forum acusticum 2026
8-12 September · Graz, Austria

SPEECH REDUCTION AND SPEECH RATE: DISENTANGLING TEMPORAL AND PROSODIC EFFECTS

Hansjörg Mixdorff[1], Hartmut R. Pfitzinger[2], Noam Amir[3]

[1] *Berlin University of Applied Sciences, Germany*
[2] *Pfitzinger Voice Design, Lübeck, Germany*
[3] *Tel Aviv University, Israel*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

An analysis of the link between speech rate and speech reduction identifies the degree to which rate alone drives reduction. Using the German part of the Bonn-Tempo corpus, we analyze vowel, consonant, and pause reduction as a function of speech rate and prosodic accentuation. Phone- and syllable-based rate measures were first computed, followed by Pfitzinger's perceptual local speech rate (PLSR), allowing a detailed characterization of rate distributions across the five corpus-defined speed levels for individual speakers. Substantial inter-speaker variability is observed, indicating that nominal rate categories only partially reflect actual production. Acoustic analyses of vowel formants show that significant differences in vowel centralization occur primarily between very slow, normal, and very fast speech. After Lobanov normalization, vowel centralization increases approximately linearly with PLSR before reaching a saturation plateau at very high rates. Temporal analyses further reveal that vowel, consonant, and pause compression and elision are not uniform processes, but depend systematically on both accent status and underlying PLSR. These findings suggest that speech reduction constitutes a structured, prosodically mediated acoustic dimension that cannot be fully derived from speech rate alone. A companion perceptual study addresses how listeners respond to these reduction patterns.

Keywords: *speech reduction, prosodic features, speech rate*

1. Introduction

Speech reduction — the systematic loss of articulatory, acoustic, or perceptual detail under prosodic and communicative pressure — is traditionally attributed to increased speech rate. Faster speaking naturally shortens segments and compresses formant targets, yet reduction is not a mechanical byproduct of tempo. It reflects prosodic organization, lexical prominence, and listener-oriented clarity adjustments. The central question is whether reduction is **fully predictable from rate**, or whether it constitutes an **independent prosodic dimension**.

Early work treated rate globally, but subsequent approaches emphasized local speech rate (LSR), capturing the density of phonetic events over time. Pfitzinger [1] demonstrated that perceived rate aligns more closely with local phone- and syllable-based metrics than with overall tempo and later introduced methods for rate normalization to reveal residual timing variation once rate effects are removed [2]. These residuals, systematic across prosodic contexts, suggest that prosody itself modulates temporal realization.

Extending this perspective, Pfitzinger [3] identified "degree of reduction" as a core prosodic dimension alongside timing, intonation, intensity, and voice quality. He noted that manipulating one dimension (e.g., timing through time-scale modification) inevitably perturbs others. Pfitzinger and Mixdorff [4] confirmed perceptually that listeners distinguish between natural and time-scaled utterances even when durations are matched, implying that temporal cues alone cannot explain clarity judgments.

Recent research reinforces this multidimensional view. Spectro-temporal analyses of vowel clarity show that intelligibility gains depend on enhanced spectral contrast, not merely slower rate (see, for instance, [5]). Rate-matched clear-speech experiments reveal that prosodic modifications change perceived tempo beyond duration scaling [6], while modeling studies highlight continued divergence between measured and perceived tempo [7]. Work on time-scale modification confirms that PSOLA-type manipulations may introduce subtle spectral artifacts that affect naturalness [8]. Together, these findings support viewing reduc-

Corresponding author: *hmixdorff@bht-berlin.de.*

tion as a complex interaction of temporal, spectral, and articulatory adjustments. The present study addresses this question acoustically. We analyzed a part of the German BonnTempo containing 12 speakers (3 male, 9 female) uttering seven sentences at five different levels of speed.

2. Speech Material and Methods

2.1. Speech Material

The acoustic analysis draws on the BonnTempo Speech Rate Corpus [9, 10], a large multilingual database designed to study timing and rhythm as a function of speech rate. The present work analyzes part of the German subset, consisting of readings of a passage from Bernhard Schlink's *Selbs Betrug* (ca. 76 syllables in 7 chunks). The corpus was recorded in a sound-treated booth at the University of Bonn with a condenser microphone (44.1 kHz, 16 bit). Of the original 15 native German speakers (mean age ca. 28 years), we included 12 (3 males and 9 females) in the present analysis. All speakers produced the text at five speech rates: very slow (*l1*), slow (*l2*), normal (*no*), fast (*s1*), and very fast (*s2*).
The conditions very slow, slow, and fast were produced with reference to the normal condition under explicit pacing instructions, while the condition very fast corresponded to "as fast as possible". This will explain the noticeable speech rate jump we witness in the following between fast and very fast.
The data set contains roughly **4500** syllables and corresponding C/V intervals, all of which were manually labeled in Praat [11]. This corpus therefore provides robust measures of **syllable, vowel, and consonant durations** across systematically varied speech rates, suitable for modeling global and local rate effects on reduction. Two speakers from this corpus were later selected to create the perceptual stimuli analyzed in a companion clarity-judgment study [12].

2.2. Segmentation and Acoustic Measures

Since we needed — besides the manually labeled syllables — phone segmentations of all utterances, we segmented them automatically by MAUS forced alignment [13] and manually verified the results in Praat [11]. From the aligned TextGrids, we extracted boundaries for all vowel, consonant and pause segments, enabling computation of segment and syllable durations and vowel-mean and center formant frequencies (F1–F3).
Derived measures include (a) syllable rate (SR, syllables s^{-1}), (b) phone rate (PR, phones s^{-1}), (c) perceptual local speech rate (PLSR) applying a 600 ms sliding window at a step of 50 ms using the formula [14, Abb. 8.3, p. 208]

$$\mathrm{PLSR}(t) = 7.834 \cdot \mathrm{SR}(t) + 4.060 \cdot \mathrm{PR}(t) - 1.41, \quad (1)$$

as well as (d) vowel reduction indices measuring F1/F2 displacement toward the schwa reference [15, 16, 17]. Consonant reduction was indexed by the mean duration of oral closures and fricative intervals.

2.3. Derived Parameters and Normalization

Durations were normalized by each speaker's mean at the normal rate to yield comparable relative factors. Local-rate trajectories were smoothed and mapped onto the utterance timeline to assess correlations between temporal density and spectral centralization. Formant values were taken at 50% of vowel duration.

2.4. Statistical Analysis

We first examined whether the five rate classes actually differed with respect to vowel centrality. Therefore, we performed statistical analysis on the Euclidean distance of each vowel token in the F1/F2 plane from F1=500Hz/F2=1500Hz. This reference corresponds to the German schwa target and, after Lobanov normalization, allows for a cross-speaker comparison.
Results showed that neighboring classes did not significantly differ, but very slow, normal and very fast rates did (Kruskal-Wallis test of independent samples, $p < .001$). When we look at the actual PLSR that speakers produced which excludes pauses (see Fig. 1), it generally rises with the speed level, but highly uneven within and between speakers. What is obvious is the conspicuous jump from fast to very fast speed. Inspection of slow and very slow speed data of Speaker 15 suggested a possible labeling inconsistency. It is obvious that speed class can only serve as a very coarse reference.

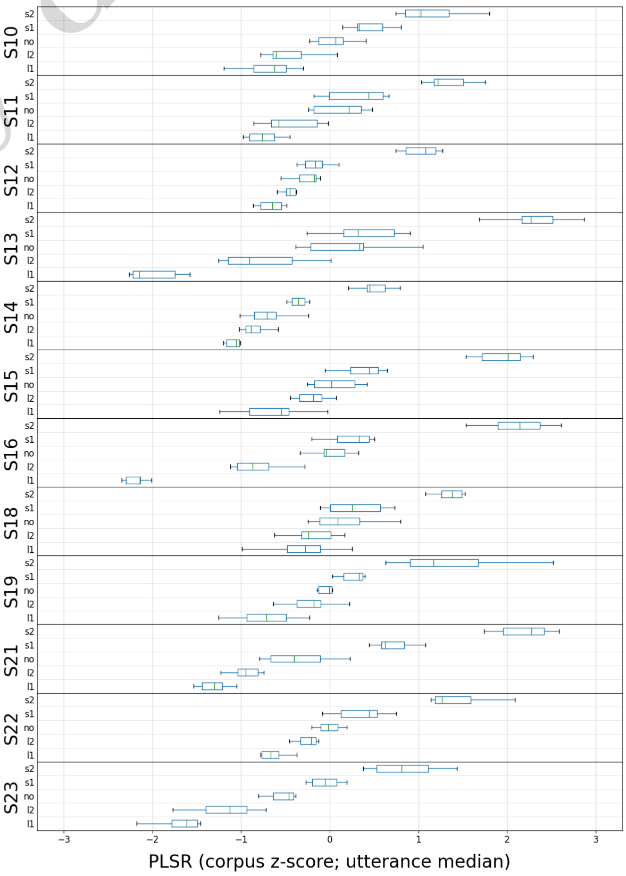

Figure 1: PLSR distributions for all 12 speakers (S10–S23), five box plots with rising speech rate (l1–s2).

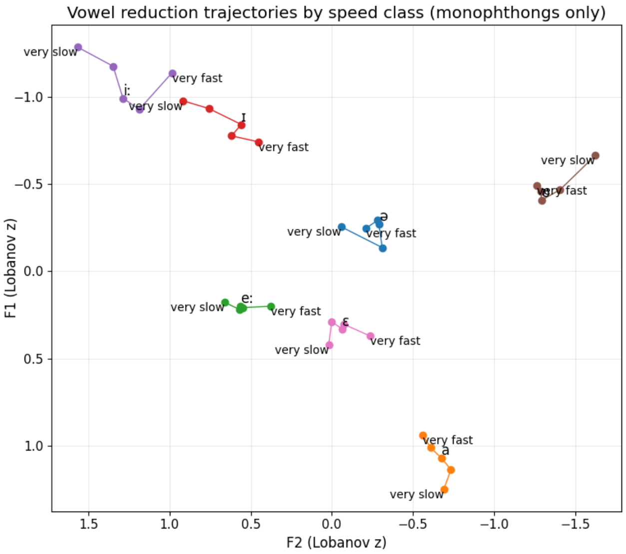

Figure 2: Average reduction paths for the tense and lax monophthongs iː, ɪ, eː, ɛ, a, ə, and o in the normalized F1/F2 plane.

3. Results of Reduction Analysis

Fig. 2 shows vowel reduction paths, applying Lobanov z normalization, calculated by taking the centroids of F1/F2 distributions for each of the five speed classes for the respective vowel and joining them with straight lines. Centralization becomes apparent visually, meaning that vowel formants approximate the center of the chart. We established the relationship between PLSR and the mean distance of vowel tokens from the center of the F1/F2.

After Lobanov normalization, vowel-space radius exhibits a delayed-onset contraction with PLSR that is approximately linear over the mid-range and saturates at $\sim 1z$-unit at high rates, indicating a bounded articulatory limit; token-weighted and vowel-balanced fits are closely aligned in the upper range (Fig. 3). This means that reduction is low at slow speeds and increases with rising PLSR and approaches for high speeds an asymptotic value (orange track), as speakers produce the fastest versions they are capable of.

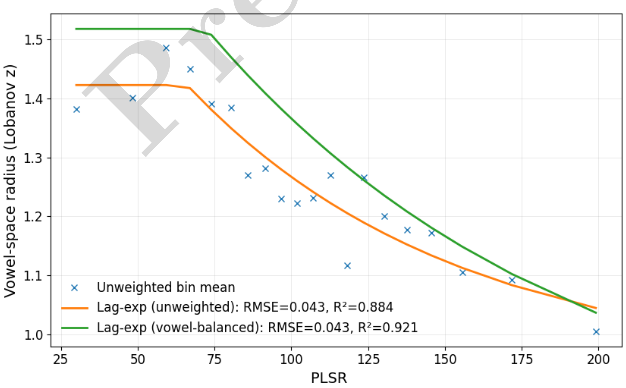

Figure 3: Vowel-space radius vs. PLSR: Mean distance of vowels from center of the F1/F2 plane.

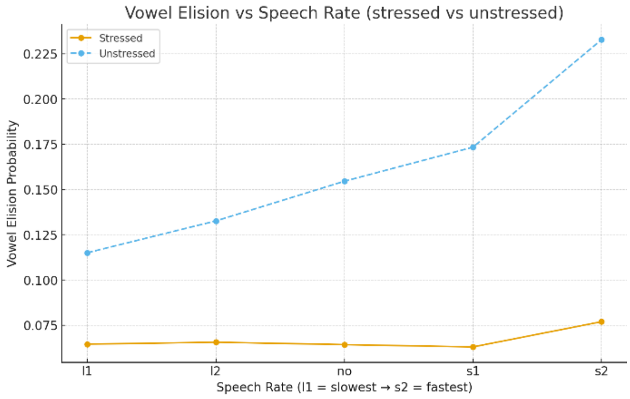

Figure 4: Probability of vowel elision for stressed and unstressed syllables.

Since different vowel types by nature of their inherent durations pertain to different PLSR classes, we also calculated a vowel-balanced version of the track (in green, vowel-balanced data points omitted for clarity). Fig. 4 shows the probability of vowel elision based on the forced alignment results for the five different levels of speed. As we can see, stressed syllables are much more protected against complete elision.

We performed the same analysis for the consonants. Fig. 5 shows the corresponding graphs, underlining that consonants in stressed syllables are better protected against elisions than those in unstressed ones. We now look at the behavior of segments that were not completely removed but rather compressed. To visualize the relation between local speech rate (PLSR) and segmental duration, the data were divided into quantile-based bins of PLSR. Mean phone durations were normalized to 100% at the slowest-rate bin and fitted with a lagged-exponential model ($y = a + b \cdot e^{-cx}$, see Fig. 6). As can be seen both vowels and consonants shorten with increasing rate, but vowels show a steeper and earlier contraction (higher b, c), whereas consonants approach a higher asymptote (≈18–20%). Stress preserves duration in both categories, shifting the curves upward (≈ +10–15%).

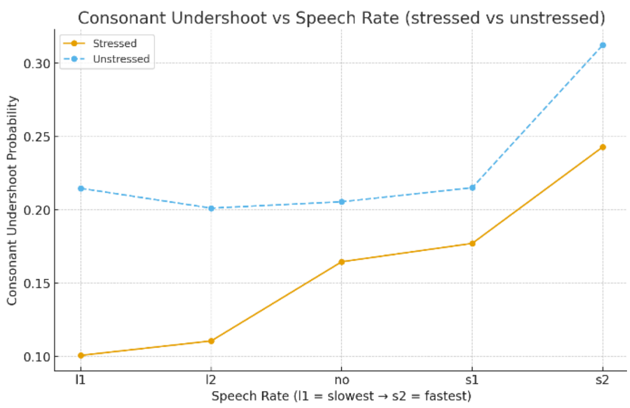

Figure 5: Consonant elision for the five different speed levels.

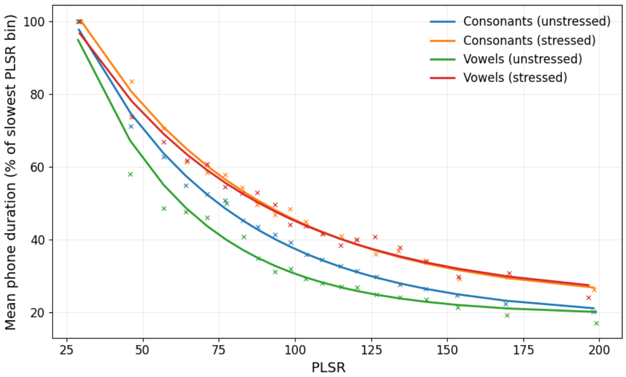

Figure 6: Percentage compression of phone duration as a function of local speech rate (PLSR) modeled using an exponential decay fit.

This procedure ensured that each bin contained approximately the same number of phone tokens, thereby equalizing statistical reliability across the rate continuum. Because the bins were defined by percentiles rather than fixed PLSR intervals, the resulting bin widths in PLSR units varied slightly depending on the density of the underlying rate distribution. For each bin, mean phone durations were computed separately for stressed and unstressed vowels and consonants, then normalized to the mean duration in the slowest-rate bin (100%) to obtain relative compression profiles.
For pause segments, local speech rate (PLSR) was estimated as the mean of the two preceding syllables' PLSR values, capturing the average tempo in the phrase-final region. Pauses were then binned by quantiles of this mean PLSR, ensuring approximately equal token counts per bin. Mean pause durations were normalized to the slowest-rate bin (100%) to express relative compression.
Each pause was linked to the mean PLSR of the two preceding syllables to represent the local phrase-final rate context. Mean pause durations were normalized to the slowest-rate bin (100%) and fitted with a lagged-exponential function (see Fig. 7). Pauses shorten steeply with increasing rate and approach minimal

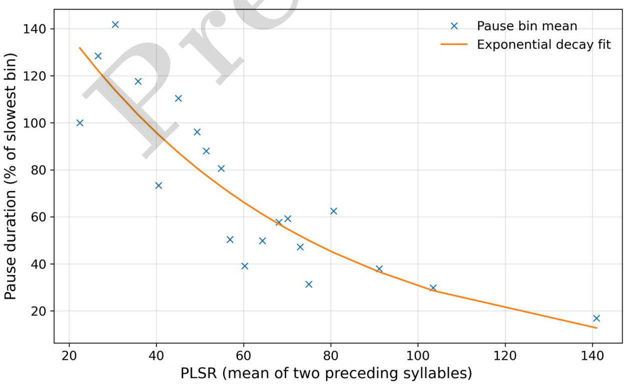

Figure 7: Pause durations as a function of local speech rate (PLSR).

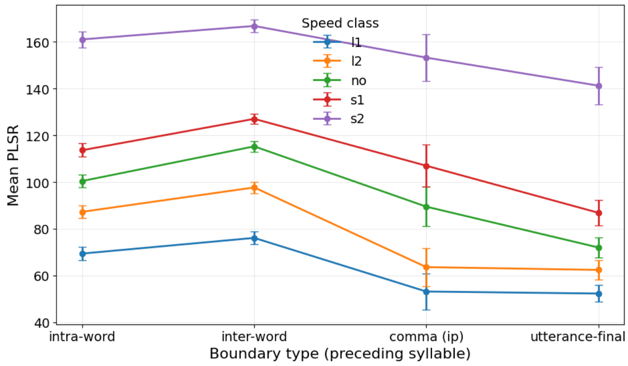

Figure 8: Mean PLSR before different types of boundary.

durations at fast tempos, consistent with their role as an extreme manifestation of phrase-final temporal expansion. The resulting curve shows that pauses shorten sharply with increasing local rate, converging toward an asymptote around 15–25% of baseline — consistent with pauses representing the extreme end of phrase-final temporal expansion.
If reduction is prosodically structured, boundary depth should systematically modulate local rate independent of global tempo. In order to examine the phrasal structure and its influence on PLSR, we first annotated syllabic boundaries as either intra-word, inter-word, comma or period boundaries and examined the PLSR of the syllable immediately preceding it. The results are displayed in Fig. 8.
Obviously PLSR is affected by the speech rate class, with deeper boundaries reducing PLSR. We would expect intra-word boundaries to be faster than inter-word boundaries. However, there are very few polysyllabic words in the text, most of which are di-syllabic with stress on the first syllable. This explains why we find lower PLSR in these places. Conversely, many inter-word boundaries occur after function words or unstressed syllables. To refine our phrasal analysis, we marked noun phrase boundaries in addition to comma and period boundaries, yielding phrases between 4 and 6 syllables. If we average over the first five syllables for all phrases we see the steadily falling PLSR in Fig. 9. This corresponds to a rallentando pattern from

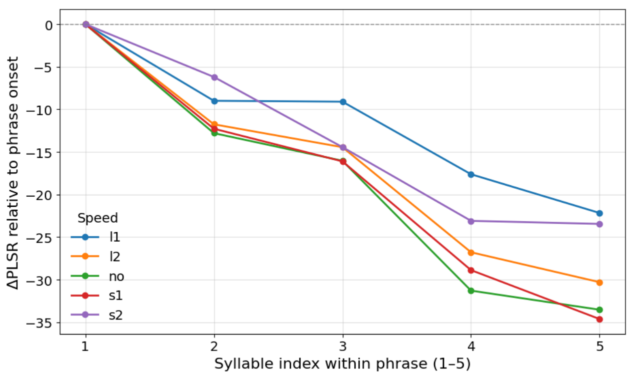

Figure 9: Relative PLSR change in first 5 syllables (from text-based phrase segmentation).

phrase start to end. We can also see that the slope is becoming shallower as speech rate increases. That means that deceleration is blocked at very high speed, which is confirmed by Kruskal-Wallis test ($p = .003$).

4. Discussion and Conclusions

The current study investigated the influence of speech rate on the degree of reduction. In order to measure speech rate we adopted Pfitzinger's **perceptual local speech rate** as a perceptually grounded measure of temporal density. Reduction reflects not only global tempo but localized timing adjustments linked to prosodic planning. We were able to show that speeding up pushes vowels towards the center of the F1/F2 plane, though this effect is almost negligible at slower speech rates, then increases almost linearly with PLSR until we reach a ceiling at the super-fast speech rate. The effect is much stronger in unstressed syllables than in stressed ones.

Similar changes were observed for vowel and consonant compressions and elisions which increased with speech rate, that is, phones in stressed syllables are more protected than in unstressed ones. Pause durations are reduced as PLSR increases, but the negative gradient becomes shallower as we approach the super-fast condition. We were also able to show how prosodic boundaries influence the reduction of syllables. Boundary depth reduces PLSR, and this effect is preserved across the different speech rate classes. We also found that the general trend that PLSR decreases along a prosodic phrase is preserved at all speech rates though the gradient is shallower in faster speech.

The close relationship between reduction measures and PLSR observed in Figs. 3, 6 and 7 is fully expected and does not in itself distinguish between competing accounts of speech reduction. The present study does not argue that reduction operates independently of local speech rate. Rather, the results suggest that both reduction and local speech rate are systematically shaped by prosodic structure. This is evident from the effects of boundary strength and phrase position on PLSR itself. An alternative interpretation is therefore that prosodically conditioned temporal allocation gives rise to both local speech rate variation and reduction. The current data do not allow a definitive distinction between these causal accounts. However, they demonstrate that reduction cannot be understood solely in terms of nominal speech rate categories and must be interpreted in relation to local prosodic organization.

In a companion study [12] we linked our acoustic result to perceptual experiments and examined how clarity ratings could be derived from phonetic and phonological parameters.

Overall, the findings support the view that speech reduction is systematically organized by prosodic structure and cannot be interpreted as a simple by-product of nominal speech rate. The observed relationships between PLSR, segmental reduction, and prosodic context suggest that reduction forms part of a broader system of temporal and phonological organization.

Taken together, these results suggest that reduction is not emerging from local speech rate but — like local speech rate itself — is a surface correlate of prosodic planning, which governs both segmental compression and boundary-related timing structure.

References

[1] H. R. Pfitzinger, "Local speech rate perception in German speech," in *Proc. of the XIVth Int. Congress of Phonetic Sciences (ICPhS)*, vol. 2, San Francisco, 1999, pp. 893–896.

[2] ——, "Reducing segmental duration variation by local speech rate normalization of large spoken language resources," in *Proc. of the third Int. Conf. on Language Recources and Evaluation (LREC '02)*, vol. 1, Las Palmas; Spain, 2002, pp. 313–320.

[3] ——, "Five dimensions of prosody: Intensity, intonation, timing, voice quality, and degree of reduction," in *Speech Prosody Abstract Book. Studientexte zur Sprachkommunikation*, R. Hoffmann and H. Mixdorff, Eds. Dresden: TUDpress, 2006, vol. 40, pp. 6–9.

[4] H. R. Pfitzinger and H. Mixdorff, "Perceiving speech rate differences between natural and time-scale modified utterances," in *Proc. Interspeech*, Lyon; France, 2013, pp. 277–281.

[5] A. Marczyk, B. O'Brien, P. Tremblay, V. Woisard, and A. Ghio, "Correlates of vowel clarity in the spectrotemporal modulation domain: Application to speech impairment evaluation," *J. of the Acoustical Society of America*, vol. 152, no. 5, pp. 2675–2691, 2022.

[6] L. Plug, Y. Zheng, and R. Smith, "The impact of clear speech modifications on perceived tempo of rate-matched English utterances," *J. of the Acoustical Society of America*, vol. 157, no. 4, pp. 2835–2846, 2025.

[7] L. Plug, R. Lennon, and R. Smith, "Measured and perceived speech tempo: Comparing canonical and surface articulation rates," *J. of Phonetics*, vol. 95, pp. 101–193, 2022.

[8] K. S. Rao and A. K. Vuppala, "Non-uniform time scale modification using instants of significant excitation and vowel onset points," *Speech Communication*, vol. 55, no. 6, pp. 745–756, 2013.

[9] V. Dellwo and P. Wagner, "Relations between language rhythm and speech rate," in *Proc. of the XVth Int. Congress of Phonetic Sciences (ICPhS)*, Barcelona, 2003, pp. 471–474.

[10] V. Dellwo, I. Steiner, B. Aschenberner, J. Dankovičová, and P. Wagner, "BonnTempo-corpus and BonnTempo-tools: A database for the study of speech rhythm and rate," in *Proc. of ICSLP '04*, Korea, 2004, pp. 1025–1028.

[11] P. Boersma and V. van Heuven, "Speak and unSpeak with PRAAT," *Glot International*, vol. 5, no. 9/10, pp. 341–347, 2001.

[12] H. Mixdorff, H. R. Pfitzinger, N. Amir, and A. Hönemann, "Speech reduction and perceived clarity: Just a matter of rate?" in *Proc. of Speech Prosody*, Philadelphia, 2026, pp. 604–608.

[13] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[14] H. R. Pfitzinger, "Phonetische Analyse der Sprechgeschwindigkeit," IPSK, Univ. München, Forschungsberichte (FIPKM) 38, 2001.

[15] M. Nadeu, "Stress- and speech rate-induced vowel quality variation in Catalan and Spanish," *J. of Phonetics*, vol. 46, pp. 1–22, 2014.

[16] E. Schulz, Y. M. Oh, B. Andreeva, and B. Möbius, "Impact of prosodic structure and information density on vowel space size," in *Proc. of Speech Prosody*, Boston, 2016, pp. 350–354.

[17] B. Siebenhaar and M. Hahn, "Vowel space, speech rate and language space," in *Proc. of the 19th Int. Congress of Phonetic Sciences (ICPhS)*, Melbourne, 2019, pp. 879–883.