect intensity trajectories provide a more informative representation of speaker behavior than single aggregated recording-level statistics alone. Rather than reducing speakers to fixed categories, temporally resolved trajectories reveal dialect variation as a dynamically evolving structure within interaction.

Limitations

First, no temporally resolved external reference measure of dialect intensity is currently available. While the D-value provides an established global reference at the recording level, it does not allow direct evaluation of local temporal variation within recordings. Consequently, the temporal trajectory analyses in the present work can only be evaluated indirectly through structural properties such as smoothness, separation, and consistency across speaking situations. At the same time, the high agreement between overlapping and non-overlapping 8-second trajectories suggests that the extracted temporal structures are not purely segmentation artifacts, but remain comparatively stable across related trajectory extraction settings.

Second, the present study focuses only on three recording situations (WSD, WSS, and FG). Additional speaking situations available in the REDE corpus may reveal further forms of situational variation and potentially different temporal realization patterns. Similarly, speaker behavior may vary across interactional settings beyond those investigated here.

References

Anonym. 2026a. Anonym.

Anonym. 2026b. Anonym.

Allan Bell. 1984. Language style as audience design. *Language in Society*, 13(2):145–204.

Hanna Fischer and Alfred Lameli. 2026. German dialects across situations, generations, and regions: The rede corpus as an oral resource for nlp. In *Dialects in NLP — A Resource Perspective (DialRes)*, pages 144–152, Palma, Mallorca, Spain. European Language Resources Association (ELRA).

Joachim Herrgen, Alfred Lameli, Stefan Rabanus, and Jürgen Schmidt. 2001. Dialektalität als phonetische distanz. ein verfahren zur messung standarddivergenter sprechformen.

Joachim Herrgen and Jürgen Erich Schmidt. 1989. Dialektalitätsareale und dialektabbau. In Wolfgang Putschke, Werner H. Veith, and Peter Wiesinger, editors, *Dialektgeographie und Dialektologie*, volume 90 of *Deutsche Dialektgeographie*, pages 304–346. Elwert, Marburg.

Eva-Maria Krech, Eberhard Stock, Ursula Hirschfeld, and Lutz-Christian Anders. 2009. *Deutsches Aussprachewörterbuch*. Walter de Gruyter.

William Labov. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.

Alfred Lameli. 2025. *Gesprochenes Deutsch in den Regionen: Eine Standortbestimmung für die Bundesrepublik Deutschland*, pages 51–80. De Gruyter, Berlin, Boston.

Alfred Lameli. 2026. Evaluating phonetically weighted and unweighted distance measures in dialectometry. In *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)*, pages 4141–4151, Palma, Mallorca, Spain. European Language Resources Association (ELRA).

Helmut Lausberg. 1993. *Situative und individuelle Sprachvariation im Rheinland*, volume 136 of *Rheinisches Archiv*. Böhlau, Köln.

Alexandra N. Lenz. 2003. *Struktur und Dynamik des Substandards: Eine Studie zum Westmitteldeutschen (Wittlich/Eifel)*, volume 125 of *Zeitschrift für Dialektologie und Linguistik – Beihefte*. Franz Steiner Verlag, Stuttgart.

Jürgen Macha. 1991. *Der flexible Sprecher*. Veröffentlichung des Instituts für Geschichtliche Landeskunde der Rheinlande der Universität Bonn. Böhlau, Köln [u.a.].

John Nerbonne and Wilbert Heeringa. 1997. Measuring dialect distance phonetically. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*.

Jürgen Erich Schmidt, Joachim Herrgen, Roland Kehrein, Alfred Lameli, and Hanna Fischer. 2020–. Regionalsprache.de: Forschungsplattform zu den modernen Regionalsprachen des Deutschen. Digitale Sprachressource. Bearbeitet von Lisa Dücker, Robert Engsterhold, Marina Frank, Heiko Girnth, Simon Kasper, Juliane Limper, Salome Lipfert, Georg Oberdorfer, Tillmann Pistor, Anna Wolańska. Unter Mitarbeit von Dennis Beitel, Lea Fischbach, Milena Gropp, Heiko Kammers, Maria Luisa Krapp, Vanessa Lang, Salome Lipfert, Nathalie Mederake, Jeffrey Pheiff, Bernd Vielsmeier. Studentische Hilfskräfte.

Adriana Rosalina Galván Torres and Tanja Müller. 2021. Die tagesschau in 100 sekunden: Die aussprache von< ä> bei nachrichtensprecher* innen. *Sincronía*, 25(80):803–828.

Georg Wenker. 2013. *Schriften zum "Sprachatlas des Deutschen Reichs". Gesamtausgabe*. Number 111.1–3 in Deutsche Dialektgeographie. Olms, Hildesheim, New York, Zürich.

A Appendix

P_{FG} Threshold	Moveless	Shifter	Hybrid	Switcher
0.7	77	178	37	15
0.8	77	200	21	9
0.9	77	216	9	5

Table 3: Sensitivity analysis for different FG polarization thresholds. As expected, stricter polarization thresholds reduce the number of *Switcher* classifications while increasing the number of *Shifter* classifications. However, the overall transition from classical *Switcher* to temporal *Shifter* behavior remains stable across thresholds.