UNIWERSYTET
IM. ADAMA MICKIEWICZA
W POZNANIU

Wydział Neofilologii

Hanna Kasperek

Numer albumu: 285671

Jaw and lip gestures as articulatory correlates of rhythmic features of Polish utterances

Gesty żuchwy oraz warg jako korelaty artykulacyjne cech rytmicznych polskich wypowiedzi

Rozprawa doktorska
napisana pod opieką naukową
prof. UAM dr hab. Katarzyny Klessy oraz prof. UW dr hab. Anity Lorenc

Poznań, 2025

Streszczenie

Niniejsza rozprawa przedstawia pierwsze systematyczne badanie ruchów żuchwy oraz warg jako artykulacyjnych korelatów rytmu wypowiedzi w języku polskim, z wykorzystaniem artykulografii elektromagnetycznej (EMA). Eksperyment oraz jego omówienie zostały osadzone w istniejącym stanie badań i porównane z wynikami uzyskanymi dla innych języków.

W rozprawie przedstawiono wybrane podejścia lingwistyczne dotyczące badania cech rytmicznych wypowiedzi, w ujęciu fonetyczno-akustycznym, a także artykulacyjnym, nakreślając wielopoziomowość badanego zjawiska. Przedstawione ujęcia są omawiane w kontekście wybranych etapów rozwoju badań nad rytmem mowy — zarysowano elementy toczącej się dyskusji naukowej oraz wskazano, w jaki sposób postęp technologiczny sprzyjał prowadzeniu coraz bardziej precyzyjnych badań lingwistycznych, w tym również nad rytmem mowy.

Następnie przedstawiono rezultaty badań z wykorzystaniem artykulografii elektromagnetycznej (EMA) dla wybranych języków: angielskiego amerykańskiego, japońskiego, hiszpańskiego, chińskiego mandaryńskiego oraz brazylijskiej odmiany portugalskiego. Na podstawie przeglądu literatury sformułowano wstępne przewidywania dla języka polskiego.

W kolejnej części omówiono charakterystykę akcentową języka polskiego, podkreślając, że wiele kwestii, takich między innymi, jak akustyczne korelaty akcentu zdaniowego, wciąż pozostaje przedmiotem debaty.

Po części zarysowującej tło badań przedstawiono szczegółowy opis metodologii. Zaprezentowano metodę badawczą — quasi-eksperyment z dwoma czynnikami w układzie z powtarzanym pomiarem, wraz z omówieniem doboru planu, zmiennych zależnych i niezależnych. Celem badania było określenie, czy i w jaki sposób dwie zmienne prozodyczne wpływają na zmiany w układzie artykulatorów i na wybrane charakterystyki fonetyczno-akustyczne realizacji samogłosek w kontekście zdaniowym. Badane zmienne prozodyczne to:

- obecność kontrastowego akcentu zdaniowego: wypowiedź neutralna oraz wypowiedź z fokusem kontrastywnym;
- pozycja w wypowiedzi: oczekiwane miejsce akcentu zdaniowego, oczekiwane miejsce fokusu kontrastywnego oraz pozostałe pozycje.

Badane korelaty artykulacyjne to **wychylenie żuchwy** oraz **rozwarcie warg**, a wybrane charakterystyki fonetyczno-akustyczne to: **czas trwania, intensywność, częstotliwość podstawowa**

(F0) i relacje między częstotliwościami formantowymi F1 i F2.

Badanie objęło sześć rodzimych użytkowniczek języka polskiego, które realizowały kontrolowane zdania zawierające samogłoski ustne polszczyzny (/a/, /e/, /i/, /o/, /u/, /I/). Każda uczestniczka zrealizowała sześć kombinacji eksperymentalnych, przy czym scenariusz nagraniowy został zaprojektowany w taki sposób, aby zapewnić wykorzystanie identycznego materiału leksykalnego we wszystkich warunkach. Do pomiaru i rejestracji zmiennych artykulacyjnych wykorzystano artykulograf Carstens AG501. Na potrzeby niniejszej rozprawy analizie poddano dane pochodzące z czujników umieszczonych na żuchwie oraz na górnej i dolnej wardze, jak również dane z czujników referencyjnych, które posłużyły do obliczenia pozycji i trajektorii ruchów wybranych artykulatorów. Rejestracji ruchów artykulatorów towarzyszył jednoczesny zapis sygnału akustycznego.

Aby ograniczyć wpływ czynników zakłócających, kontrolowano treść leksykalną (jednakowe zdania we wszystkich warunkach), warunki nagraniowe (jednolite ustawienia laboratoryjne i kalibracja sprzętu). W badaniu uczestniczyły wyłącznie kobiety, co pozwoliło dodatkowo zminimalizować wpływ potencjalnych źródeł zmienności związanych z różnicami płci.

Analiza statystyczna z wykorzystaniem modeli mieszanych wykazała hierarchiczny gradient artykulacyjny prominencji: samogłoski w pozycji akcentu wykazywały znacząco większe przemieszczenie żuchwy (–0,383 SD) oraz szersze rozwarcie warg (+0,513 SD) w porównaniu z innymi pozycjami. Najsilniejsze efekty obserwowano w pozycji fokusu kontrastywnego, gdzie przemieszczenie żuchwy wynosiło –1,172 SD, a rozwarcie warg +1,119 SD (wszystkie efekty $p < 0,001$). Kluczowe odkrycia obejmują: (1) systematyczne wzmacnianie artykulacyjne w pozycjach akcentu i fokusu, (2) akustyczne wzbogacenie fokusu kontrastywnego poprzez podwyższenie F0 i intensywności, (3) globalny wpływ obecności fokusu kontrastywnego na dynamikę wypowiedzi zarówno na poziomie artykulacyjnym, jak i akustycznym, oraz (4) efekt wydłużający ostatnią samogłoskę w wyrazie z fokusem kontrastywnym.

Dodatkowo podczas badań zaobserwowano dryf żuchwy, polegający na stopniowo malejącym wychyleniu żuchwy w trakcie wypowiedzi. Zjawisko to jest w stopniu minimalnym opisywane w literaturze. Możliwe, że pozostaje ono w związku z deklinacją, czyli systematycznym obniżaniem się wartości F0 w przebiegu wypowiedzi. Jest to jeden z potencjalnie istotnych obszarów do dalszego zbadania.

Wyniki potwierdzają hipotezę, że **żuchwa oraz rozchylenie warg funkcjonują jako artykulacyjne korelaty rytmu mowy w polszczyźnie, przy czym wzorce przemieszczenia odzwierciedlają hierarchiczną strukturę metryczną wypowiedzi**. Tym samym rozprawa wnosi istotny wkład w rozwój fonetyki artykulacyjnej języka polskiego. Metodologicznie rozprawa ustanawia ramy analityczne łączące dane artykulacyjne z miarami rytmu mowy, oferując narzędzia przydatne w przyszłych badaniach dla języka polskiego oraz analizach porównawczych międzyjęzykowych. Należy podkreślić, że jest to ujęcie nowatorskie, które dzięki zastosowaniu elementów automatyzacji w znacznym stopniu przyspiesza i ułatwia pracę z danymi artykulograficznymi, umożliwiając zwiększanie skali korpusu oraz powtarzalne prowadzenie analiz.

Słowa kluczowe: artykulografia elektromagnetyczna (EMA), fonetyka artykulacyjna, rytm mowy, akcent zdaniowy, fokus kontrastywny, wychylenie żuchwy, rozwarcie warg

Abstract

This dissertation presents the first systematic study of jaw and lip movements as articulatory correlates of speech rhythm in Polish, using electromagnetic articulography (EMA). The experiment and its discussion are placed in the context of existing literature and compared with results obtained for other languages.

The dissertation presents selected approaches to studying rhythmic features of speech from phonetic-acoustic and articulatory perspectives, outlining the multilevel nature of the phenomenon. These approaches are discussed in the context of selected stages in the development of speech rhythm research, outlining elements of the ongoing academic debate and demonstrating how technological progress has enabled increasingly precise linguistic investigations, including research on speech rhythm.

Next, the results of studies using electromagnetic articulography (EMA) on several languages — American English, Japanese, Spanish, Mandarin Chinese, and Brazilian Portuguese — are presented. Based on a review of the literature, preliminary predictions for the Polish language are formulated.

The following section discusses the stress characteristics of the Polish, emphasising that many issues, such as the acoustic correlates of utterance-level accent, remain a subject of ongoing debate.

After outlining the research background, the dissertation provides a detailed description of the methodology. A repeated-measures quasi-experimental design with two within-subject factor is described, along with discussion of the rationale for this design and the selection of dependent and independent variables. The aim of the study was to examine whether and how two prosodic variables affect articulatory configuration and selected phonetic-acoustic characteristics during the production of vowels embedded in sentence contexts. The prosodic variables studied were:

- the presence of contrastive focus: neutral utterance vs. utterance with contrastive focus;
- position in the utterance: expected utterance-level *Accent* position, expected contrastive *Focus* position, and *Other* positions.

The articulation correlates examined were **vertical jaw displacement** and **lip aperture**, while the selected phonetic–acoustic characteristics included **duration, intensity, fundamental frequency**

(F0), and the relationship between frequency formants F1 and F2.

The study involved six native female Polish speakers who produced controlled utterances containing all six oral vowels of Polish (/a/, /e/, /i/, /o/, /u/, /I/). Each participant completed all six experimental conditions, with the recording protocol designed to ensure identical lexical material across all conditions. A Carstens AG501 articulograph was used to measure and record articulatory variables. For this dissertation, data from sensors placed on the jaw and upper and lower lips were analysed, as well as data from reference sensors used to calculate the position and trajectory of movements of selected articulators. Articulatory movements were recorded simultaneously with the acoustic signal.

To limit the influence of confounding factors, the lexical content (identical utterances in all conditions) and recording conditions (uniform laboratory settings and equipment calibration) were controlled. Only women participated in the study, which further minimized the influence of potential sources of variability related to gender differences.

Statistical analysis using mixed models revealed a hierarchical articulatory gradient of prominence: vowels in stressed positions showed significantly greater mandibular displacement (–0.383 SD) and wider lip aperture (+0.513 SD) compared to other positions. The strongest effects were observed in the contrastive focus position, where mandibular displacement was –1.172 SD and lip aperture was +1.119 SD (all effects $p < 0.001$). Key findings include: (1) systematic articulatory strengthening in stressed and focused positions, (2) acoustic enhancement of contrastive focus through elevated F0 and intensity, (3) global influence of contrastive focus presence on utterance dynamics at both articulatory and acoustic levels, and (4) lengthening of the final vowel in words with contrastive focus.

Additionally, jaw drift was observed during the study, consisting of a gradual reduction in vertical jaw displacement over the course of an utterance. This phenomenon has received minimal attention in the literature. It may be related to declination, i.e., the systematic lowering of F0 values throughout an utterance. This represents a potentially important area for further investigation.

The results confirm the hypothesis that **vertical jaw displacement and lip aperture function as articulatory correlates of speech rhythm in Polish, with displacement patterns reflecting the hierarchical metrical structure of utterances**. Thus, the dissertation makes a significant contribution to the development of Polish articulatory phonetics. Methodologically, the dissertation establishes an analytical framework combining articulatory data with speech rhythm measures, offering tools useful for future research on Polish and cross-linguistic comparative analyses. It should be emphasized that this is an innovative approach that, through automation, significantly accelerates and facilitates work with articulographic data, enabling scalable data processing and replicable analyses.

Keywords: electromagnetic articulography (EMA), articulatory phonetics, speech rhythm, utterance-level accent, contrastive focus, vertical jaw displacement, lip aperture

To patience, which makes the hidden visible

Contents

Introduction

Extensive research on prominence has resulted in some terminological inconsistencies. The phenomenon has been referred to by various terms such as stress, accent, focus, or emphasis, and analysed across different domains. Another source of complexity is the cross-linguistic variation: languages differ substantially in how prominence is manifested, which may obscure the overall understanding of the concept. In this dissertation, **prominence** is understood in a broad but clearly delimited sense. The term refers to those speech elements that stand out relative to their context due to systematic enhancement, achieved through articulatory or acoustic means. As such, prominence involves:

1. **Articulatory correlates**;
2. **Acoustic correlates**;
3. **Levels of analysis** — two prominence types are distinguished:
    - Utterance-level accent, i.e. in Polish the expected place in an utterance, typically placed on the penultimate syllable of the last word;
    - Contrastive focus, which highlights new or contrastive information in an utterance.

Although acoustic correlates of prominence have been investigated in Polish, this dissertation presents **the first systematic account of the articulatory correlates of prominence in Polish**. It is hypothesised to be expressed through articulatory adjustments and the central question is whether these prominence levels rely on similar patterns or diverge in their strategies. By analysing jaw and lip movements alongside acoustic parameters, the study aims to clarify the basis of prominence in Polish and situate it within broader typological frameworks. The dissertation is outlined as follows:

- Chapter 1 introduces the theoretical background, with special attention to speech rhythm, and reviews previous findings on articulatory and acoustic correlates of prominence.
- Chapter 2 establishes a theoretical distinction between lexical stress and utterance-level accent in Polish, reviewing their acoustic correlates and functional roles in speech.
- Chapter 3 describes the methodology, including the participants, materials, experimental design, and procedures of data collection with electromagnetic articulography, as well

as the preprocessing and analytical steps. To address the complexity of the multimodal data, the chapter presents a comprehensive analytical framework with corresponding processing pipeline.

- Chapter 4 integrates the *Results* and *Discussion* sections. This approach reflects the nature of the data: each statistical result and visualisation is presented alongside interpretative commentary to enhance clarity.
- Chapter 5 concludes the dissertation with a summary of the key findings, limitations, and avenues for future research.

The author is aware of the importance of research on the perception of prominence; however, perceptual aspects are not the focus of this dissertation and are only briefly addressed in the literature review.

Understanding prominence requires consideration of its relationship to speech rhythm. Rhythm provides the temporal and metrical scaffold of speech, while prominence marks specific positions within this scaffold — and is hypothesised to emerge through systematic articulatory and acoustic modifications. The following section examines the theoretical foundations of speech rhythm research, thereby providing the necessary framework for the subsequent analyses.

Chapter 1

Speech rhythm — a state-of-the-art overview

Rhythm in speech is a complex, multidimensional phenomenon and remains challenging to analyse comprehensively, despite decades of research across many languages and from various perspectives (e.g., Abercrombie, 1967; Dauer, 1983, 1987; Gibbon, 2003; James, 1940; Jassem et al., 1984; Lehiste, 1977; Pike, 1945). **Given this complexity, it is important to briefly outline the diverse landscape of rhythm research to provide context for the subsequent discussion**. A chronological approach has been adopted here as the most natural way to trace the historical development of rhythm studies.

Languages exhibit specific phonological properties which govern the organisation of speech over time and the emergence of rhythmic patterns. These patterns are subject to variation not only between languages but also within varieties of the same language (Frota & Vigário, 2001; Giordano & D'Anna, 2010) and individual speakers (Arvaniti, 2012; Gibbon & Gut, 2001).

To further understand these rhythmic phenomena, it is important to consider the main theoretical perspectives that have shaped the research. The earliest systematic attempts to analyse rhythm were impressionistic in nature.

1.1 From perceptual impressions to quantitative studies

Early research on speech rhythm concentrated on classifying languages into distinct rhythmic categories, relying primarily on **perceptual impressions**. This reliance is unsurprising, given the practical unavailability of acoustic measurement tools and, consequently, instrumental evidence at the time. Lloyd James (1940) famously compared the impression of syllable-based languages, such as Spanish or Italian, to the sound of a machine gun, while stress-based languages like English or Dutch reminded him of Morse code. He attributed these differences in auditory impression to systematic differences in temporal organisation. Building on this line of thought, Pike (1945) distinguished between stress-timed and syllable-timed languages, noting that

> Many non-English languages (Spanish, for instance) tend to use a rhythm which is more closely related to the syllable than the regular stress-timed type of English; in this case, it is the syllables, instead of the stresses, which tend to come at more-or-less evenly recurrent intervals — so that, as a result, phrases with extra syllables take proportionately more time,

> and syllables or vowels are less likely to be shortened and modified (Pike, 1945, p. 35).

Abercrombie (1967) expanded on these ideas by claiming that speech rhythm is based either on the isochrony of syllables or on the isochrony of interstress intervals, and that these constitute the only two rhythmic patterns available for the languages of the world. Japanese was later framed as mora-timed (Han, 1962 after: Bloch, 1950) and constituted the third option in the classic proposal dividing languages into rhythmic classes (cf. Turk & Shattuck-Hufnagel, 2013) — syllable-timed, stress-timed, and mora-timed. In this temporal view:

- syllable-timed languages have approximately equal syllable durations;
- stress-timed languages have approximately equal foot durations;
- mora-timed languages have approximately equal mora durations.

These views significantly shaped the perception of speech rhythm and have left a lasting impact that persists to this day. The question of whether languages can be meaningfully classified into distinct rhythmic categories remains an active topic of debate. In research focusing on phonetic-acoustic aspects, there is ongoing discussion regarding the validity of such categorical classifications based on rhythmic properties. While the stress-timed, syllable-timed, and mora-timed distinction had a lasting impact on rhythm research, its reliance on isochrony as a defining principle was increasingly called into question.

1.1.1 Considerations on isochrony

The underlying assumption of isochrony has been repeatedly criticised: to date, empirical support for equal-duration units appears to be insufficient (cf. Arvaniti, 2012; Lehiste, 1977; Roach, 1982). Dauer posits that stress patterns can be discerned in every language, but the differences in perception arise not from the temporal intervals themselves, but from variations in **prominence**. Hence, languages can be positioned along a continuum ranging from least to most stress-timed. Accordingly, Dauer also proposed that rhythm arises by grouping elements into larger units (e.g., syllables into feet), with syllable prominence playing a crucial role. In addition, Dauer created an inventory of phonetic and phonological features shaping a language-specific pattern of utterances; these include the duration of prominent syllables, syllable structure, pitch, and vowel quality (1987). Key arguments include:

- Languages with stress-timed rhythm allow consonant clusters in onset (CCVC) and coda (CVCC) positions, with vowels often undergoing reduction;
- Languages closer to the syllable-timed end of the continuum implement processes like epenthesis or liaison to avoid consonant clusters.

These phenomena affect the temporal organisation of utterances, and thus the rhythm, as duration is one of the acoustic correlates of rhythm. Dauer's critique of the stress-timing versus

syllable-timing dichotomy has been particularly influential, and many later accounts have drawn on her insights.

Reconsidering isochrony shifted the focus from equal timing units to prominence and phonological structure. This shift laid the groundwork for later quantitative rhythm metrics. Their development was further facilitated by the increasing use and computational power of personal computers in the 1990s.

While much of the discussion focused on phonetic and acoustic evidence, a parallel line of research within phonology developed under the label of metrical theory (Hayes, 1995; Liberman & Prince, 1977). Although not directly aimed at rhythm typology, metrical theory provides an influential background for the analysis of stress and prominence, to which this dissertation will return in a later section.

The limitations of impressionistic and phonological accounts therefore motivated a move towards quantitative metrics, designed to capture temporal organisation in a more objective way.

1.1.2 Quantitative rhythm metrics

As pointed out by Ramus et al. (1999), Dauer's inventory neither explains to what extent a given phonological feature contributes to the perception of rhythm, nor determines the interaction of these features. These observations led further to a proposal for quantitative rhythm measurements based on **rhythm metrics**. Ramus et al. (1999) proposed methods to quantify timing parameters of utterances through three primary metrics:

- %V — the percentage of vocalic intervals;
- ΔV — the standard deviation of vocalic intervals;
- ΔC — the standard deviation of consonantal intervals.

Vocalic intervals (V) are defined as segments extending from the onset to the offset of a vowel or group of consecutive vowels or vocalic segments (e.g., diphthongs, triphthongs, sonants). Similarly, a consonant interval (C) spans from the onset to the offset of a consonant or groups of consonants. The duration of vowel and consonant intervals sums to the total duration of the utterance.

%V and ΔC have been regarded as the most reliable indicators of rhythmic type. However, alternative measures such as rPVI (*Raw Pairwise Variability Index*) and nPVI (*Normalised Pairwise Variability Index*) — which quantify variability in the duration of consecutive successive acoustic-phonetic intervals — have produced different classifications, positioning the same languages differently along the *syllable-timed / stress-timed* continuum. For instance, %V and ΔC distinguish Japanese from other languages, whereas PVIs analyses already place it closer to syllable-timed languages (Grabe & Low, 2002). **This discrepancy has been attributed to the lack of control for speech rate in the speech samples.**

Because of this limitation, further rhythm metrics were developed to control for tempo effects. One example is VarcoC (*Variation Coefficient*), i.e. the coefficient of variation (Barry et al., 2003; Dellwo, 2006; Dellwo et al., 2003); VarcoC is calculated by dividing the standard deviation of duration of consonant intervals, by the average value of the duration of these intervals. Its analogue, VarcoV, was also proposed by Ferragne and Pellegrino (2004), amongst others. P. Wagner and Dellwo (2004) introduced YARD (an acronym for a tongue-in-cheek full name *Yet Another Rhythm Determination*); to enable objective comparison of speech rhythm regardless of speaking rate and individual differences, YARD standardises syllable durations by applying a z-score transformation, thereby allowing objective comparisons of rhythm across speakers and speech rates.

Nevertheless, the purely quantitative nature of these interval-based approaches has been criticised for reducing rhythm to timing alone (cf. Arvaniti, 2009). Alternatively, qualitative aspects of rhythm such as the marking of prominent events in the speech stream as well as different speaking situations, intentions and interaction should be taken into account (cf. Auer et al., 1999).

Taken together, the above classification problems illustrate that rhythm metrics do not provide an unambiguous basis for assigning languages to rhythmic classes on the grounds of duration properties alone. Furthermore, there are languages with ambiguous rhythmic nature, such as Catalan (cf. Payne, 2021; A. Wagner, 2017), Polish (Malisz, 2013; A. Wagner, 2017; P. Wagner, 2007, 2008), or Romanian (Payne, 2021). To some extent rhythmic category labels remain in use, although in the context of trends observed in the temporal patterns of languages, rather than in the sense of strict and non-overlapping rhythmic categories/classes. This makes Polish a particularly suitable test case: while its classification remains ambiguous, rhythm metrics have proven informative in distinguishing contrastive focus versus neutral utterances, as will be demonstrated in the present study (see Chapter 4, Section *Rhythmic metrics*).

At the same time, the difficulties outlined above do not nullify the usefulness of rhythm metrics (Ramus et al., 2003), as they continue to provide valuable insights into the global temporal properties of utterances. It should be noted, however, that rhythm metrics capture only one dimension of rhythmic structure and hence offer an indirect perspective on the mechanisms underlying speech rhythm (A. Wagner, 2017).

In short, rhythm metrics help to describe timing patterns, but they cannot account for the processes that generate them. To gain a deeper understanding of rhythm, it is necessary to go beyond descriptive measures. In the present study, therefore, rhythm metrics were complementary tools: they capture global timing tendencies and, as will be shown later, they also reveal differences between focus and neutral contexts in Polish. At the same time, the main analytical perspective is shifted towards articulatory correlates of rhythm and prominence as they can be investigated as jaw and lips movements rather than abstract intervals.

1.1.3 Beyond surface temporal patterns

Given the above limitations, alternative theoretical perspectives have been developed to account for speech rhythm beyond surface timing. These include, but are not limited to, two lines of

research: *localized lengthening*, and dynamical systems models based on *hierarchical coupling of oscillators*.

The first approach argues that speech timing is shaped mainly by segmental properties and local coarticulation rather than overarching temporal constraints from higher-level units (cf. White & Malisz, 2021). From this perspective, prosodic structure influences timing through localized lengthening at specific domains — for instance, at phrase edges or stressed syllables — which serves communicative functions rather than compensatory timing effects (White, 2014).

The second line of research based on the concept of *coupled oscillator*[1], proposes that cyclic processes such as syllable or foot oscillators are hierarchically or non-hierarchically coupled. Their interaction is assumed to generate the temporal organisation of speech without requiring surface isochrony (cf. Cummins & Port, 1998; Malisz et al., 2017; O'Dell & Nieminen, 2009). Some models are more complex and include non-hierarchically coupled subsystems (cf. Barbosa, 2007). In these theories, rhythm is the outcome of coupled oscillatory processes operating across multiple temporal scales.

Another line of research is Time Group Analysis (TGA; Yu et al., 2014), which conceptualises rhythm through the analysis of local timing relations between successive intervals. In contrast to global rhythm metrics, TGA focuses on the dynamic unfolding of temporal organisation across utterances and enables systematic exploration of large speech corpora. Cross-linguistic applications, including English, Mandarin, and Polish, have revealed robust differences in syllable duration patterns.

Together, these perspectives extend the understanding of rhythm beyond simple durational metrics, showing that rhythmic organisation can arise from multiple interacting factors. These insights open the way to broader perspectives on speech rhythm: some rooted in Metrical Phonology, others treating rhythm as an epiphenomenon or framing it as a descriptive metaphor.

1.1.4 Speech rhythm as structure, epiphenomenon, and metaphor

Within the structural tradition — already introduced earlier — metrical theory (Hayes, 1995; Liberman & Prince, 1977) proposes a new approach to stress; no longer was it conceptualized as an inherent property of syllabic segments. Instead, prominence is considered an emergent property of hierarchical prosodic structures. In **metrical trees**, syllables branch into weak (w) and strong (s) nodes, resulting in scalar stress levels for each syllable or, on higher organisational levels — morphosyntactic (or prosodic) units; the order can be either w-s or s-w. Another way of implementing metrical structure is through **metrical grids** which map the tree onto a layered grid. This results in an observable pattern of prominence in utterances. Metrical grids layer prominence tiers (syllable, word, foot, utterance), marking each with beats whose sum produces a stress value (cf. A. Wagner, 2017).

In contrast to these structural approaches, other theories treat rhythm as a metaphor rather than

[1]Any process that tends to repeat itself regularly can generally be described as an oscillator. Speech examples of oscillators include such basic phenomena as syllables or vocal fold vibration during voicing (O'Dell & Nieminen, 2009, p. 180).

a fixed property. Nolan and Jeon (2014), building on Dauer's critique of the stress-timing versus syllable-timing dichotomy (Dauer, 1983), argue that rhythm is best understood as emergent patterns of prominence alternations rather than as true temporal isochrony. The concept of rhythm in speech remains elusive, with evidence suggesting that neither strict coordination with an external clock nor clear-cut alternations of prominence fully capture natural speech patterns. Instead, rhythm emerges as a metaphorical phenomenon shaped by language-specific lexical, phonetic, and pragmatic factors rather than as a fixed universal structure.

Since indeed languages differ substantially in how prosody is manifested, one final theoretical framework discussed here is Jun's model of prosodic typology (Jun, 2014); it provides a convenient way to compare prosody across languages and has been employed in such works (e.g., Erickson & Kawahara, 2016; Smith et al., 2019). In Jun's model languages are distinguished by:

- prominence marking at lexical level: stress, tone, pitch accent, or none, meaning lack of distinctive marking;
- prominence marking at phrasal level: head-, head/edge-, or edge-prominence;
- macro-rhythm: a tonal rhythm, perceived by changes in F0; though gradual, it can be quantified on a three-point scale (Strong, Medium, Weak) based on the degree of regularity in alternating low and high tones.

By adopting a typological rather than definitional perspective, Jun's model complements theories discussed in this chapter by providing a practical framework for broader cross-linguistic comparison.

The approaches outlined above, represent only a subset of the many perspectives on speech rhythm, have significantly contributed to our understanding of the phenomenon. Nevertheless, across these perspectives — be they metrical, structural, metaphorical, or multidimensional — the analysis of speech rhythm has remained grounded primarily in acoustic evidence. While valuable, this reliance often overlooks the articulatory mechanisms underlying speech production. The following section therefore shifts the focus to articulatory aspects of prosody, including rhythm, which were the subject of research by Erickson (1998, 2004) for American English, Japanese (Erickson, 1998; Kawahara et al., 2014, 2015), Spanish (Erickson et al., 2015), Mandarin Chinese (Erickson et al., 2016), French (Smith et al., 2019), and Brazilian Portuguese (Erickson et al., 2024).

1.2 Towards articulatory accounts of speech rhythm

1.2.1 Limitations of acoustic approaches

Acoustic analyses, while invaluable for examining the speech signal, exhibit various considerable limitations. First, as noted earlier, **speech rate control** poses a challenge: acoustic data capture the effects of temporal variation but do not reveal the reason behind these fluctuations

in rate. The causes might include motor planning (e.g., cognitive load due to low-frequency words), emotional state, and overall well-being (e.g. stress, arousal, fatigue), or other constraints. These can mostly only be inferred indirectly. Secondly, even though considerable information about movement of articulators might be gained from acoustic analysis, its otherwise substantial interpretive power is limited by complex many-to-one relations between articulatory and acoustic events (Kochetov, 2020; Stone & Shadle, 2016). It is also because the actual articulator movements remain, mostly, unobservable (Kochetov, 2020). This limitation is particularly problematic because it prevents the investigation of **the coordination of articulatory gestures**, which is crucial for the analysis of speech rhythm, coarticulation, and prosody — acoustic signals do not capture the synchronisation and sequencing of articulator movements (Browman & Goldstein, 1992). It also limits practical applications in speech-language pathology and pronunciation training, where information about specific articulatory movements is essential but unattainable from the acoustic signal (Fuchs, 2019). Finally, there is **a need to complement acoustic data with articulatory measurements** to obtain a comprehensive understanding of speech dynamics and the relationship between articulatory movements and the resulting acoustic signal, which is particularly critical for research on speech rhythm and prosodic variation.

1.2.2 Brief history of articulatory research

The history of articulatory research extends back to antiquity, with the earliest phonetic descriptions — such as those in the Sanskrit grammar of Pāṇini — based on somatosensory perception and anatomical inference. Significant advances occurred in the 18th and 19th centuries, when general anatomical knowledge improved the accuracy of descriptions of the vocal tract (Spreafico & Vietti, 2022). However, it was not until the 20th century that imaging technologies enabled empirical research on the relationship between vocal tract configurations and acoustic output.

Development of various new techniques was driven by need. One such method was **electromagnetic articulography (EMA)**, which estimates the position and orientation of sensors in an electromagnetic field. It was introduced in response to a challenge in collecting extensive data on mandibular movements during speech. The problem was due to the lack of real-time tools for tracking mandibular movement (Hixon, 1971). During the same period another point tracking technique was being developed; **X-ray microbeam system (XRMB)**, which allowed for the dynamic tracking of articulator movement based on a different principle — focused radiation beams. The X-ray microbeam was, however, problematic due to its limited accessibility, complex system configuration, high cost, and radiation exposure. As a result, it was soon complemented by more accessible techniques. EMA provided a less invasive, safer, and more flexible alternative to X-ray systems and has since become one of the most widely used articulatory tools in laboratories worldwide. Research comparing EMA with the X-ray microbeam has demonstrated no significant differences in the accuracy of tracking articulator movement (Byrd et al., 1999), validating EMA's scientific reliability.

EMA is now employed in a wide range of linguistic contexts (cf. Kochetov, 2020) and has been applied to languages such as English, French, and Japanese, amongst others. In contrast,

articulatory phonetics of Polish has only recently begun to develop, with initial EMA-based studies in the early 2000s (e.g., Pompino-Marschall & Żygis, 2003) and more systematic applications appearing since 2009 (cf. Lorenc, 2016). Poland's only currently active EMA facility is now located in Warsaw using the Carstens AG501 model, supported by a multidisciplinary EMA research team.

1.2.3 Limitations of articulatory approaches

Despite the significant contributions of articulatory methods to phonetic science, their use in empirical studies remains relatively limited compared to acoustic approaches. This is largely due to the higher cost, technical complexity, and invasiveness of articulatory methods (Spreafico & Vietti, 2022), contrasted with the relative ease, speed, and scalability of acoustic data collection and analysis. As a result, researchers frequently opt for acoustic data alone — particularly when the research focus allows for inferences about articulation without direct observation — given that articulatory methods tend to be more time-consuming, resource-intensive, and may interfere with natural speech production by restricting speaker movement, or altering posture (Lin, 2021).

In practice, articulatory research tends to concentrate on a narrow subset of available techniques; the majority of articulatory studies rely on just three — electropalatography (EPG), ultrasound, and electromagnetic articulography (EMA) — while all remaining methods account for only 40% of usage (Kochetov, 2020). However, this imbalance reflects pragmatic considerations, including equipment availability, analytical complexity, and participant comfort rather than inferiority of less frequently used techniques.

Simply put, EMA data collection and analysis are time-consuming and technically demanding; for a more detailed account of the technique's limitations, see Rebernik et al. (2021), who provides an extensive literature review of 905 publications employing EMA.

Beyond these rather logistical concerns, articulatory methods also raise methodological and interpretative challenges. Instruments serve to collect data, yet they also actively shape the research process by influencing what can be observed and how experiments are structured. Limitations such as sensor placement variability, calibration issues, or user-induced error can introduce both systematic and random inaccuracies (Lin, 2021; Spreafico & Vietti, 2022). As a result, there is a growing interest in the development of portable and less intrusive tools. Advances in miniaturization and integration of multimodal systems offer the possibility of maintaining data quality without the above-mentioned limitations.

Nevertheless, all the listed issues do not diminish the value of articulatory research. On the contrary, articulatory data remain essential in addressing questions that cannot be resolved through acoustic analysis alone — some examples have already been provided in the previous chapter sections. They have logical applications in research focusing on the relationship between speech articulation and acoustics (Lin, 2021).

Taken together, the limitations discussed above highlight the importance of integrating both acoustic and articulatory data to obtain a more comprehensive and accurate account of speech production. Articulatory measurements are essential for uncovering the constellations of

movement patterns related to speech production — particularly within research areas where acoustic cues are insufficient or ambiguous. One such area concerns the role of individual articulators in shaping segmental and prosodic structure, with special attention to the jaw. The following section examines its articulatory and prosodic functions.

1.3 Articulatory approaches to speech rhythm: theoretical and cross-linguistic insights

1.3.1 Theoretical frameworks of articulatory rhythm

To understand the role of the jaw in prosodic organisation, it is necessary to situate empirical observations within broader theoretical approaches. In the following section three of them are outlined: Frame/Content theory (MacNeilage, 1998), Articulatory Phonology (Browman & Goldstein, 1992; Browman & Goldstein, 1986), and the Converter and Distributor (C/D) model (Fujimura, 2000).

1.3.1.1 The Frame/Content theory

The Frame/Content theory (MacNeilage, 1998) aimed to explain articulatory patterns in early speech development from an evolutionary perspective. It proposes that rhythmic, mandibular oscillations, rooted in ingestive cyclicities (e.g., chewing, sucking) create a syllable-sized frame which is sequentially layered with content — consonants and vowels. In other words, jaw opening and closing are seen as providing timing and structure for speech, within which specific articulatory gestures for consonants and vowels are implemented. MacNeilage refers to the syllable as a universal unit in speech.

Although this approach has been subject to critique, the debate is not directly relevant here, as they do not bear directly on the questions addressed in this dissertation. What matters for the present purposes is that The Frame/Content theory provides a useful foundation for linking jaw dynamics with syllable structure.

1.3.1.2 Gestural coordination and Articulatory Phonology

This approach stemmed from the need to bridge the gap between the linguistic and physical structure of speech within a unified phonological framework (Browman & Goldstein, 1992; Browman & Goldstein, 1986). In the principles of Articulatory Phonology, the basic units of phonological contrast are gestures; they are also abstract characterisations of articulatory events, each with its intrinsic time or duration. Within this framework, utterances are modeled as organised patterns, referred to as constellations, of such gestures that may overlap in time. Articulatory actions are structured around global vocal tract configurations, with gestures specified using related tract variables, such as constriction location and its degree in the oral tract, rather than as isolated movements of individual articulators.

In relation to the Frame/Content theory, Articulatory Phonology can be seen as a framework specifying how the segmental content is organised through gestural coordination. Neither of these accounts, however, provide explicit accounts on temporal organisation of speech gestures. This dimension was addressed in the Converter and Distributor (C/D) model (Fujimura, 2000), discussed below.

1.3.1.3 The Converter and Distributor (C/D) model

The Converter and Distributor (C/D) model offers a distinctive account of how prosodic structure interacts with articulation — it proposes that speech gestures are coordinated by syllable pulses. Such an approach enables embedding rhythmic structure into articulatory planning and, therefore, conceptualising rhythm at the articulatory level.

The Converter and Distributor (C/D) model is a phonetic theory for representing utterances in conversational speech, developed within the framework of nonlinear phonology. The model rests on the assumption that speech is not a simple string of phonemic segments; this view is supported by articulatory evidence and by psycholinguistic phenomena such as spoonerisms, where initial sounds in a phrase are swapped (e.g., *pack of lies* → *lack of pies*). Such errors suggest that speech sounds are represented as discrete and movable units, rather than as a continuous chain.

In the C/D model, similarly to the Frame/Content theory, the basic segment of speech is a syllable. The first stage of the phonetic implementing process in the C/D model — the Converter — is responsible for transforming prosodic input into parameters that guide the coordination of speech gestures. **The parameters are assigned numerical values indicating the prominence of specific metrical units to reflect utterance's prosodic structure**. The Distributor selects elemental speech gestures for the syllable margins corresponding to each component (onset, coda, prefixes or suffixes). Once the base function is specified, the process enters its final phase: Actuators retrieve impulse response functions (IRFs), which are shaped by Control Function Generators and mapped onto the base function. Then these control signals are sent to the Signal Generator, completing the phonetic implementation process.

It is crucial to highlight that within the C/D model **the prominence of each syllable is specified in advance as part of the metrical structure of an utterance and is encoded in the magnitude of a syllable pulse**. This magnitude is reflected articulatorily as relative jaw opening. Amplified speech gesture leads to enhanced vocalic gestures and, subsequently, acoustic changes including an increase in sonority/intensity as well as changes in vowel quality.

The predictions of the C/D model, particularly concerning the link between jaw opening and prominence, have been explored empirically in studies by Erickson and colleagues.

1.3.2 Empirical evidence from English

While numerous articulatory studies of English have documented jaw displacement patterns in speech, this section presents some of these investigations that illustrate gradual change from

general articulatory observations to rhythm-specific theoretical frameworks.

Oshimat and Gracco (1992) examined jaw movements during the production of vowels and consonants at normal and fast speaking rates with the assumption that articulation is coordinated. They used electromyography (EMG) and registered jaw movement via optoelectronic transduction. Their results demonstrated that the mandible is integral to vowel production in conjunction with the tongue and assists other articulators in consonant formation. Jaw movements not only coordinate positions of other articulators but also set the size and geometry of the oral cavity, affecting overall sound production. The amplitude of jaw opening systematically differentiates vowel qualities, with greater openings for low vowels and smaller openings for high vowels. However, these differences diminish at faster speech rates, suggesting that jaw actions are organised around syllabic movement cycles — coordinated opening and closing patterns for syllable production — rather than specific articulatory targets. When jaw displacement is reduced, the tongue compensates to maintain vowel quality. Interestingly, in summary of their findings Oshimat and Gracco used the term *phonetic gesture*, even though they have not cited works of Browman and Goldstein (1986), suggesting independent convergence on similar theoretical concepts.

Erickson (1998, 2002) and Erickson and Fujimura (1996) investigated the relationship between jaw opening and contrastive emphasis which later constituted a starting point for further analyses of the jaw's role in speech rhythm organisation. The initial works demonstrated that in American English jaw opening is greater for syllables that are more prominent, no matter their position, low or high. An interesting observation on relation on articulation and vowel frequency formants was also reported — **to emphasise a vowel, speakers exaggerated speech gestures**. A joint jaw-tongue movement was observed:

- in case of low vowels, the jaw's aperture was greater while the tongue dorsum position was lower and more back;
- in case of high vowels, **in contrast**, the tongue dorsum position was higher and more to the front.

Such more pronounced movements were reflected acoustically, in change in F1 and F2 formants, since these are closely related to tongue dorsum position, F1 to its vertical (high-low) position and F2 — horizontal (front-back) respectively. Emphasis for high vowels is associated with low F1 and a higher F2, producing a **greater separation** between the formants. In contrast, for low vowels formant coordinates are opposite with higher F1 and a lower F2, resulting in a **closer, more compact, spacing** between these formants. The increased jaw opening observed may contribute to enhanced sonority, although this was not directly measured in the study (Erickson, 2002). The authors offer possible interpretation of the exaggerated articulation within the framework of hyperarticulation (cf. Lindblom, 1990) and within the Converter and Distributor (C/D) model (Fujimura, 2000).

Building upon findings on jaw opening and prominence, Erickson et al. (2012) investigated jaw kinematics by linking patterns of mandibular oscillations with the metrical structure of English rhythm. Rhythm in that particular approach was reflected as temporal patterns of syllable prominences (cf. Kohler, 2008) distributed within prosodic grouping such as foot or

phrase. The notion of rhythm was differentiated from the melody/intonational patterns of an utterance: the former relates to variation in stress and the latter — in pitch.

The study reports findings obtained from 3 speakers recorded using EMA (Carstens AG500) and 1 person from previous sessions recorded with XRMB. The results support the conclusion that, indeed, jaw displacement seems to be an articulatory organiser of rhythm. Moreover, it is significantly correlated with F1: lower jaw positions are associated with higher F1 values. Finally, patterns of jaw movement — and, to some extent, F1 — appear to align with the metrically generated syllable stress levels. These results suggest that the rhythmic structure of English might be described in terms of Metrical Phonology as the magnitude of metrical prominence correlates well with the magnitude of jaw displacement.

These findings have motivated similar research examining whether such articulatory patterns are observable in other languages that exhibit distinct phonetic realisations of metrical organisation compared to English. Although Metrical Phonology has provided a robust framework for describing cross-linguistic prosodic patterns, the articulatory realisation of metrical structure remains less well studied; even more so in languages other than English.

1.3.3 Cross-linguistic insights

An extension of this research has been to examine whether similar articulatory patterns emerge in languages with a fundamentally different prosodic structure. Cross-linguistic research included Japanese (Kawahara et al., 2014), Spanish (Erickson et al., 2015), Mandarin Chinese (Erickson et al., 2016), French (Smith et al., 2019), and Brazilian Portuguese (Erickson et al., 2024).

1.3.3.1 Japanese

Erickson and Kawahara (2016) report that Japanese is often considered a pitch-accent language which lacks stress. This makes it particularly interesting for studying jaw displacement — this raises the question of whether one should expect its significant lowering.

In their study, Kawahara et al. (2015) explain that Japanese is characterised by pitch accent that is primarily manifested through a fall in F0, rather than via intensity or duration. The accented vowel carries a high tone (H) followed by a low tone (L) vowel. Although the presence of pitch accent in Japanese may influence jaw displacement due to its association with higher F0 peaks — as accentual H tones have been shown to reach greater F0 values than non-accentual ones — its absence would not be unexpected. The results of an EMA study indicated that while Japanese pitch accent does not yield robust articulatory effects, **it does display patterns of jaw displacement**; the jaw opening is greater at phrase-edge syllables (Kawahara et al., 2014) which cannot be explained by domain-edge lengthening (Erickson & Kawahara, 2016). Moreover, similarly to English, the increased jaw opening in these syllables was reflected acoustically, in higher F1 (Kawahara et al., 2015).

1.3.3.2 Spanish

Spanish has been classified as a head prominence language with an initial rising pitch pattern (Jun, 2014), suggesting that the strongest phrasal stress tends to occur on the initial syllable of a phrase, while the final syllable is typically the weakest. At the same time, nuclear stress is generally realised on the final content word of the phrase. Studies on Spanish articulatory patterns remain comparatively limited; Erickson et al. (2015) recorded 3 speakers of Salvadorian Spanish using an ultrasound probe under the chin and Erickson and Niebuhr (2023) recall on EMA session collected from 1 speaker for Mexican Spanish. Due to a very limited sample, the results remain inconclusive, although jaw displacement was the most pronounced in the beginning of the phrases in the EMA study and observed in some of the ultrasound sessions. This, tentatively, leads to the conclusion that Spanish may display its own articulatory patterns following its metrical prosodic structure organisation. Such a claim, however, needs further and more robust research.

1.3.3.3 Mandarin Chinese

Mandarin Chinese represents another language with a prosodic structure that differs substantially from English. As noted by the authors in the introduction to their EMA study (Erickson et al., 2016), it is a tonal language in which lexical contrasts are primarily realised through F0. Mandarin has four lexical tones (T1 to T4) and a neutral tone, produced by modulating vocal fold tension, resulting in systematic variations in fundamental frequency (F0); in particular, T3, the low rising tone, is often associated with a vocal fry–like voice quality. In their framing of the study, the authors also emphasise phrase-level stress patterns, with duration emerging as the most important acoustic cue: the longest syllable is the final one, initial syllable is shorter, and medial ones are the shortest. Aligning with duration-based cues, structuralist accounts have proposed an iambic pattern in Mandarin, with the final syllable as the most prominent. However, authors point that more recent theoretical work suggests a trochaic, left-headed foot structure, applicable across both phrases and compounds. EMA research results indicated that Mandarin Chinese, though a tonal language with lexical contrasts realised primarily through F0, displays jaw displacement patterns aligned with phrasal stress structure. Phrase-final syllables show the greatest jaw displacement, and sometimes phrase-initial syllables as well. These articulatory patterns are acoustically reflected in increased F1 and duration, but not in F0 or intensity, suggesting that **jaw displacement encodes post-lexical (phrasal) prominence independently of tone** (Erickson et al., 2016).

1.3.3.4 French

French, in contrast to English, lacks word-level prominence and nuclear stress (cf. Jun, 2014), although the status of the latter is still debated in the literature (cf. Cole et al., 2019). The Accentual Phrase is the smallest domain of accent assignment and typically contains one content word along with preceding clitics; there are typically multiple Accentual Phrases in an intonational phrase (cf. Cole et al., 2019). Prominent syllables occur immediately before an Accentual Phrase boundary and the two are closely related — French is described as an

head/edge prominence language (cf. Jun, 2014). This phrase-level boundary is marked both acoustically and articulatorily on the final non-schwa syllable preceding a boundary, and — optionally — at the beginning of a phrase. Acoustic marking is displayed by changes in F0 patterns and increased duration at the end of each Accentual Phrase, and the end of an Intonational Phrase which has also greater intensity (Smith et al., 2019). Importantly, this final accent does not signal focus or information-structural prominence. Rather, it is structure-marking and serves to delimit prosodic constituents. Thus, prosodic prominence in French is not used to highlight new or contrastive information, unlike in English or Spanish (Cole et al., 2019).

EMA research (Smith et al., 2019) indicated that L1 French speakers typically produce syllables in final Accentual Phrase (AP) positionwith expanded articulation — characterised by greater jaw displacement, higher F1 values, and longer durations — which aligns with the classification of French having head/edge prominence (Jun, 2014). Smith et al. (2019) also reported increased F1 and longer duration for the mid and low vowels /ɛ/ and /a/ in final syllables at the end of AP and IP.

1.3.3.5 Brazilian Portuguese

Brazilian Portuguese is a stress-accent language in which lexical stress can fall in final, penultimate or antepenultimate position; the most frequent, canonical stress pattern being the penultimate stress. The main acoustic correlates of lexical stress are increased duration of the stressed syllable (or syllable-sized unit) and an intensity drop from the stressed to the post-stressed position (Barbosa & Albano, 2004). Nuclear stress usually falls on the last word in the utterance. When a clash of lexical stresses in the same phonological phrase occurs, usually the first one becomes deaccented. F0 tends to rise at the end of each phrase, except in declarative utterances, where it typically falls on the last stressed syllable of the final word. According to Jun's (2014) prosodic typology, Brazilian Portuguese is a head-prominence language.

For Brazilian Portuguese the jaw displacement data have been collected from 37 speakers using the MARRYS (Manibular-Action Related Rhythm Signals) cap, a portable helmet with two bending sensors for bidirectional mandible movement measurement (Erickson et al., 2024). The results suggest the following pattern of mandible movement for Brazilian Portuguese speakers: generally, for each stressed syllable, the mandible is in a lowered position followed by a rapid closing movement. The degree and timing of mandibular lowering vary depending on the location of lexical stress within the word (final, penultimate or antepenultimate position). Each phonological phrase typically contains at least one stressed word which tends to occur towards the end of the phonological phrase. The nuclear accent is marked by **increased lowering on the stressed syllable and by a steep, strong jaw closing**. Moreover, mandible movement differs between non-terminal phrases and final nuclear accent, with the former involving distinct jaw opening and the latter characterised by distinct jaw closing.

1.3.3.6 Summary of research on different languages

As presented in previous sections, each researched language displayed its own distinctive patterns of utterance-level prominence, which are reflected in jaw displacement patterns. These cross-linguistic differences in prosodic typology, acoustic correlates, and articulatory realisations are summarised in Table 1.1.

In English, such a pattern follows a prominence hierarchy, with the greatest mandible opening occurring on nuclear stress syllables, followed by phrasal and foot-level stress. Mandarin Chinese and French demonstrate a phrase-final prosodic alignment, with jaw lowering most pronounced at the end of each phrase, and the largest displacement often observed at the utterance-final syllable. Japanese, by contrast, displays edge-related patterns of jaw displacement, with increased jaw opening observed at phrase boundaries rather than restricted to phrase-final position. In Brazilian Portuguese, jaw displacement is characterised by jaw lowering during each stressed syllable, followed by a rapid closing movement.

The above findings support the view that the jaw functions as the articulatory "beat" of speech, while the vocal folds provide the "melody" through pitch modulation — a perspective consistent with Fujimura's (2000) Converter/Distributor (C/D) model of articulation. The model highlights the coordination of prosodic structure with articulatory implementation, with rhythmic timing grounded in oscillatory jaw movements.

Cross-linguistic comparisons further reveal systematic differences in how jaw opening is prosodically anchored: French and Mandarin Chinese tend to show phrase-final expansion, reflecting right-edge prominence, whereas Spanish appears to favour phrase-initial opening, suggesting left-edge anchoring. These distinctions emphasise that jaw dynamics not only encode rhythmic structure in a biologically grounded manner but are also shaped by language-specific prosodic systems. Extending this line of research to Polish — where the rhythmic structure is still contested — offers the possibility of uncovering articulatory patterns that both enrich cross-linguistic typology and clarify the prosodic organisation of Polish itself.

1.3.4 Polish: phonetic and phonological background

The case of Polish might be particularly relevant given that it differs from the previously examined languages in both its phonotactic structure and phonological inventory. At the same time, it shares a predominantly paroxytone stress pattern with Brazilian Portuguese. Polish, however, allows phonotactically complex words with dense consonant clusters (e.g., *pstryk*, *pstrąg*, *wstrząs*, *wstręt* with a CCCCVC structure[2]), which distinguishes it from many other Indo-European languages and gives it a markedly consonantal profile. In contrast, Brazilian Portuguese favours open CV syllables, with only limited consonant clusters and a broader vowel inventory. These structural properties, combined with its prosodic organisation, suggest that Polish might exhibit unique articulatory correlates of prominence and rhythm.

At the same time, the rhythmic classification of Polish remains debated. Previous studies

[2]Such structure is present in common, everyday words and not limited to exotic or infrequently used ones — the examples given being Eng. 'snap', 'trout', 'shock', 'disgust' respectively

Table 1.1. Cross-linguistic comparison of prosodic typology (based on Jun, 2014), main acoustic correlates of prominence, and jaw displacement patterns described in the research.

Language	Prosodic typology	Acoustic correlates	Jaw displacement
English	Word prosody: stress Prominence type: head Macro-rhythm: medium	F0 (intonation), duration, intensity	Greatest opening on nuclear stress; correlated with F1. No strong correlation was found between jaw displacement and syllable duration
Japanese	Word prosody: tone/lexical pitch accent Prominence type: head/edge Macro-rhythm: strong	F0 (high tone followed by a low tone); little effect of duration/intensity	Limited effect of pitch accent on jaw displacement; greater opening at phrase edges; correlated with F1
Spanish	Word prosody: stress Prominence type: head/edge Macro-rhythm: strong	Duration, intensity, nuclear stress at phrase-final	Tentative: greater displacement phrase-initially; evidence limited (small samples)
Mandarin	Word prosody: stress + tone/lexical pitch accent Prominence type: head Macro-rhythm: weak	Lexical tones via F0; phrase prominence mainly by duration	Largest displacement phrase-final (sometimes initial); correlates with F1 and duration, not F0/intensity
French	Word prosody: none Prominence type: head/edge Macro-rhythm: strong	F0 (boundary tones), duration, intensity at AP/IP boundaries	Displacement increases towards sentence end; stronger on final syllables; correlates with F1 and duration
Portuguese	Word prosody: stress Prominence type: head Macro-rhythm: strong	Duration, intensity drop after stressed syllable; F0 rising/falling phrase-final	Lowering on stressed syllable with rapid closing; nuclear stress marked by strong lowering + steep closing

variably position Polish between syllable-timed and stress-timed prototypes, often describing it as ambiguous or mixed (Malisz, 2013; A. Wagner, 2017; P. Wagner, 2008). A systematic articulatory investigation therefore has the potential to clarify this classification and to contribute new evidence to cross-linguistic accounts of rhythm.

1.3.5 Theoretical predictions for Polish

Building on cross-linguistic results and theoretical frameworks of articulatory rhythm, one of the central hypotheses adopted in this doctoral dissertation is that jaw position correlates with the metrical structure of Polish speech. Prominent syllables are expected to be associated with stronger jaw displacement relative to the occlusal plane. Trajectories of jaw movement should coincide with utterance-level prominence.

Given the complex phonotactic profile of Polish, its ambiguous rhythmic classification, and yet highly predictable utterance-level accent (see the next Chapter 2 for more details), it is further expected that observable articulatory patterns may emerge — patterns that go beyond what can be inferred from acoustic evidence alone. Quantitative analyses combining articulatory (EMA) and acoustic measures might make it possible to identify these hypothetical features and situate Polish within the broader typology of prosodic rhythm.

To sum up, articulatory investigation of Polish prosody offers the potential to shed new light on long-standing debates about its rhythmic status, while also providing valuable data for comparative articulatory studies. The next chapter addresses another debatable issue in Polish prosody — the nature of its stress and accent. Since there is a close relationship between accent and rhythm, it is crucial to outline Polish stress and accent prior to turning to the details of the methodological framework of the present study.

Chapter 2

Stress and accent in Polish

This chapter draws a distinction between word-level prominence, referred to as **stress**, and phrase- or utterance-level prominence, referred to as **accent**. Both these phenomena play a key role in the organisation of speech operating at their respective levels, yet they might also overlap, affecting the meaning, rhythm, and intonation of utterances (cf. Gordon, 2014; Kohler, 2008).

Introducing such a distinction is essential as, although both relate to the relevance of certain elements in speech, they serve different functions. Word-level stress bears lexical or morphological relevance whereas accent — particularly nuclear accent — operates at the utterance level, serving pragmatic and informational functions, such as highlighting focus, contrast, or new information (cf. Gussenhoven, 2004; Jassem & Gibbon, 1980; Ladd, 2008; van der Hulst, 2002).

> The main distinction to be made is between accent as a textual, concrete, observable category and stress as an abstract, possibly lexical, analytic category (Jassem & Gibbon, 1980, p. 8–9).

Moreover, depending on the elicitation method, lexical stress patterns may be skewed for various reasons (cf. Warner, 2021). One such case involves words uttered in isolation: when a word constitutes a complete utterance, the resulting stress patterns might, in fact, reflect phrase-level prominence rather than purely lexical stress (Gordon, 2014). Another issue arises when target words are analysed without regard to their position within the utterance, which may yield inconsistent conclusions, as their prosodic realisation can vary significantly depending on discourse placement.

The rhythmic structure of speech emerges from the recurrent, cyclic pattern of prominence across speech units, and this prominence operates and is expressed at multiple levels (Kraska-Szlenk, 1995; Rubach & Booij, 1985). Lack of distinction between stress and accent might complicate the analysis of the interaction between prosody and meaning — especially in languages like Polish, with a fixed stress and yet robust accent placement possibilities depending on discourse context or pragmatic function (cf. Francuzik et al., 2004). Moreover, some previous studies attempting to identify the correlates of stress in Polish have not taken into account for higher-level prominence phenomena (cf. Ćwiek & Wagner, 2018).

2.1 Word-level prominence: lexical stress in Polish

2.1.1 Position

Lexical-level prominence in Polish is highly regular: the general primary stress pattern is penultimate, with some well-described exceptions with mostly antepenultimate stress (Dłuska, 1950; Dogil, 1999; Jassem, 1962; Newlin-Łukowicz, 2012; Ostaszewska & Tambor, 2000; Wierzchowska, 1971). Stress in Polish is regarded as quantity-insensitive — unlike quantity-sensitive systems, prominence-level stress position is conditioned neither by the presence of long vowels, consonant clusters in coda position, nor other factors that contribute to syllabic heaviness (Dogil, 1999; Kraska-Szlenk, 1995). Longer words might employ secondary and tertiary stress (Kraska-Szlenk, 1995; Łukaszewicz, 2018; McCarthy & Prince, 1993; Rubach & Booij, 1985) manifesting rhythmic repetitions.

> Secondary and tertiary stress, usually not formally distinguished in the literature, as in Rubach and Booij (1985) and McCarthy and Prince (1993), falls on every other syllable of the word, provided it is not the primarily stressed syllable neither the syllable immediately preceding it (Kraska-Szlenk, 1995, p. 13).

Such a system is described as bidirectional with internal lapses (McCarthy, 2003; McCarthy & Prince, 1993). However, this approach was challenged by Newlin-Łukowicz (2012), who found no consistent acoustic evidence of prominence for either secondary or tertiary stress in Polish, implying that they may either not exist at all, or — if they do — might be marked by non-acoustic parameters (e.g., articulatory features), though no strong evidence supports this. These findings are in line with perception experiments by Steffen-Batóg (2000), who demonstrated that native Polish listeners cannot reliably recognise a secondary stress. The very existence of secondary and tertiary stresses thus remains a debated issue in Polish. In this regard, Łukaszewicz (2018) provided empirical evidence that Polish does indeed exhibit iterative secondary and tertiary stresses, showing that they are cued not by vowel-based correlates but by consonantal rhythm, in particular by lengthening of onset consonants and/or shortening of the preceding vowel. This strongly supports the view of Polish as a bidirectional trochaic system with internal lapses, contrary to the earlier conclusions of Newlin-Łukowicz (2012).

2.1.2 Markers

A range of acoustic features have been examined as potential correlates of prominence in Polish in many theoretical and experimental works. Traditionally, Polish word-level stress has been described as dynamic (Dłuska, 1950; Wierzchowska, 1967, 1971) — stressed syllables are articulated with greater vocal effort resulting in more acoustic energy being placed onto them. Therefore, **intensity** has frequently been identified as a key acoustic correlate of prominence, both in early (Dłuska, 1950, 1976; Wierzchowska, 1967, 1971) and more recent studies, where it was found to be a more salient correlate than F0 or duration (Łukaszewicz & Rozborski, 2008). Newlin-Łukowicz (2012) observed that in simple words, both the initial and penultimate

syllables exhibit significantly higher maximum intensity. A similar pattern was found in compounds.

Jassem (1962; 1968) considered **pitch movement** the most salient correlate of word-level stress in Polish and postulated that stress is melodic or tonal, as opposed to the earlier notion of its dynamic nature. Building on these findings, Dogil (1999) further investigated prominence markers in Polish. In his analysis, he took into consideration earlier observations on metrical relabelling (Dogil, 1980): namely, under narrow focus, the prominence values of primary and secondary stresses may be reversed, allowing a non-default syllable (e.g., the initial one) to assume the role of the most prominent syllable in the prosodic structure. Therefore, Dogil (1999) presented target words in utterances under narrow, broad, and no focus position in his subsequent study. Upon completing the study, the main observation was that in no focus position, the primary stress in the target word is characterised by the highest F0 with a sharp pitch slope, which aligns with Jassem's (1962) findings. Nevertheless, considering all the contexts, Dogil concluded that **Polish lexical stress does not exhibit consistent phonetic cues of its own**, but rather serves as a structural placeholder for intonational pitch accents, marking prosodic positions where nuclear or pre-nuclear accents may be realised depending on the utterance's information structure. More recent analyses suggest a nuanced interaction of pitch with intensity and duration, especially when data is controlled for prosodic structure (cf. Newlin-Łukowicz, 2012).

Another phonetic-acoustic measure that has been considered a prominence correlate was **duration** (Jassem, 1962; Malisz & Wagner, 2012; Malisz & Żygis, 2018; Newlin-Łukowicz, 2012; Wierzchowska, 1967). Jassem (1962), however, emphasises that its role may only be secondary. In a more recent account, Ćwiek and Wagner (2018) added further nuance to this perspective by demonstrating that syllable duration is only increased if both lexical stress and phrase accent occur together, contradicting previous findings by Newlin-Łukowicz (2012). In general, vowel **duration** appears to have a minor, albeit measurable, effect — ratios of stressed to unstressed vowel durations range from 1.17 to 1.45 (see Table 2.1):

Table 2.1. Stressed to unstressed vowel duration ratio in selected studies

Author	Stressed to unstressed vowels ratio
Jassem (1962)	1.17
Nowak (2006)	1.22
Klessa (2006)	ca. 1.2
Rojczyk (2019)	1.45

Malisz and Wagner (2012) demonstrated that duration variability, along with pitch and intensity, covaries with perceptual prominence in Polish. Malisz and Żygis (2018) found that both primary and secondary stress significantly affect overall syllable duration compared to unstressed syllables. It is worth noting that secondary stress in Polish falls on the **initial syllable** of words consisting of at least three syllables (Kraska-Szlenk, 1995; Rubach & Booij, 1985). Newlin-Łukowicz (2012) identified a robust secondary stress in compound words, correlated with higher values of maximum pitch, maximum intensity, and onset length. This effect, however, may result from **word-initial strengthening** rather than stress per se.

Other potential prominence correlates propositions include **F0 range** (Dogil, 1999; Jassem, 1962; Newlin-Łukowicz, 2012); **F0 maximum** (Newlin-Łukowicz, 2012; Wierzchowska, 1967), and **spectral tilt** (Crosswhite, 2003). Malisz and Wagner (2012) found no evidence for the latter remarking, however, this should be investigated further.

Overall, earlier studies indicate that prominence is most reliably marked by intensity, pitch movement, and duration. However, the acoustic manifestation of prominence in Polish is comparatively subtle — Ćwiek and Wagner (2018) found that a reliable prominence marking occurs on phrase-level accents only; these findings are in line with Dogil's results (1999). Consequently, it can be assumed that prominence is realised through a set of acoustic correlates that operate in a cumulative or complementary way, a perspective further supported by more recent research.

2.1.3 Function

Polish stress is typically considered as parsed into a trochaic foot alignment: strong-weak syllable, and this foot is always aligned with the right edge of Prosodic Word (Dłuska, 1976; Kraska-Szlenk, 1995; Rubach & Booij, 1985). However, since each syllable in Polish can only have one vowel and there are no syllables without a vowel (with some degree of exception in case of fast and lax pronunciation):

> To account for the stress pattern of Polish words it suffices to know only how many vowels they contain. Vocalic peaks are the sole stress-bearing units and the only units (Kraska-Szlenk, 1995, p. 7–8).

From the functional perspective, the main role of the primary stress in Polish is **delimitation**, signaling an upcoming end of a phrase (Jassem, 1962; Ostaszewska & Tambor, 2000; Wierzchowska, 1967) or enabling interpretation of the meaning of a speech stream. Consider the following example: [POraDNIA] (= *pora dnia*, Eng. 'time of day') vs [poRADnia] (= *poradnia*, Eng. 'clinic, counseling centre'). In the first case, the prominence falls on PO which is the penultimate syllable of the word *pora*, whereas in the second case it falls on RAD, the penult of the word *poradnia*. The **culminative function** — signaling the number of lexical units — is also present, but it is not of such importance as in languages with less strict prominence placement rules.

Psycholinguistic studies indicate that speakers of fixed-accent languages, such as Polish, are less sensitive to variations in stress placement compared to speakers of free-accent languages. This reduced sensitivity — known as **stress deafness** (Peperkamp et al., 2010) — stems from the expectation that stress placement is predictable and non-contrastive in function. Rather than, as already mentioned, serving to distinguish lexical meaning, stress in Polish primarily plays a delimitative role, signalling the boundaries of prosodic or lexical units. Several studies have indeed shown that Polish listeners exhibit some degree of stress deafness (cf. Domahs et al., 2012; Peperkamp et al., 2010; Steffen-Batóg, 2000).

2.2 Phrase-level prominence: accent in Polish

The body of literature addressing accent in Polish remains noticeably less extensive than that addressing lexical stress. Ćwiek and Wagner (2018) pointed out that numerous studies attempting to establish correlates of stress in Polish did not account for higher levels of prominence. Moreover, publications on this topic often fail to clearly distinguish between these two phenomena, leading to an intertwining. Despite the limited availability of literature, the subsequent sections aim to extract and synthesise the insights it offers regarding utterance-level prominence in Polish.

In perceptual terms, accent functions as a salient landmark for the listener, marking a focal point within the informational structure of an utterance (Cutler, 1984). In neutral, no focus context, the utterance accent in Polish is typically placed on the penultimate syllable of the last word — the rightmost stressed syllable in a word or phrase tend to carry the strongest stress (Kraska-Szlenk, 1995; Rubach & Booij, 1985). While these observations primarily derive from elicited and isolated speech material, rather than fully spontaneous production, it is noteworthy that such patterns have also been documented in analyses of a robust semi-spontaneous speech represented e.g., in the PoInt database (Karpiński & Kleśta, 2001). Francuzik et al. (2002) demonstrated that the nuclear accent typically fell on the penultimate syllable of the last word in the intonational phrase. Nonetheless, exceptions can be observed. Depending on the information structure of the utterance, as well as speaker-specific factors and the speaker's communicative stance, phrase-level prominence may occur on different parts of the phrase. An example of this flexibility is narrow focus, where the accent shifts to the word or phrase that is in focus, regardless of its position within an utterance, as shown by Eschenberg (2008); the focused element consistently carries prosodic prominence.

Research by Ćwiek and Wagner (2018) demonstrates that in Polish, acoustic prominence is reliably marked only at the phrase-level, while lexical stress lacks independent acoustic correlates. F0, intensity, and spectral balance are used primarily to express phrase-level accents, and **increased syllable duration occurs only when accent and stress coincide**. These findings contribute to the view that in fixed-stress languages, lexical stress functions as an anchor point for phrase-level prominence. Authors propose explanations within the framework of Hyper- and Hypo-articulation theory, known as H&H theory (Lindblom, 1990) according to which this lack of prominence marking may be explained by speakers' need to modulate articulatory effort and adjust between hyper- and hypospeech. Since **Polish lexical stress is not distinctive**, as it serves mostly a deliminatory function, its marking is not needed for comprehension; it might have an independent purpose such as signaling phrase accent and information structure. Redundant acoustic cues may be a result of local hyperarticulation or from speakers adapting prominence marking to the communicative context, while still maintaining production economy.

Chapter 3

Method

Building on the discussion in Chapter 2, where speech rhythm was described as emerging from cyclic patterns of prominence, the present chapter advances this perspective by investigating, for the first time in Polish, their articulatory correlates using electromagnetic articulography (EMA). The present study combines complementary phonetic methodologies, integrating articulatory data with phonetic-acoustic indicators. Specifically, the study draws on the following methodological components:

- **Analysis of articulatory data** — acquired by using EMA, an advanced technique for the recording of speech gesture movements. The EMA setup allows for precise collection of the position of articulators, including the jaw, lips, and the degree of mouth opening in real time.
- **Phonetic-acoustic analysis of audio recordings** — in addition to articulatory data, this part focuses on acoustic components of prominence, contextualising the articulatory findings within the framework of selected aspects of state-of-the-art speech rhythm research.

Such an integrated approach enables a multidimensional analysis in which correlates of speech rhythm are examined primarily from the perspective of physiological speech production, using EMA, and are complemented by acoustic-phonetic properties of rhythmic patterns in utterances. This may provide a more comprehensive picture of the mechanisms underlying Polish speech rhythm, representing a step towards a better understanding of their interaction.

For the sake of simplicity, the terms *utterance* and *sentence* are used interchangeably in this and following chapters, although these concepts are not strictly synonymous. The present material consists of short, controlled sentences functioning as single prosodic units, which justifies this terminological simplification.

3.1 Experimental plan

This section introduces the overall experimental framework. The study was designed to investigate whether articulatory parameters potentially correlated to speech rhythm — particularly

jaw displacement and lip aperture — align with the realisation of prominence in Polish. The research questions focus on whether utterance-level stress and contrastive focus have measurable effects on articulatory displacement and whether these effects are consistent across different vowel categories and speakers.

The design is based on connected factors:

1. Polish stress and accent are highly predictable, which provides a controlled environment for comparisons between stressed/accented and unstressed/unaccented positions in an utterance. However, predictability alone may not be sufficient; its value here lies in reducing variability that could obscure articulatory effects. This reduction was ensured through the use of uniformly constructed target words (see Section *Material* for more details).
2. Electromagnetic Articulography (EMA) allows for detailed observation of jaw and lip positions, and their movements, which may help to reveal underlying rhythmic mechanisms.

By combining these two, the study addresses both aspects of prosodic typology and the practical challenges of precise measurement.

3.2 Research questions and hypotheses

The primary objective was to examine whether jaw displacement for a vowel in the penultimate syllable of the final word of an utterance — i.e., the expected position of sentence accent in Polish — is significantly greater than the displacement of the same vowel in other positions within the utterance (see Figure 3.1). A further objective was to determine whether jaw displacement increases in contexts of focus, and whether lip aperture and overall mouth opening patterns show systematic variation under these prominence conditions.

Jaw displacement was measured as the distance (in mm) of the sensor attached immediately below the lower incisors relative to the occlusal plane. The upper lip sensor and lower lip sensors were attached immediately above the vermilion border (more details are in Section *Apparatus*).

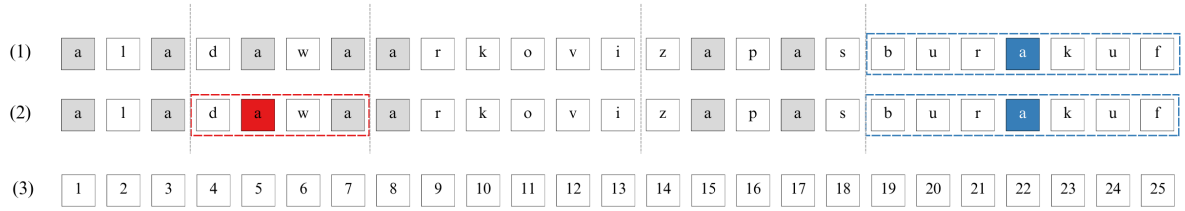

Figure 3.1. In the above examples, the vowel /a/ is realised eight times throughout the utterance *Ala dała Arkowi zapas buraków* /ala dawa arkovi zapas burakuf/ (Eng. 'Ala gave Arek a supply of beets'). Statistical tests are run to determine whether the jaw displacement is increased (1 & 2) in the predicted *Accent* position in **neutral** condition (marked in blue) and (2) in the *Focus* position in **focus** condition (marked in red). (3) Each phonemic segment is indexed. **Note on transcription:** The transcription of all the utterances in the dissertation reflects the predominant manner of realisation among all speakers. That is, in some cases, the presented transcript does not reflect the expected manner of realisation described in Polish phonetics textbooks. E.g., in the above example, backward interword voicing would be expected — /zapaz burakuf/ (cf. e.g., Madelska & Witaszek-Samborska, 1991).

3.2.1 Research questions

1. To what extent does vertical jaw displacement differ for the penultimate vowel of the final word — the expected placement of utterance accent — compared to other positions within the utterance?
2. To what extent does lip aperture increase for the penultimate vowel of the final word — the expected placement of utterance accent — compared to other positions?
3. To what extent does the magnitude of jaw displacement in stressed/accented vs. unstressed/unaccented vowels vary in presence of contrastive focus?
4. To what extent does the magnitude of lip aperture in stressed/accented vs. unstressed/unaccented vowels vary in presence of contrastive focus?

Together, these questions define the framework for the analyses that follow. Based on the research questions outlined above, the following **main research hypotheses** have been proposed.

3.2.2 Main research hypotheses

Based on research question 1, which seeks to quantify the articulatory correlate of utterance-level accent through jaw displacement measurements, two competing hypotheses were formulated:

H_0: Utterance-level accent does not affect vertical jaw displacement.

H_1: Utterance-level accent increases vertical jaw displacement.

Building on the framework established in research question 1, research question 2 examines a different possible articulatory correlate — lip aperture. Two competing hypotheses were developed to test whether lip aperture shows systematic variation in response to utterance-level accent:

H_2: Utterance-level accent does not affect lip aperture.

H_3: Utterance-level accent increases lip aperture.

Research question 3 addresses the issue of articulatory correlates in a different type of prosodic prominence situation — presence of contrastive focus. Two competing hypotheses were formulated for examining this research question:

H_4: Contrastive focus does not affect vertical jaw displacement.

H_5: Contrastive focus increases vertical jaw displacement.

Research question 4 directly parallels research questions 2 and 3. Drawing on their framework, the following competing hypotheses were established:

H_6: Contrastive focus does not affect lip aperture.

H_7: Contrastive focus increases lip aperture.

In addition to the main research hypotheses presented above, specific hypotheses were formulated and are presented in the corresponding analytical Section 4.4). **The study tested a total of 15 hypotheses**: 8 main hypotheses (H_0–H_7) and 7 specific hypotheses (H_8–H_{14}).

3.3 Experimental design

The general plan of the experiment consisted of controlled elicitation of 266 utterances, each embedding a target vowel in medial and final positions, produced under **neutral** and contrastive **focus** conditions. Participants were recorded using EMA, accompanied by simultaneous acoustic recordings, aligned with EMA. This framework is described in greater detail in subsequent sections: *Research design*, *Participants*, *Material*, *Apparatus*, *Recording session*, and *Data preparation and processing*.

3.3.1 Research design

This study employed a repeated-measures quasi-experimental design with two within-subject factors to investigate articulatory and, in a complementary manner, phonetic-acoustic patterns

in **neutral** and contrastive **focus** conditions. While the articulographic research studies reported in previous sections have typically focused on discussion of the findings and not on explicit methodological framing, the present research adopts a more rigorous approach, as discussed, for example, by Rogers and Revesz (2019) and defined as follows:

> Factorial designs include more than one independent variable; that is, factorial designs are employed to investigate the effects of two or more independent variables on the dependent variable. The independent variables in a factorial design are also referred to as factors (Rogers & Revesz, 2019, p. 139).

The two factors manipulated in this study were:

- **Condition** (2 levels): **neutral** vs contrastive **focus**;
- **Position** (3 levels): *Accent*, *Focus*, and *Other* positions

As Rogers and Revesz (2019) explain

> Factorial designs allow researchers to examine not only the impact of each independent variable separately but also the combined effects of the independent variables on the dependent variable. The separate effects of the independent variables are described as main effects and their combined effects are referred to as interaction effects (Rogers & Revesz, 2019, p. 139).

However, because the *Focus* position occurred only under the contrastive **focus** condition, the design is best described as partially nested rather than fully factorial. This structure nevertheless allows examination of:

- **Main effect of condition**: whether articulatory and phonetic-acoustic patterns differ between **neutral** and contrastive **focus** conditions;
- **Main effect of position**: whether articulatory and phonetic-acoustic patterns differ across *Accent*, *Focus*, and *Other* positions;
- **Interaction effect**: whether the impact of **focus** condition depends on syllable position within the utterance.

Such a design is particularly valuable for prosodic research, as it enables investigation of how global prosodic conditions (**neutral** vs **focus**) interact with local prosodic position (*Accent*, *Focus*, and *Other*), providing insight into the predicted hierarchical organization of speech production.

Regarding the within-participants structure, each participant was exposed to all relevant levels of the independent variables.

> Repeated-measures designs, also known as within-participants designs, are characterized by a single group of participants who take part in all the different treatment conditions and/or are measured at multiple times (Rogers & Revesz, 2019, p. 138).

This approach addresses several methodological challenges inherent in articulographic research, particularly controlling for the technical complexity and invasiveness of EMA data collection. The within-participants manipulation also allows for sufficient statistical power despite the relatively small sample size, as each participant serves as their own control across conditions. The experimental manipulation involves systematic variation of contrastive focus presence, establishing a clear causal direction where focus type influences articulatory patterns, as detailed below.

Finally, following Rogers and Revesz (2019), the quasi-experimental nature stems from the absence of random assignment to experimental conditions and the lack of a true control group:

> The main feature that distinguishes non-experiments from true experiments is the lack of random assignment. Quasi-experiments are a subtype of non-experiments that attempt to mimic randomized, true experiments in rigor and experimental structure but lack random assignment (Rogers & Revesz, 2019, p. 134).

The following section provides a detailed overview of the independent, dependent, and controlled variables used in the study.

3.3.1.1 Variables

3.3.1.1.1 Independent variables

This study employed a **partially nested repeated-measures design** with two within-subject independent variables (factors):

Condition (categorical variable with two levels):

- **neutral**: utterances produced naturally (see *Appendix A* for precise instructions)
- **focus**: utterances produced under contrastive focus, i.e., with prosodic prominence placed on certain indicated words

Position (categorical variable with three levels):

- *Accent* position: syllables carrying primary lexical stress;
- *Focus* position: syllables targeted for contrastive focus manipulation;
- *Other* positions: remaining syllables in the utterance.

The *Focus* position occurred only under the contrastive **focus** condition, making the design partially nested rather than fully factorial. Each participant produced all relevant combinations of **Condition** and **Position**, using identical lexical material to control for segmental effects.

3.3.1.1.2 Dependent variables

The study measures articulatory and phonetic-acoustic parameters as dependent variables:

Articulatory measures:

- *vertical jaw displacement*: amplitude of jaw lowering expressed as relative displacement of sensor from reference points (referred in mm and SD);
- *lip aperture*: distance between upper and lower lips during articulation expressed as relative displacement of sensors from reference points (referred in mm and SD).

Acoustic measures:

- *duration*: temporal extent of target vowels (ms);
- *intensity*: sound level measured in decibels (dB);
- *fundamental frequency (F0)*: pitch (Hz);
- *F2-F1 frequency distance*: vowel space parameter indicating degree of vowel diffuse/compactness.

The combination of articulatory and acoustic measures allows for the examination of the relationship between articulatory coordination and the resulting acoustic output under different conditions (**neutral** vs **focus**).

3.3.1.1.3 Controlled variables

Several variables were controlled to minimize confounding effects:

- *lexical content*: identical lexical material across all conditions with systematic stimulus presentation patterns (see Subsection *Material* on materials);
- *recording conditions*: standardized laboratory environment and equipment calibration maintained consistently across all participants and sessions;
- *gender*: only women participated in the study, which further minimized the influence of potential sources of variability related to gender differences;
- *vowel category*: lexical stimuli represented the Polish oral vowel inventory (/a/, /e/, /i/, /o/, /u/, and /I/).

3.3.2 Participants

The selection of study participants was carefully considered and strategically planned, taking into account the limitations inherent in electromagnetic articulography (EMA) described in

earlier sections (see Chapter 1, Section 1.2.3). Due to these constraints, it is difficult to include a larger participant pool; according to Rebernik et al. (2021), five participants appear to be accepted as a sufficient sample size in EMA research. Therefore, clear and well-thought-out participant selection criteria are particularly important to ensure the reliability of the sample for addressing the research questions.

The inclusion criteria required that participants had no malocclusion, other occlusal defects, or ongoing orthodontic treatment; this restriction was adopted for methodological reasons. Malocclusion can lead to compensatory jaw movements (such as anterior or lateral displacements, or partial substitution of articulatory function by the tongue), which substantially impact the natural trajectories of mandibular motion. In addition, individuals with malocclusion or orthodontic appliances may exhibit limitations in articulatory motor control and modified interrelations between the movements of the jaw, tongue, and lips (Lorenc, 2016). Such factors would introduce additional, difficult-to-control variability into the dataset. Excluding these cases allowed the analyses to reflect typical articulatory mechanisms and maintain internal consistency.

Additional exclusion criteria were (1) being a multilingual speaker, (2) having spent extended periods of time, defined as longer than a month, in an environment where a foreign language was the primary means of communication, and (3) extensively using a foreign language in everyday life, which might, for example, be motivated by professional reasons. The rationale behind these exclusion criteria was to mitigate the risk of cross-linguistic transfer effects, which could act as a confounding factor (cf. De Leeuw et al., 2010, 2023).

A total of six adult female native speakers of Polish participated in the study, which slightly exceeds the typical sample size reported in comparable EMA research (cf. Rebernik et al. 2021). The participants were between 21 and 61 years of age and were recruited through local networks and held higher education degrees; none of the speakers presented obvious regional characteristics in their speech. They all self-reported no history of speech or hearing impairments.

Experimental research with human subjects was carried out after obtaining written consents from the subjects to participate in the study and to use the results in scientific research, as well as approval from the *Rector's Committee on Ethics in Research with Human Participation at the University of Warsaw* no. 179/2022.

Before taking part, participants were fully informed about the voluntary nature of their participation, and their right to withdraw their consent and discontinue participation at any time. They also received a small remuneration for their time and effort.

3.3.3 Material

The set of stimuli was selected with the aim of differentiating between sentences produced with and without contrastive focus. The process of selecting research material for the recording scenarios was guided by both segmental and formal criteria.

- **Segmental criteria** aimed to reflect the Polish vowel inventory — stimuli needed to represent each oral vowel (/a/, /e/, /i/, /o/, /u/, and /I/[1]) and account for vowel position. Polish nasal vowels, which are often characterized as *nasalized* rather than fully nasal due to their complex articulatory timing, were excluded from this study given their intricate phonetic structure (cf. Lorenc et al., 2018).
- **Formal criteria** were intended to ensure semantic neutrality and reproducibility of the stimuli. Sentences avoided strong emotional or cultural connotations and excluded long consonantal clusters that could cause excessive strain or pronunciation difficulties (cf. Milewski & Binkuńska, 2024). Certain speech sounds, such as /S/ and /s/[2], were limited as they may present adaptation difficulties for participants during EMA data collection (cf. Dromey et al., 2018).

Because different vowels differ in their inherent qualities, including variation in their articulatory aperture, the stimuli were designed to include all six oral vowels of Polish. The selection process incorporated methodologies as described, inter alia, in Smith et al. (2019) and Malisz and Żygis (2018). Furthermore, Polish lexical stress is highly predictable — primary stress typically falls on the penultimate syllable (Dogil, 1999), with some exceptions that are largely regularised (cf. Osowicka-Kondratowicz, 2021). In semi-spontaneous speech produced in neutral-focus contexts, sentence accent is generally placed on the penultimate syllable of the last word (Francuzik et al., 2004). Therefore, target words with selected vowels were placed both in medial and final positions of sentences. Such design allowed comparison between neutral intonation contexts and contrastive focus contexts, as well as analysis of unit prominence without contrastive focus.

The language material consisted of **six short sentences** (see Table 3.1 for more details), each containing target words with a different oral vowel. They were similar in length, ranging between 11 and 15 syllables. The sentences were designed to include words familiar to most Polish speakers, commonly used in everyday communication, and which are also stylistically and emotionally neutral. Moreover, since the recording protocol included elicited speech production, the stimuli could not be long. It was crucial that the target words had a uniform number of syllables and a similar structure.

At the same time, the utterances were intended not only to be comparable in length and structure but also to resemble natural Polish utterances rather than rigid carrier phrases often used in elicited speech research. This was particularly important given the study's focus on speech rhythm, which could not be adequately examined if the stimuli were restricted to carrier phrases that, by their structure, do not allow for the investigation of phrasal rhythm.

Consequently, the selected material consisted of words that are short – maximally three syllables long — and repetitive, in order to minimise fatigue-related errors such as speech disfluencies, changes in tempo, or lapses in memory. This was particularly important in the context of EMA recordings, where stability of sensor attachment and speaker comfort were crucial. The relatively small number of syllables per word represents a limitation of the study, as certain

[1]All the phonetic transcriptions in this dissertation are in SAMPA (see more in 3.3.8.1)

[2]The presence of sensors at the tip and middle of the tongue, in the place where the fricatives /s/ and /S/ are formed directly impacts their spectral properties.

phenomena can only be observed in words with more complex syllabic structures (cf. e.g., Łukaszewicz, 2018; Mołczanow et al., 2018); however, this choice was a necessary compromise given the methodological constraints of EMA procedures.

Before the recording sessions began, each sentence was pre-tested on a small pilot group of non-participant readers, who were asked to read the items aloud and provide feedback on potential difficulties, such as pronunciation challenges or unintended associations. This preliminary evaluation was conducted to ensure that the stimuli were easy to produce and free of distracting connotations before being used in the main experiment. The only comment from the pilot group related to the use of bold instead of capital letters, since the latter might, to some extent, be associated with screaming as per internet communication conventions. This, however, was not possible due to software constraints (EMA stimuli presentation application developed within the project *Examination of disordered speech and primary functions using articulograph CARSTENS AG501 and Acoustic Field Distribution analyser*, see *Funding*).

3.3.4 Speech elicitation procedure

To elicit focus, participants were instructed to emphasise designated words following a controlled protocol. A standard dialogue-based (e.g., *wh*-question–answer contexts) elicitation method, reported in the phonetic literature on acoustic studies (cf. Hamlaoui et al., 2019), was not applicable here due to technological limitations and constraints inherent in the articulographic research process. The instruction-based method was selected instead of a dialogue-based elicitation, as the latter could have significantly prolonged recording sessions and, consequently, increased the risk of sensor detachment, induced more micro-movements, and contributed participant fatigue. Instead, participants produced the stimuli from memory after stimulus presentation and responding to a visual cue, which helped maintain recording consistency. This approach also represented an optimum balance between the desired research outcomes and the available resources: each EMA session generates large quantities of data (Lorenc, 2016), which requires careful design of procedures to maximise analytical yield while maintaining methodological rigour and efficiency.

The research protocol also incorporated procedures for handling common issues occurring during EMA sessions, such as already mentioned sensor detachment and participant errors. When the earlier happened, recording was paused, and immediate corrective actions were taken — including reattaching sensors. The latter was addressed either by repeating measurements — to minimise data loss and ensure the integrity and continuity of the collected data. Inconsistent recording, e.g. substituting words were excluded from subsequent analyses. One example is pronouncing *wuja* (/vuja/, Eng. 'uncle') instead of *wujka* (/vujka/, Eng. diminutive of 'uncle') in /u/ vowel stimuli utterance. Even though it is a small divergence from the script, it affects phrase dynamics and, consequently, slightly changes articulator trajectories.

The quasi-experimental design described in the previous section (see Section 3.3.1) ensured that each vowel was represented in controlled contexts, both neutral and contrastive, while limiting the influence of extraneous semantic or pragmatic factors. In addition, the target words were trisyllabic, with the expected utterance accent falling on the penultimate syllable. In utterances eliciting contrastive focus, the emphasised words occurred in different positions

Table 3.1. Target sentences used as stimuli in **neutral** and **focus** conditions. Target words for neutral context are each 3-syllable-long and target words for contrastive focus 2-syllable-long with varying placement in sentence (see Figure 3.1 for more details).

Target vowel	Utterance	SAMPA transcription	English translation
/a/	Ala dała Arkowi zapas buraków	ala dawa arkovi zapas burakuf	'Ala gave Arek a supply of beets'
/e/	Edek jedzie do Łeby, jeszcze bez adresu	edek jedˆz'e do webI jeStˆSe bes adresu	'Edek is going to Łeba, no address yet'
/i/	Irek widzi kilka irysów na stoliku	irek vidˆz'i kilka irIsuf na stoliku	'Irek sees a few irises on the table'
/o/	Ogon kota opadł po skoku na Karola	ogon kota opadw po skoku na karola	'A cat's tail has dropped after jumping on Karol'
/u/	Uraz barku wujka ustał w południe	uraz barku vujka ustaw fpowudn'e	'Uncle's shoulder injury cleared up at noon.'
/I/	Solimy ryby i dajemy trzy łyżki cytryny	solimI rIbI i dajemI tˆSI wISki tˆsItrInI	'Season the fish with salt and add three tablespoons of lemon juice.'

within the utterance, which allowed the design to disentangle potential position effects from genuine focus effects. The use of short sentences with a single word carrying the target vowel also facilitated direct comparisons between **neutral** and **focus** versions of the same item. Although the full dialogic context was absent, this limitation was compensated by objective verification of prominence realisation through measurable articulatory effects such as articulator displacements and apertures, and by statistical control for word position within the utterance.

3.3.5 Instructions and stimuli presentation

Each participant underwent an onboarding session during which they became accustomed to the study framework, signals, and overall flow of sessions. **Set 0** also helped to ensure that participants were sufficiently comfortable, and to determine whether any adjustments were needed (e.g., moving the display screen closer). In addition, it allowed speakers to accommodate their speech productions to the articulatory sensors (cf. Dromey et al., 2018).

For the Set 0 (pilot test), participants were shown a series of short phrases, presented one at a time on a TV screen. The task of the speakers was to:

- read lexical units presented for 2 seconds on a screen located at eye level at a distance of approximately 1.5 m by default; the font size used in the experiment was confirmed in previous research to be adequate for comfortable reading (Lorenc, 2016); the screen was mounted on a movable stand, allowing the distance to be adjusted if necessary to accommodate individual participants' needs;
- recall and pronounce each phrase aloud in a manner typical of their everyday speech, at an agreed light signal (green screen bar).

A message stating 'The next phrase will appear shortly' was displayed in between the stimuli. Participants were reminded they could opt for a break at any moment by signaling the experimenter, which helped maintain their comfort throughout the session.

Upon completing the onboarding, **Set I** (**neutral** utterances) was presented (see *Appendix A* for more details on recording procedure). It comprised the list of phrases included in the recording scenario, each presented three times in random order. The procedures and instructions were the same as for the pilot test.

In the next stage, participants were presented with **Set II** of the stimuli (contrastive **focus**), participants were instructed to emphasise specific words printed in uppercase letters. Set I and Set II contained the same phrases with visual adjustments specifying which words were to be produced with emphasis. The phrase list included three repetitions of each stimulus, presented in randomized order.

During the recording sessions, the experimenter remained in the same sound-proof room as the participant in order to respond promptly to any questions or difficulties. At the same time, the experimenter was able to observe the adjoining control room through a glass partition, where an engineer operated the Carstens AG 501 system and recording equipment along with other researchers present. This configuration facilitated efficient two-way communication: the

engineer and researchers could notify the experimenter of pause e.g., due to technical problems impacting recording sessions or give signal to resume recording after a break. In this way, data collection proceeded in a coordinated manner and with minimal interruptions.

3.3.6 Apparatus

All the articulographic measurements were carried out using the only Carstens AG501 electromagnetic articulograph (EMA) available in Poland, located at the Maria Przybysz-Piwko Applied Phonetics Laboratory, at the *Logopedic Centre of the Institute of Applied Polish Studies* at the University of Warsaw. The key piece of equipment, **Carstens AG501 electromagnetic articulograph** (Articulograph, n.d.), enables precise tracking of speech articulator movements in real time.

The AG501 operates on the principle of electromagnetic induction: six transmitter coils generate an alternating magnetic field, which induces voltage in small receiver sensors attached to the articulators of the speaker. Measuring this voltage allows the software to calculate three-dimensional coordinates (X, Y, Z) for each sensor, as well as two angular measures (ϕ, θ). Raw data were managed, visualised, and pre-processed using the manufacturer's dedicated software.

To ensure efficiency and data re-usability, the sessions that constitute the empirical basis of this dissertation were recorded as part of a larger data collection effort. This broader context shaped the recording protocol in terms of the number of sensors used along with stimuli presentation software.

A total of 16 sensors were used in each recording session, which is the maximum supported by the system used in Warsaw University[3]. For the purposes of this dissertation, however, only three measurement sensors are analysed in detail:

- **upper lip:** on the vermilion border;
- **lower lip:** on the vermilion border;
- **jaw:** at the gingival margin of the lower incisors.

[3]The maximum capacity is 24 sensors

Figure 3.2. Example of EMA sensor placement – the upper lip sensor is clearly visible.

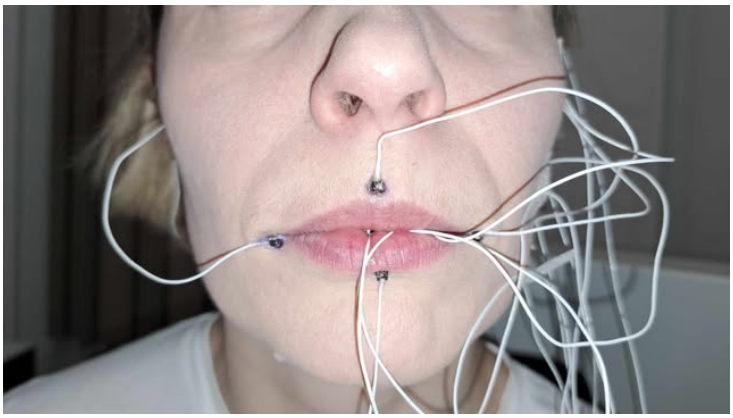

Figure 3.3. Example of EMA sensor placement – the upper lip and lower lips sensors, and part of lower incisor sensor, are clearly visible.

The remaining sensors either served calibration and reference functions or were not essential to the present research objectives, and are therefore not reported here. The full specification of sensor is available in *Appendix C*. Nevertheless, the data obtained from the remaining sensors may be used in future analyses to examine the trajectories of other articulators in the recorded utterances, including aspects of tongue configuration and movement. See Figure 3.4 for placement of the above mentioned sensors along with other used within the broader scope of the recording sessions.

In addition, acoustic data were simultaneously collected using the dedicated Carstens AG 501 equipment and software, as well as with an external Zoom recorder running as a backup in case of any malfunction.

The initial sampling rate for articulatory data were 1250 Hz which was normalised to 250 Hz and audio sampling frequency was 44 kHz.

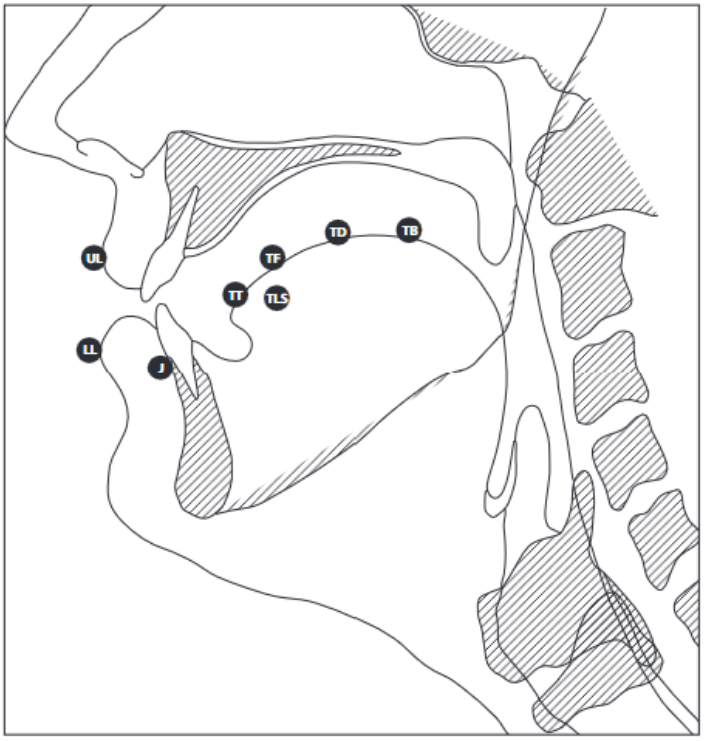

Figure 3.4. Schematic representation of EMA sensor placement in the midsagittal plane. The sensors are positioned on the following articulators: **upper lip (UL), lower lip (LL), jaw (J)**, tongue back (TB), tongue tip (TT), tongue dorsum (TD), tongue front (TF), and tongue left side (TLS). The hatched areas represent hard anatomical structures (teeth, hard palate, and pharyngeal wall). Bolded sensors were crucial for this dissertation. Adapted from Lorenc (2016).

3.3.7 Recording session

Preparation for the recording sessions begins up to 30 hours in advance. The articulograph device requires approximately **3-4 hours of warm-up time** to achieve optimal temperature stabilization, which is essential for maintaining measurement accuracy. Sensor calibration is then performed using the manufacturer's *cs5cal* software, which executes a full rotation of the calibrator to ensure precise positioning data. The switching on sequence of individual devices and software is also according to protocol, following the most optimal route since there are multiple programs and pieces of equipment running at the same time (see Table 3.2 for the list of basic software used during sessions).

Figure 3.5. A speaker prepared for an Electromagnetic Articulography (EMA) recording session using the Carstens AG501 system. Receiver sensors are already attached. The subject is wearing the reference coil headband, with transmitter coils visible in the background, which generate the alternating magnetic field used for tracking sensor coordinates.

To prepare the sensors for attachment to participants, a comprehensive sterilization protocol is followed. The sensors undergo 12 hours of disinfection in laboratory-grade solution, after which each sensor is coated with liquid latex to create a protective barrier against the tissue adhesive used during attachment. All procedures including preparation and attachment of sensors were performed using sterile latex surgical gloves (see Figure 3.6).

Table 3.2. List of software used during the recording sessions

Program	Source / Provider	Main Function
mc5recorder	Carstens	Recording control, session parameters, audio test
cs5view	Carstens	3D visualization of sensor positions (real-time / stored data)
cs5cal	Carstens	Sensor calibration
NormPos	Carstens	Normalization of data w.r.t. head movements
CalcPos	Carstens	Computation of sensor positions in 3D
Bin2ASCII	Carstens	Conversion of binary data into ASCII format
Just View	Carstens	Visualization of recorded 3D sensor data
VisArtico	External software	Advanced visualization: trajectories, sagittal view, spectrogram, segmentation
phoneEMAtool	Custom Matlab (Mik & Lorenc, 2024b)	Trajectory visualization, position extraction, velocity-based segmentation

Figure 3.6. Preparation for placing biteplane in the participant's mouth. All the procedures involving physical contact with participants and equipment were performed in sterile latex surgical gloves.

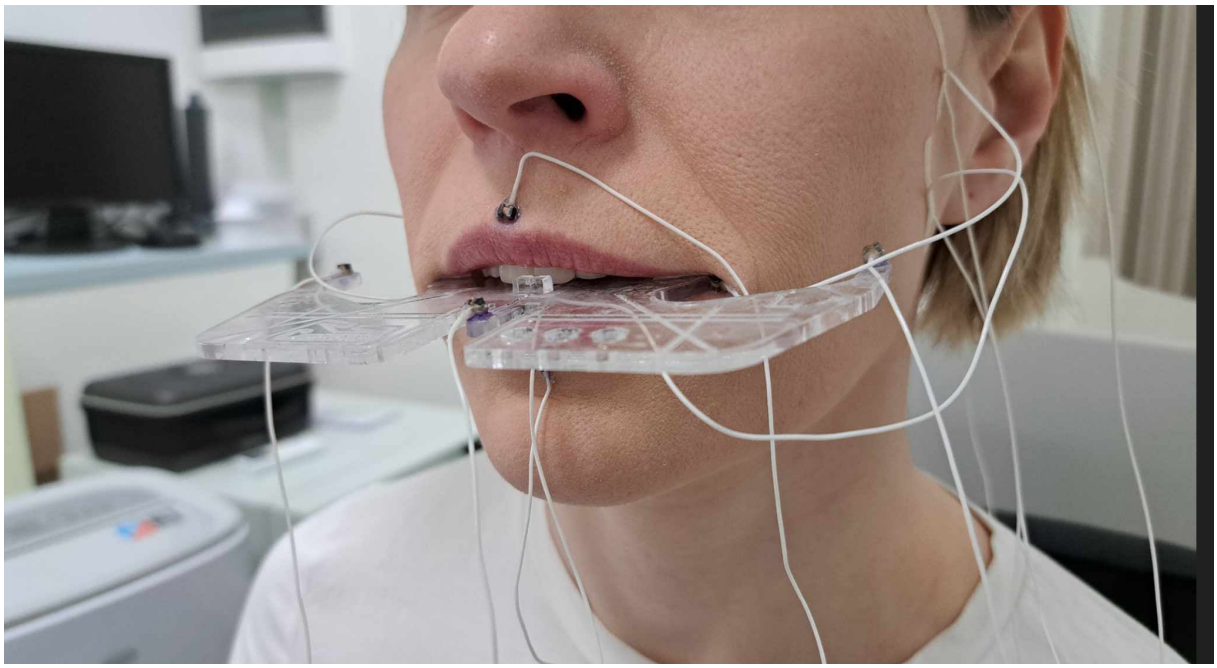

Figure 3.7. The biteplane, held between the participant's teeth, serves multiple functions, including providing a fixed reference point for the measurement system.

The experimental setup involves organizing 16 sensors (see Figures 3.8- 3.9), with an additional backup set maintained for contingency purposes. Cable connections are carefully routed to replicate the exact pathways used during actual data collection, as maintaining consistent conditions between calibration and measurement phases is critical for data quality. Each sensor is marked to ensure the right one is plugged in during the recording session (see Figure 3.8). Calibration accuracy is subsequently verified through recording and analysis of test samples before proceeding with participant testing.

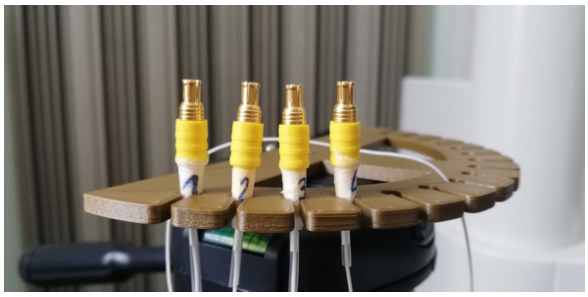

Figure 3.8. Electromagnetic sensors mounted in a wooden calibration holder. Four sensors are shown with their cable connections and marking, enabling precise use in the AG501 articulograph system.

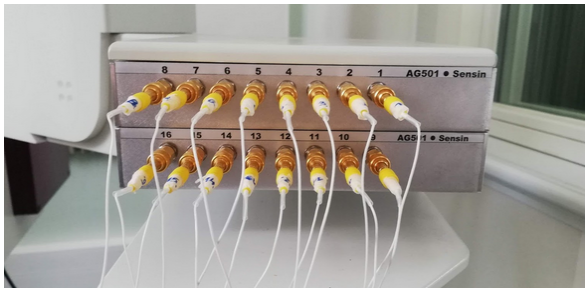

Figure 3.9. Electromagnetic sensors connected to the AG501 Sensin receiver unit. Each sensor is plugged into its numbered port according to the calibration protocol, ready for articulographic data acquisition.

Preparation of each speaker took 60-90 minutes; during this time participants were familiarized with the procedure, signed written consents, and removed all metallic elements from their body (e.g., earrings, glasses, necklaces etc.). Once that step was completed, the sensor attachment process begun which inclued the following steps:

1. degreasing and cleansing facial skin using an isopropyl alcohol;
2. sensors were attached using non-toxic tissue adhesive, PeriAcryl®90, commonly used in dental surgery procedures. Place of attachment is specified accoridng to a detailed protocol, not discussed fully here (see e.g., Lorenc (2016) for more details);
3. part of the sensor cables were secured using special adhesive patches on the speaker's face, neck, and nape, selecting neutral locations that would not interfere with sensor attachment or video marker registration (see Figure 3.5).

3.3.8 Data preparation and processing

3.3.8.1 Audio recordings segmentation

Recordings of utterances were segmented into words, syllables, and phones using AnnotationPro (Klessa et al., 2013) with the ANNPRO plugin (Klessa et al., 2022). For 26 recordings, automatic segmentation was not feasible due to low signal level; in these cases, the signal was amplified using the Amplify effect in Audacity (Audacity Team, 2023). Automatic annotations were subsequently verified and manually corrected by the author to account for inaccuracies resulting from the limitations of automatic segmentation algorithms, following the phonetic segmentation guidelines outlined by Machač and Skarnitzl (2009).

All phonetic segments were transcribed using the SAMPA alphabet (*Speech Assessment Methods Phonetic Alphabet*, cf. Wells, 1997). In the case of phone level Polish SAMPA labels are used in the tool, for the syllable level, the standard Polish SAMPA coding was extended by including affricate labels ˆ as a tie bar and /y/ label instead of /I/ vowel label. Pauses were marked with the dollar symbol — $. The threshold value for distinguishing a pause was set at 300 ms, a

boundary adopted following auditory inspection of the recordings; pauses longer than this value were perceived as separate units, rather than transitional junctures between phrases. The adopted threshold is consistent with corpus-based findings for length of a perceivable pause in Polish (Francuzik et al., 2002). These files provided reliable segment boundaries for linking acoustic and articulographic measurements in the subsequent analyses.

For the needs of data analysis and processing with tools dedicated to the analysis of articulographic data such as *phoneEMAtool* (Mik & Lorenc, 2024b), the annotation files including time-aligned segmentations and transcriptions were exported from Annotation Pro (Klessa & Gibbon, 2014) ANTx file format to Praat (Boersma & Weenink, 2022) TextGrid file format.

3.3.8.2 Extraction of articulatory data

One of the major initial challenges in working with articulatory data were the preparation of the dataset for subsequent stages of analysis. For the purposes of segmentation and articulatory gesture analysis, custom software, developed in the Matlab environment, was used in the initial stage of the study. The *phoneEMAtool* application (Mik & Lorenc, 2024b) enabled dynamic visualisation of sensor trajectories (excluding reference sensors) along the X-axis (front–back) and Z-axis (up–down), as well as the analysis and extraction of positional information of individual sensors over time, including angular displacements. In addition, the software allowed for the calculation of sensor velocities and the identification of their minima and maxima, information crucial for detecting whether an articulator reached its target position for a particular phonemic segment (visible in Figure 3.10). In the earlier stages, the analyses focused only on jaw displacement, and for that purpose, data from three speakers were processed using *phoneEMAtool*.

3.3.8.3 Articulatory gesture annotation

Articulatory gesture boundaries are defined based on sensor velocity profiles, not acoustic boundaries — therefore, articulatory and acoustic segmentation do not overlap. Gesture boundary points are determined using velocity-based landmarks, most commonly at a 20-percent threshold of peak values (PVEL1, PVEL2) (cf. Lorenc, 2016). Depending on application used to collect articulatory data, the gesture nucleus phase is defined differently:

- *phoneEMAtool* allows for the identification of velocity minima (V_{min});
- *MVIEW* uses the Mid-C point (midpoint closure), which does not necessarily coincide with the velocity minimum. It should be noted that this is the primary software package used by most authors cited in Chapter 1, Sections 1.3.2– 1.3.3.

For consistency, articulatory landmarks were identified within intervals defined by acoustic segmentation, which allowed the velocity-based annotation of gestures to remain anchored in phonemic units while still reflecting articulatory dynamics.

Manual annotation was used to determine the position of the jaw at the point of minimal velocity (V_{min}) for each phonetic segment (within segment boundary positions derived from the

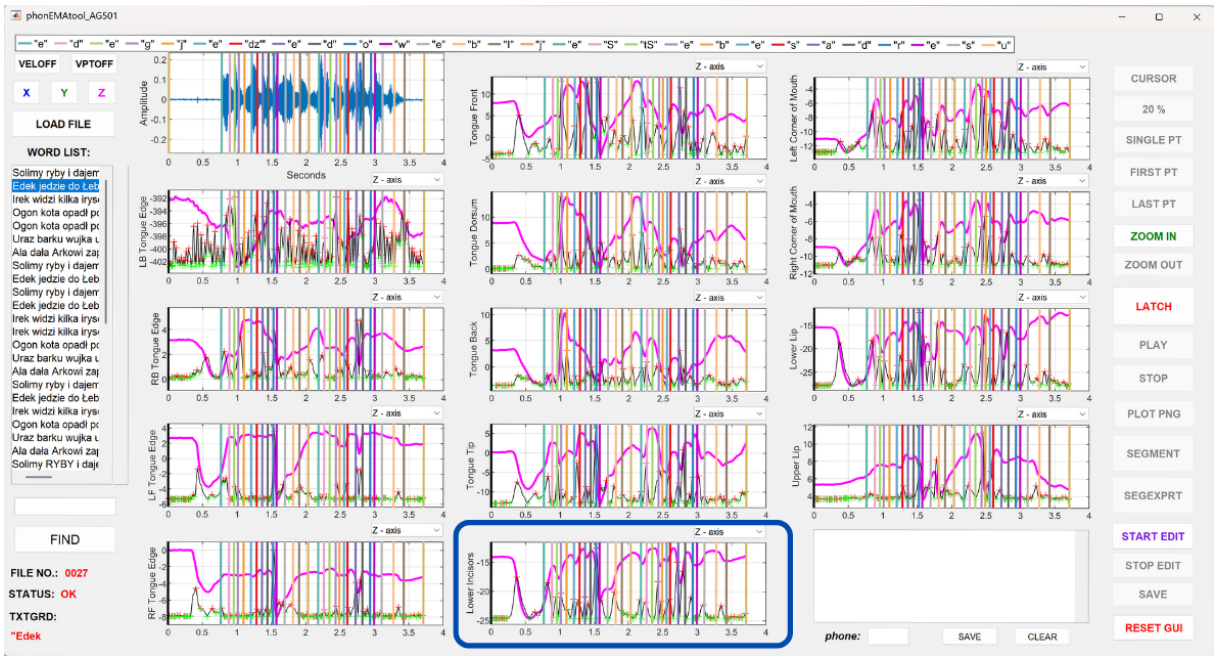

Figure 3.10. Exemplary screenshot from the *phonEMAtool_AG501* software (Mik & Lorenc, 2024b) — data from the lower incisor sensor on Z-axis marked in **blue**. The pink line represents jaw trajectory, black line represents velocity, and the vertical coloured bands indicate phonetic segment boundaries in the recording *OKPC_027* — realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet') (author's illustration).

TextGrid files), serving as an indicator of articulator stabilization and target position achievement. Given the scale of the dataset, manual annotation was subsequently replaced by automated data extraction using custom Python scripts (Python Software Foundation, 2023) developed by the author and deposited in the author's profile in the AMU Research Data Repository as supplementary material to this dissertation. This decision was motivated by the following considerations:

- the task followed a clear and replicable procedure;
- the dataset was extensive, comprising 7,449 phonetic segments from 266 recordings;
- as the study progressed, positional data from additional sensors were incorporated, which required segment-level extraction for every newly included sensor for each recording and position anew;
- developing the script ensures corpus scalability and the compatibility of future datasets processed with the same data types and supports replicability of the procedure, making it less prone to human error.

The *phoneEMAtool* application continued to be used as supplementary support to the automated procedures, serving primarily as a visual aid for data inspection, facilitating initial insights into

phonetic material. It is particularly effective for scaling work with smaller datasets and is well-suited for preliminary visualisation of data trends. Moreover, its output files are compatible with the *EMAviewer* (Mik & Lorenc, 2024a) software tool enabling visualisation of articulators in cross-sectional views of the vocal tract.

3.3.8.4 Automation of articulatory data extraction — procedures

3.3.8.4.1 Input files and configuration

Automated data processing required a series of preparatory steps. Raw sensor position data were stored in POS files in plain-text (.txt) format. Each recording session corresponded to a single file, consisting of tab-delimited rows and columns. The following transformations were performed:

- adding a header row specifying the sensor and positional axis (up–down, left–right, front–back);
- retaining only selected columns, namely jaw displacement (*X, Y, Z*) and vertical positions of the upper and lower lips, in order to reduce data volume and improve script efficiency;
- adding a timestamp column, with each row corresponding to a 250 Hz sample (i.e., 4 ms increments);
- adding a filename column to allow for reliable merging of datasets for statistical analysis.

As a result, the processed files contained positional data of selected EMA sensors annotated with recording sessions and time values. The reliability of automated extraction was verified by comparing results with manual annotations from *phoneEMAtool* across randomly selected recordings. Both jaw trajectories and minima of velocity were compared, calculating the proportion of matching landmarks on a sample of files, yielding over 95% match. To ensure methodological consistency, all recordings that had previously been annotated manually were reprocessed through the automated pipeline.

3.3.8.4.2 Integration of articulatory gesture data

For subsequent analyses, articulatory data were enriched with additional layers of information listed below:

- metadata;
- phonemic labels from annotations;
- indices of phonetic segments within each recording;
- results of derived calculations;
- phonetic-acoustic parameters.

The procedures of obtaining, processing, and integrating these layers are described in more detail in the following sections.

3.3.8.4.3 Metadata preparation

Metadata were stored in a tabular format (.tsv), including filename, target phoneme (*Vowel*), and presence or absence of contrastive focus condition (*Focus*). Using filenames as primary keys, *Vowel* and *Focus* were added to the master dataset. In addition, the first four characters of each filename were extracted to identify the speaker.

3.3.8.4.4 Annotation file preparation

Previously obtained .TextGrid files with segment boundaries were converted to align with articulatory data. A custom Python script identified the *Phone* tier in each file and extracted start and end times of segments. Temporal conversion was carried out in two stages:

1. segment boundaries were converted from seconds to milliseconds, rounded up to avoid excessive decimal places, and
2. boundaries were aligned to a time grid with a step size of 4 ms, consistent with the 250 Hz sampling rate of articulatory dataset.

In .TextGrid files, the end value of one segment is immediately the beginning value of the next segment following it. Due to the greater granularity of EMA data, the phonetic segmentation had to be adjusted to 4 ms intervals — this also meant that the end of one phonetic segment and the beginning of the next had to have different values, because **a given articulator movement could only be assigned to one phonetic interval**. The rules for assigning a given spatial point to a sound were as follows: rounding was performed to the nearest grid point, so that, for example, a time of 23 ms was assigned to the 24 ms grid point, and not to 20 ms; for each subsequent phonetic segment, the script retrieved information about the end time of the previous segment and added 4 ms. This ensured the continuity of intervals and prevented temporal overlaps or gaps.

For each line, both values in ms and values adjusted to the 4 ms grid were recorded, along with the phonemic label, the duration of the phonemic segment with and without projection onto the 4 ms EMA grid, and the source file name. The latter information is particularly important, as each recording session has its own unique name, which serves as a key to merging data from various sources.

Data extracted from the TextGrid files was saved in tabular format (.tsv), allowing for further integration with articulatory material.

After integrating the information from the EMA with the phonemic labels, a process of organising and standardising the data were implemented by removing all rows without phonemic labels. This resulted in a clean and consistent **dataset that contained only those**

fragments for which both kinematic and phonemic data were available, and eliminated fragments that did not contribute to further analysis and research (e.g., background noise).

In the next step, a column with new time values was added for each separate file, assigning 0 to the first point with a phonemic label and incrementing subsequent values by 4 ms. The column with the original time values was retained as it might serve as an important reference point if data from additional sensors are to be added to the masterfile. Each phonemic segment was assigned an index indicating its position within the utterance relative to its beginning. This was important for at least two reasons. Firstly, it allows for the use of orderly operations within a session, which were essential for further analysis. The indices are unique for each phonemic segment within an utterance, whereas labels are not; this uniqueness makes it possible to trace, select, and compare corresponding segments across files. They also enable aggregation procedures, such as averaging values for a given segment or aligning trajectories relative to segmental boundaries, which would not be reliable if only phonemic labels were used. Secondly, it can be useful in case the dataset would be used by other researchers in their studies, as it might encode positional information for each segment[4]. Position within the utterance is a factor known to affect the realisation of speech sounds (cf. e.g., Łukaszewicz, 2018; Newlin-Łukowicz, 2012).

3.3.8.4.5 Derived calculations

As emphasised earlier, it is crucial to determine the position of the sensors at the moment of minimum velocity of an articulator movement. Since raw EMA data contain only positional values, velocities were calculated as the derivative of the difference in position over time ($v = \Delta x/\Delta t$). Velocity curves were computed using a dedicated Python script (Python Software Foundation, 2023) developed specifically for the purposes of the present dissertation, for the jaw, upper lip, and lower lip along the vertical (Z) axis separately. Velocity for the horizontal (X) axis was calculated only for the jaw, as it was needed for supplementary analyses for this articulator. For each phonetic segment, the V_{min} point was identified. If multiple minima occurred, consonants were assigned the first minimum within the interval, whereas vowels were assigned the one closest to the segment midpoint.

Lip aperture was computed as the absolute difference in vertical position between the upper and lower lip sensors. In addition, auxiliary measures were extracted, including:

- three columns with V_{min} values in Z axis — for jaw, upper lip, and lower lip;
- one column with V_{min} value in X axis — for jaw;
- three columns with V_{min} time in Z axis — for jaw, upper lip, and lower lip;
- one column with V_{min} time in X axis — for jaw;
- three columns position at V_{min} in Z axis — for jaw, upper lip, and lower lip;
- one column position at V_{min} in X axis — for jaw,

[4]Which can be easily remapped to match needs of a given research objective.

- maximal lip aperture;
- timestamp of maximal lip aperture at V_{min};
- temporal offset between maximal lip aperture and maximal jaw displacement.

Also, detrending and z-score normalisation were applied separately for jaw, upper lip, and lower lip and lip aperture. This step is further discussed in the section *Normalisation of the dataset*.

The result was a masterfile comprising 187,839 rows (individual position samples from EMA) and 48 columns (including both crucial values and auxiliary measurements), serving as a centralized, structured repository for subsequent analyses.

3.3.8.4.6 Phonetic-acoustic parameters

In addition to articulatory measures, the following phonetic-acoustic parameters were extracted in order to provide a complementary perspective on the realisation of prominence: fundamental frequency (F0), the first two frequency formants (F1, F2), and intensity. The extraction was performed in Praat (Boersma & Weenink, 2022) using two custom scripts developed by the author specifically for this research. The following Praat software settings were used:

- Fundamental frequency (F0) — extracted using automatic time step with pitch range set to 75–600 Hz;
- Intensity — calculated with minimum pitch threshold of 100 Hz[5], automatic time step, and quadratic interpolation;
- Frequency formants (F1, F2) — extracted using the Burg algorithm with automatic time step, 5 formants up to 5500 Hz, 25 ms window length, and 50 Hz pre-emphasis.

Output data were saved in tabular format (.tsv), which allowed integration with the EMA data.

For the purposes of visualisation of change in time, the parameters were obtained with Praat's *Get value at time function* (default analysis settings). Measurements were sampled at 4 ms timestep to ensure compatibility with the temporal grid of the EMA data; articulatory and acoustic files were then aligned for each recording individually using a dedicated Python script (Python Software Foundation, 2023).

For statistical analyses, a separate Praat script was used to extract average values based on the Get mean function. These summary statistics included data for vowels only and were then linked to the metadata through an additional manual alignment procedure in spreadsheets.

[5]An adjustment to female voice, the standard Praat setting is 75 Hz.

3.3.9 Data for visualisation — signal normalisation

For the purposes of graphical presentation of the trajectory of speech articulators, it was necessary to normalise the data. The procedure included two-step time normalisation and centring the vertical positional values around 0.

3.3.9.1 Temporal normalisation

The duration of phonetic segments was first normalised to enable direct comparison of articulatory signals over time, as segment length varied across recordings and conditions. While pauses were present in the annotations for the **focus** conditions, they were excluded from the visualisations to emphasise continuous articulatory motion. Starting from the master dataset (sensor positions with metadata and segment indices), representative durations were established by computing the median duration for each combination of *Segment Index* × *Vowel* × *Focus*. This means that typical duration was calculated separately for every unique type of vowel in every specific position within an utterance and under each condition (**neutral** vs **focus**. The median was chosen for its robustness against outliers, providing a stable central tendency measure.

To assess variability, the Coefficient of Variation (CV%) was calculated as the ratio of the standard deviation to the mean, expressed as a percentage. Most CV% values fell within a moderate to high range; interpretatively, this means considerable variation in segment durations across realisations and suggesting that the articulation process was not entirely stable. This variability supports using the median rather than the mean to better represent central tendency in the data.

Normalisation proceeded in two stages:

1. intra-segment scaling, where each time point was stretched or compressed according to the scale factor k (defined as the ratio of median duration to observed duration), and
2. global temporal normalisation, in which consecutive segments were aligned such that the start of each new segment followed the scaled endpoint of the preceding segment plus 4 ms scaled by the k-factor for that following segment (see Figure 3.11). This ensured continuity and preserved proportional timing.

The intra-segment scaling allows phonemic segments to be represented with adjusted lengths along the x-axis. Such a step enables direct comparison of articulatory movements across different speakers in normalised temporal space. The global temporal normalisation then concatenates these readjusted segments together, like beads on a string, ensuring continuity of the signal and preserving proportional timing across the entire utterance.

3.3.9.2 Vertical centring of trajectories

To enhance comparability across speakers and recording sessions, the raw articulatory trajectories were normalised through mean-centring, so that they oscillated around zero rather

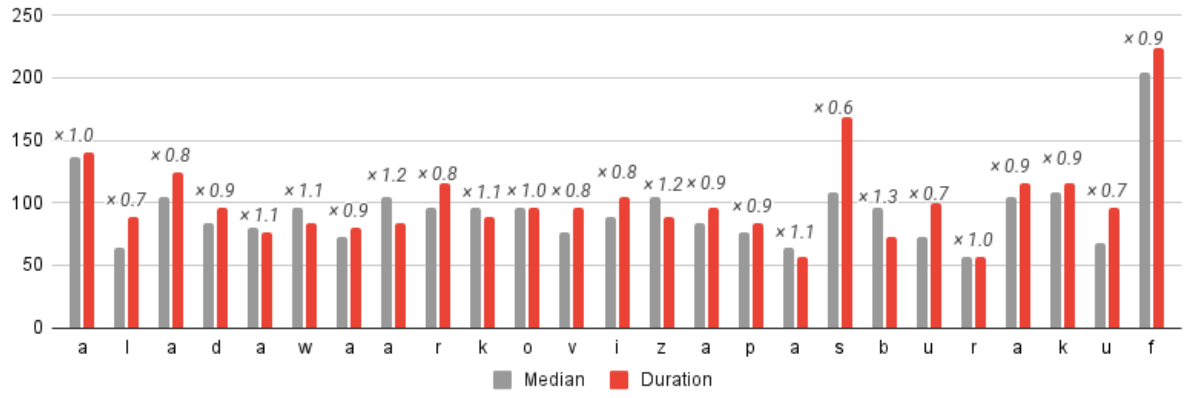

Figure 3.11. Example of intra-segment scaling for the *RHAK_532* session. Actual segment durations were compared according to *Segment Index* × *Vowel* × *Focus* conditions. The *k*-factor for scaling each segment is indicated above the plot bars. In this case, the initial segment /a/ will be shortened by a factor of 0.97, while the segment /w/ will be lengthened by a factor of 1.14.

than absolute positional values. The **mean** was used instead of the median, as the goal was to visualise EMA trajectories rather than to suppress minor variations. Since the data consist of densely sampled signals with no substantial skew, the mean provided a natural and mathematically consistent reference level. Mean-centring was performed by subtracting the mean displacement value $\bar{x}$ from each observed value x_i, as shown in Equation 3.1:

$$x_i' = x_i - \bar{x} \tag{3.1}$$

where x_i denotes the observed displacement at time point i, $\bar{x}$ – the mean displacement across the interval, and x_i' – the centred value. Plots are presented in Chapter 4, Section 4.1.

3.3.9.3 Adjustment of lip sensor positions for visualisation

To facilitate the inspection of articulatory dynamics, custom scripts were implemented to generate visualisations of both jaw displacement and lip aperture trajectories. The pipeline, written in Python (Python Software Foundation, 2023), was based on sensor coordinates exported from the Carstens AG 501 articulograph and aligned with the acoustic signal. Because the lip sensors in EMA are placed slightly above the vermilion border, their absolute positions in the raw coordinate space appear farther apart than in actual articulation. For the purpose of visualisation, the upper- and lower-lip sensor trajectories were therefore shifted relative to each other so that their vertical distance more accurately reflected lip aperture. This adjustment was computed separately for each target sentence, ensuring that the alignment reflected vowel-specific articulation within individual utterances. The steps were completed as follows:

- **Identification of the minimum aperture** — for each target utterance, the point of minimum lip aperture was identified. This value served as an offset reference for

alignment;

- **Vertical adjustment** — the trajectories of the upper- and lower-lip sensors were shifted vertically by the offset so that the distance between them at this point corresponded to zero aperture, i.e. complete lip closure;
- **Computation of relative values** — after this adjustment, the offset value was subtracted from the entire trajectory, which made the measurements directly interpretable: zero meant that the lips were fully closed, and larger values meant a wider opening. These relative values were calculated separately for each recording;
- **Averaging for visualisation** — finally, the relative trajectories were averaged across each target sentence in a given condition, which allowed for clear visual comparison of articulatory patterns between the prominence contexts.

Importantly, this procedure was applied for the subset of data for the visualisation only. Data used for statistical analyses were prepared separately, as described below.

3.3.10 Data for statistical analyses

For statistical modelling, a condensed dataset of segment-level measures was created. This step served to optimise the process, as working with smaller, precisely defined subsets of data is much more efficient and less computationally intensive than working with the entire masterfile (see Table 3.3). Moreover, since the first phonetic segments of each utterance were excluded for methodological reasons (see Section 4.2.1 for more details), an additional filter column was added to indicate whether a segment should be included in the analysis.

Table 3.3. Summary of data points included in the EMA master dataset per speaker — each EMA sample equals a point in time (every 4 ms). Each data point = all the data available at given time, including metadata, positions, phonetic-acoustic parameters, etc.

Speaker	EMA time sample	Data points
OKPC	23,659	1,230,268
RHAK	47,842	2,487,784
UOKV	33,656	1,750,112
VTVK	22,754	1,183,208
SLDT	36,619	1,904,188
RLRG	23,309	1,212,068
Sum	**187,839**	**9,767,628**

EMA data (sampled every 4 ms) were condensed into one representative record per phonetic segment. Each record included interval boundaries, duration, jaw displacement, lip aperture, and velocity minima, together with metadata (*Recording, Vowel, Focus, SegmentIndex*). This process transformed almost two hundred thousand individual measurements into a structured, segment-level dataset suitable for statistical analysis.

The final dataset, enriched with metadata for statistical analyses describing the segment index with expected position of *Accent* and position of contrastive *Focus* (see *Appendix B* for more details), constituted the foundation for subsequent quantitative analyses.

3.4 Statistical analyses

The decision regarding the statistical method employed in this study was preceded by a thorough evaluation of the data structure and research questions to ensure the chosen approach appropriately addressed the hierarchical and unbalanced nature of the dataset.

3.4.1 Choice of the statistical method

The statistical analysis was conducted using linear mixed-effects models (LMEMs), which included random intercepts to account for individual speaker differences. Two dependent variables — mandibular displacement and lip aperture — were analysed separately. The choice of the method was based on considerations presented as follows.

3.4.1.1 Repeated measures and the assumption of independent observations

The selection of LMEMs over more traditional methods, such as ANOVA or linear regression, was motivated by the hierarchical structure of the data. Data collection, all 266 recording sessions were produced by six people; because repeated measurements from the same speaker would violate **the assumption of independent observations**[6], classical between-group ANOVA or regression analyses were not feasible since this assumption is a *sine qua non* for both these methods.

To be more specific, recordings obtained from a single speaker may[7] violate the principle of measurement independence, since successive observations are typically interrelated. Such dependencies may result from intra-individual factors, including habitual articulatory patterns, distinct strategies when emphasising words, and anatomical differences, just to name a few.

3.4.1.2 Data structure complexity

On one hand, nonparametric tests, which bypass this limitation such as Friedman test or Wilcoxon signed-rank test, do not perform well with a large number of factors and interactions. This is because they only compare the overall distribution of data — medians or ranks between groups and, on this basis, can answer the question of whether the groups differ from each other but not how much they differ.

[6] In other words, assumption that each data point in a sample is unrelated to and unaffected by every other data point, meaning one measurement does not influence another.

[7] Depending on specific objectives of research, independence of observations might be operationalised differently; e.g. when only overall group-level differences are of interest.

Since data structure in this study has a certain degree of complexity, such as three fixed factors (*Focus, Position, Vowel*) and random factors (*Speaker*), these tests would not enable in-depth analyses. On the other hand, the above mentioned factors also interact with each other, which is another issue these tests cannot reliably account for.

In contrast, LMEMs explicitly account for clustering by grouping observations within speakers, allowing each speaker to have their own baseline (random intercepts) and individual effect slopes. This modelling structure effectively separates general effects — such as the influence of presence of contrastive focus — from inter-speaker variability by incorporating dedicated variance components.

3.4.1.3 Data imbalance

Data imbalance occurs when certain categories are represented by many more observations than others, leading to uneven distributions in the dataset. In other words, some factors contribute disproportionately more data than others, which can bias statistical analyses if not accounted for.

The dataset was fairly balanced considering the number of observations per speaker and vowel (Table 3.4). The source of imbalance, however, is the representation of stress/accent position — each utterance contains multiple tokens of a given vowel, but typically only one token occurs in a stressed / accented position. Therefore, the dataset includes fewer stressed / accented tokens relative to unstressed ones, which could lead to unequal statistical power across conditions.

Although linear mixed-effects models do not eliminate such uneven sampling across conditions, they are better suited to this design because they explicitly model random variation associated with individual speakers and items.

Table 3.4. Number of collected sessions per *target vowel* and *condition*

Condition	Speaker	/a/	/e/	/i/	/o/	/u/	/I/	Sum
neutral	OKPC	3	3	3	3	3	4	**19**
	RHAK	7	8	7	7	7	8	**44**
	UOKV	3	5	3	3	3	4	**21**
	VTVK	3	3	3	3	3	3	**18**
	SLDT	4	4	5	4	3	4	**24**
	RLRG	3	3	3	3	3	3	**18**
focus	OKPC	3	3	3	2	3	3	**17**
	RHAK	4	4	4	4	4	4	**25**
	UOKV	4	3	3	4	2	3	**19**
	VTVK	3	3	3	3	3	3	**18**
	SLDT	4	4	4	4	4	4	**24**
	RLRG	4	4	3	3	3	3	**20**
Sum		**45**	**47**	**44**	**43**	**41**	**46**	**266**

3.4.1.4 Normalisation of the dataset

3.4.1.5 Detrending

During data analysis, a consistent gradual trend present across all the speakers was observed, inexplicable by articulatory variability alone. These slow changes, hereafter referred to as **drift**, have not been extensively addressed in the literature on jaw displacement patterns being the theoretical background for this study; Kawahara et al. (2014) reflected that in the Japanese material for /a/ vowel jaw opening decreased gradually. The authors hypothesised the pattern may arise from the foot structure of Japanese. Kawahara et al. (2015) discussed the topic using term *declination*, in the context of Japanse articulatory patterns. There, however, are also few mentions of this phenomenon in the literature; such observations were discussed in terms of possible strategies for articulatory adjustment to a changing speaking rate (cf. Sonoda & Nakakido, 1986). More detailed analysis of this phenomenon was present in a rather technical and procedural literature (Richmond, 2002) where it is considered a possible effect of equipment temperature changes, subtle sensor shifts, or speaker adaptation to wearing recording devices. Overall, a definitive explanation is still unknown.

Since there is no agreed upon or standardised method for addressing the issue, in this study, the detrending procedure was implemented as a linear regression fit and subtraction. For each recording, the jaw displacement trajectory $y(t)$ was approximated by a **linear trend** (3.2):

$$y_{trend}(t) = at + b \tag{3.2}$$

Here a denotes the slope, b the intercept, and t — time. The detrended values were obtained by subtracting this fitted line from the original trajectory (3.3)::

$$y_{detrended}(t) = y(t) - y_{trend}(t) = y(t) - (at + b) \tag{3.3}$$

Figure 3.12 presents the procedure in a more visual way.

3.4.1.6 Z-score normalisation

The choice of normalisation method is non-trivial, as various procedures have been proposed for EMA samples, each addressing slightly different sources of variation. Examples include proportional measures of jaw opening (see e.g., Erickson et al. (1998)) and utterance-wise scaling (cf. Al Bawab et al. 2008).

In this study, articulatory data were normalised using a z-score transformation applied separately to each speaker, based on all recordings covering the annotated phonemic segments. This type of transformation is widely used in linguistics and is considered amongst the best normalisation approaches, exemplified by Lobanov's vowel formant normalisation method (Lobanov, 1971) and the YARD rhythm measurement procedure (P. Wagner & Dellwo, 2004). The z-score method is regarded as one of the most reliable vowel normalisation techniques as

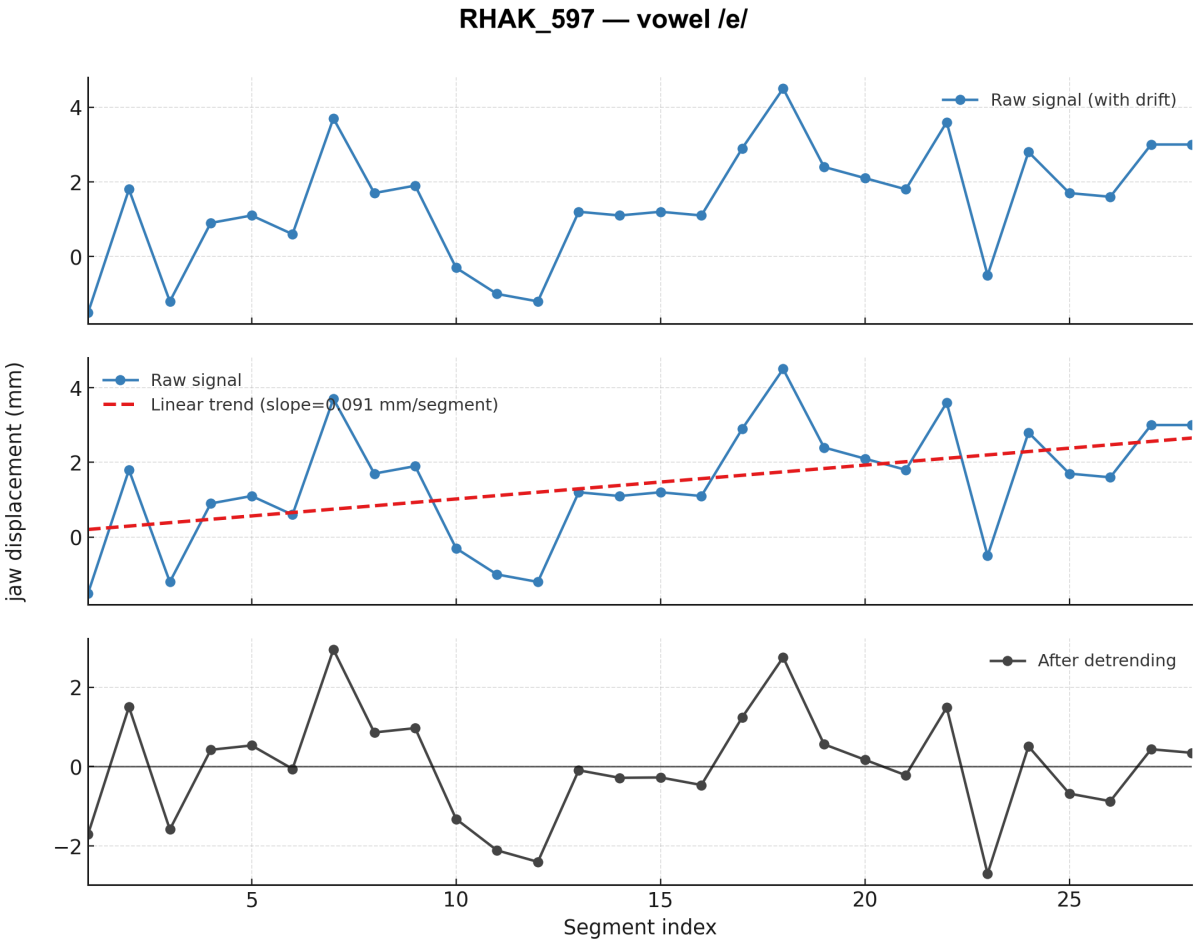

Figure 3.12. Illustration of the detrending procedure applied to vertical jaw displacement values for one recording (*RHAK_597*, vowel /e/). The upper panel shows the raw values with a slow drift. The middle panel presents the fitted linear trend (red dashed line), reflecting a gradual upward shift of jaw position across segments. The bottom panel displays the detrended signal, oscillating around zero and only reflecting articulatory variation.

it minimises anatomical variation across speakers while preserving phonologically relevant contrasts (cf. Adank et al. 2004).

Simply put, the z-score expresses how far a given data point deviates from the mean of that speaker's dataset, measured in standard deviation units rather than absolute millimeters. In other words, it creates a relative scale indicating the number of standard deviations a value lies above or below the mean.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{3.4}$$

As shown in Equation (3.4), each value x_{ij} is transformed by subtracting the speaker-specific mean μ_j and dividing by the speaker-specific standard deviation σ_j.

This method reduces individual differences, such as jaw movement range and calibration

settings, while preserving variation related to prominence. As a result, this procedure standardised all speakers onto a common scale, enabling meaningful cross-speaker comparisons and facilitating the use of linear mixed-effects models with generalization to the population level. Based on all the above mentioned, the per-speaker z-score transformation was selected as the optimal method.

Chapter 4

Results and discussion

This chapter reports the results obtained using the articulatory and acoustic methods and procedures described in Chapter 3. The main objective of the study is to examine the relationships between selected articulatory features including measures of jaw displacement, lip aperture, and segment duration, and utterance-level accent, contrastive focus, and vowel type.

The following sections provide visualisations of jaw and lip trajectories in two experimental conditions: **focus** and **neutral** conditions, along with descriptive statistics, and the outcomes of the linear mixed-effects models introduced in Chapter 3. These are also supplemented by acoustic analyses focusing on the rhythmic properties of the utterance and additional analyses extending beyond the main hypotheses.

For clarity, the following conventions are adopted throughout this chapter in all figures and graphical visualisations: utterances produced under the **neutral** condition are shown in blue, and those produced under the contrastive **focus** — in red. Three types of accent markings are distinguished throughout this chapter:

- Vowel positions are labelled as: *Accent* indicating the place of expected utterance-level accent; *Focus*, indicating the place of expected contrastive focus; and *Other* referring to all remaining positions within the utterance.
- Colour coding for the vowel positions is as follows: pale blue for *Accent*, pale red for *Focus*, and grey for *Other*.
- bold marking for conditions **focus** — utterances produced under constrastive focus condition — and **neutral** — utterances produced in neutral conditions.

4.1 Trajectories of vertical jaw movement under neutral and contrastive focus conditions

The following plots illustrate average vertical jaw movement trajectories during vowel production under two conditions: **neutral** and contrastive **focus**.

Please note, the trajectories are raw values only adjusted to phonetic segment durations and

average height (see Chapter 3, Section 3.3.9). The pauses annotated in the texgrids i.e. longer than the accepted threshold (cf. Chapter 3, Section 3.3.8.1) were cut out.

The transcription applied in all the plots within this chapter reflects the predominant pronunciation pattern observed in the majority of utterances, which occasionally contradicts the rules described in Polish phonetics textbooks. Examples of such being e.g.:

- word *zapas* /zapas/ (Eng. 'a supply') in the utterance *Ala dała Arkowi zapas buraków* (Eng. 'Ala gave Arek a supply of beets');
- word *Irek* /irek/ (Eng. 'Irek'[1]) in the utterance *Irek widzi kilka irysów na stoliku* (Eng. 'Irek sees a few irises on the table');
- word *bez* /bes/ (Eng. 'without') in the phrase *jeszcze bez adresu* /jeSt^Se bes adresu/ (Eng. 'no address yet') in utterance *Edek jedzie do Łeby, jeszcze bez adresu* (Eng. 'Edek is going to Łeba, no address yet')

The textbook descriptions state that any voiced consonant should trigger voicing assimilation in the preceding consonant, regardless of pronunciation variety, whether Warsaw-based or Cracow-Poznań-based.

In Figures 4.1– 4.12, thin curves represent individual recordings, while the bold curve marks the trajectory averaged across all speakers. Colour-shaded areas indicate vowels with expected utterance-level *Accent* (blue) and contrastive *Focus* stress (red). The horizontal x-axis represents time values, and the scale on this axis has been adjusted to represent median duration of phonemic segments for a given *Vowel* $\times$ *Condition*. Jaw displacement values are shown on the vertical y-axis, and the jaw positions have been centred to allow for comparison (see Chapter 3, Section 3.4.1.4 for more details). The y-axis values are given in millimetres (mm) and are scaled consistently across all vowels to maintain comparability.

The following conventions apply to the plots presented in Figures 4.1–4.12:

- vertical grey lines indicate segment boundaries;
- the red rectangles — only in **neutral** condition — which mark where the vowel in the *Focus* position is in the **focus** condition. Such marking was applied to facilitate comparison across these two conditions;
- the blue shading highlights the *Accent* position within the utterance;
- the red shading highlights the *Focus* position within the utterance — it is present only in utterances with contrastive **focus**.

[1]Polish given name.

Vowel /a/: under the contrastive **focus** condition, a pronounced downward jaw movement is observed, affecting not only the target vowel but also the preceding /d/. As a result, the entire word /dawa/ is longer than under the **neutral** condition, with a notable lengthening of the final /a/ (compare Figure 4.1 and Figure 4.2).

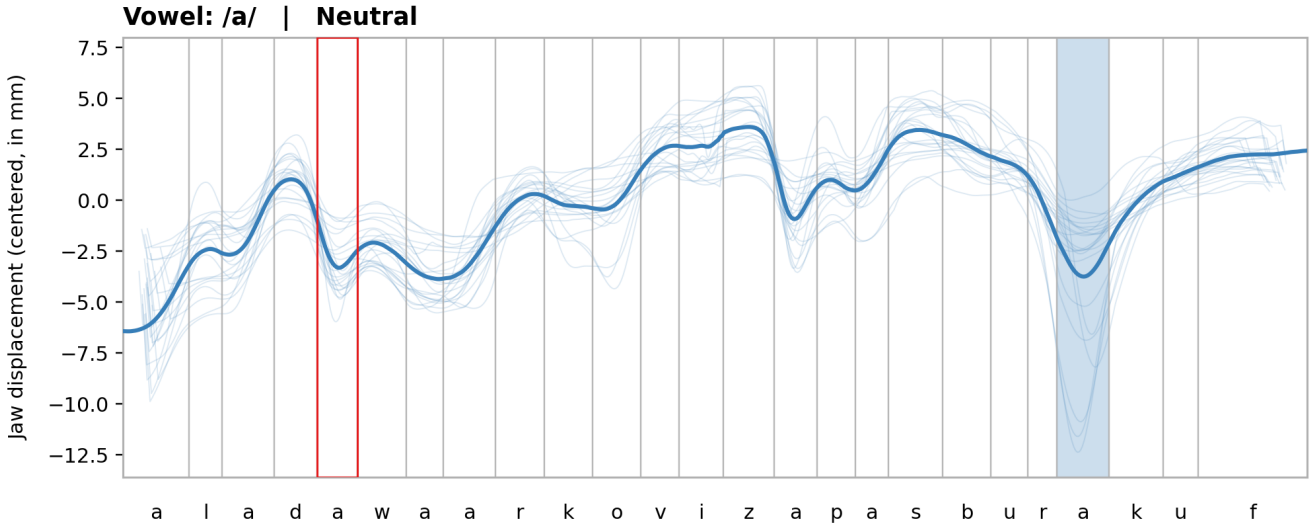

Figure 4.1. Mean trajectory of jaw displacement for the vowel /a/ across the realisations of the Polish utterance *Ala dała Arkowi zapas buraków* (/ala dawa arkovi zapas burakuf/, Eng. 'Ala gave Arek a supply of beets'). Bold curve: average over speakers. Thin curves: individual recordings. Neutral condition.

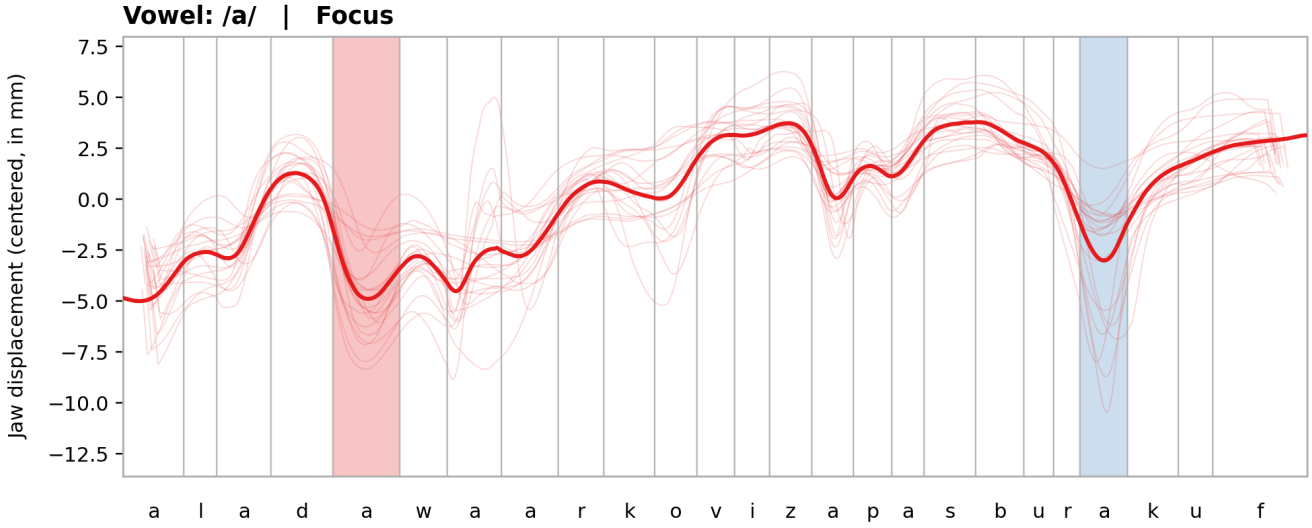

Figure 4.2. Mean trajectory of jaw displacement for the vowel /a/ across the realisations of the Polish utterance *Ala dała Arkowi zapas buraków* /ala dawa arkovi zapas burakuf/ (Eng. 'Ala gave Arek a supply of beets'). Bold curve: average over speakers. Thin curves: individual recordings. Focus condition. Contrastive focus on the word *dała* /dawa/ (Eng. 'gave').

Vowel /e/: under the contrastive **focus** condition, a definitely sharper and more lowered gesture is observed. Duration of /e/ is greater, however, the most standing out change is the prolonged final /I/ in /webI/ (compare Figure 4.3 and Figure 4.4). This would be discussed in detail in the end of the current section.

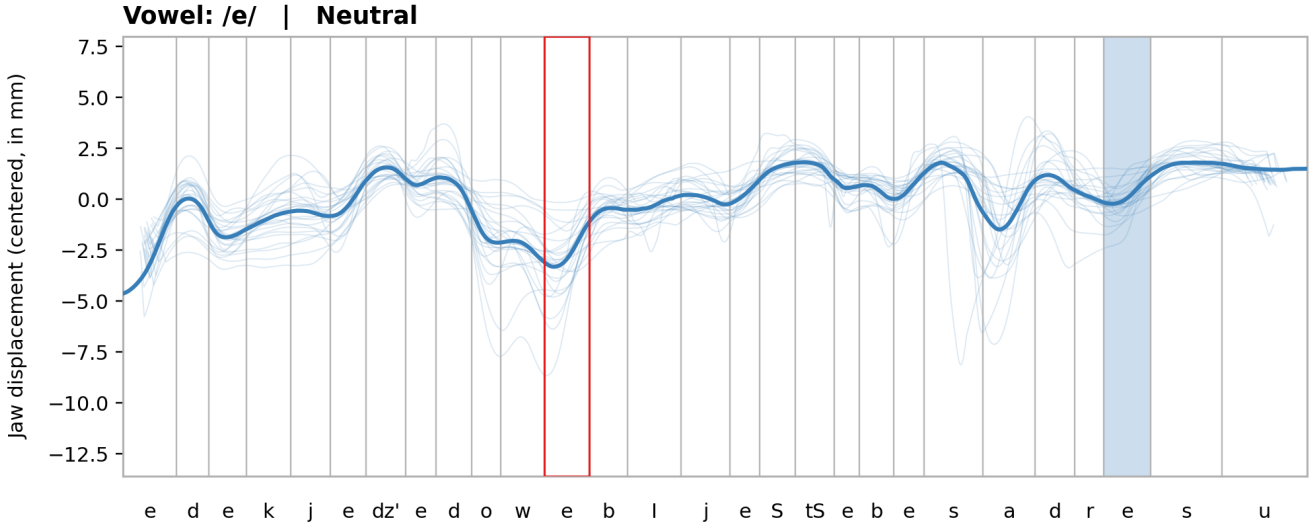

Figure 4.3. Mean trajectory of jaw displacement for the vowel /e/ across the realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet'). Bold curve: average over speakers. Thin curves: individual recordings. Neutral condition.

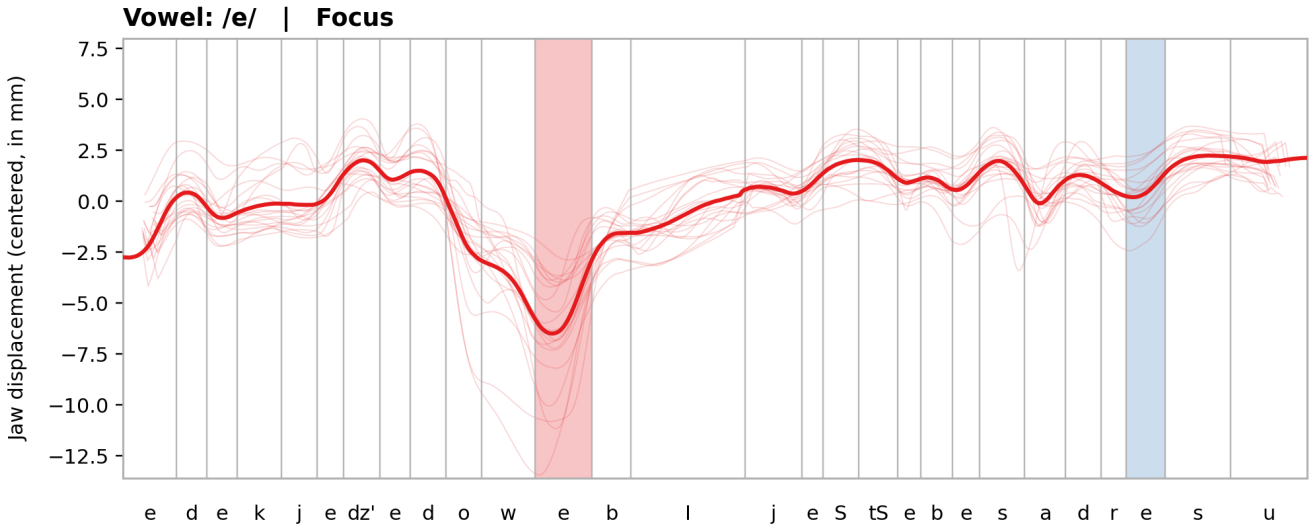

Figure 4.4. Mean trajectory of jaw displacement for the vowel /e/ across the realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet'). Bold curve: average over speakers. Thin curves: individual recordings. Focus condition. Contrastive focus on the word *jedzie* /jedˆSz'e/ (Eng. 'is going').

Vowel /i/: Overall, the jaw movements remain constrained, which is consistent with the high vowel articulation. Nevertheless, presence of contrastive **focus** introduces slightly broader excursions and, similarly to /e/, there is a definite lengthening of the final /a/ in /kilka/ and also, greater jaw displacement (compare Figure 4.5 and Figure 4.6).

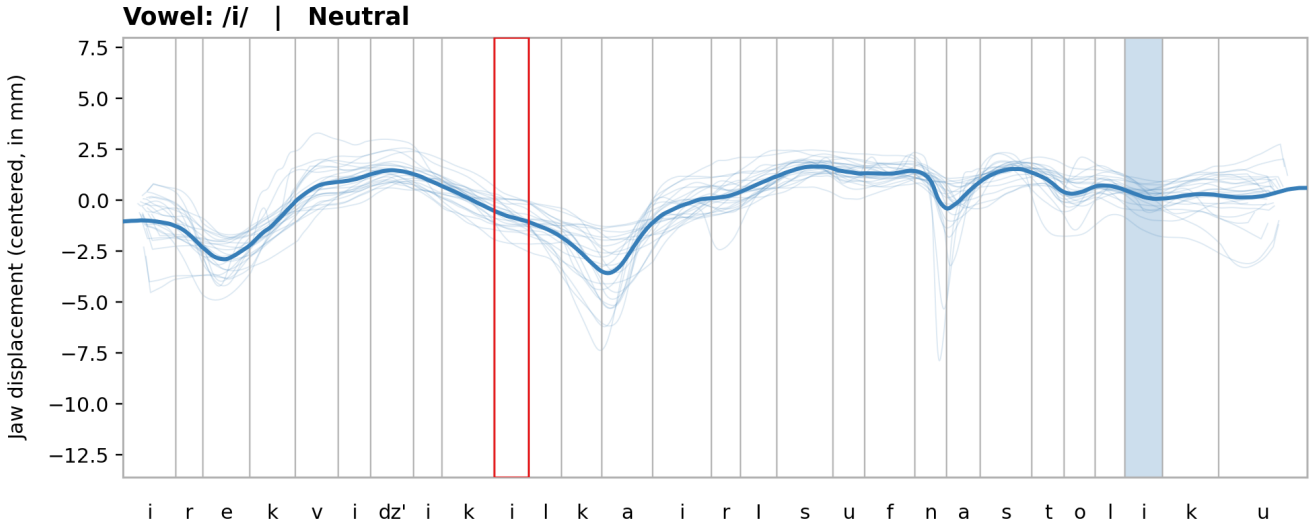

Figure 4.5. Mean trajectory of jaw displacement for the vowel /i/ across the realisations of the Polish utterance *Irek widzi kilka irysów na stoliku* /irek vidˆz'i kilka irIsuf na stoliku/ (Eng. 'Irek sees a few irises on the table'). Bold curve: average over speakers. Thin curves: individual recordings. Neutral condition.

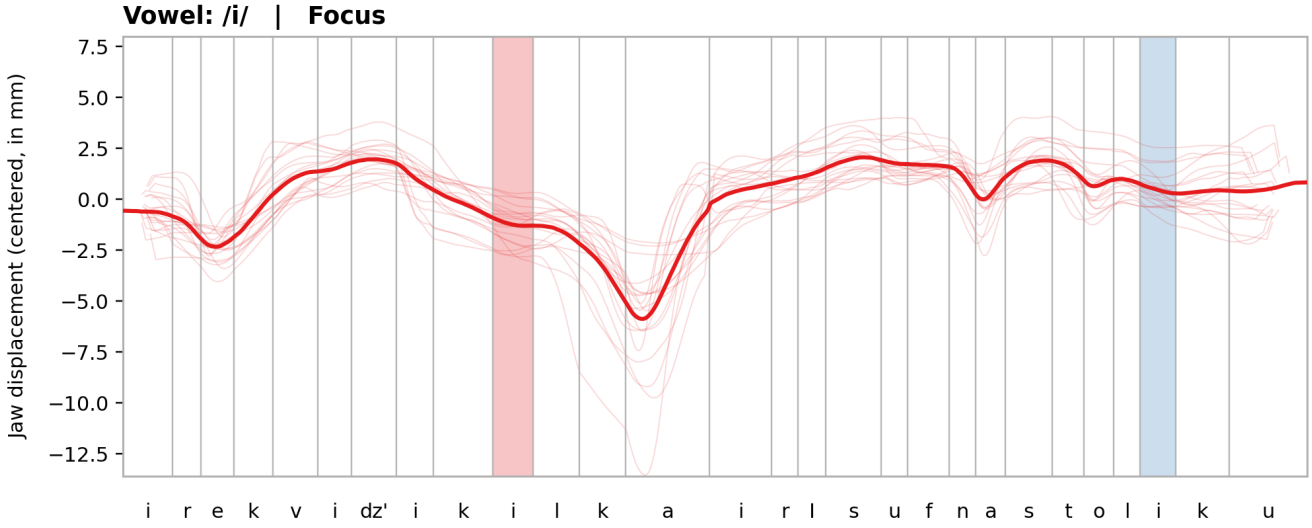

Figure 4.6. Mean trajectory of jaw displacement for the vowel /i/ across the realisations of the Polish utterance *Irek widzi kilka irysów na stoliku* /irek vidˆz'i kilka irIsuf na stoliku/ (Eng. 'Irek sees a few irises on the table'). Bold curve: average over speakers. Thin curves: individual recordings. Focus condition. Contrastive focus on the word *kilka* /kilka/ (Eng. 'a few').

Vowel /o/: Patterns follow the emerging prominence effect: contrastive **focus** leads to expanded jaw displacement, though less markedly than for low vowels, and lengthening of the final /a/ in the word /kota/ (compare Figure 4.7 and Figure 4.8).

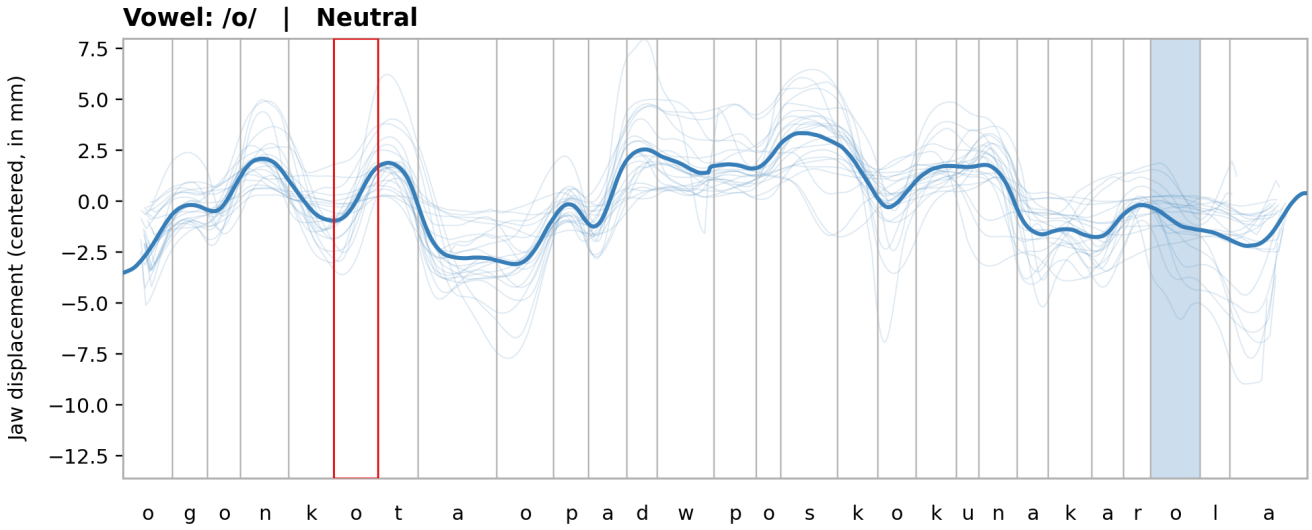

Figure 4.7. Mean trajectory of jaw displacement for the vowel /o/ across the realisations of the Polish utterance *Ogon kota opadł po skoku na Karola* /ogon kota opadw po skoku na karola/ (Eng. 'A cat's tail has dropped after jumping on Karol'). Bold curve: average over speakers. Thin curves: individual recordings. Neutral condition.

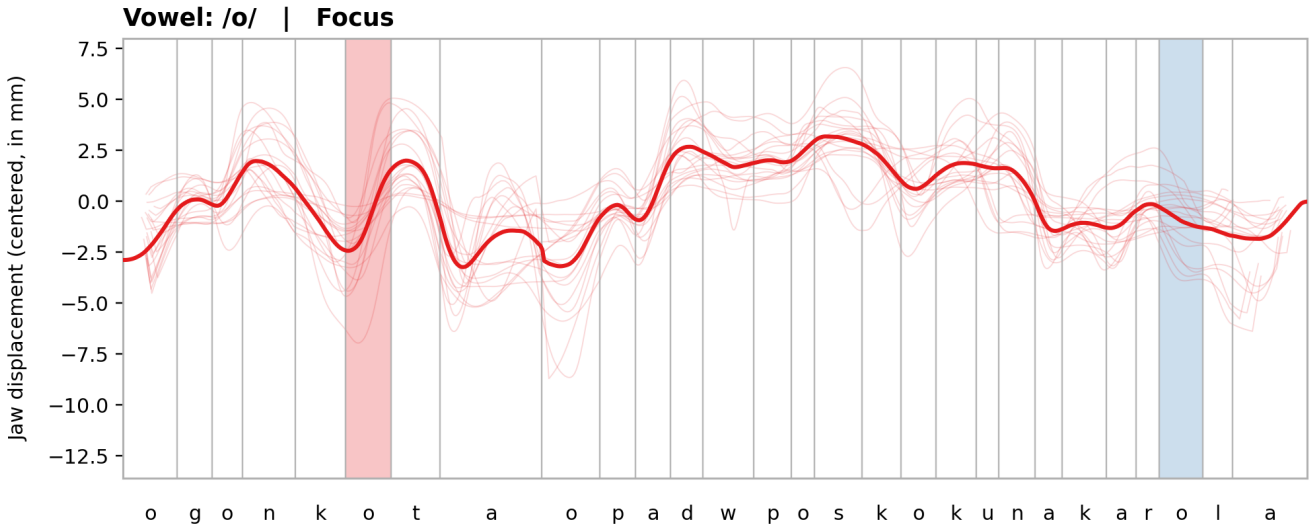

Figure 4.8. Mean trajectory of jaw displacement for the vowel /o/ across the realisations of the Polish utterance *Ogon kota opadł po skoku na Karola* /ogon kota opadw po skoku na karola/ (Eng. 'A cat's tail has dropped after jumping on Karol'). Bold curve: average over speakers. Thin curves: individual recordings. Focus condition. Contrastive focus on the wor on the word *kota* /kota/ (Eng. 'cat's').

Vowel /u/: The neutral trajectories are compact, while the **focus** condition triggers subtle increases in amplitude; the highlighted word has final vowel lengthening — /a/ in /vujka/ (compare Figure 4.9 and Figure 4.10).

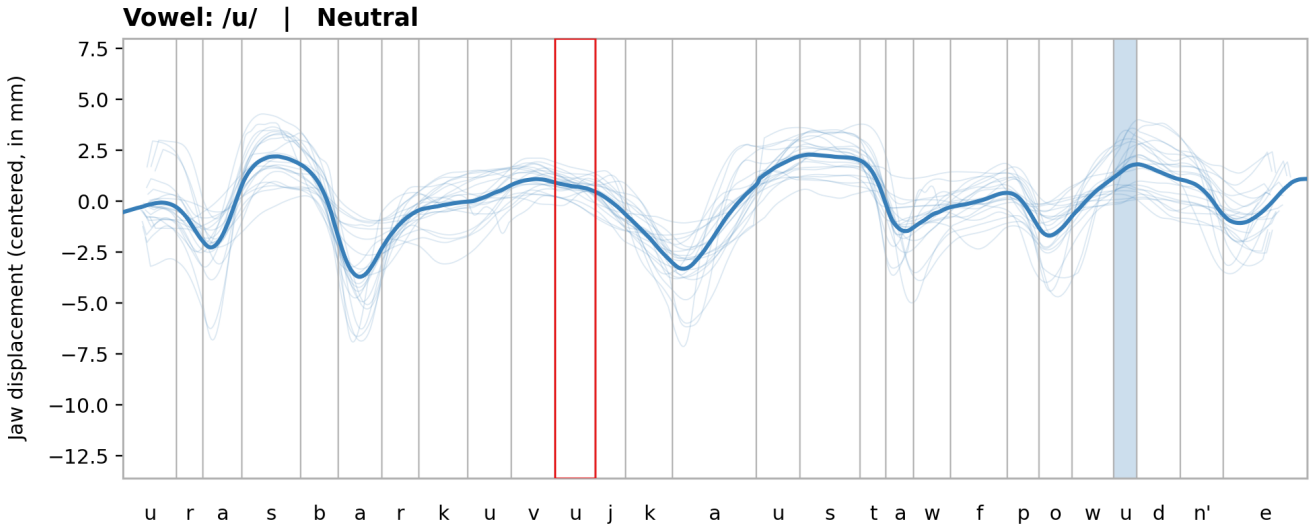

Figure 4.9. Mean trajectory of jaw displacement for the vowel /u/ across the realisations of the Polish utterance *Uraz barku wujka ustał w południe* /uras barku vujka ustaw fpowudn'e/ (Eng. 'Uncle's shoulder injury cleared up at noon.'). Bold curve: average over speakers. Thin curves: individual recordings. Neutral condition.

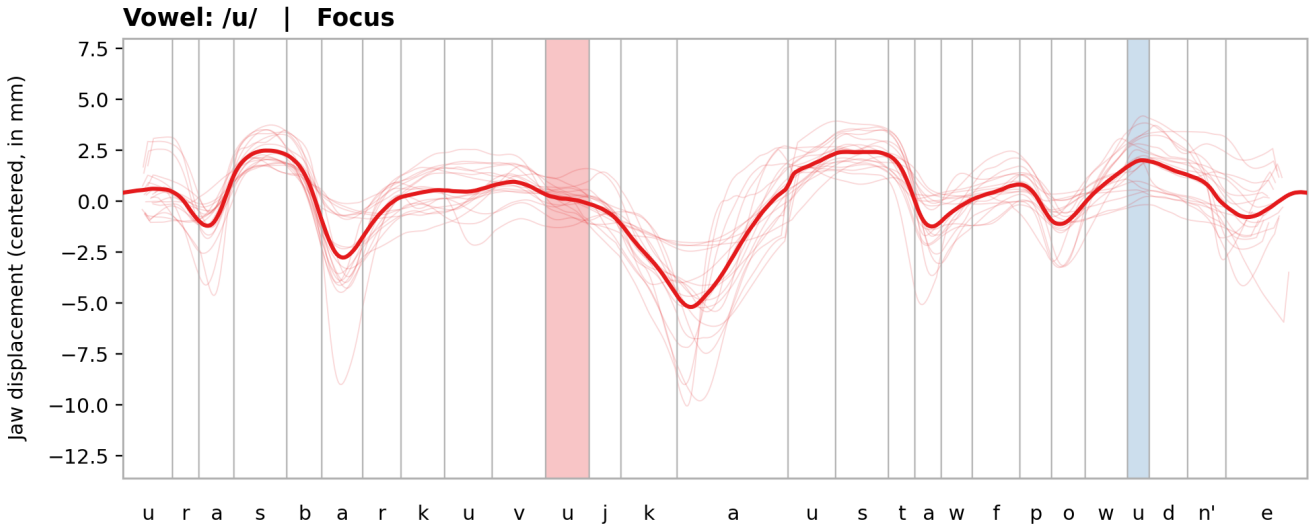

Figure 4.10. Mean trajectory of jaw displacement for the vowel /u/ across the realisations of the Polish utterance *Uraz barku wujka ustał w południe* /uras barku vujka ustaw fpowudn'e/ (Eng. 'Uncle's shoulder injury cleared up at noon.'). Bold curve: average over speakers. Thin curves: individual recordings. Focus condition. Contrastive focus on the word *wujka* /vujka/ (Eng. 'uncle's).

Vowel /I/: Jaw movement is compacted, minimal in both conditions. Similarly as in /i/, this is consistent with the high vowel articulation. In this case, the word under contrastive **focus** displays a modest downward shift and final vowel lengthening (compare Figure 4.11 and Figure 4.12).

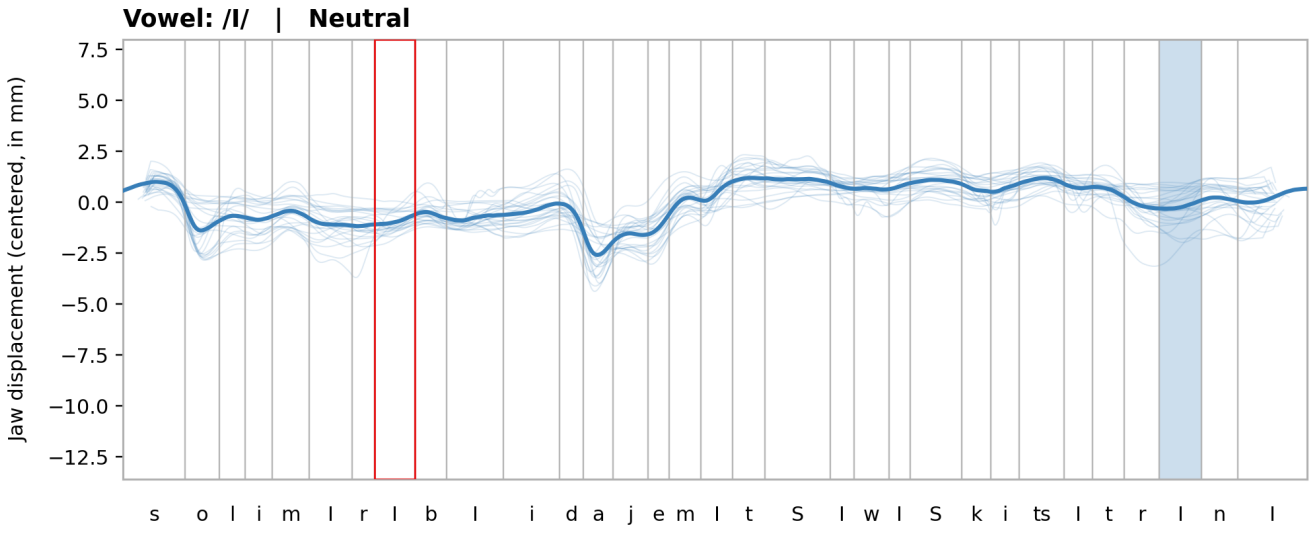

Figure 4.11. Mean trajectory of jaw displacement for the vowel /I/ across the realisations of the Polish utterance *Solimy ryby i dajemy trzy łyżki cytryny* /solimI rIbI i dajemI tSI wISki tˆsItrInI/ (Eng. 'Season the fish with salt and add three tablespoons of lemon juice.'). Bold curve: average over speakers. Thin curves: individual recordings. Neutral condition.

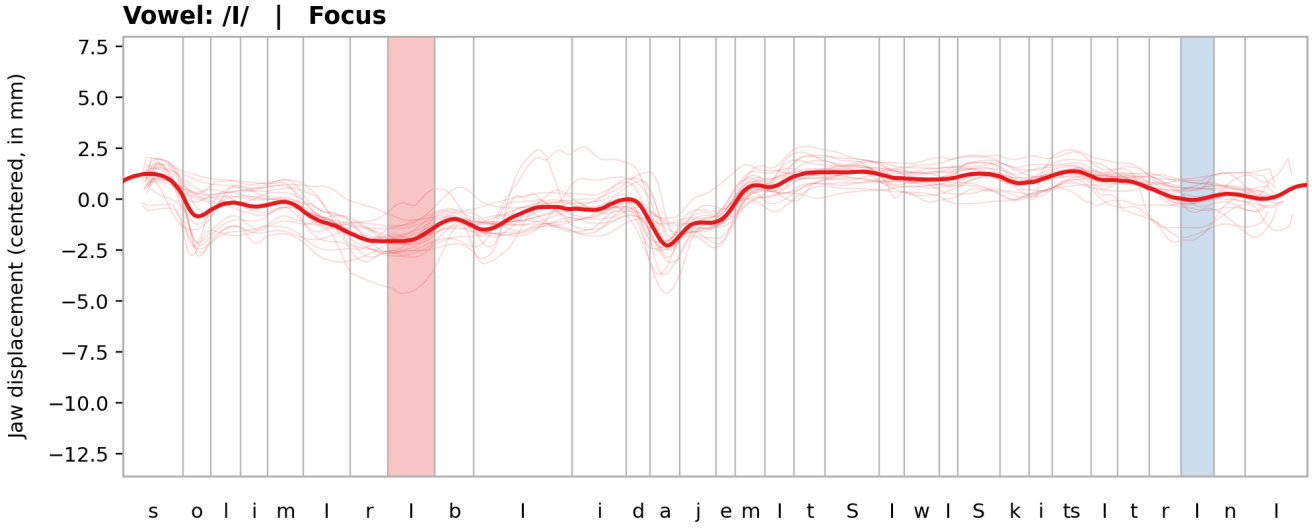

Figure 4.12. Mean trajectory of jaw displacement for the vowel /I/ across the realisations of the Polish utterance *Solimy ryby i dajemy trzy łyżki cytryny* /solimI rIbI i dajemI tSI wISki tˆsItrInI/ (Eng. 'Season the fish with salt and add three tablespoons of lemon juice.'). Bold curve: average over speakers. Thin curves: individual recordings. Focus condition. Contrastive focus on the word *ryby* /rIbI/ (Eng. 'fish').

Some trailing artefacts are visible at trajectory boundaries — they should be disregarded as an effect of normalisation and do not affect the interpretation of these plots.

The trajectories exhibit systematic patterns of jaw displacement across vowels, with **consistent** differences between **neutral** and contrastive **focus** conditions. In the **neutral** context, movements appears to be more uniform, while in utterances with contrastive **focus** trajectories show larger jaw displacements and sharper transitions, reflecting greater articulatory dynamics.

In case of /e/, the prolonged final /I/ might be related to the syntactic structure of the phrase *Edek jedzie do Łeby, jeszcze bez adresu* (Eng. 'Edek is going to Łeba, no address yet'). The part *Edek jedzie do Łeby* /edek jedˆz'e do webI/ might function as a full sentence on its own and the second part — *jeszcze bez adresu* /jeSˆtSe bes adresu/ — is an adverbial phrase. There is also a comma graphically separating the two phrases. Such structure may encourage speakers to highlight the boundary between the two intonational phrases, resulting in an extra lengthening of the final /I/ in *Łeby* /webI/, which serves both as a focus marker and as a cue to syntactic segmentation (cf. Francuzik et al., 2002).

Another possible explanation is employing a focus marking strategy by prolonging the highlighted word, here expressed in **lengthening of the final vowel in focus word**. This would be supported by a noticeable trend from the other vowel examples — consistent prolongation of the final vowel in the highlighted word is present across all the examined vowels.

4.2 Descriptive statistics

4.2.1 General characteristics of the dataset

As described in Chapter 3, Section 3.3.8, the dataset comprised 266 recording sessions, from which 7,449 phonetic segments and 187,839 EMA sensor position samples were obtained.

For the statistical modelling purposes, the data were condensed into single records, each corresponding to a single phonemic segment and containing measures of duration (in ms), jaw position, lip aperture (both in two parallel forms: raw values in millimetres and normalised values in standard deviation units, referred to as SD). The distribution of factors across speakers, vowel categories, and **focus** conditions is summarised in Table 3.4 (see Chapter 3, Section 3.4).

As discussed earlier in the context of data imbalance, (i.e., those labelled as *Accent* occurring in the expected stress position and those labelled as *Focus* produced with contrastive focus) relative to unstressed ones (labelled as *Other* and occurring in all remaining positions), since only one vowel per utterance can occur in either the *Focus* or *Accent* position. This asymmetry is an inherent property of the design and was taken into account in the statistical modelling through the use of linear mixed-effects models.

Descriptive statistics for the dependent variables (duration, jaw displacement, and lip aperture) are presented in the sections below. Each variable is analysed on its own, with tabulated data accompanied by boxplots and other visualisations to illustrate the distributions. Although **vowel duration was not part of the main hypotheses**, it is reported descriptively due to a consistent pattern of final vowel lengthening reported in the previous section. This may

reflect an additional focus marking strategy and provides relevant context for interpreting prominence-related articulatory patterns.

Because the contrastive **focus** condition affects the entire utterance by modulating the duration of phonemic segments and slightly altering their movement trajectories (see the plots in the previous section, the plots in upcoming Section 4.7, as well as the commentary in Section 4.6), descriptive statistics are presented separately for each condition — **neutral** and with contrastive **focus**. Additionally, as shown in the above-mentioned trajectory plots, at the onset of some utterances the jaw was lowered, since the lips were not closed. This may result partly from the experimental laboratory context, certainly having an impact on the preference to have one's mouth open (see Figure 4.13), and partly from phrase boundries (cf. Mołczanow et al., 2018). Given that both lips and jaw positions at utterance onset may not be representative, these segments were excluded from the statistical analyses in order to minimise potential artefacts less likely to occur under natural speech conditions.

Figure 4.13. A speaker prepared for an Electromagentic Articulography (EMA) recording session using the Carstens AG501 system. Receiver sensors are already attached. The speaker is adjusting cables in her mouth.

With these exclusions in place, the analysis proceeds to the presentation of descriptive statistics for individual articulatory dimensions.

4.2.1.1 Vertical jaw displacement

Normalised minimum, maximum, and mean jaw displacement (in SD) for each vowel, broken down by target vowel position are presented in Table 4.1.

Mean values averaged across all speakers for jaw displacement for *Accent* positions tend to be consistently lower across vowels with the exception of /u/ in which the difference is minimal. The most pronounced displacement is observed in /a/ where *Accent* positions are markedly lower on average (-2.2) compared to *Other* (-0.9), indicating stronger jaw displacement under accent. A similar tendency, though weaker, is observed for /o/ (-0.9 vs -0.1). For /e/ and /I/, means are close to zero in both positions, showing minimal positional effects.

The boxplot in Figure 4.14 illustrates a strong effect of *Accent* position for /a/, reflecting a markedly lowered jaw; a similar but weaker tendency occurs for /o/. In contrast, high vowels /i/, /I/,/u/, and a mid vowel /e/ cluster closer to zero, with *Accent* positions only slightly lower than *Other* ones.

Table 4.1. The displacement range and mean of normalised vertical jaw displacement for each vowel averaged across all speakers, measured at the time of minimum velocity, by position (*Accent* = utterance-level accent; *Other* = not accented)

Vowel	*Accent*				*Other*			
	N	**Max**	**Min**	**Mean**	**N**	**Max**	**Min**	**Mean**
/a/	23	-5.6	-0.5	-2.2	138	-2.9	1.1	-0.9
/e/	26	-1.3	0.4	-0.3	182	-2.5	1.2	0.0
/i/	24	-0.5	4.6	0.2	96	-0.9	1.5	0.4
/o/	20	-2.4	0.0	-0.9	118	-3.1	1.6	-0.1
/u/	22	-0.5	2.8	0.6	66	-0.7	2.1	0.5
/I/	26	-0.8	0.4	-0.2	208	-1.1	1.1	0.1

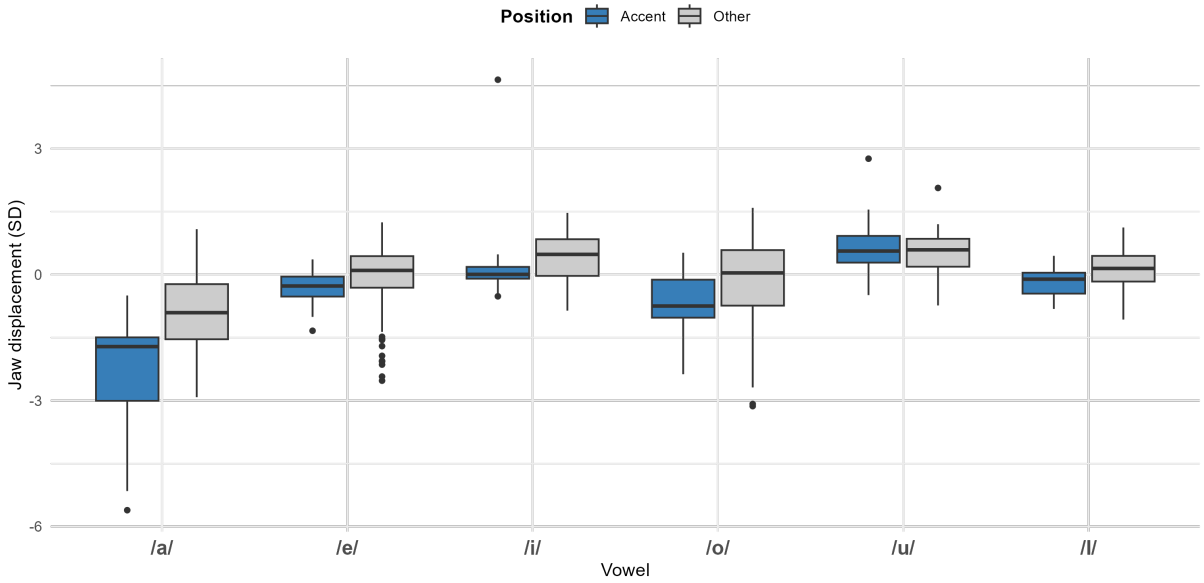

Figure 4.14. Normalised jaw displacement (SD) across vowels in *Accent* (blue) and *Other* (grey) positions averaged across all speakers.

The role of vowel height effects on jaw displacement has been discussed in Erickson and Kawahara (2016) who tackled the issue by implementing a neutralisation procedure (Williams et al., 2013). Here, the experiment controlled for vowel quality, given that vowel height inherently influences the degree of jaw opening. Therefore, an additional neutralisation procedure was not applied. A neutralisation readjusted for Polish could certainly offer new insights, however this type of analysis is beyond the scope of the present dissertation.

Summing up, *Accent* positions tend to produce more extreme articulatory values, while *Other* positions remain closer to the neutral baseline.

The table reports jaw displacement values in SD units. A consistent effect is observed for the low vowel /a/, which shows the most negative means in both *Accent* (-2.0 SD) and *Focus* (-2.1 SD) positions, indicating substantial jaw lowering compared to the *Other* position (-0.8).

Table 4.2. The displacement range and mean of normalised vertical jaw displacement for each vowel averaged across all speakers, measured at the time of minimum velocity, by position (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

Vowel	*Accent*				*Focus*				*Other*			
	N	**Max**	**Min**	**Mean**	**N**	**Max**	**Min**	**Mean**	**N**	**Max**	**Min**	**Mean**
/a/	22	-4.9	-0.4	-2.0	22	-4.3	0.0	-2.1	110	-4.2	1.6	-0.8
/e/	21	-0.9	0.6	-0.1	21	-5.7	-1.8	-3.0	126	-0.9	1.6	0.3
/i/	20	-0.5	0.7	0.0	20	-1.4	0.3	-0.5	60	-0.9	1.4	0.5
/o/	16	-2.1	0.3	-0.9	20	-3.0	2.0	-0.6	84	-3.5	1.5	-0.1
/u/	19	0.0	5.2	1.1	19	-0.7	1.7	0.2	38	-1.2	4.0	0.7
/I/	20	-1.1	0.6	-0.1	20	-2.0	0.0	-0.9	140	-1.3	1.0	0.1

Notably, for vowel /e/, the *Focus* position stands out with a more negative mean (-3.0 SD) relative to *Accent* (-0.1 SD) and *Other* (+0.3 SD) positions. This was also clearly visible in the plots presented in the Section *Trajectories of vertical jaw movement under neutral and contrastive focus conditions* where jaw lowering for *Focus* position in /e/ was very pronounced. High vowels /i/, /u/, and /I/ show smaller positional contrasts, with values clustering closer to zero. Vowel /o/ displays modest negative displacement in both *Accent* and *Focus* positions.

In summary, low vowel /a/ and mid vowel /e/ show greater jaw lowering in both *Accent* and *Focus* positions, while high vowels remain relatively stable across contexts.

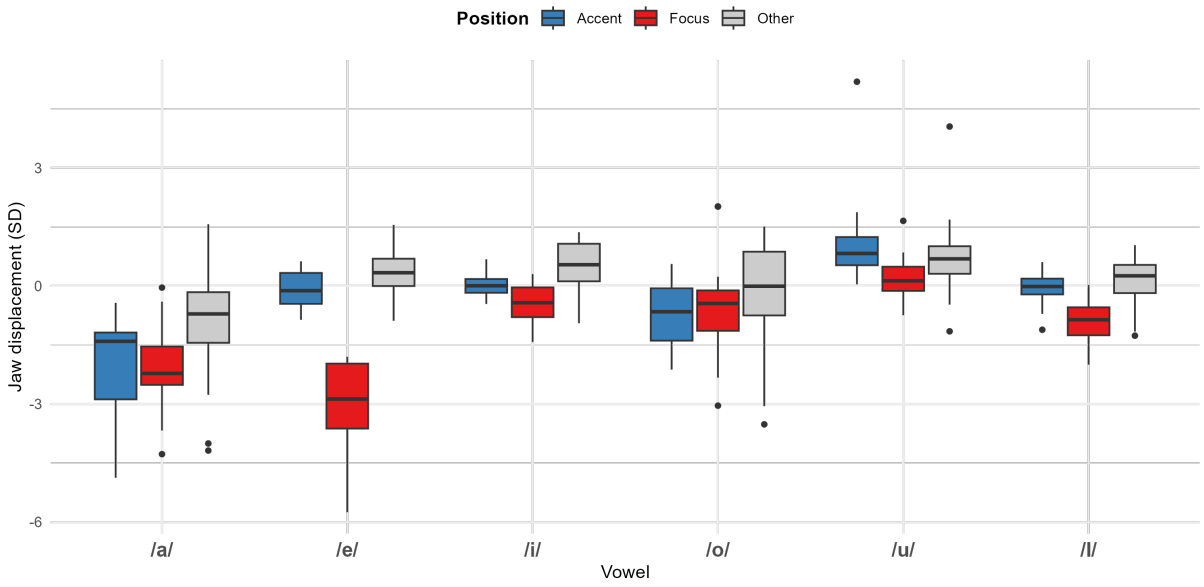

Figure 4.15. Normalised jaw displacement (SD) across vowels in *Accent* (blue) and *Other* (grey) positions averaged across all speakers.

By comparing mean jaw displacement in **neutral** and **focus** conditions, a broadly consistent pattern emerges (see Figure 4.16). Solid lines indicate mean jaw displacement values in *Accent* positions; both **neutral** and **focus** condition utterances follow similar trends, though with subtle shifts. In **focus** condition, vowels in both *Accent* and *Other* positions tend to show less displacement relative to utterances produced under the **neutral** condition.

These results suggest that the presence of contrastive **focus** reduces positional contrasts in terms of jaw displacement (with the exception of /o/ and /i/). In other words, presence of contrastive focus tends to bring the largest displacement while decreasing this displacement in non-*Focus* positions.

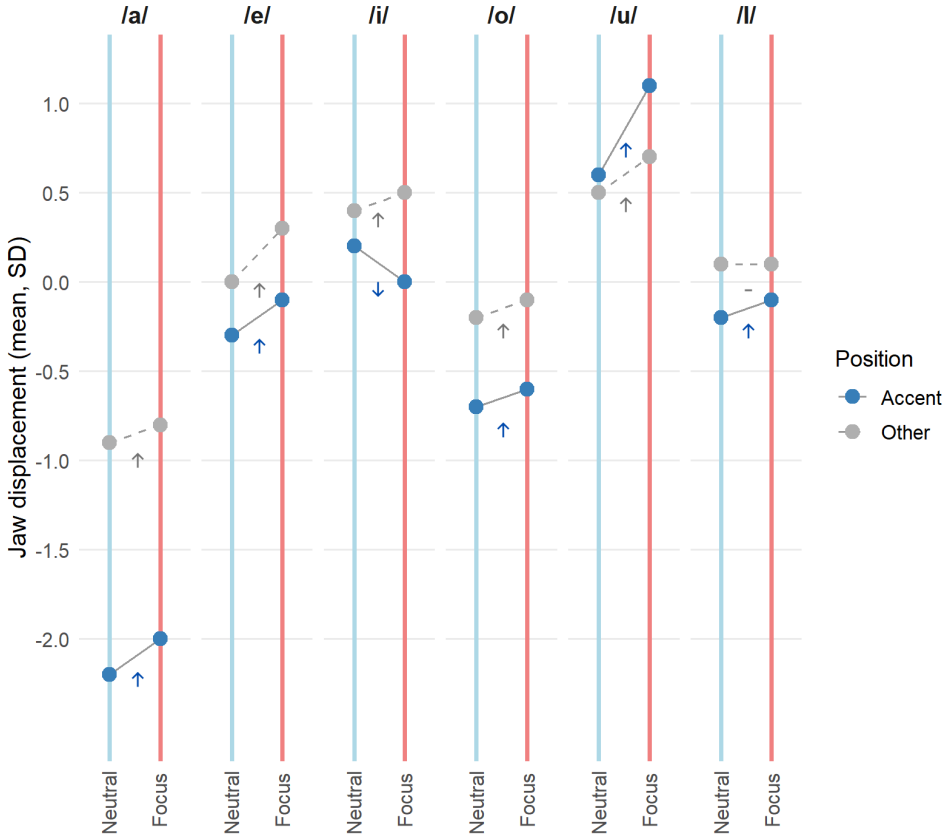

Figure 4.16. Normalised jaw displacement (SD) across vowels in **neutral** and **focus** utterances, by *Accent* and *Other* positions. The arrows indicate the direction of change. In **focus** condition vertical jaw displacement for non-*Focus* positions, both *Accent* and *Other*, tend to be less pronounced.

4.2.1.2 Lip aperture

Normalised maximum, minimum, and mean lip aperture (in SD) for each vowel, broken down by target vowel position are presented in Table 4.3 and Table 4.4.

For **/a/**, *Accent* positions show substantially higher means (2.9 SD) compared to *Other* (0.8 SD), implying a much wider lip aperture in prominent positions; this is also visible in Figure 4.17. **/e/** has a less pronounced but similar pattern, with vowels in *Accent* positions averaging 0.7 SD versus 0.1 SD in *Other* positions.

High vowels display a different tendency; both /i/ and /u/ show negative or near-zero means, with no clear contrast between *Accent* and *Other* positions. This might be explained by their generally narrower aperture. /o/ shows a slightly positive mean in *Accent* (0.2 SD) versus a negative mean in *Other* (-0.3 SD). And for /I/, vowels in *Accent* positions are more open (0.4

Table 4.3. Mean lip aperture (SD) across vowels in neutral utterances, by position (*Accent* = utterance-level accent; *Other* = not accented)

Vowel	*Accent*				*Other*			
	N	**Min**	**Max**	**Mean**	**N**	**Min**	**Max**	**Mean**
/a/	23	1.0	6.4	2.9	138	-1.3	2.4	0.8
/e/	26	-0.3	2.0	0.7	182	-1.3	2.4	0.1
/i/	24	-1.0	2.3	-0.1	96	-1.4	1.8	-0.1
/o/	20	-0.4	1.5	0.2	118	-1.4	2.1	-0.3
/u/	22	-1.4	0.4	-0.7	66	-1.3	0.8	-0.7
/I/	26	-1.4	2.3	0.4	208	-1.4	1.8	-0.2

SD) than in *Other* (-0.2 SD) positions, though the effect is modest.

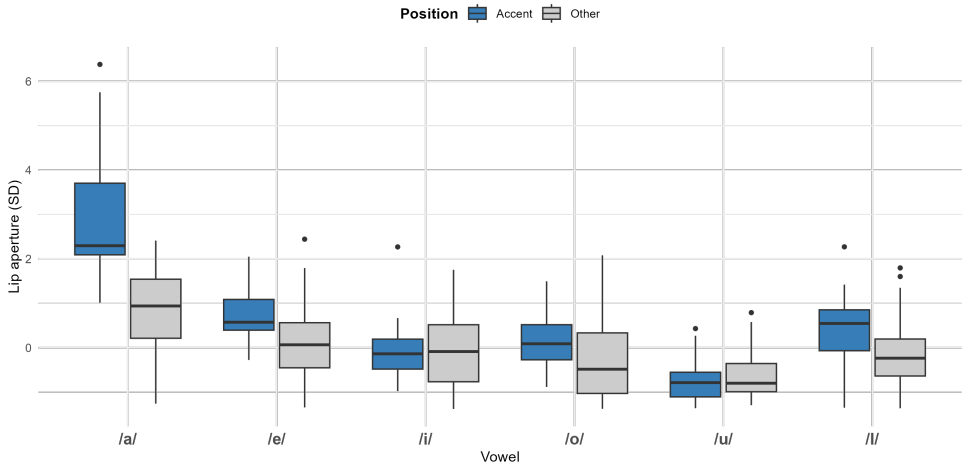

Figure 4.17. Normalised lip aperture (SD) across vowels in *Accent* (blue) and *Other* (grey) positions.

These results might imply that ***Accent* position enhances articulatory contrasts in lip aperture**. The strongest effect is displayed for the low vowel /a/ which aligns with the physiological expectation that low vowels involve greater jaw and lip opening, which is further exaggerated under contrastive *Focus* position. In contrast, high vowels show little or no change in aperture depending on their position, consistent with their inherently constrained aperture.

In utterances produced under contrastive **focus** condition, the greatest lip aperture in *Focus* position is displayed in /a/ and /e/ (see Figure 4.18). Even high vowels show distinct tendencies: for /i/, *Focus* position is linked to greater lip opening (+1.1 SD), whereas in *Accent* and *Other* positions the change remains very moderate. Similarly, /u/ and /I/ also display this change for *Focus* vs the rest of positions. For vowel /o/, a small negative mean in *Accent* (-0.1 SD), becomes positive in *Focus* (+0.6 SD), and turns negative again in *Other* (-0.4 SD).

As for the overall pattern for *Accent* and *Other* positions, it remains the same — vowel /a/ displays the greatest aperture (2.6 SD vs 1.0 SD), while high vowels /i/ and /u/ show very moderate change in opening.

Table 4.4. Mean lip aperture (SD) across vowels in utterances with contrastive focus, by position (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

Vowel	*Accent*				*Focus*				*Other*			
	N	**Min**	**Max**	**Mean**	**N**	**Min**	**Max**	**Mean**	**N**	**Min**	**Max**	**Mean**
/a/	22	0.0	5.4	2.6	22	0.1	3.2	2.1	110	-1.3	4.1	1.0
/e/	21	-0.5	1.7	0.6	21	0.2	5.6	2.2	126	-1.4	1.5	-0.1
/i/	20	-1.2	0.5	-0.4	20	-0.3	2.1	1.1	60	-1.4	1.8	-0.2
/o/	16	-1.0	0.7	-0.1	20	-0.9	2.7	0.6	84	-1.3	2.5	-0.4
/u/	19	-1.3	4.3	-0.5	19	-1.0	2.3	0.1	38	-1.3	2.7	-0.6
/I/	20	-0.7	1.4	0.1	20	-0.1	1.6	0.6	140	-1.4	1.7	-0.2

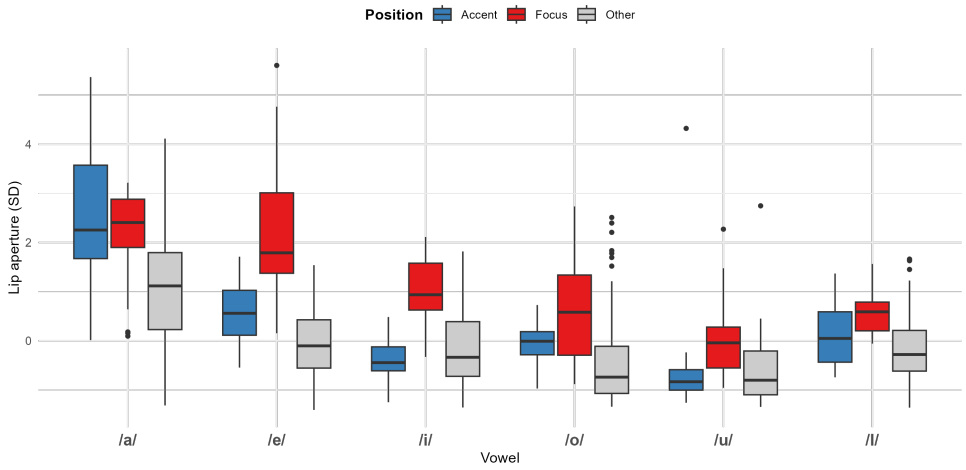

Figure 4.18. Normalised lip aperture (SD) across vowels in *Accent* (blue), *Focus* (red), and *Other* (grey) positions. As already discussed in Section 4.2.1.1, presence of contrastive focus in an utterance decreases effect in non-*Focus* positions. Here it reduces lip aperture relative to utterances produced under neutral condition.

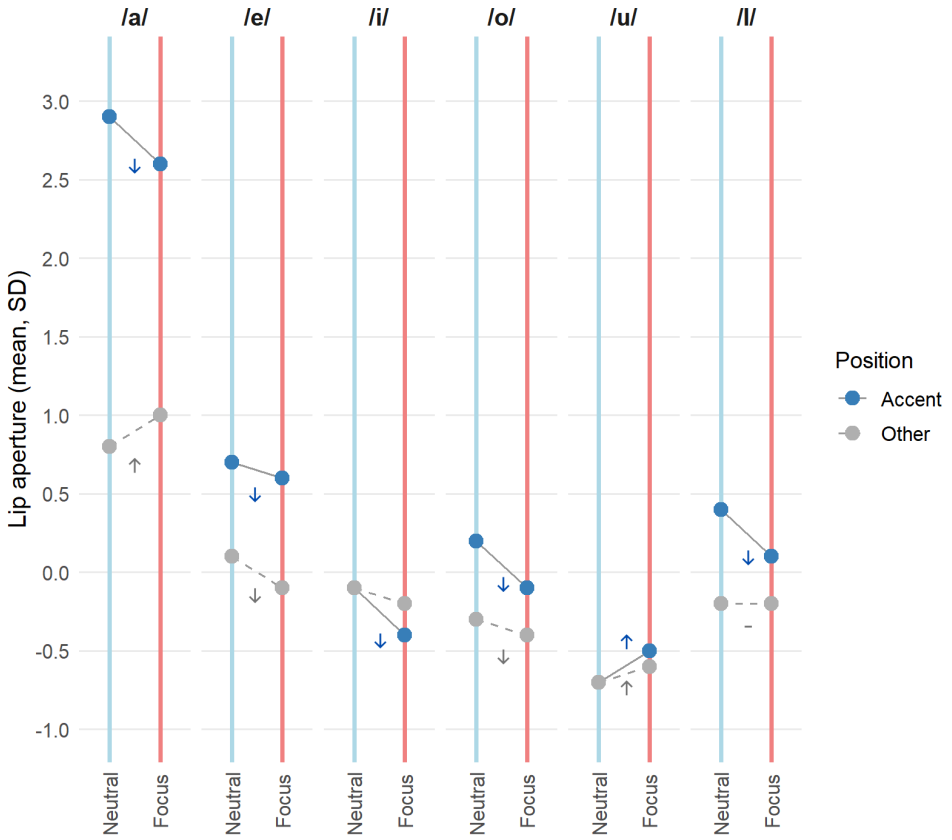

Figure 4.19. Mean lip aperture (SD) across vowels in utterances produced under **neutral** and **focus** conditions, in non-*Focus* positions. The arrows indicate the direction of change — for both *Accent* and *Other* positions pattern emerges, with the exception of the vowel /u/ (see Section 4.2.1.3 for more details on realisation of /u/). The pattern is as follows — the lip aperture for vowels in *Accent* position under **focus** condition is **smaller** than in vowels in *Accent* position under **neutral** condition. Same applies for vowels in *Other* positions.

Although the original plan focused on vertical jaw displacement and lip aperture, the trajectory analysis of vertical jaw displacement presented in the previous section (see Section 4.1) revealed that vowel duration may also play a relevant role. This measure was therefore incorporated into the analysis, and the following section is devoted to its examination.

4.2.1.3 Vowel duration

The number of cases, mean durations, and standard deviations for each vowel averaged across all speakers, broken down by prosodic condition are presented in Table 4.5 for utterances under **neutral** condition and Table 4.7 for utterances under **focus** condition. In Table 4.5, it is clearly observable how vowel duration depends on their cate and articulatory features — it decreases significantly for high vowels.

Table 4.5. Mean vowel durations (in ms) and standard deviations for neutral utterances by position (*Accent* = utterance-level accent; *Other* = not accented)

Vowel / Position	*Accent*			*Other*		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
/a/	23	104	18	138	88	36
/e/	26	106	27	182	74	35
/i/	24	77	25	96	79	29
/o/	20	111	19	118	88	33
/u/	22	57	26	66	95	36
/I/	26	98	34	208	98	55

In **neutral** condition utterances, vowels in *Accent* positions do not display a consistent pattern in terms of duration; the differences between *Accent* and *Other* positions vary across vowel types. Notably, /a/, /e/, and /o/ show clear durational enhancement in *Accent* position, whereas the vowel /u/ in the *Accent* position is in fact shorter than in *Other* positions. Vowels /i/ and /I/ reveal only minimal differences.

Regarding the duration of the vowel /u/, it was observed to be shorter in the *Accent* position. It is not a segmentational artefact, as the the same procedure was applied to all phonemic segments: the segmentation was first performed automatically and then corrected manually — the boundaries of each phonemic segment were determined based on both auditive and visual input, using the spectrogram and the oscillogram. The reduced duration of /u/ in this context can possibly be attributed to the fact that it is preceded by /w/, a labio-velar consonant articulated at the back of the oral cavity with lip rounding and other articulatory characteristics very similar to /u/.

Standard deviations are generally higher in the *Other* position, indicating greater variability when no prosodic prominence is present. However, this effect is not uniform across vowel types with /a/, /o/, and /I/ displaying more variability.

The interaction between vowel quality and duration reveals interesting patterns: vowels situated more centrally **in the formant space**, such as /a/, /e/, and /o/, were lengthened in *Accent* and *Focus* position, while peripheral vowel /u/ displayed reduction. In contrast, the high front vowels /i/ and /I/ remained temporally stable, likely due to their articulatory compactness, which could limit their susceptibility to change under prosodic variation. The observed duration patterns to some extent align with the formant distribution seen in the F1–F2 vowel space, which might indicate a potential interaction between vowel height/backness and prosodic prominence effects.

Table 4.6. Vowel-specific duration changes across prosodic conditions in relation to formant-based vowel classification. Modified from Jassem (1992) by adding new data — changes in duration.

Vowel	Δ Duration	F1 (height)	F2 (backness)
/a/	longer	low	central
/e/	longer	mid	front
/i/	no change	high	front
/o/	longer	mid	back
/u/	shorter	low	back
/I/	no change	mid	front

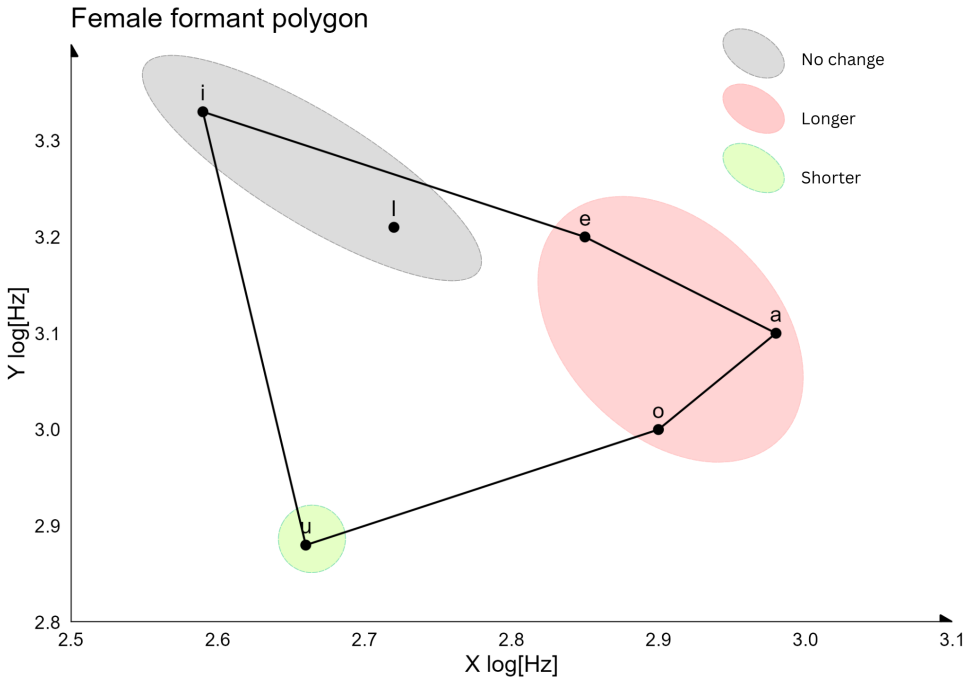

Figure 4.20. Vowel space of Polish (log-transformed F1 × F2). Coloured ellipses show duration differences between *Accent* and *Other* positions. Modified from Jassem (2003) by adding new data — colour-coded changes in the *Accent* position relative to *Other* position.

Table 4.7. Mean vowel durations (ms) and standard deviations for utterances with contrastive **focus** by position (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

Vowel	*Accent*			*Focus*			*Other*		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
/a/	22	101	17	22	147	38	110	115	82
/e/	21	91	16	21	122	27	126	57	21
/i/	20	78	24	20	87	21	60	107	54
/o/	16	93	16	20	101	23	84	82	37
/u/	19	48	19	19	100	19	38	106	26
/I/	20	91	32	20	119	22	140	114	95

The table 4.7 shows that overall, **the vowels realized in the *Focus* position are lengthened relative to those realized in the *Accent* position**. The effect is the strongest for /a/, /e/,

and /u/. For /i/ and /o/, duration differences are minor, whereas /u/ displays an atypical pattern, with very short realisations in the *Accent* position and considerably longer ones in the *Focus* position. The *Other* position is usually intermediate, though in /I/ it yields the highest and most variable values.

The boxplot in Figure 4.21 shows that vowels in *Focus* positions are mostly the longest. On average, mean duration in *Focus* position is 140% longer than in *Accent* position. Particularly strong effects are observed for /a/, which frequently exceeds 200 ms and shows multiple outliers above 400 ms. In contrast, vowels in *Accent* positions display much shorter and more stable durations, typically clustering between 70–110 ms. The realisations in the *Other* position tend to occupy an intermediate range, often with greater variability.

This outlier distribution aligns with observation from the previous section (see Section 4.1 for more details), where it was noted that a commonly observed focus marking strategy in the study involved prolonging the final vowel, which is classified as *Other* positions.

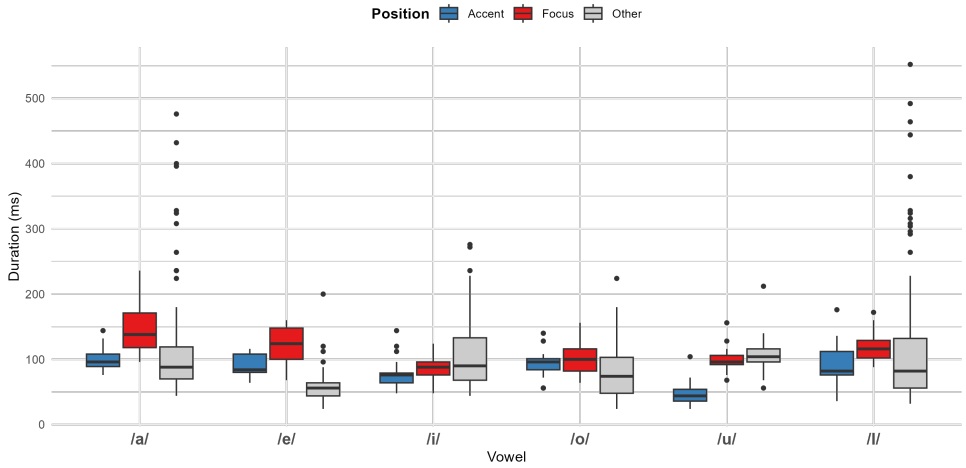

Figure 4.21. Vowel duration distributions by position in utterances with contrastive focus condition across different vowels and positions (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

By comparing boxplots of average duration across both **neutral** and contrastive **focus** conditions for all speakers, a generally consistent pattern emerges in the relationship between the duration ratios for vowels realised in the *Accent* and *Other* positions (see Figure 4.22). The blue dots represent mean value for *Accent* positions. These in utterances produced under **neutral** condition are shorter relative to *Accent* position in utterances produced under **focus** condition.

These results may suggest that, in terms of duration, the presence of focus shortens vowels in *Accent* position (with the exception of /i/) compared to their **neutral** sentence counterparts (compare Table 4.5 and Table 4.7), and, at the same time, modulates length of vowels in *Other* positions, especially last phonemic segment in the highlighted words.

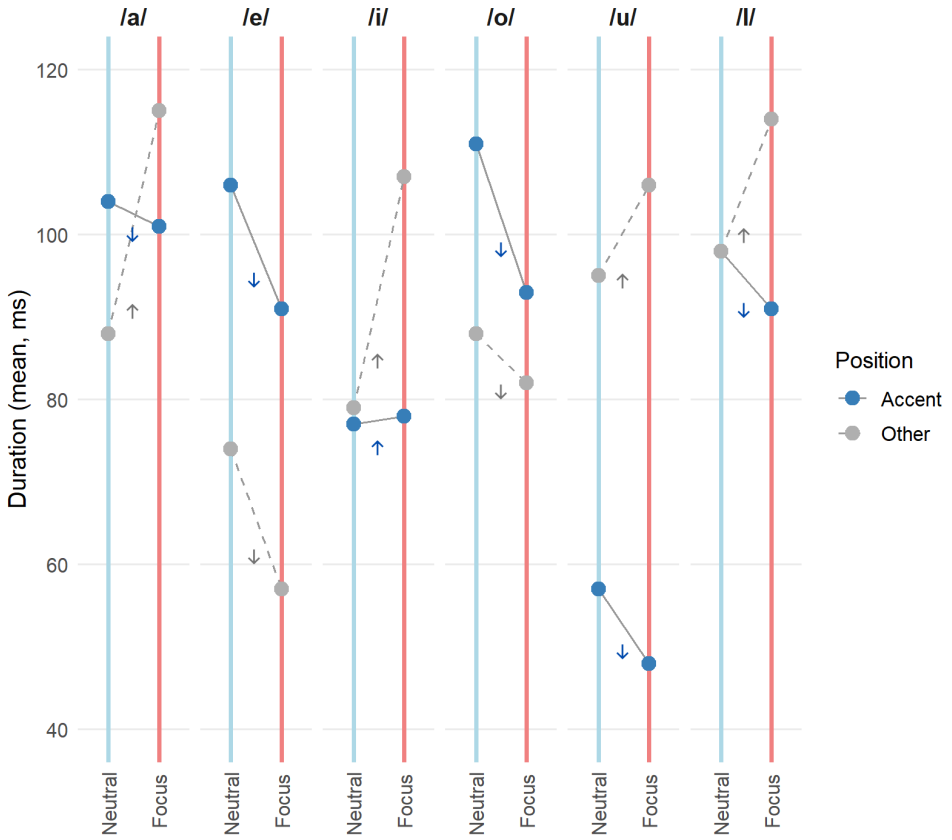

Figure 4.22. Mean vowel durations (ms) across in utterances produced under **neutral** and **focus** conditions, in non-*Focus* positions. The arrows indicate the direction of change — with the exception of the vowel /i/ — the duration of vowels in *Accent* position under **focus** condition is **shorter** than in *Accent* position under **neutral** condition. For *Other* positions there is no such pattern which might be ascribed to lengthening of the final vowel in words pronounced with contrastive **focus** as these vowels are considered *Other* position.

.

In summary, vowel duration is strongly modulated by vowel position. The **most consistent and robust lengthening is associated with *Focus* position**, whereas *Accent* position results in shorter and more uniform realisations. The effect is less substantial for high vowel /i/, while the high vowel /I/ shows limited variability, which may indicate that the effect of prominence on duration may depend on vowel quality.

While these descriptive statistics have illustrated systematic effects of prosodic prominence on articulatory parameters, both in presence of contrastive **focus** and **neutral** utterances, they do not account for potential speaker-level variability, unbalanced data distribution, or statistical significance of observed patterns. To formally test these effects while controlling for individual variation, linear mixed-effects models (LMEMs) were applied.

4.3 Prominence effects on vertical jaw displacement and lip aperture: mixed-effects analysis

4.3.1 Introduction to modelling approach

The reasons for choosing linear mixed-effects models (LMEMs) is discussed in detail in the methodological chapter (see Chapter 3, Section 3.4.1). To summarise briefly, LMEMs were selected to address the non-independence of repeated measures from individual speakers, as well as to match the complex structure of the data, which includes both fixed effects (systematic effects of experimental conditions, such as producing utterance under **neutral** or **focus** condition, that are hypothesised to systematically influence articulation) and random effects (speaker-level variability that might represent individual differences rather than experimental manipulations).

In this section jaw position always refers to the z-score normalised jaw position measured at the point of minimal velocity within each phonemic segment. The same applies for lip aperture.

The goal of the current modelling is to test whether these patterns — such as increased vertical jaw displacement or greater lip in contrastive *Focus* position — are systematic and consistent across speakers.

Each dependent variable — **vertical jaw displacement** and **lip aperture** — was analysed in a separate model. The key predictor is Vowel Position, with three levels:

- *Accent* (position of utterance-level prominence);
- *Focus* (position of contrastive focus);
- *Other* (neither accented nor focused positions) — used as a baseline.

All models include random intercepts for speakers, i.e. assessment of possible background speaker-specific differences, and the outcome measures are reported in standard deviation units (SD), as the data were normalised before the modelling (see Chapter 3, Section 3.4.1.4). The results are interpreted from both statistical and phonetic perspectives.

4.3.1.1 Computational details

All linear mixed-effects models were fitted using Restricted Maximum Likelihood (REML) estimation in R (version 4.3.2) with the lme4 package (Bates et al., 2015). Statistical significance was assessed using the lmerTest package (Kuznetsova et al., 2017), which provides p-values via Satterthwaite's method for degrees of freedom approximation.

4.3.2 Modelling vertical jaw displacement

4.3.2.1 Model specification

This model examines the influence of prominence on vertical jaw displacement. The dependent variable is the jaw displacement. The model formula can be expressed as:

```
Z_norm_at_Vmin_jaw ~ Vowel Position + (1 | Subject)
```

where $\sim$ denotes *is modeled as* meaning that vertical jaw displacement is predicted by Vowel Position (the fixed effect, independent variable) and $(1|\text{Subject})$ — the random effect (possible background speaker-specific differences).

4.3.2.1.1 Fixed effects

Table 4.8 presents the statistical results for how prosodic prominence affects jaw displacement. The table shows three key pieces of information for each condition:

- **Estimate**: the size and direction of the effect (negative values = greater jaw displacement);
- **Standard Error (SE)**: how precise the measurement is (smaller = more precise);
- **t-value and p-value**: the t-value incorporates both effect magnitude (estimate) and precision (SE), indicating the strength and reliability of each effect. The p-value determines whether the effect is statistically significant.

Intercept represents the baseline category (vowels in *Other* positions). Here the estimate is close to zero (-0.017 SD) and not statistically significant, meaning that these vowels do not show any systematic deviation in jaw displacement. The effects of prominence are interpreted relative to this neutral baseline, i.e. *Accent* vs *Other*, *Focus* vs *Other*.

Table 4.8. Fixed effects for jaw displacement model

Predictor	Estimate	SE	df	t	p-value	Sig.
Intercept (*Other*)	-0.017	0.042	5.91	-0.40	0.704	
Vowel Position: *Accent*	-0.383	0.065	1739.57	-5.92	< 0.001	***
Vowel Position: *Focus*	-1.172	0.090	1740.36	-12.99	< 0.001	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

All effects show statistical significance. The baseline condition (*Other*) shows a small negative jaw position displacement -0.017 SD (t = -0.40, p = 0.704), representing the typical jaw position for non-prominent (neither in *Accent* nor in *Focus* positions) vowels, though this baseline displacement does not differ significantly from zero.

Both prominence conditions show significant negative effects, meaning they reduce jaw displacement compared to the baseline:

- *Accent* displaces jaw position by 0.383 SD (t = -5.92, p < 0.001);
- *Focus* shows the strongest effect, displacing the jaw by 1.172 SD (t = -12.99, p < 0.001).

Both effects were highly significant (p < .001), indicating systematic articulatory adjustments linked to prosodic prominence. The jaw displacement pattern suggests graded articulatory enhancement, increasing proportionally to the degree of prominence marking: lowest in non-prominent (*Other*) positions, greater under *Accent* position, and highest in phrase prominence (*Focus* position), reflecting increasing articulatory effort with prominence level.

4.3.2.1.2 Random effects and model fit

To ensure the reliability and validity of the statistical results, several diagnostic procedures were conducted to verify that the model assumptions were met.

4.3.2.1.3 Normality of residuals

The distribution of model residuals was examined visually using histogram and Q-Q plots to assess whether the differences between observed and predicted values follow a normal distribution — a key assumption of mixed-effects models. Note that this diagnostic focuses on the normality of model errors rather than the raw data distribution, as mixed-effects models assume normally distributed residuals and random effects, not necessarily normal data.

Both diagnostic plots indicated approximately normal distribution (Figure 4.23), with minor deviations at the extreme values that are typical in large datasets. No substantial departures from normality were detected (Figure 4.24), supporting the validity of the statistical tests.

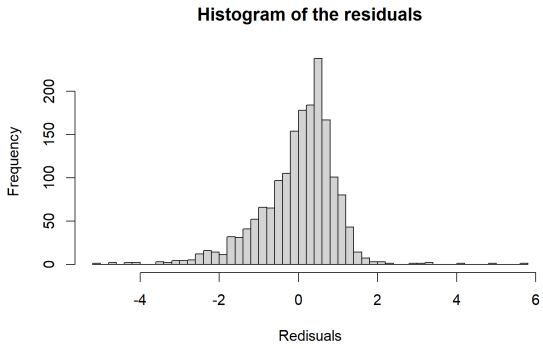

Figure 4.23. Distribution of `Z_norm_at_Vmin_jaw` residuals.

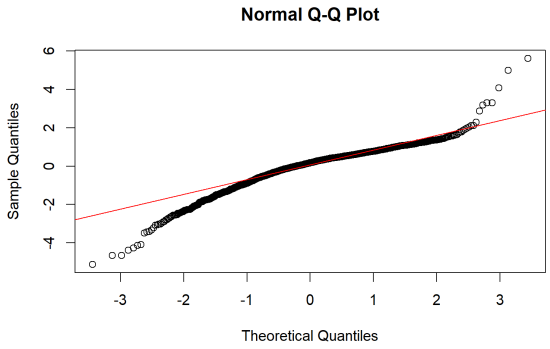

Figure 4.24. Normality of `Z_norm_at_Vmin_jaw` residuals.

4.3.2.1.4 Homoscedasticity assessment

A diagnostic plot (Figure 4.25) tested the homoscedasticity assumption by examining whether the model makes equally reliable predictions across all jaw displacement values. The residuals were approximately symmetrically distributed around zero for all prominence conditions, supporting this assumption.

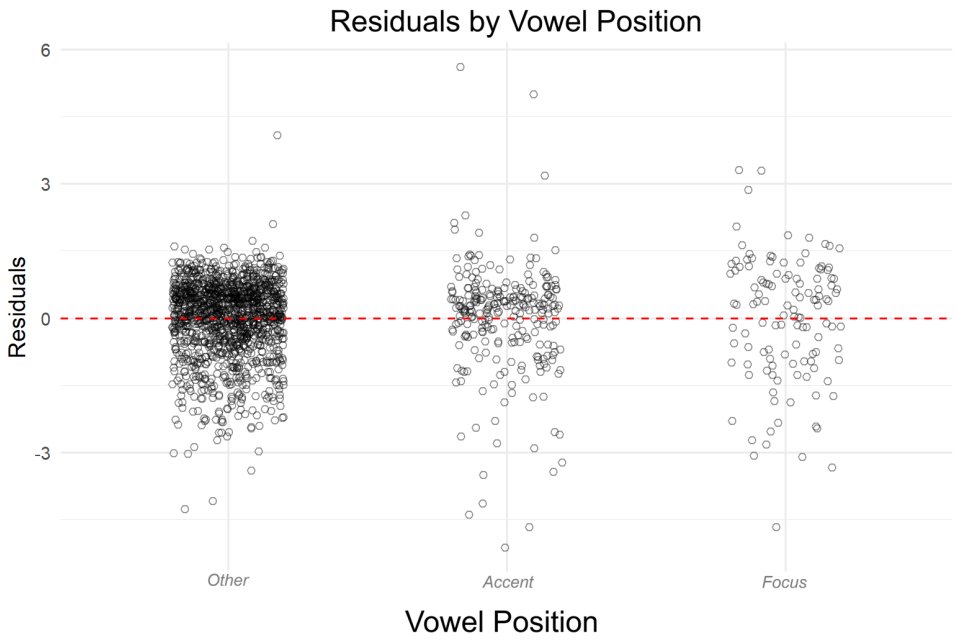

Figure 4.25. Residuals plotted against fitted values for the linear mixed-effects model predicting jaw displacement. Each vertical cluster corresponds to one level of the Vowel Position predictor (*Accent*, *Focus*, *Other*).

4.3.2.1.5 Outlier and influence analysis

Using the `influence.ME` package (Nieuwenhuis et al., 2012), influence measures including Cook's distance (Figure 4.26) was calculated to identify individual speakers whose data might disproportionately affect the model estimates. This analysis revealed that one participant (UOKV) exerted unusually strong influence on the fixed-effect estimates. To ensure that the observed effects were not driven by a single outlier, an additional model was fitted with the most influential speaker (UOKV) excluded (`model_jaw_noUOKV`). The overall pattern of results remained robust, and all prominence effects remained statistically significant, confirming that the prominence effects were not driven by a single individual's atypical patterns. The results of this model are not discussed here as they would not provide additional information; its summary is a part of *Appendix D*.

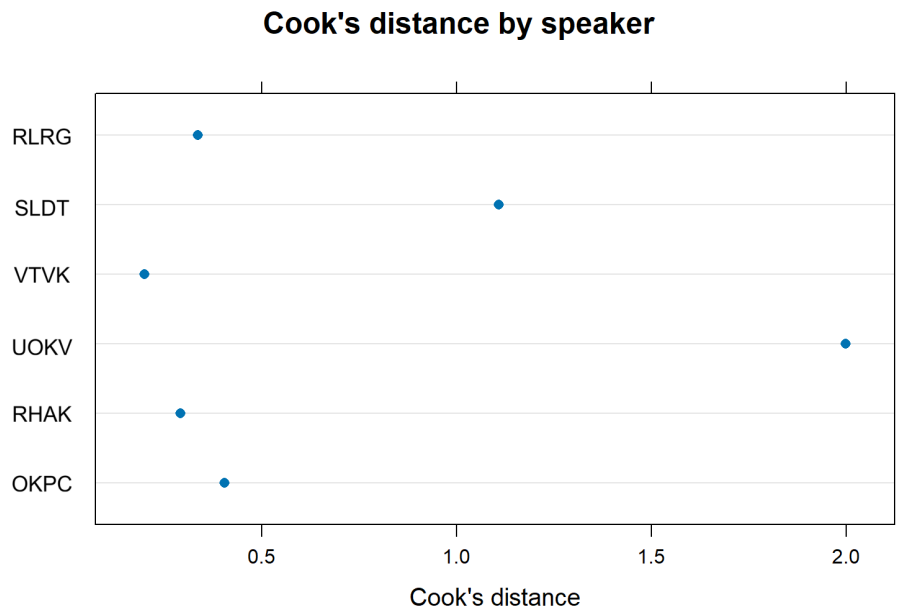

Figure 4.26. Cook's distance by speaker. Higher values indicate speakers with greater influence on the model estimates, potentially due to outlier behaviour or strong internal consistency.

4.3.2.1.6 Random effects structure validation

The model's random-effects structure was tested for singularity — a condition in which random effects account for negligible variance and should be excluded. The `isSingular()` function confirmed that speaker-specific differences were substantial enough to be kept in the model.

4.3.3 Modelling lip aperture

4.3.3.1 Model specification

The following model examines the influence of prosodic prominence on lip aperture. The dependent variable is the z-score normalised lip aperture measured at the point of minimal velocity of lower lip within each phonemic segment. The model formula can be expressed as:

```
LipAperture_max_norm ~ Vowel Position + (1 | Subject)
```

where $\sim$ denotes *is modeled as* meaning that lip aperture is predicted by Vowel Position (the fixed effect) and random effect (1|Subject)

4.3.3.1.1 Fixed effects

Table 4.9 displays the statistical findings for how lip aperture responses to different prominence positions. The results reveal three essential components for each experimental condition:

- **Estimate**: the magnitude and polarity of the observed effect (positive values = increased lip opening);
- **Standard Error (SE)**: the precision level of the measurements (lower values = higher precision);
- **t-value and p-value**: the t-value combines effect size (estimate) with measurement precision (SE), reflecting both the magnitude and reliability of each observed effect. The p-value indicates statistical significance.

Table 4.9. Fixed effects for lip aperture model

Predictor	Estimate	SE	df	t	p-value	Sig.
Intercept (*Other*)	0.011	0.042	5.50	0.27	0.799	
Vowel Position: *Accent*	0.513	0.068	1739.00	7.49	< 0.001	***
Vowel Position: *Focus*	1.119	0.095	1740.00	11.73	< 0.001	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Both levels of prominence position show significant positive effects, meaning they increase lip aperture compared to the baseline. The baseline condition (*Other*) shows a small positive lip aperture displacement 0.011 SD (t = 0.27, p = 0.799), representing the typical lip aperture for *Other* position (neither in *Accent* nor in *Focus* positions) vowels, though this baseline displacement does not differ significantly from zero.

Both levels of prominence position conditions show significant positive effects:

- *Accent* increases lip aperture by 0.513 SD (t = 7.49, p < 0.001);
- *Focus* shows the strongest effect, increasing lip aperture by 1.119 SD (t = 11.73, p < 0.001).

Both effects were highly significant (p < .001), indicating systematic articulatory adjustments linked to the level of prosodic position. The lip aperture pattern reveals a hierarchical articulatory response that scales with prominence level: minimal opening in non-prominent (*Other*) syllables, moderate expansion for utterance-level *Accent*, and maximum aperture in *Focus* position, indicating **progressively greater articulatory investment corresponding to the level of prominence strength**.

4.3.3.1.2 Random effects and model fit

The lip aperture model has undergone the same diagnostic procedures as the jaw displacement model. The results are presented below.

4.3.3.1.3 Normality of residuals

The distribution of residuals was inspected visually using a histogram and a normal Q–Q plot to assess model fit. The histogram (Figure 4.27) revealed a right-skewed distribution, indicating

that most model predictions were fairly accurate (concentration of small errors), but a subset of cases showed substantially larger prediction errors (longer tail towards higher values). The Q-Q plot (Figure 4.28) revealed marked deviations from normality, especially in the upper quantiles where points curve upward, confirming the right-skewed pattern observed in the histogram. These large residuals likely reflect the vowel-specific articulatory patterns described in previous sections (see Section 4.2.1.2), where certain vowels respond more strongly to prominence effects than the model predicted. For instance, low vowels like /a/ show greater lip opening under prominence, and typically constrained high vowels such as /i/ undergo unexpected aperture expansion in *Focus* positions (+1.1 SD).

These patterns suggest that prominence type alone cannot fully predict lip aperture behaviour — the interaction between vowel category and prominence creates more complex articulatory responses than a simple additive model can capture. However, the observed deviations from normality were not deemed severe enough to compromise the overall model validity.

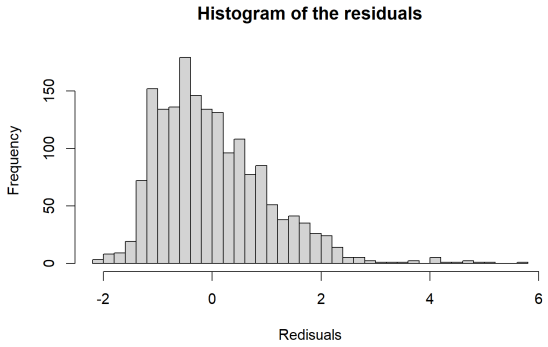

Figure 4.27. Distribution of `LipAperture_max_norm` residuals.

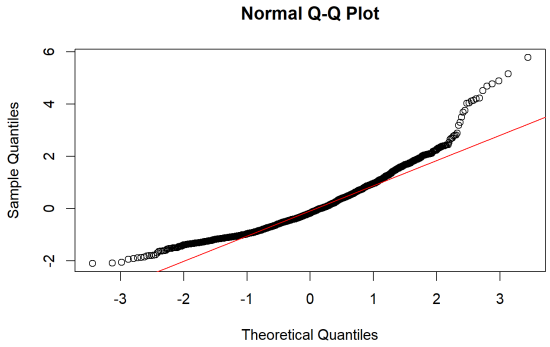

Figure 4.28. Normality of `LipAperture_max_norm` residuals.

4.3.3.1.4 Homoscedasticity assessment

Homoscedasticity was assessed by plotting model residuals against prominence type (Figure 4.29). The residual distribution revealed no systematic patterns related to the predictors, though variance differed across conditions. The *Other* position category displayed residuals densely clustered around zero with an extended positive tail, indicating cases with larger prediction errors. Conversely, the accent and *Focus* positions contained fewer observations and exhibited somewhat reduced variance.

The plot suggests mild heteroskedasticity, primarily attributed to higher residual dispersion in the *Other* position. Given the substantial sample size and the robustness of linear mixed-effects models to moderate variance heterogeneity, these deviations were not deemed problematic for model validity.

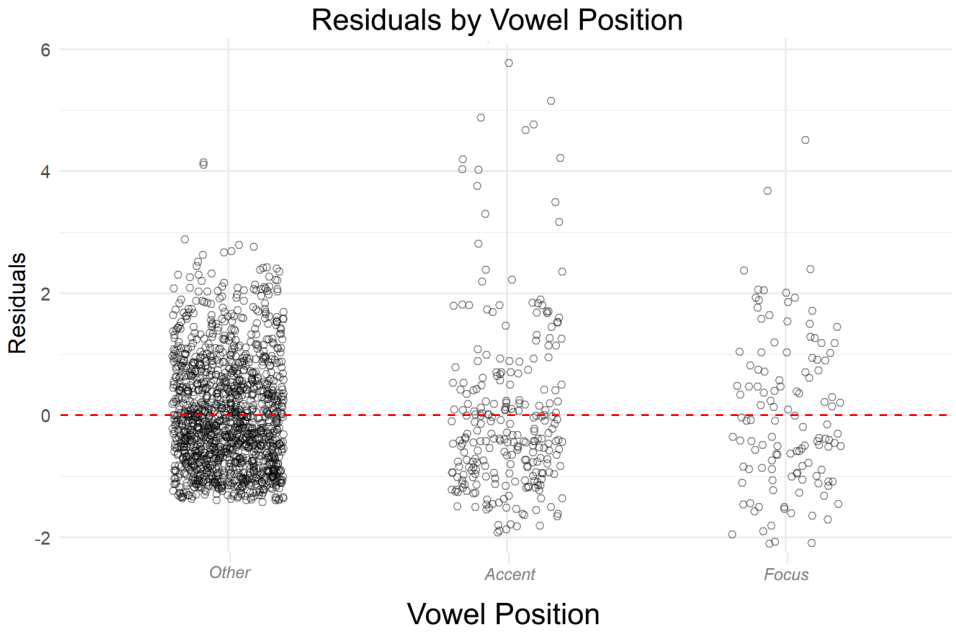

Figure 4.29. Residuals plotted against fitted values for the linear mixed-effects model predicting lip aperture. Each vertical cluster corresponds to one level of the Vowel Position predictor (*Accent*, *Focus*, *Other*).

4.3.3.1.5 Outlier and influence analysis

To assess whether the model estimates were disproportionately influenced by individual speakers, influence diagnostics were performed using the `influence.ME` package (Nieuwenhuis et al., 2012). Cook's distance was calculated for each speaker (Figure 4.30). One speaker (SLDT) exhibited notably high Cook's distance, suggesting that their data may have disproportionately influenced the model's fixed-effect estimates.

To test the robustness of the prominence level effects, the model was re-run excluding speaker SLDT (`model_lip_noSLDT`). The revised model (n = 1434) yielded the same pattern of results: both accent (t = 8.20, p < .001) and *Focus* positions (t = 8.28, p < .001) significantly increased lip aperture compared to the *Other* positions. This confirms that the observed effects were not driven by a single influential speaker. More details on speakers are in section *Speaker-specific variation*.

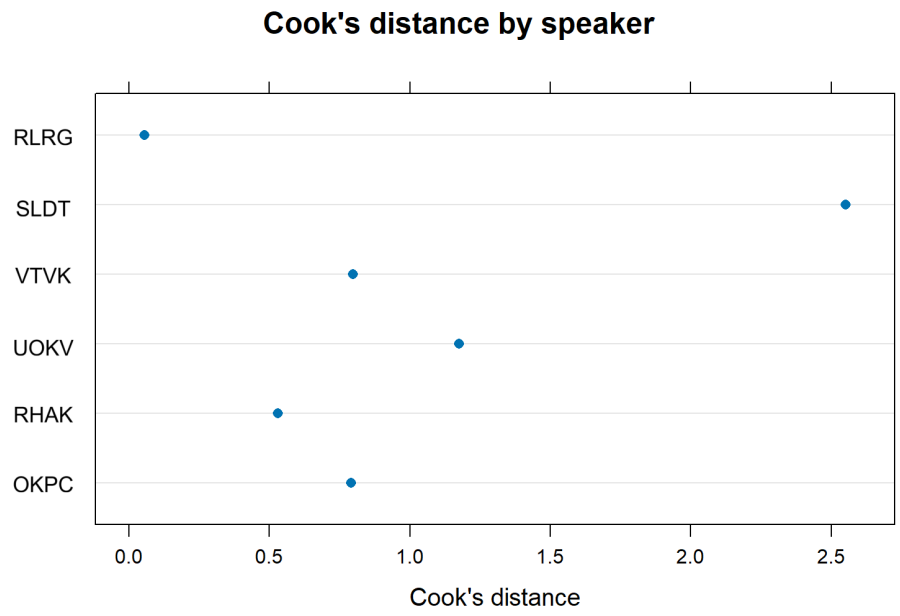

Figure 4.30. Cook's distance by speaker for the lip aperture model. A higher Cook's distance indicates greater influence on the model's fixed-effect estimates

4.3.3.1.6 Random effects structure validation

Singularity testing confirmed that individual speaker variations were meaningful enough to justify keeping them in the model.

4.4 Supplementary analyses

As mentioned earlier in the *Descriptive statistics*, the presence of contrastive focus appeared to affect articulatory displacement not only in the *Focus* position, but across the utterance, exerting a global prosodic effect. Specifically, utterances produced with contrastive **focus** displayed reduced vertical jaw displacement for in non-*Focus* positions, compared to the **neutral** counterparts. This pattern — observed consistently across most vowels with the exception of /o/ and /i/ — mirrored the distributional effects reported also for lip aperture (see Section 4.2.1.2) and vowel duration (see Section 4.2.1.3). Based on these insights, supplementary hypotheses were formulated:

H$_8$ The presence of contrastive focus in an utterance has no effect on jaw displacement in non-*Focus* position vowels.

H$_9$ The presence of contrastive focus in an utterance reduces jaw displacement in non-*Focus* position vowels.

H$_{10}$ The presence of contrastive focus in an utterance has no effect on lip aperture in non-*Focus* position vowels.

H$_{11}$ The presence of contrastive focus in an utterance reduces lip aperture in non-*Focus* position vowels.

To test this observation statistically, a series of supporting linear mixed-effects models was constructed. The analyses targeted phonemic segments in both *Accent* and *Other* position excluding all positions bearing phrasal stress and their counterparts in **neutral** utterances (Figure 4.31).

Condition	Stimuli	1	2	3	4	5	6	7	...	25
neutral	Ala dała Arkowi zapas buraków	a	l	a	d	a	w	a	...	f
focus	Ala DAŁA Arkowi zapas buraków	a	l	a	d	a	w	a	...	f

Figure 4.31. Phonemic segment with index 5 is excluded from the analyses to compare articulatory behaviour across matching segment positions in both **neutral** and **focus** utterances, for non-*Focus* positions only.

The aim of this supplementary analysis was to determine if **presence of contrastive focus**, aside from its immediate effects, causes gradual changes in articulatory patterns across the entire utterance, following an earlier noted tendency.

Another aspect taken into consideration and not included in the hypotheses stated in Section *Main research hypotheses* in Chapter *3*, is vowel duration. It was examined descriptively as a possible focus marking strategy. **However, there might also be a chance that increased articulatory jaw displacement and lip aperture in *Focus* position may partially reflect lengthening, rather than prominence per se**. Similarly, greater lowering in *Accent* position might be secondary to increased vowel duration; as in a longer segment the articulators have more time allotted to lower/lift themselves. Given these concerns, an additional set of models was constructed to assess whether articulatory displacement remains significant when controlling for segmental duration. Additional hypotheses are:

H$_{11}$ When vowel duration is controlled for, contrastive focus has no significant effect on jaw displacement.

H$_{12}$ Contrastive focus affects jaw displacement independently of vowel duration.

H$_{13}$ When vowel duration is controlled for, utterance-level accent has no significant effect on jaw displacement.

H$_{14}$ Utterance-level accent significantly affects jaw displacement, beyond what is explained by vowel duration.

Detailed analyses are in the following sections.

4.4.1 Focus influence on global prosodic pattern

As the structure and interpretation of linear mixed-effects models have already been discussed in detail for the main scope of the dissertation, the following sections report additional models in a more compact format.

4.4.1.1 Jaw displacement in neutral vs focus condition — model specification

This model examines the effect of contrastive **focus** presence on jaw positioning. The dependent variable is the z-score normalised jaw position measured at the point of minimal velocity during vowel production. The key predictor is Focus (presence of contrastive **focus** in the utterance, here marked with uppercase for more clarity), with two levels:

- **neutral** utterances produced under **neutral** condition;
- **focus** utterances produced under contrastive **focus** condition.

The model formula can be expressed as:

```
Z_norm_at_Vmin_jaw ~ Focus + Vowel + (1 | Subject)
```

where $\sim$ denotes *is modeled as*, meaning that jaw position is predicted by Focus condition and Vowel category (the fixed effects) and random intercepts for Subject.

Vowel category is included as a control variable, given that jaw displacement is known to vary systematically due to vowel height: high vowels (e.g., /i/, /u/) require less jaw lowering than low vowels (e.g., /a/) due to their articulatory nature. Including vowel category in the model helps separate the effects caused by the sounds themselves from those caused by the presence of Focus.

4.4.1.1.1 Fixed effects

All vowels show significant positive effects relative to the reference vowel /a/, indicating higher jaw positions with the hierarchy ranking from most jaw displacement to least: /a/ $<$ /o/ $<$ /I/ $<$ /e/ $<$ /i/ $<$ /u/ (Table 4.10).

Crucially, the utterance under **neutral** condition displayed a significant (at $p < 0.05$) negative effect (-0.082 SD), indicating greater jaw lowering compared to the Focus factor (presence of contrastive **focus** in an utterance). This supports the hypothesis that contrastive **focus** suppresses articulatory expansion outside the focal segment.

Table 4.10. Fixed effects for jaw displacement neutral vs focus condition model

Predictor	Estimate	SE	df	t	p-value	Sig.
Intercept (**focus**, vowel /a/)	-0.967	0.069	17.56	-14.07	< 0.001	***
Focus: **neutral**	-0.082	0.041	1473.51	-2.02	0.044	*
Vowel: /e/	1.184	0.063	1469.15	18.73	< 0.001	***
Vowel: /i/	1.453	0.075	1469.34	19.47	< 0.001	***
Vowel: /o/	0.733	0.070	1469.24	10.42	< 0.001	***
Vowel: /u/	1.702	0.084	1469.37	20.32	< 0.001	***
Vowel: /I/	1.132	0.062	1469.40	18.33	< 0.001	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.4.1.1.2 Random effects and model fit diagnostics

Model fit diagnostics were performed to assess residual distribution, variance homogeneity, presence of influential data points, and adequacy of the random effects structure. The summary of these checks is presented in Table 4.11. A few extreme values were observed as outliers, but there was no evidence of systematic influence.

Table 4.11. Summary of diagnostics for jaw displacement in neutral vs focus condition model

Diagnostic for	Assessment
Normality of residuals	Slight right skew in the upper quantiles
Homoscedasticity	No substantial heteroscedasticity detected
Outlier analysis	Scaled residuals range from –5.73 to 5.81
Random effects	No singularity detected

4.4.1.2 Lip aperture in neutral vs focus condition — model specification

The model examines the effect of contrastive focus presence on lip aperture. The key predictor is Focus, with two levels:

- **neutral** utterances produced under **neutral** condition;
- **focus** utterances produced under contrastive **focus** condition.

```
LipAperture_max_norm ~ Focus + Vowel + (1 | Subject)
```

where $\sim$ denotes *is modeled as*, meaning that lip aperture is predicted by Focus condition (i.e. presence of contrastive **focus** in an utterance) and Vowel category (the fixed effects) and random intercepts for Subject.

4.4.1.2.1 Fixed effects

All vowels show significant negative effects compared to vowel /a/, indicating smaller lip aperture with the hierarchy ranking from most lip opening to least: /a/ > /e/ > /I/ > /i/ > /o/ > /u/. The Focus effect is non-significant (p = 0.656), indicating that presence of contrastive **focus** does not systematically alter lip aperture during vowel production. This might indicate that changes in lip opening are mainly influenced by the specific needs of each vowel rather than by prosodic emphasis.

Table 4.12. Fixed effects for lip aperture neutral vs focus condition model

Predictor	Estimate	SE	df	t	p-value	Sig.
Intercept (**focus**, vowel /a/)	1.213	0.066	36.65	18.37	< 0.001	***
Focus: **neutral**	0.020	0.045	1472.90	0.45	0.656	
Vowel: /e/	-1.125	0.071	1469.26	-15.91	< 0.001	***
Vowel: /i/	-1.427	0.083	1469.63	-17.10	< 0.001	***
Vowel: /o/	-1.478	0.079	1469.43	-18.78	< 0.001	***
Vowel: /u/	-1.879	0.094	1469.68	-20.06	< 0.001	***
Vowel: /I/	-1.368	0.069	1469.73	-19.81	< 0.001	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.4.1.2.2 Random effects and model fit diagnostics

The summary of model fit diagnostics with interpretation is presented in Table 4.13. A few high positive values observed in outlier assessment, but no indication of systematic influence.

Table 4.13. Summary of diagnostics for lip aperture in neutral vs focus condition model

Diagnostic for	Assessment
Normality of residuals	No substantial deviations observed
Homoscedasticity assessment	No clear signs of heteroscedasticity
Outlier and influence analysis	Scaled residuals range from –2.96 to 5.87
Random effects structure validation	No singularity detected

4.4.2 Duration as a potential confounding factor

4.4.2.1 Jaw displacement in duration-controlled model

This model examines the effect of prosodic prominence on vertical jaw displacement while controlling for segmental duration.

```
Z_norm_at_Vmin_jaw ~ Vowel Position + Vowel
        + Duration + (1 | Subject)
```

where ∼ denotes *is modeled as*, meaning that jaw position is predicted by Vowel Position, Vowel category, and vowel Duration (the fixed effects) and random intercepts for Subject.

Duration is added to control for the possibility that prominence-related jaw displacement effects might be secondary to vowel lengthening rather than reflecting prosodic prominence per se. Including duration in the model allows to determine whether prominence affects jaw displacement directly or whether the effects are mediated by vowel lengthening.

4.4.2.1.1 Fixed effects

Both *Focus* and *Accent* position show significant negative effects (Table 4.14), indicating greater jaw lowering compared to the baseline (*Other*). The *Accent* position is linked to moderate jaw lowering, while *Focus* position produces the most pronounced jaw displacement. All vowels show significant positive effects with the same hierarchy as previous models (/a/ < /o/ < /e/ < /I/ < /i/ < /u/); in other words, the effect is present for all vowel categories, with vowel /a/ requiring the greatest jaw lowering and vowel /u/ the least.

Most importantly, duration shows a significant negative effect, indicating that longer vowels are associated with greater jaw lowering. However, both prominence effects remain highly significant even when controlling for duration, which might imply that prominence affects jaw positioning independently of vowel lengthening effects.

Table 4.14. Fixed effects for jaw displacement model (duration-controlled)

Predictor	Estimate	SE	df	t	p-value	Sig.
Intercept (*Other*, vowel /a/)	-0.542	0.068	49.84	-7.94	< 0.001	***
Vowel Position: *Accent*	-0.480	0.053	1734.00	-9.04	< 0.001	***
Vowel Position: *Focus*	-1.152	0.074	1735.00	-15.48	< 0.001	***
Vowel: /e/	0.861	0.060	1735.00	14.24	< 0.001	***
Vowel: /i/	1.383	0.069	1734.00	20.13	< 0.001	***
Vowel: /o/	0.783	0.066	1734.00	11.95	< 0.001	***
Vowel: /u/	1.747	0.075	1734.00	23.15	< 0.001	***
Vowel: /I/	1.110	0.058	1734.00	19.08	< 0.001	***
Duration	-0.004	0.0004	1702.00	-10.33	< 0.001	***

Note: *** p < 0.001, ** p < 0.01, * p < 0.05

4.4.2.1.2 Random effects and model fit diagnostics

The summary of model fit diagnostics with interpretation for duration-controlled model for jaw displacement is presented in Table 4.15. Some moderate skew (range: –5.70 to 6.08 SD) was noted for normality of residuals, however, not substantially deviant. Ouliers' impact appears

minimal.

Table 4.15. Summary of diagnostics for jaw displacement in duration-controlled model

Diagnostic for	Assessment
Normality of residuals	Residuals approximately normally distributed
Homoscedasticity	Residual spread consistent across predicted values
Outliers and influential points	Few extreme residuals (outside ± 5 SD)
Random effects structure	Random intercepts for Subject; no singular fit.

4.4.2.2 Lip aperture in duration-controlled model

The model inspects the effect of prosodic prominence on lip aperture while controlling for segmental duration. The dependent variable is normalised lip aperture.

```
LipAperture_max_norm ~ Vowel Position + Vowel
          + Duration + (1 | Subject)
```

where $\sim$ denotes *is modeled as*, meaning that lip aperture is predicted by Vowel Position, Vowel category, and vowel Duration (the fixed effects) and random intercepts for Subject.

4.4.2.2.1 Fixed effects

Both *Accent* and *Focus* position significantly affect lip aperture, with the latter one yielding the strongest increase relative to the baseline (Table 4.16). All vowel categories show significant negative estimates, following the gradient /a/ > /e/ > /I/ > /i/ > /o/ > /u/, consistent with their articulatory constraints. Duration emerges as a significant positive predictor, indicating that longer vowels tend to co-occur with increased lip aperture. Importantly, the prominence-related effects remain robust when controlling for vowel duration, suggesting that prosodic prominence contributes to lip aperture independently, i.e. beyond lengthening alone.

Table 4.16. Fixed effects for lip aperture model (duration-controlled)

Predictor	Estimate	SE	df	t	p-value	Sig.
Intercept (*Other*, vowel /a/)	0.596	0.068	69.19	8.75	< 0.001	***
Vowel Position: *Accent*	0.590	0.055	1733.00	10.68	< 0.001	***
Vowel Position: *Focus*	1.073	0.077	1735.00	13.87	< 0.001	***
Vowel: /e/	-0.860	0.063	1735.00	-13.68	< 0.001	***
Vowel: /i/	-1.237	0.071	1734.00	-17.32	< 0.001	***
Vowel: /o/	-1.375	0.068	1734.00	-20.19	< 0.001	***
Vowel: /u/	-1.837	0.078	1734.00	-23.41	< 0.001	***
Vowel: /I/	-1.308	0.061	1734.00	-21.62	< 0.001	***
Duration	0.005	0.0004	1661.00	11.85	< 0.001	***

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

4.4.2.2.2 Random effects and model fit diagnostics

The summary of model fit diagnostics with interpretation for duration-controlled model for lip aperture is presented in Table 4.17. Normality of residuals was assessed visually, indicating approximate normality with slight right skew; central values aligned well in Q–Q plot. There is also no evidence of influential outliers and random intercept for Subject had SD = 0.069, confirming non-zero inter-speaker variability.

Table 4.17. Summary of diagnostics for lip aperture in duration-controlled model

Diagnostic for	Assessment
Normality of residuals	Residuals ranged from –3.26 to +5.80 (z-scores)
Homoscedasticity	Residuals were symmetrically distributed.
Outliers and influence	Only a few values exceeded ± 3 SD
Random effects structure	Model was not singular

While the overall effects were consistent across the group, individual speakers exhibited variability in the magnitude of displacement, particularly in the jaw data. The minimum, maximum, and mean displacement (in mm) for each speaker are presented (Table 4.18). Since the variability amongst them was substantial, normalisation was applied to reduce inter-individual differences (see Chapter 3 Section 3.4.1.4 for more details).

Table 4.18. The range of vertical jaw displacement for each speaker, measured at the time of minimum velocity (not limited to target vowels)

Subject	Minimum	Maximum	Mean
OKPC	-11.32	-28.1	-15.96
RHAK	4.71	-22.93	-7.69
UOKV	-9.43	-29.05	-14.24
VTVK	-8.97	-15.96	-11.43
SLDT	-7.93	-23.28	-10.79
RLRG	-11.1	-26.5	-15.34
Grand Total	**4.71**	**-29.05**	**-11.95**

To assess the robustness of the main prominence effects, control analyses were conducted excluding the most influential speakers (UOKV for jaw displacement, SLDT for lip aperture). UOKV displayed greater than average jaw displacement, whereas visual inspection of the lip trajectories revealed that SLDT displayed a fairly narrow lip aperture relative to other speakers. In both cases, all prominence effects remained highly significant ($p < 0.001$) with only minor changes in parameter estimates, confirming that the results were not driven by individual outlying speakers.

4.4.3 Jaw protrusion (horizontal displacement)

In the study by Erickson et al. (2017), it was observed that while producing target sound /aI/ under contrastive emphasis the jaw was not only lowered but also protruded. This displacement was interpreted as a compensatory strategy allowing wide jaw opening to co-occur with a high F0 peak. Authors noted that 4 out of 6 speakers demonstrated statistically significant protrusion.

Given that jaw protrusion was identified there as a novel finding, it was considered relevant to examine whether such a protrusion would be observable in Polish, a language with different phonological structure. This was done for the target vowels in all positions (*Accent*, *Focus*, *Other*) and in an exploratory way, without formal statistical testing, in order to gain a preliminary insight into potential tendencies. The results are visible in the plot in Figure 4.32. Plots are shown without outliers employing IQR criterion[2].

[2]Outliers were removed using a standard interquartile range (IQR) criterion (values outside $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$)

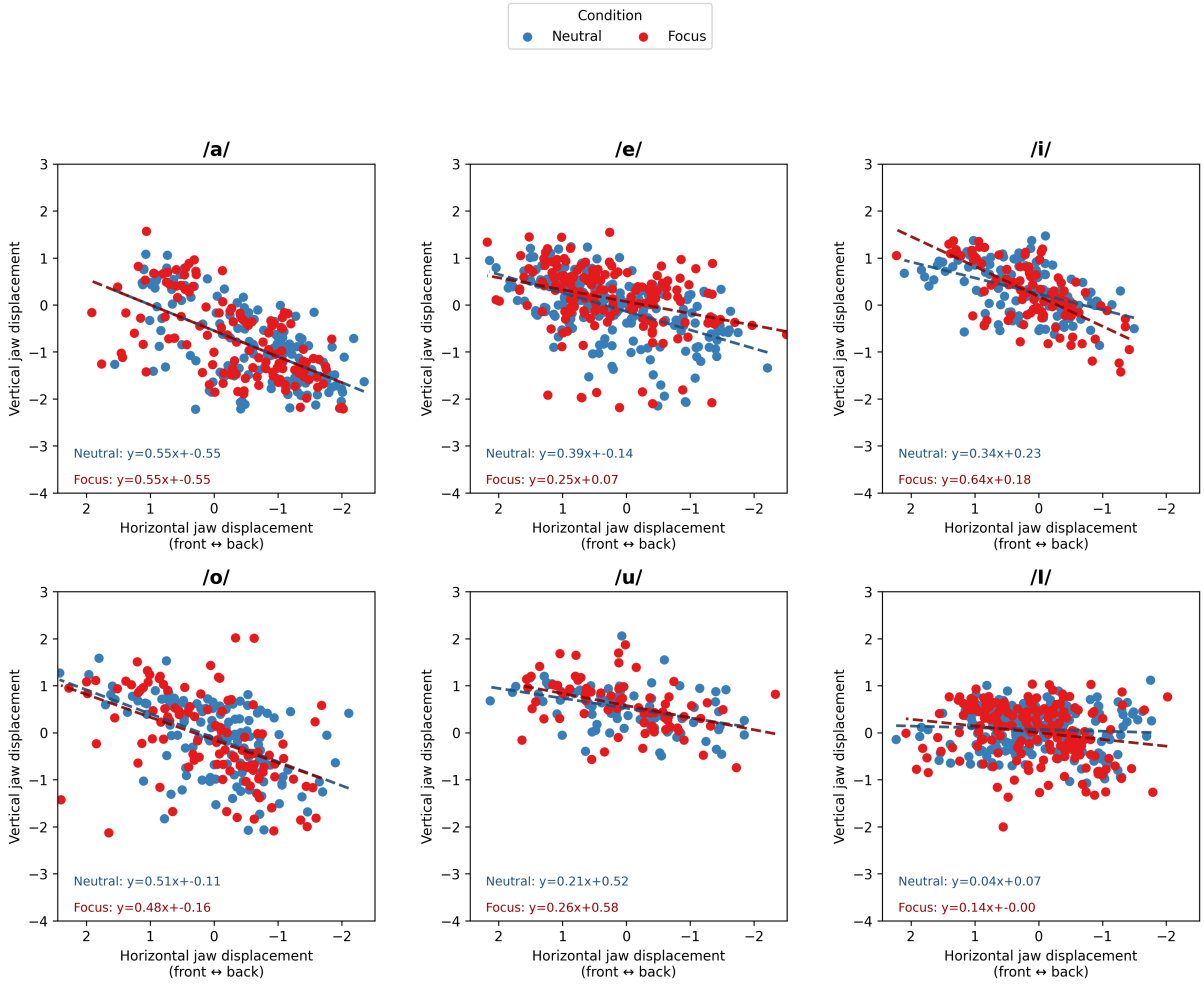

Figure 4.32. Scatterplots of horizontal and vertical jaw displacement for six Polish oral vowels. Blue points indicate **neutral** condition and red points indicate **focus** condition. Linear regression lines are shown with fitted equations.

The scatterplots indicate that vertical lowering of the jaw is generally accompanied by forward displacement, as reflected in the overall negative slopes. The effect of focus was not uniform across vowels; a clearer difference between conditions was observed for /i/, where the slope was steeper under **focus**, which might implicate stronger coupling of vertical and horizontal displacement. In contrast, no such effect was found for /a/, which exhibited equally steep slopes in both conditions, while /I/, /u/, /e/, and /o/ showed only minor or inconsistent differences. Moreover, the examination of individual speakers (see *Appendix E*) shows that while some speakers displayed markedly steeper slopes in **focus**, for others the difference was negligible or even reversed.

This asymmetry across vowel height may reflect biomechanical constraints: in high vowel /i/ where vertical jaw opening is limited, forward displacement might be an additional resource to achieve prominence-related adjustments. In English /aI/, such adjustment may be driven by the need for a rapid transition between the low vowel /a/ and the high vowel /i/, where jaw protrusion facilitates simultaneous wide opening and high F0 production. In Polish, however, the /i/ in focus was produced as a steady monophthong, without the need for fast articulatory transitions, and yet protrusion was also observed.

This insight therefore points to a possible role of jaw protrusion in prominence marking. However, since the results are inconsistent, further research might be applied and for now — albeit tentatively — it might be concluded that such protrusion patterns are displayed in Polish for /i/.

4.5 Phonetic-acoustic statistics

To further complement the articulatory analyses, selected phonetic–acoustic parameters were also analysed, including intensity, fundamental frequency (F0), the first two frequency formants (F1, F2), and the difference between F2 and F1. Specifics of parameters extraction are detailed in Section 3.3.8.4.6.

This F2-F1 distance captures the compactness/diffuseness of formant structure, with lower F2-F1 values characteristic for low vowels, in Polish represented by /a/. Production of low vowels involves greater mouth opening and, hence, more jaw displacement; this is reflected in bringing F1 and F2 closer together. For high vowels, represented in Polish by /i/, /I/, and /u/, where a more closed jaw and raised tongue dorsum increase the separation of the two frequency formants, higher values are observed. Therefore, the F2–F1 difference can be taken as a simplified acoustic proxy of vertical jaw movements relative to tongue dorsum gestures. Since the latter are not within the scope of the present research, F2-F1 allows into these gesture patterns without directly analysing tongue dorsum movement. In particular, prominence has been associated with greater jaw opening and more extreme tongue gestures, which can enhance these spectral differences by making low vowels even more compact and high vowels even more diffuse. Analyses of American English have indeed shown that F2–F1 distance is sensitive to prominence, serving as an acoustic correlate of jaw and tongue displacement (cf. Erickson, 2004, 2006). Such observation motivates investigating whether a comparable pattern can be observed in Polish, given the language's different prosodic organisation, especially a fairly standardised fixed penultimate stress and utterance-level accent as opposed to the variable stress placement that characterises English. These raise the question of whether in Polish such spectral adjustments emerge only under contrastive focus or are also present in utterance-level accent.

These phonetic–acoustic measures, however, are considered supplementary rather than central, aiming to provide additional validation for the prominence effects identified in jaw displacement and lip aperture. For the descriptive statistics, a set of 26 recordings was excluded, since they had been amplified in Audacity (Audacity Team, 2023) to correct low intensity due to microphone missetting. Such processing could bias estimates of phonetic–acoustic parameters.

The distributions and descriptive summaries by vowel and prominence condition are presented in the following section.

4.5.1 Intensity

The results (see Table 4.19 and Table 4.20) show that vowels in *Accent* and *Other* positions display similar values, with differences rarely exceeding 1–2 dB. In contrast, *Focus* position

has consistently raised intensity, with mean values about 7–12 dB higher than in non-*Focus* position. This might imply that intensity functions as a correlate of contrastive focus rather than utterance-level accent.

Table 4.19. Descriptive statistics (N, Mean, SD) of vowel intensity [dB] across vowels in neutral utterances, by position (*Accent* = utterance-level accent; *Other* = not accented)

Vowel	*Accent*			*Other*		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
/a/	22	68	6	242	69	7
/e/	24	68	6	288	69	7
/i/	23	65	5	276	67	5
/o/	22	69	6	264	67	6
/u/	21	68	5	210	68	7
/I/	25	68	5	350	67	6

Table 4.20. Descriptive statistics (N, Mean, SD) of vowel intensity [dB] across vowels in utterances with contrastive **focus**, by position (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

Vowel	*Accent*			*Focus*			*Other*		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
/a/	18	67	6	18	79	4	180	69	6
/e/	18	67	6	18	77	4	198	68	7
/i/	17	61	6	17	73	4	187	66	6
/o/	16	68	5	16	74	4	176	67	6
/u/	17	66	5	17	75	3	153	68	7
/I/	17	65	4	17	76	4	221	67	7

Although the present analysis concentrated on articulatory measures, the accompanying acoustic data suggest that intensity remains relatively stable across *Accent* and *Other* positions, but increases systematically in focus positions (by about 7–12 dB). This pattern, consistent with previous findings for Polish and Czech (Hamlaoui et al., 2019; Malisz & Żygis, 2018), indicates that intensity—together with F0—functions primarily as a correlate of focus rather than of utterance-level accent.

4.5.2 Fundamental frequency F0

The results (see Table 4.21 and Table 4.22) reveal a clear difference in F0. Vowels in *Focus* positions display markedly higher mean F0 values (e.g., /i/ rises from 194 Hz in *Other* positions to 281 Hz in *Focus*). On the other hand, vowels in *Accent* position are consistently lower than in *Other* positions; instead, F0 is strongly enhanced under contrastive **focus**, where the increase often exceeds 70–90 Hz relative to *Accent* positions (see Figure 4.33). Such a pattern implies pitch might be amongst primary correlates of contractive **focus**. This findings are in line with observations by Dogil, 1999.

Table 4.21. Descriptive statistics (N, Mean, SD) of fundamental frequency (F0) [Hz] across vowels in neutral utterances, by position (*Accent* = utterance-level accent; *Other* = not accented)

Vowel	*Accent*			*Other*		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
/a/	22	167	25	242	200	35
/e/	24	165	29	288	202	40
/i/	23	172	28	276	212	37
/o/	22	152	36	264	207	36
/u/	21	183	28	210	205	41
/I/	25	166	29	350	212	36

Table 4.22. Descriptive statistics (N, Mean, SD) of fundamental frequency (F0) [Hz] across vowels in utterances with contrastive focus, by position (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

Vowel	*Accent*			*Focus*			*Other*		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
/a/	18	147	36	18	238	34	180	188	38
/e/	18	181	114	18	220	36	198	197	45
/i/	17	189	50	17	281	46	187	194	42
/o/	16	161	32	16	265	51	176	196	40
/u/	17	179	41	17	244	48	153	195	43
/I/	17	162	50	17	238	38	221	202	35

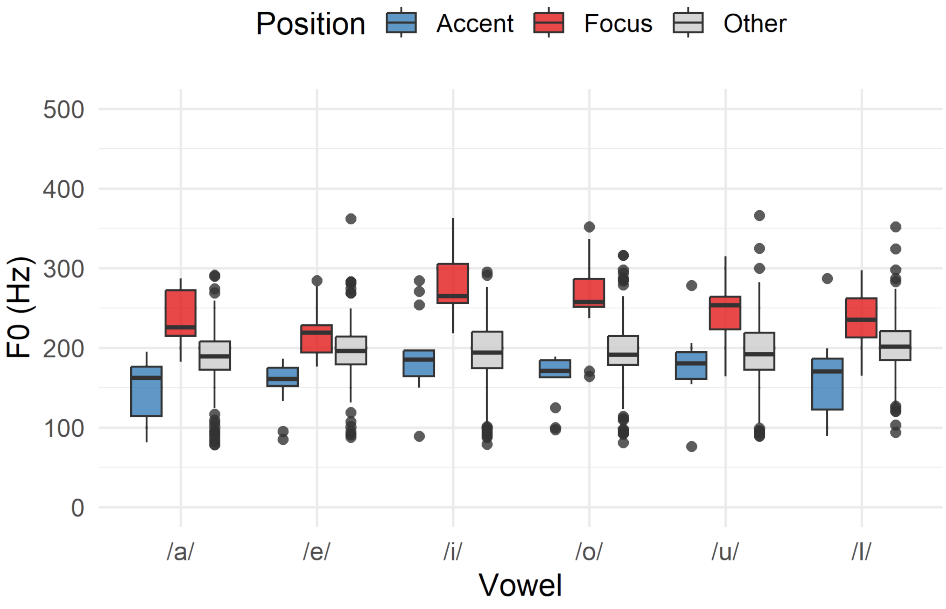

Figure 4.33. Fundamental frequency (F0) [Hz] across vowels in utterances with contrastive focus, by position (*Accent* = utterance-level accent; *Focus* = contrastive focus; *Other* = neither accented nor focused)

The next section addresses spectral parameters (F1 and F2), providing additional evidence on

the relation between vowel quality and prominence marking.

4.5.3 F1 and F2 frequency formants and their relation

The plot in Figure 4.34 shows mean F1 and F2 frequencies each vowels under two experimental conditions (**neutral** vs **focus**), averaged across all speakers, and illustrates that formant structure varies systematically with prominence. In utterances produced under **neutral** condition, vowels in both *Accent* and *Other* positions tend to be of similar levels, which may indicate that utterance-level *Accent* has little to no effect on formant dispersion. The case of /i/ stands out, showing an expanded F2–F1 difference in *Accent* position, a possible result of prominence-induced hyperarticulation.

In the case of utterance with presence of contrastive **focus** /e/ become more compact in *Focus* position and /i/ is clearly more diffuse. Other vowels, /a/, /o/, /u/, and /I/ do not display systemic changes. Based on these results, it might be tentatively concluded that in Polish neither utterance-level *Accent* nor contrastive *Focus* significantly affect formant structure. Given the limited sample size these observations should be addressed with caution and regarded as preliminary, being more as an indication of the need for further analysis than as firm evidence.

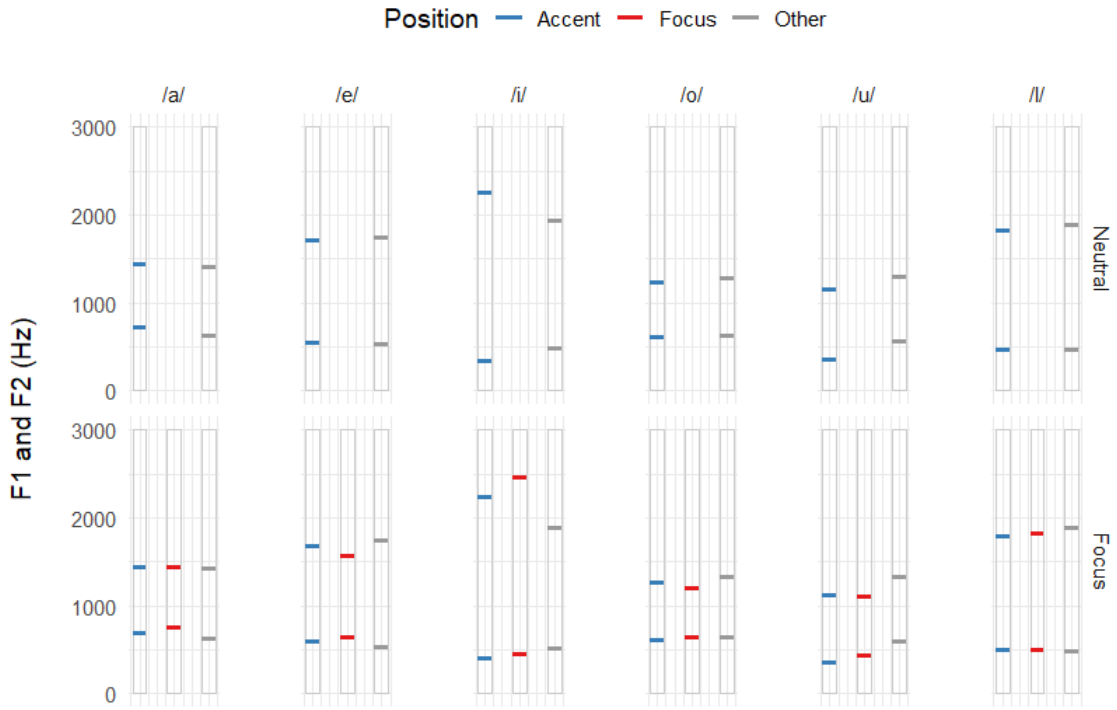

Figure 4.34. Mean values of F1 and F2 (Hz) for six Polish oral vowels across three prominence positions (*Accent*, *Focus*, *Other*), shown separately for the utterances realised under **neutral** (upper row) and contrastive **focus** condition (bottom row).

4.6 Rhythmic metrics

In addition to above phonetic-acoustic analyses, selected rhythm metrics were also calculated in order to examine the temporal organisation of the utterances. The measures used were already discussed in more detail in the section *Quantitative rhythm metrics* (Chapter 1). A detailed discussion of these acoustic metrics is beyond the scope of this dissertation; for a comprehensive overview with commentary, see Malisz (2013).

The following metrics were used:

- normalised Pairwise Variability Index (nPVI; Grabe & Low, 2002),
- VarcoC (Dellwo, 2006),
- Time Group Analysis (TGA; Gibbon, 2013)).

nPVI and VarcoC were selected as they are amongst the most widely used quantitative indices and supplement each other: nPVI (Equation 4.1) captures pairwise variability between successive phonetic segments (in the present case: vowels), while VarcoC reflects overall dispersion of consonant durations. Time Group Analysis (TGA) was applied to visualise grouping tendencies and temporal regularities beyond the segmental level, within interpausal time groups.

Additionally, VarcoV (Ferragne & Pellegrino, 2004) and %V (Ramus et al., 1999) were calculated, to present the two experimental conditions (**neutral** vs contrastive **focus**) in rhythmic space (cf. A. Wagner, 2014).

All the metrics were first derived at the level of individual utterances (per recording) and subsequently averaged across vowels and conditions (**neutral** vs **focus**). The nPVI and TGA were calculated using a plugin to AnnotationPro (Klessa & Gibbon, 2014), VarcoV, VarcoC, and %V measures were carried out using custom R scripts (Posit Software, PBC, 2023) developed by the author and deposited in the author's profile in the AMU Research Data Repository as supplementary material to this dissertation.

$$\text{nPVI} = \frac{100}{n-1} \sum_{i=1}^{n-1} \left| \frac{d_i - d_{i+1}}{(d_i + d_{i+1})/2} \right| \tag{4.1}$$

Normalized Pairwise Variability Index (nPVI) calculation formula (Equation 4.1), where: n = total number of intervals; d_k = duration of the k-th interval; d_{k+1} = duration of the subsequent interval; $|d_k - d_{k+1}|$ = absolute difference between consecutive durations; $(d_k + d_{k+1})/2$ = mean of the pair used for normalization. The formula sums normalized pairwise differences across all consecutive intervals and scales the result to provide a measure of rhythmic variability.

$$\text{VarcoC} = \frac{\sigma_C}{\mu_C} \times 100 \tag{4.2}$$

VarcoC (Coefficient of Variation for Consonants) formula (Equation 4.2). The metric quantifies rhythmic variability by calculating the ratio of standard deviation (σ_C) to mean (μ_C) of consonant

durations, multiplied by 100 to express the result as a percentage. Where: σ_C = standard deviation of all consonant durations in the analysed speech sample; μ_C = arithmetic mean of all consonant durations.

Table 4.23. Mean nPVI values (mean) and standard deviations (SD) across utterances containing target vowels produced in **focus** and **neutral** conditions

Vowel	Utterance	Focus		Neutral	
		Mean	SD	Mean	SD
/a/	Ala dała Arkowi zapas buraków	36	6	30	6
/e/	Edek jedzie do Łeby, jeszcze bez adresu	49	10	48	7
/i/	Irek widzi kilka irysów na stoliku	54	7	48	7
/o/	Ogon kota opadł po skoku na Karola	50	10	48	7
/u/	Uraz barku wujka ustał w południe	54	7	54	8
/I/	Solimy ryby i dajemy trzy łyżki cytryny	42	7	41	7
Total		47	10	45	10

The results (see Table 4.23) indicate consistently higher average nPVI values for the utterances realised in the **focus** condition (overall mean = 47) compared to the **neutral** condition (overall mean = 45). This suggests that the presence of contrastive focus may slightly increase temporal variability in vowel timing. Such an effect is plausible in light of the previously described lengthening of final vowels under contrastive focus (see Section 4.1). At the same time, the standard deviations remain comparable across conditions, which indicates that the effect is relatively stable and not driven by specific vowel categories.

Table 4.24 summarises VarcoC values, derived as the standard deviation of vowel durations normalised by their mean (see Equation 4.2).

Table 4.24. Mean VarcoC values (%) and standard deviations (SD) across utterances containing target vowels produced in **focus** and **neutral** conditions

Vowel	Utterance	Focus		Neutral	
		Mean	SD	Mean	SD
/a/	Ala dała Arkowi zapas buraków	40	9	43	8
/e/	Edek jedzie do Łeby, jeszcze bez adresu	43	16	34	7
/i/	Irek widzi kilka irysów na stoliku	47	26	35	12
/o/	Ogon kota opadł po skoku na Karola	39	17	33	15
/u/	Uraz barku wujka ustał w południe	51	32	35	9
/I/	Solimy ryby i dajemy trzy łyżki cytryny	36	13	40	8
Total		43	20	37	11

VarcoC values provide an index of how evenly or unevenly consonants are timed, capturing the relative degree of their temporal dispersion within a sequence of vowels. Higher VarcoC values reflect greater irregularity in duration, whereas lower values indicate more even timing. In the present data, mean VarcoC values were observed in the **focus** condition (overall mean = 43) compared to the **neutral** condition (overall mean = 37), which implies that presence

of contrastive focus tends to increase variability in consonant timing. In the section *Focus influence on global prosodic pattern* it was already discussed that presence of contrastive focus in an utterance impacts vowel, here such greater variability is also displayed for consonants, supplementing previous observations.

The effects, however, are not uniform across the stimuli: the strongest increases occur for utterances with /i/ and /u/, whereas those with /a/ remain stable and those with /I/ show reduced consonant variability under **focus**. It should be noted that consonant composition was not controlled in the stimulus design, apart from ensuring that no nasal consonants /m/ or /n/ occurred in the immediate vicinity of the target vowels and avoiding consonantal clusters. Therefore, part of the observed differences may reflect both natural variation in consonant environments and effect of the stimuli construction rather than an influence of presence of contrastive focus.

To visualise how the above mentioned differences between conditions and vowels translate into broader rhythmic organisation of the utterances, the results are plotted in rhythm spaces defined by VarcoV–VarcoC and %V–nPVI (see Figure 4.35). Such multidimensional mapping helps comparing variability across conditions.

In the **focus** condition, mean VarcoC values are higher with the exception of /a/ and /I/. VarcoV values also shift upward for all the vowels. These points towards increased variability in phonemic segments, both consonants and vowels, timing in the presence of contrastive focus. Moreover, **focus** condition displays greater dispersion than **neutral** in terms of these two metrics.

Mean nPVI values are also mostly higher in the **focus** condition (with the exception of /u/ which has already been discussed in the Section 4.2.1.3), confirming a tendency towards greater temporal irregularity in vowel timing. Differences in %V remain fairly small ($\sim$1.1 pp of difference on average), indicating that the overall proportion of vocalic material in the utterance does not change substantially between conditions.

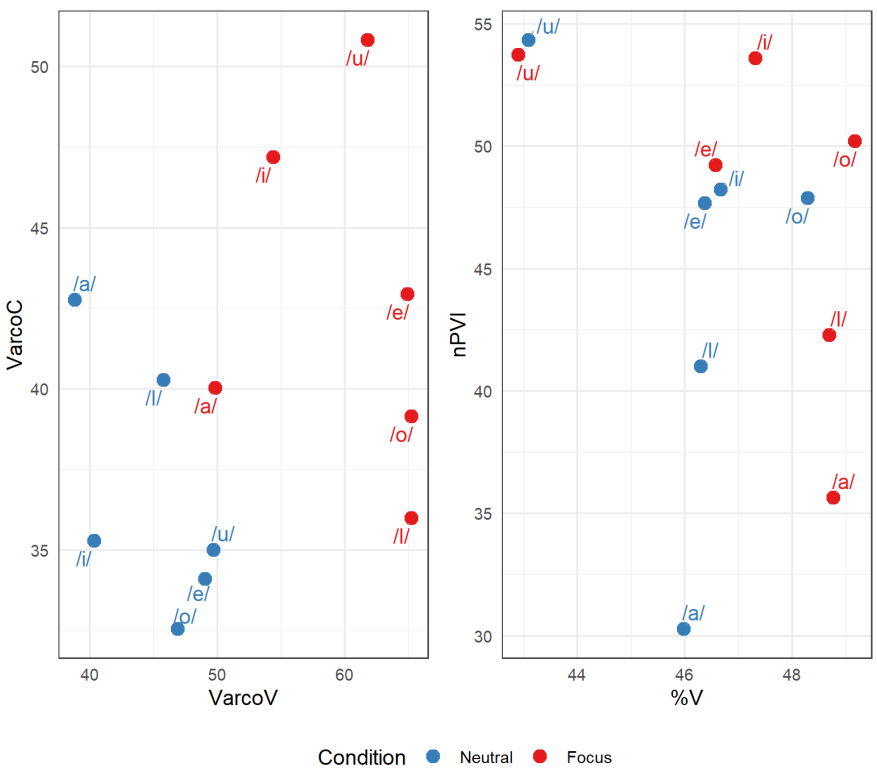

Figure 4.35. Rhythm space determined by VarcoV–VarcoC (left) and %V–nPVI (right) for Polish vowels under **neutral** (blue) and **focus** (red) conditions. Each point represents the mean value for a given *Vowel* × *Condition*, averaged all speakers (adapted from A. Wagner (2014) for cross-linguistic comparison; here applied to contrast two prosodic conditions).

While Varco, nPVI, and %V capture overall variability, they do not reveal how timing differences unfold across utterances. To address this limitation, the analysis was extended with **Time Group Analysis (TGA), which provides insights into the unfolding dynamics of entire sentences**. In this respect, TGA complements interval-based rhythm metrics by offering a perspective on how rhythmic structure emerges over time.

A modified version of the approach described in Klessa and Gibbon (2014) was employed: TGA plots are based on averaged durations per *Vowel* × *Condition* and the resulting patterns are presented in the figures below. In Figures 4.36- 4.47, the duration variability tendencies are shown in the form of duration differences chart (top), and a top-suspended bar chart (bottom). Following the solution implemented in the original TGA tool (Gibbon, 2013), the y-axis values in the bottom chart are inverted to better visualize acceleration-deceleration patterns.

All the TGA figures below refer to the realisation of utterances containing the vowel specified in the headings. Utterances with pauses were excluded from the visualization (for pause detection methodology, see Section 3.3.8.1).

For utterances featuring /a/ as the target vowel, the regression slope under **focus** is noticeably flatter (0.02) than in the **neutral** utterances (0.051), see Figure 4.36 and Figure 4.37. In **neutral** condition syllables are fairly evenly distributed across the utterance, whereas in the **focus** condition, the word in *focus* position *dała* /dawa/ (Eng. 'gave') is longer.

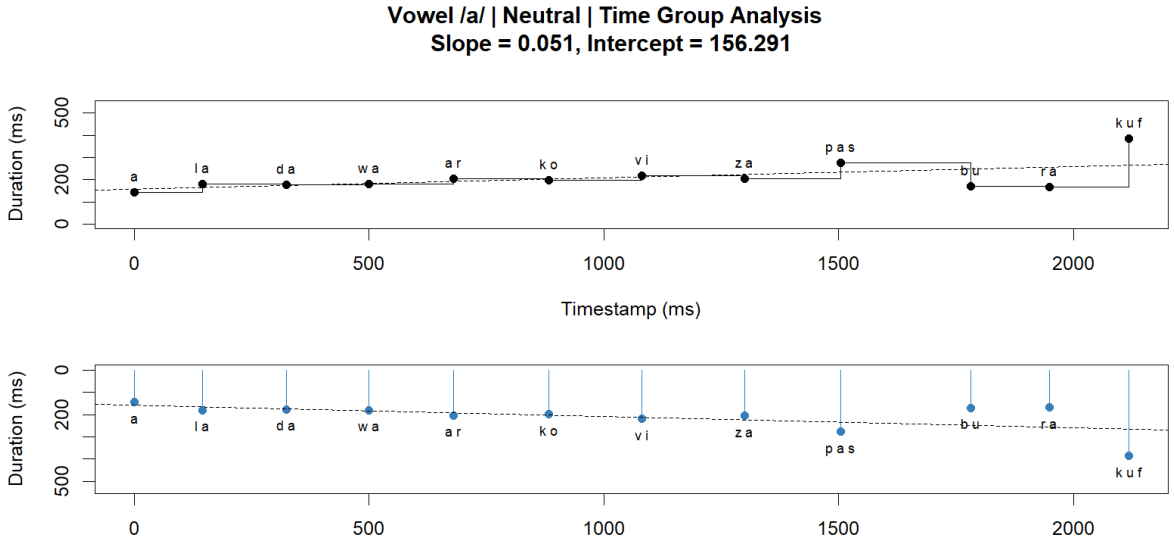

Figure 4.36. Time Group Analysis (TGA) across the realisations of the Polish utterance *Ala dała Arkowi zapas buraków* (/ala dawa arkovi zapas burakuf/, Eng. 'Ala gave Arek a supply of beets'), featuring /a/ as the target vowel. Neutral condition.

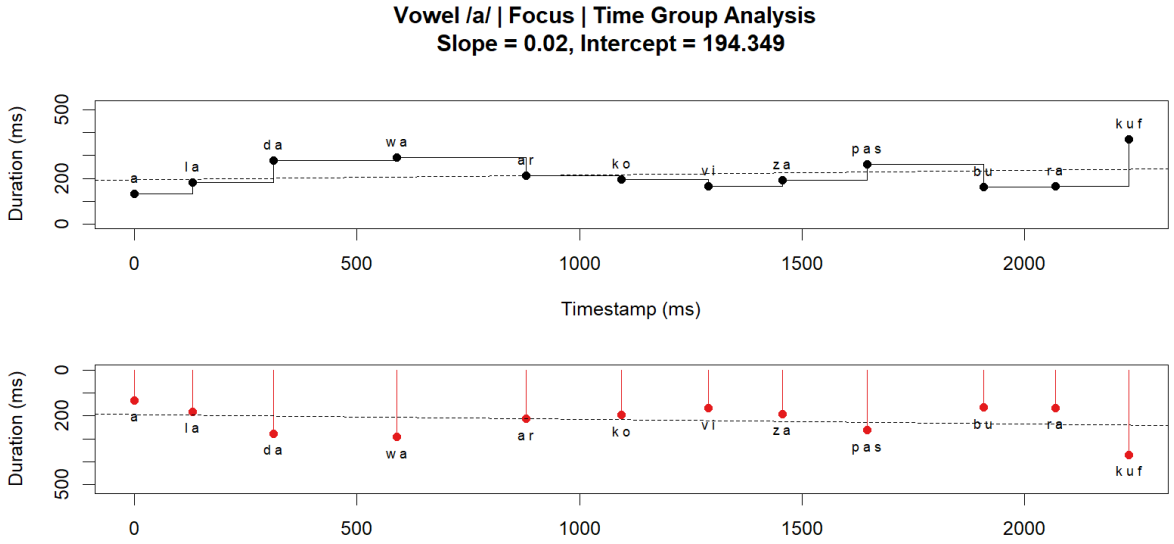

Figure 4.37. Time Group Analysis (TGA) across the realisations of the Polish utterance *Ala dała Arkowi zapas buraków* /ala dawa arkovi zapas burakuf/ (Eng. 'Ala gave Arek a supply of beets'), featuring /a/ as the target vowel. Focus condition. Contrastive focus on the word *dała* /dawa/ (Eng. 'gave').

For utterances featuring /e/ as the target vowel, the regression slope is flatter as well under **focus** condition (0.023) than under **neutral** condition (0.039) — see Figure 4.38 and Figure 4.39. This pattern mirrors the results for utterances with the target vowel /a/: the presence of contrastive focus tends to suppress cumulative lengthening and stabilise timing, while neutrality permits a stronger temporal drift. For /a/ and /e/, the contrast is clear: **neutral** slopes (0.039–0.051) are more than double those in **focus** (0.02–0.023). In short, the presence of contrastive focus constrains rhythmic expansion.

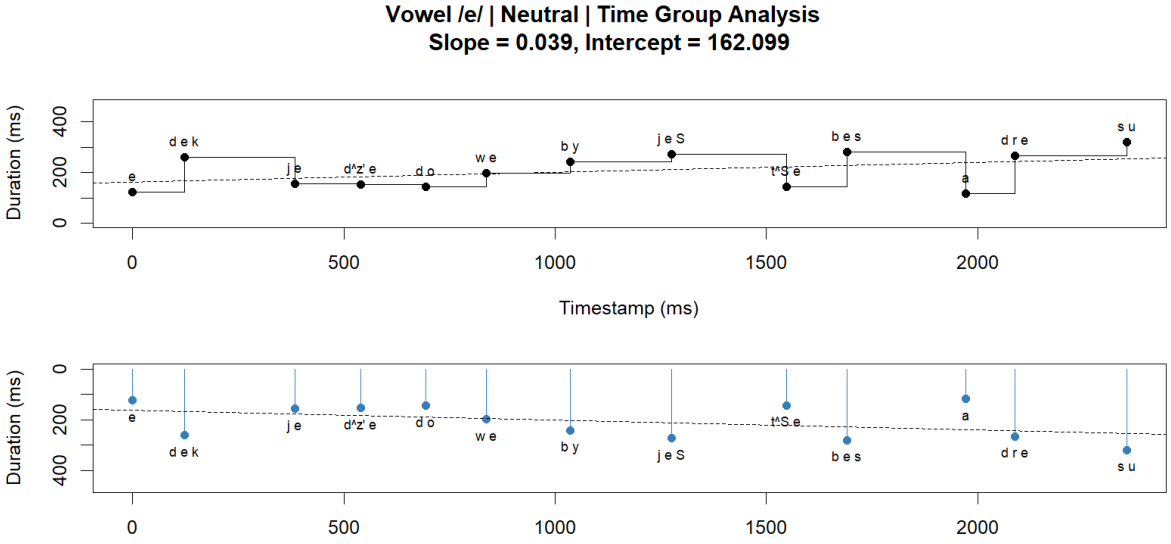

Figure 4.38. Time Group Analysis (TGA) across the realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet'), featuring /e/ as the target vowel. Neutral condition.

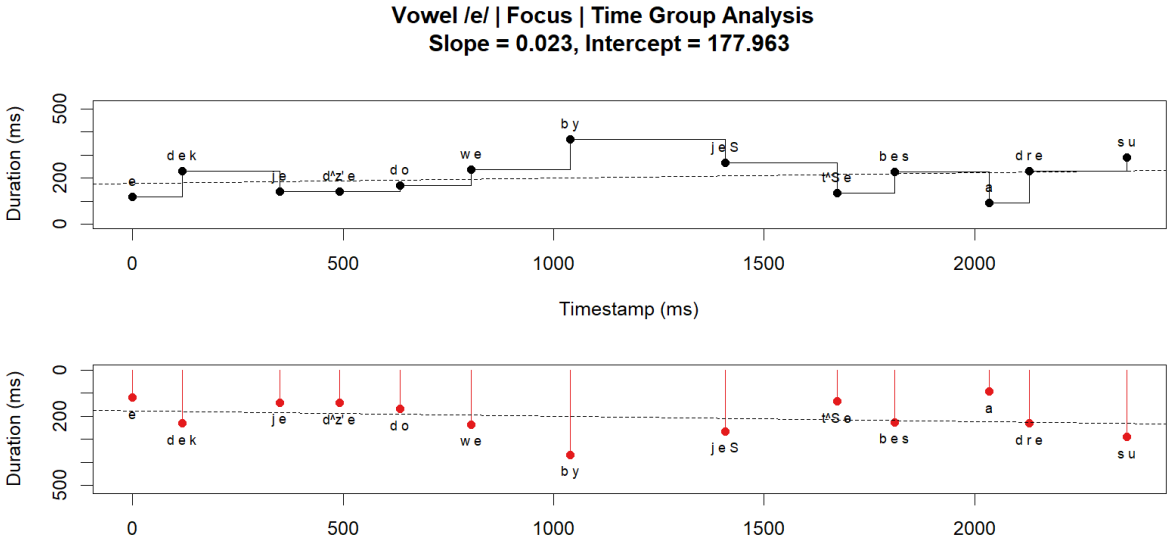

Figure 4.39. Time Group Analysis (TGA) across the realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet'). Focus condition, featuring /e/ as the target vowel. Contrastive focus on the word *jedzie* /jedˆSz'e/ (Eng. 'is going').

For utterances featuring /i/ as the target vowel, the effect of rhythmic compression in the presence of contrastive focus is the strongest: **focus** condition slope is essentially flat (see Figure 4.41) compared to a rising slope in **neutral** condition (0.035) — as depicted in Figure 4.40.

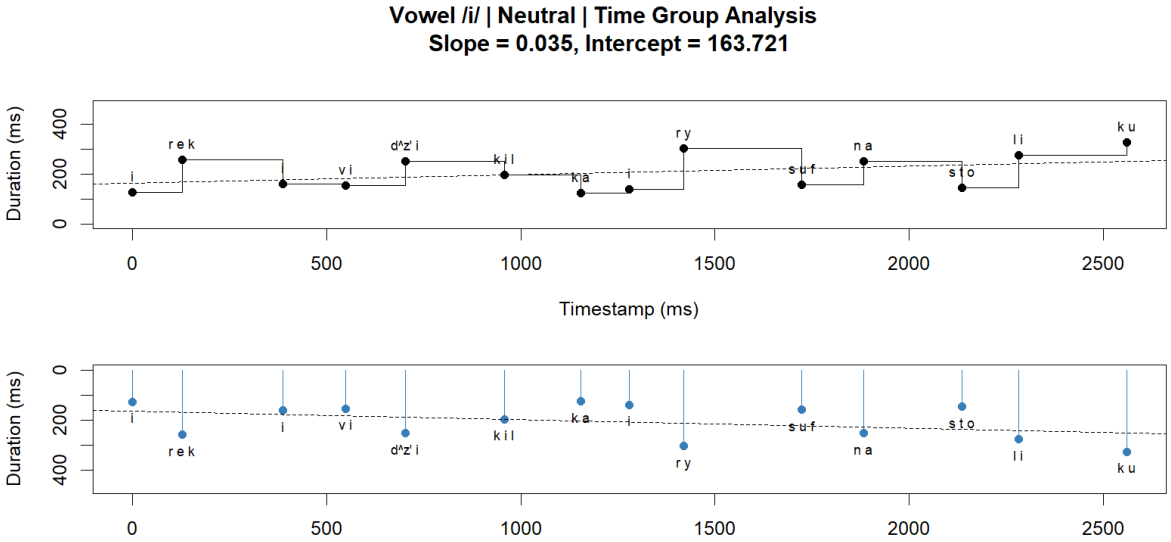

Figure 4.40. Time Group Analysis (TGA) across the realisations of the Polish utterance *Irek widzi kilka irysów na stoliku* /irek vidˆz'i kilka irIsuf na stoliku/ (Eng. 'Irek sees a few irises on the table'), featuring /i/ as the target vowel. Neutral condition.

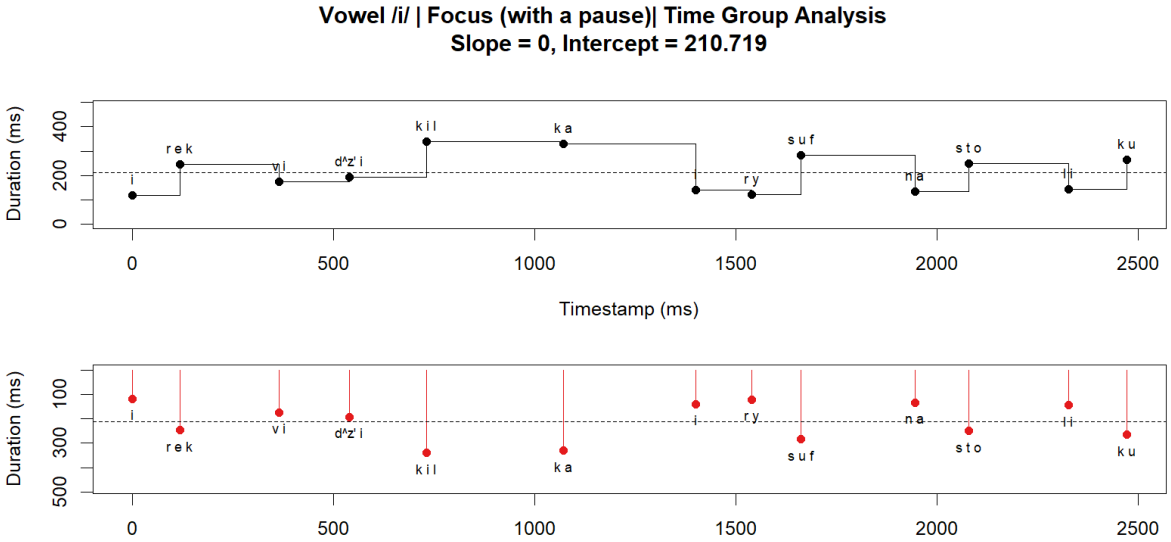

Figure 4.41. Time Group Analysis (TGA) across the realisations of the Polish utterance *Irek widzi kilka irysów na stoliku* /irek vidˆz'i kilka irIsuf na stoliku/ (Eng. 'Irek sees a few irises on the table'), featuring /i/ as the target vowel. Focus condition. Contrastive focus on the word *kilka* /kilka/ (Eng. 'a few').

For utterances featuring /o/ as the target vowel, slopes are negative in both conditions. Nevertheless, the slope is more flat in **focus** condition(−0.032 vs −0.011) than in **neutral** condition, as presented in Figure 4.42 and Figure 4.43.

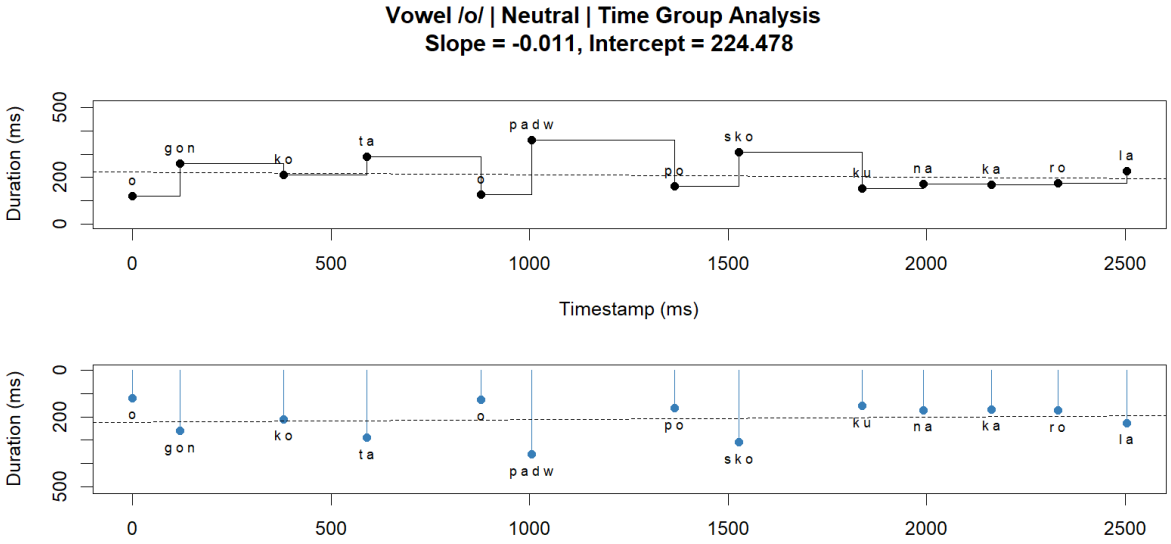

Figure 4.42. Time Group Analysis (TGA) across the realisations of the Polish utterance *Ogon kota opadł po skoku na Karola* /ogon kota opadw po skoku na karola/ (Eng. 'A cat's tail has dropped after jumping on Karol'), featuring /o/ as the target vowel. Neutral condition.

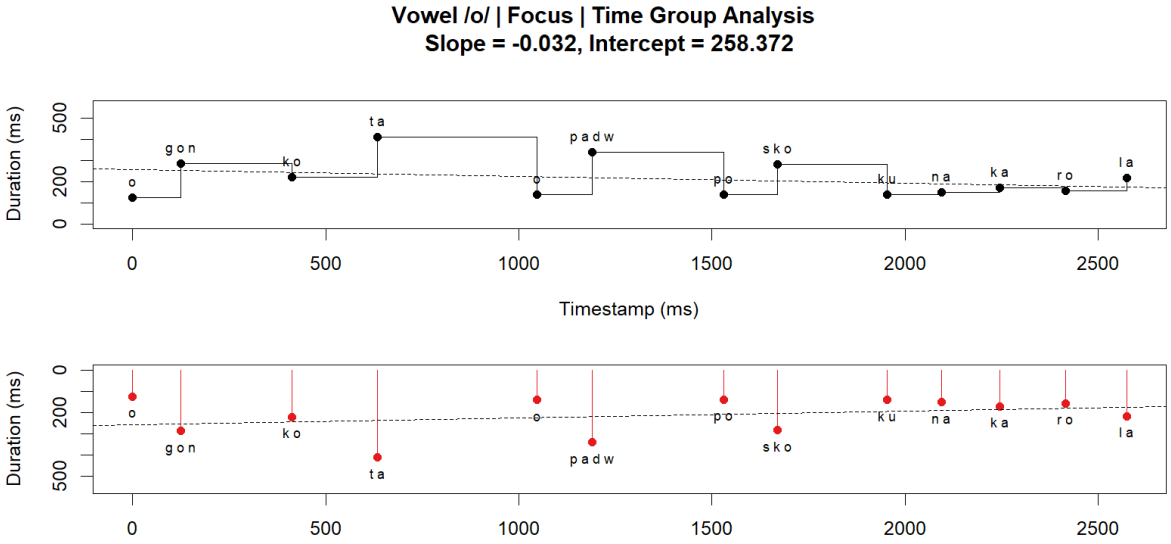

Figure 4.43. Time Group Analysis (TGA) for the vowel /o/ across the realisations of the Polish utterance *Ogon kota opadł po skoku na Karola* /ogon kota opadw po skoku na karola/ (Eng. 'A cat's tail has dropped after jumping on Karol'). Focus condition. Contrastive focus on the wor on the word *kota* /kota/ (Eng. 'cat's').

For utterances featuring /u/ and /I/ as the target vowels, slopes remain positive in both conditions, but tend to be less steep in **focus** conditions, as presented in Figures 4.44–4.47.

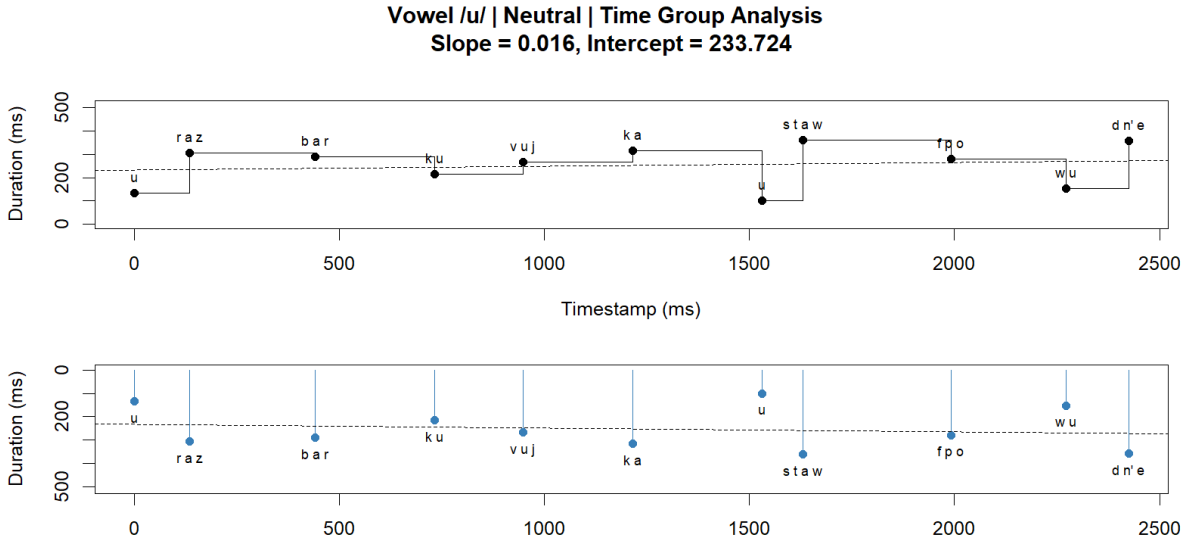

Figure 4.44. Time Group Analysis (TGA) across the realisations of the Polish utterance *Uraz barku wujka ustał w południe* /uras barku vujka ustaw fpowudn'e/ (Eng. 'Uncle's shoulder injury cleared up at noon.'), featuring /u/ as the target vowel. Neutral condition.

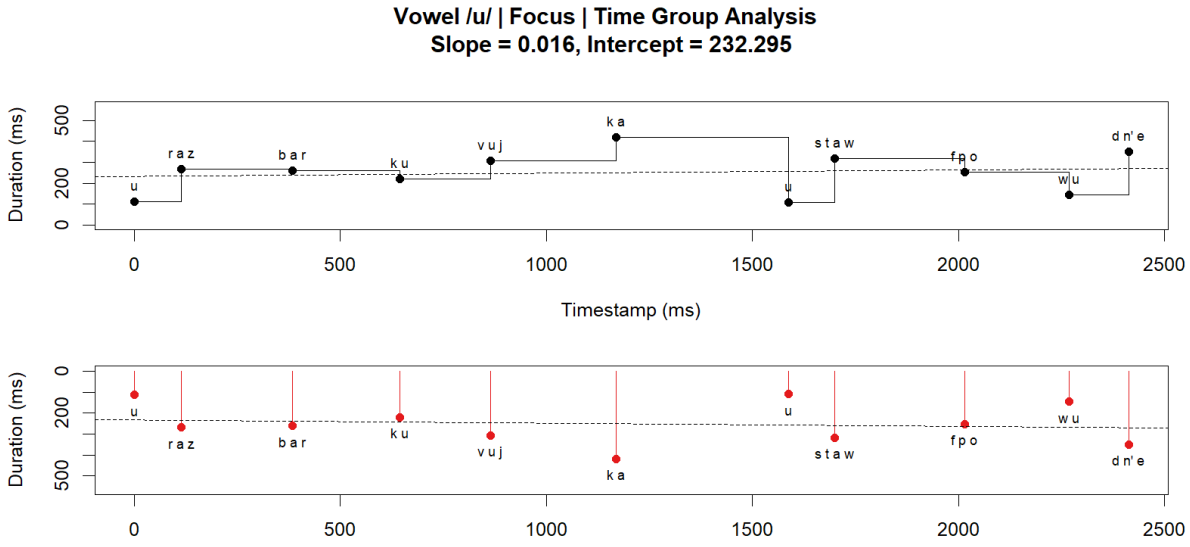

Figure 4.45. Time Group Analysis (TGA) across the realisations of the Polish utterance *Uraz barku wujka ustał w południe* /uras barku vujka ustaw fpowudn'e/ (Eng. 'Uncle's shoulder injury cleared up at noon.'), featuring /u/ as the target vowel. Focus condition. Contrastive focus on the word *wujka* /vujka/ (Eng. 'uncle's).

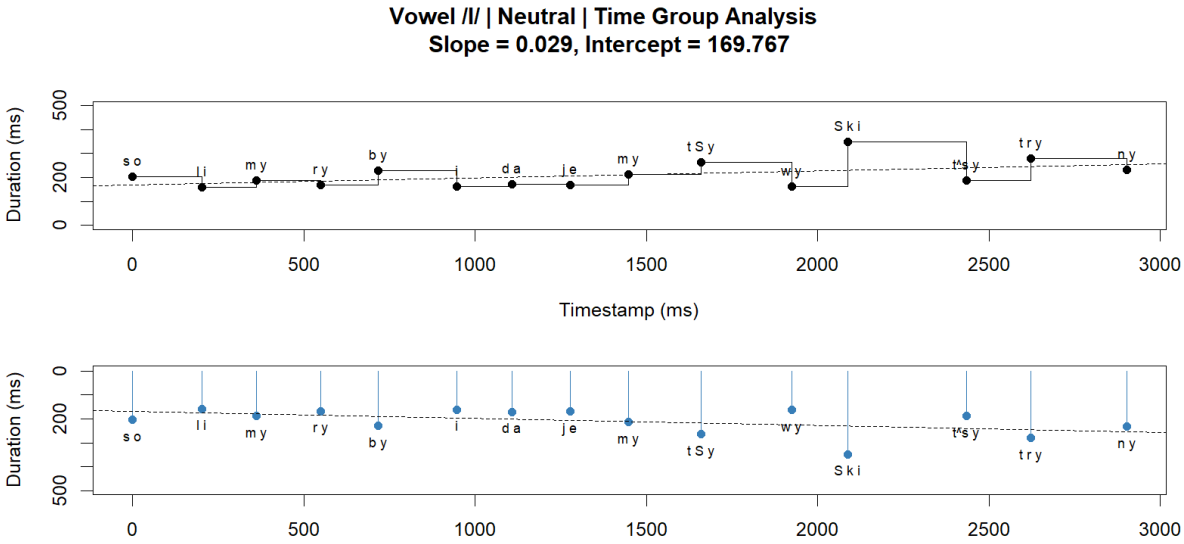

Figure 4.46. Time Group Analysis (TGA) across the realisations of the Polish utterance *Solimy ryby i dajemy trzy łyżki cytryny* /solimI rIbI i dajemI tSI wISki tˆsItrInI/ (Eng. 'Season the fish with salt and add three tablespoons of lemon juice.'), featuring /I/ as the target vowel. Neutral condition.

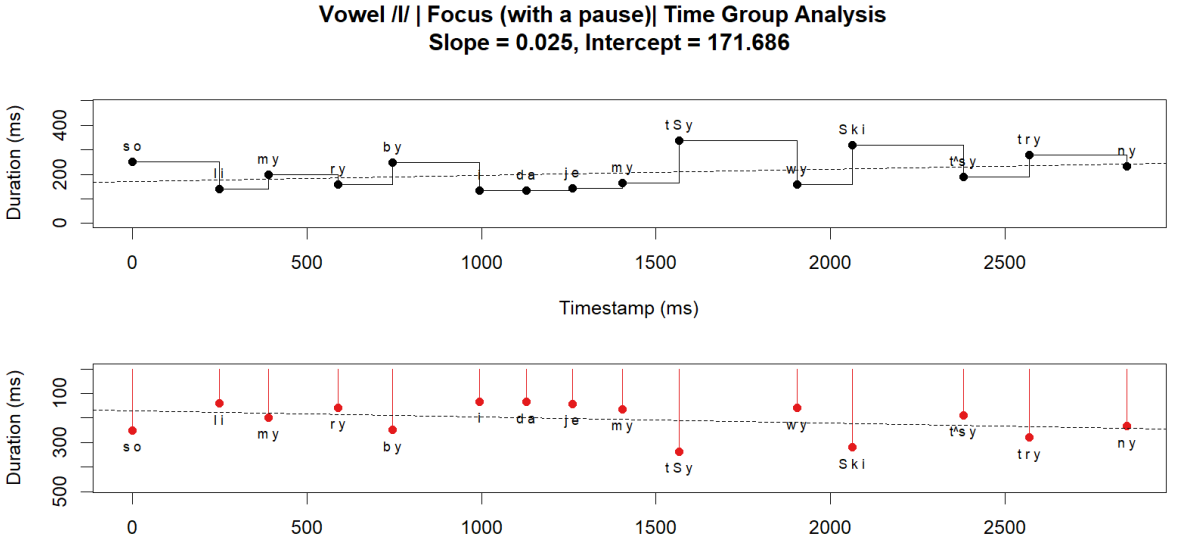

Figure 4.47. Time Group Analysis (TGA) across the realisations of the Polish utterance *Solimy ryby i dajemy trzy łyżki cytryny* /solimI rIbI i dajemI tSI wISki tˆsItrInI/ (Eng. 'Season the fish with salt and add three tablespoons of lemon juice.'), featuring /I/ as the target vowel. Focus condition. Contrastive focus on the word *ryby* /rIbI/ (Eng. 'fish').

In summary, the presence of contrastive focus consistently reduces the slope of the TGA regression lines, indicating a stabilising effect on temporal organisation. Utterances produced under **neutral** condition, by contrast, exhibit stronger cumulative lengthening (or shortening in the case of /o/), suggesting that timing drifts more freely. **Duration intercept values are generally higher in focus contexts, which means the initial phonemic segments are already a bit longer, while neutral utterances tend to start from lower values but increase more sharply over time**.

Taken together, the rhythm metrics discussed in this section point towards those already highlighted in the previous parts in this chapter. The presence of contrastive focus influences overall governance of an utterance and exerts a dual influence: it increases local variability in phonemic segments timing, while simultaneously constraining the global drift of durations across the utterance. In other words, in **focus** conditions some segments become longer at the expense of others, which become shorter, while, at the same time, equalising the overall duration of an utterance. Metaphorically, in **neutral** condition, the rhythm of an utterance resembles a calm sea surface with a rather steady current that gradually carries the sounds away from the starting point, whereas in **focus** utterances the current is largely suppressed but the rough, uneven surface with local fluctuations more pronounced.

These findings, together with the phonetic–acoustic analyses presented earlier, conclude the supplementary set of analyses. The following section summarises the acoustic and articulatory profiles of vowels across the two conditions.

4.7 Jaw and lip movements with corresponding F0 and intensity profiles

This section illustrates the effects of experimental conditions on the realisation of vowels using trajectory plots presented separately for the **neutral** and **focus** conditions. For each case, variability patterns of both articulatory parameters (jaw displacement, upper and lower lip movements) and phonetic–acoustic features (F0 and intensity) are presented in time-aligned multi-panel trajectory plots. The time axis was normalised to segment duration, and articulatory trajectories were centred according to the procedure detailed in Chapter 3 (Section 3.3.9).

In all panels in **neutral** condition, a red marking indicates where the vowel in *Focus* position would occur, to make comparisons easier. Also, for the same reason, in the plots of mean F0 and intensity include red and blue dashed lines, representing the maximum and minimum values for **focus** (red lines) and **neutral** (blue lines) conditions.

There are several limitations in interpreting the F0 contour plots. Firstly, due to aggregation some artefacts are present, especially in the /e/ vowel plot (see Figure 4.50) where high inter-speaker variability in F0 contour production was observed. In addition, part of the F0 data is missing, due to the Praat algorithm returning undefined values whenever the signal quality does not allow for reliable pitch estimation. These discontinuities and gaps were controlled for in the statistical analyses and should not be interpreted as genuine intonational movements. That being said, there are indeed some tendencies present which will be outlined below.

The general shape of the trajectories for vowel /a/ is similar in both the **neutral** (see Figure 4.48) and **focus** (see Figure 4.49) conditions. However, the latter is marked by greater articulatory displacements, particularly in the jaw and the lower lip. Both F0 and intensity peak in *Focus* position.

For /e/, the **focus** condition results in a deeper jaw lowering and a more pronounced lip aperture compared to the **neutral** condition. Such articulatory change was paralleled by an increase in F0 and higher intensity peaks (see Figure 4.50 and Figure 4.51).

For /i/, the overall articulatory differences between conditions were more subtle due to the inherently narrow oral configuration of this vowel (see Figure 4.52). Nevertheless, the **focus** condition induced a slightly deeper jaw lowering and lip aperture. Acoustically, realisations produced under the **focus** condition displayed a distinct F0 rise and stronger intensity peaks (see Figure 4.53).

The vowel /o/ showed the most pronounced reorganisation of jaw–lip coordination under the **focus** condition (see Figure 4.55). While in **neutral** condition the jaw initiated the movement followed by the lips (see Figure 4.54), in **focus** the three articulators acted nearly simultaneously, resulting in a wider opening. This articulatory synergy coincided with a marked F0 peak and higher, sharper intensity maxima.

In /u/, the **focus** condition elicited a stronger and faster jaw lowering (see Figure 4.57) as well as an almost synchronous response of the lower lip. The acoustic correlates reflected this strengthening: F0 was raised and intensity peaks were clearly amplified. As in the case of other vowels, intensity displayed a cyclic alternation, with conspicuous drops on adjacent consonants (see Figure 4.56).

The articulatory effects of focus for /I/ were comparatively modest. The jaw movement remained shallow, though slightly more dynamic, and the lips exhibited somewhat greater responsiveness, with the lower lip reacting more quickly and the upper lip contributing small additional adjustments (see Figure 4.59). Acoustically, focus realisations were associated with a mild F0 increase and moderately higher intensity. In both conditions an interesting F0 pattern is visible with the F0 maximal value reached at the boundary of /i/ and /I/ in **neutral** utterances (see Figure 4.58). Similarly, this peak is present in contrastive **focus** condition, but less pronounced. Since both these are high vowels, such an effect might be a way of boundary marking, however such an insight would need further analysis.

Across all vowels, the presence of focus systematically enhanced the jaw displacement and the lip aperture which was demonstrated earlier in statistical modelling part of this chapter. Here these findings were complimented with visual input. These articulatory adjustments were mirrored acoustically: F0 was consistently raised in *Focus* position vowels with sharp, abrupt declination after the peak value. Intensity also clearly rises in *Focus* position. Importantly, intensity showed a robust cyclic alternation, with maxima on vowels and immediate declines on consonants, a pattern that remained consistent across both conditions but was more pronounced under **focus**.

As for articulatory gesture coordination, the jaw and lips constitute partly independent systems; the jaw provides the global opening–closing frame, initiating the movement before the new

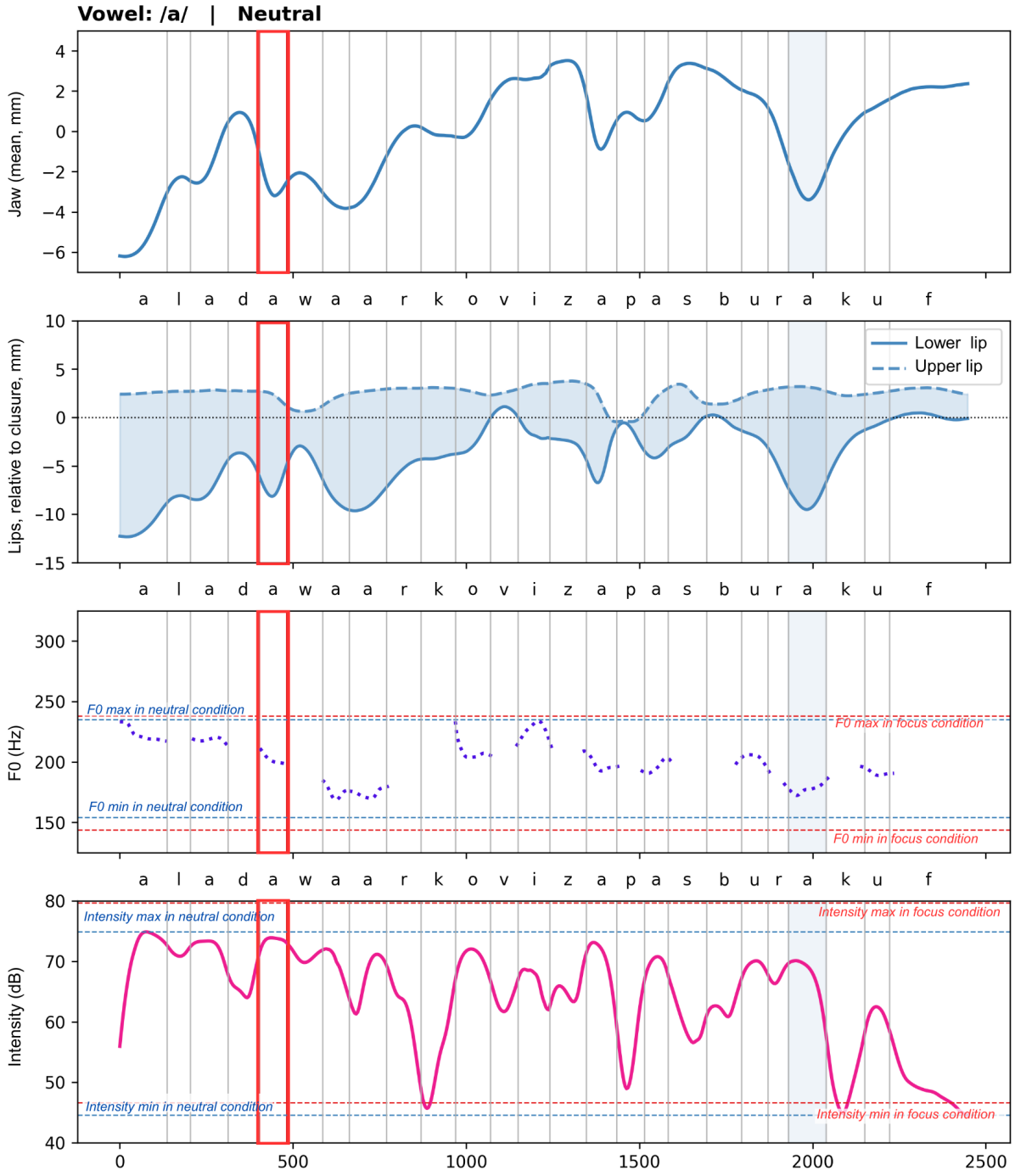

Figure 4.48. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Ala dała Arkowi zapas buraków* (/ala dawa arkovi zapas burakuf/, Eng. 'Ala gave Arek a supply of beets'), featuring /a/ as the target vowel. Neutral condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

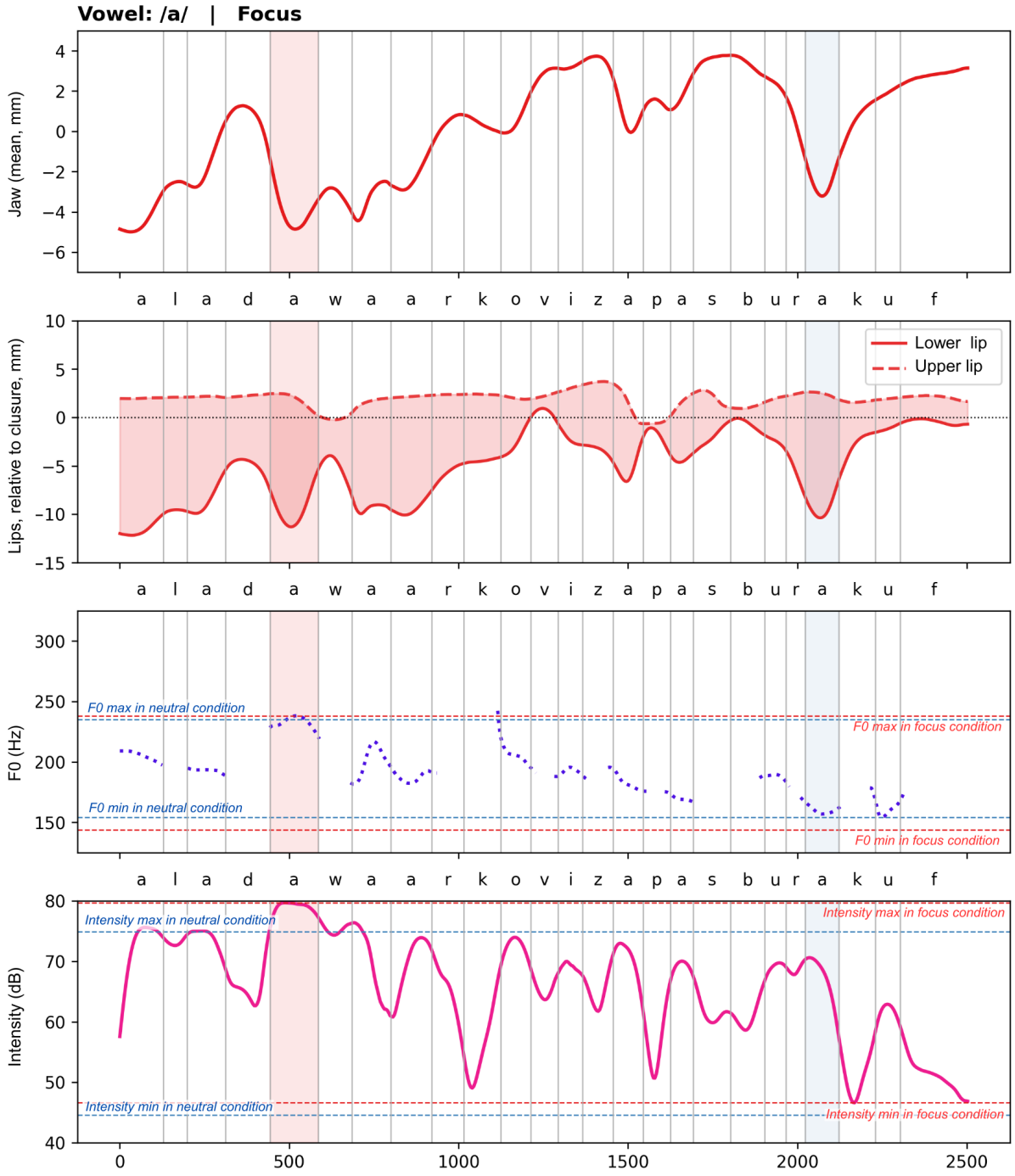

Figure 4.49. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Ala dała Arkowi zapas buraków* (/ala dawa arkovi zapas burakuf/, Eng. 'Ala gave Arek a supply of beets'), featuring /a/ as the target vowel. Focus condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

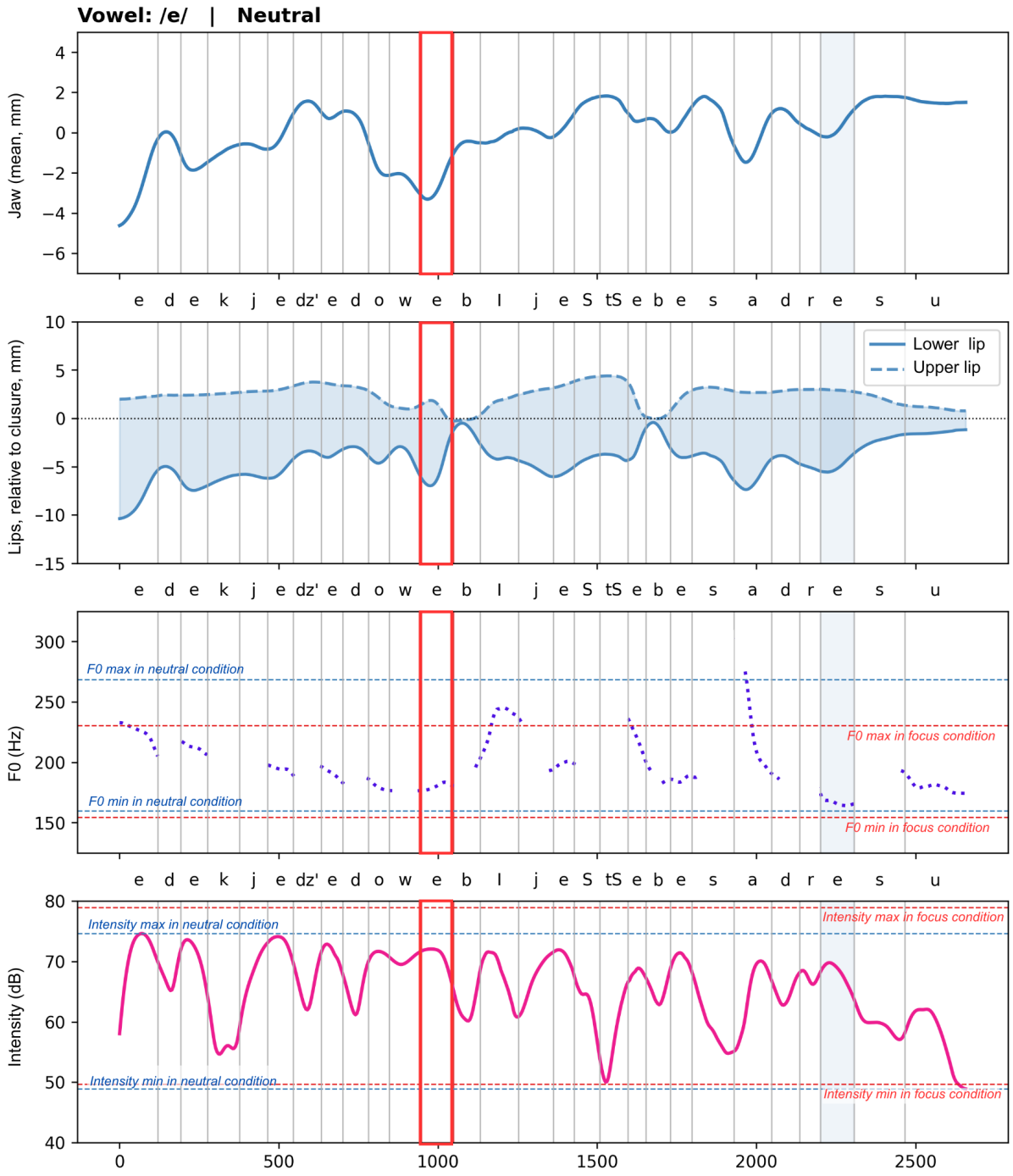

Figure 4.50. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet'), featuring /e/ as the target vowel. Neutral condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

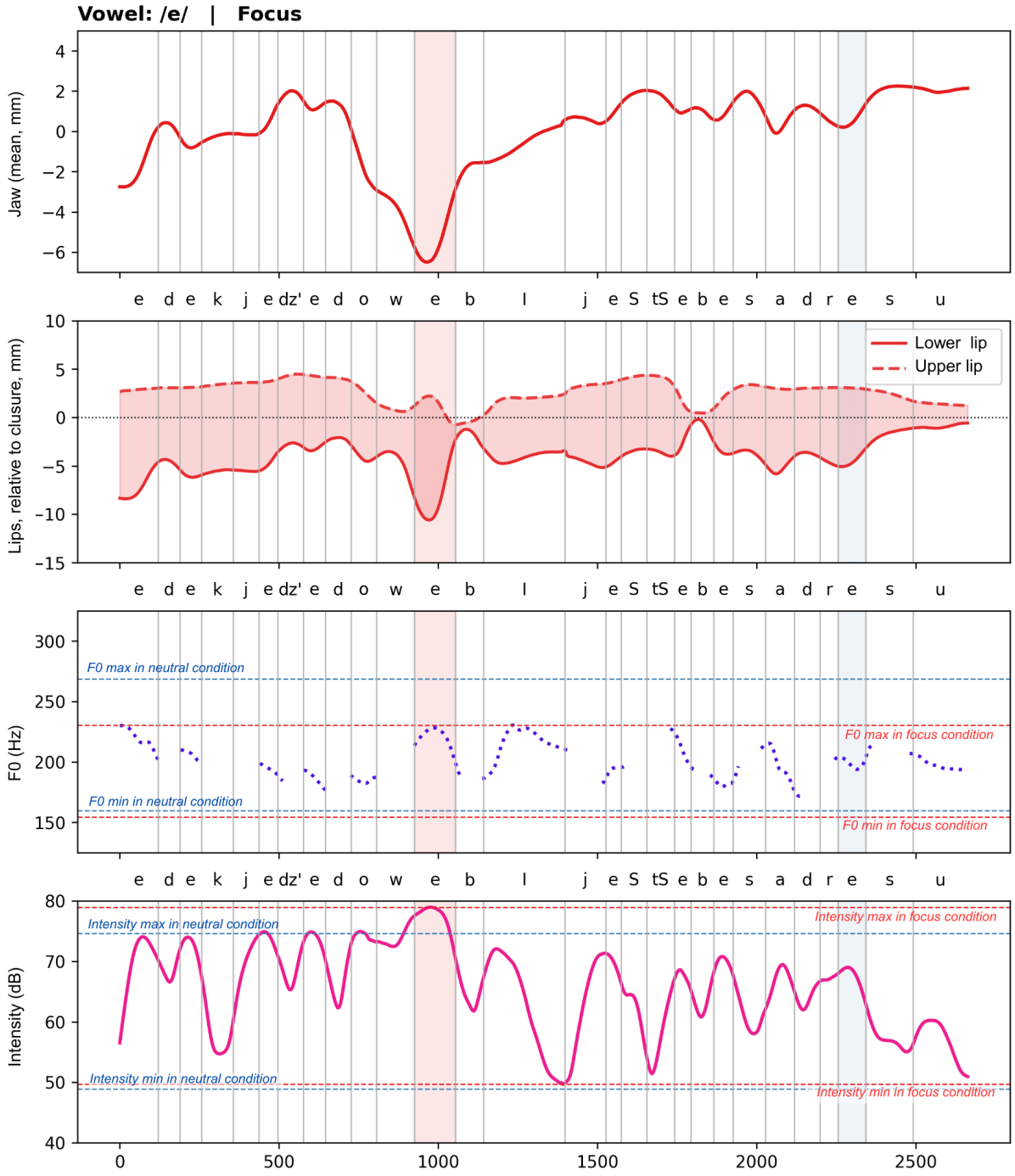

Figure 4.51. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Edek jedzie do Łeby, jeszcze bez adresu* /edek jedˆz'e do webI jeStˆSe bes adresu/ (Eng. 'Edek is going to Łeba, no address yet'), featuring /e/ as the target vowel. Focus condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

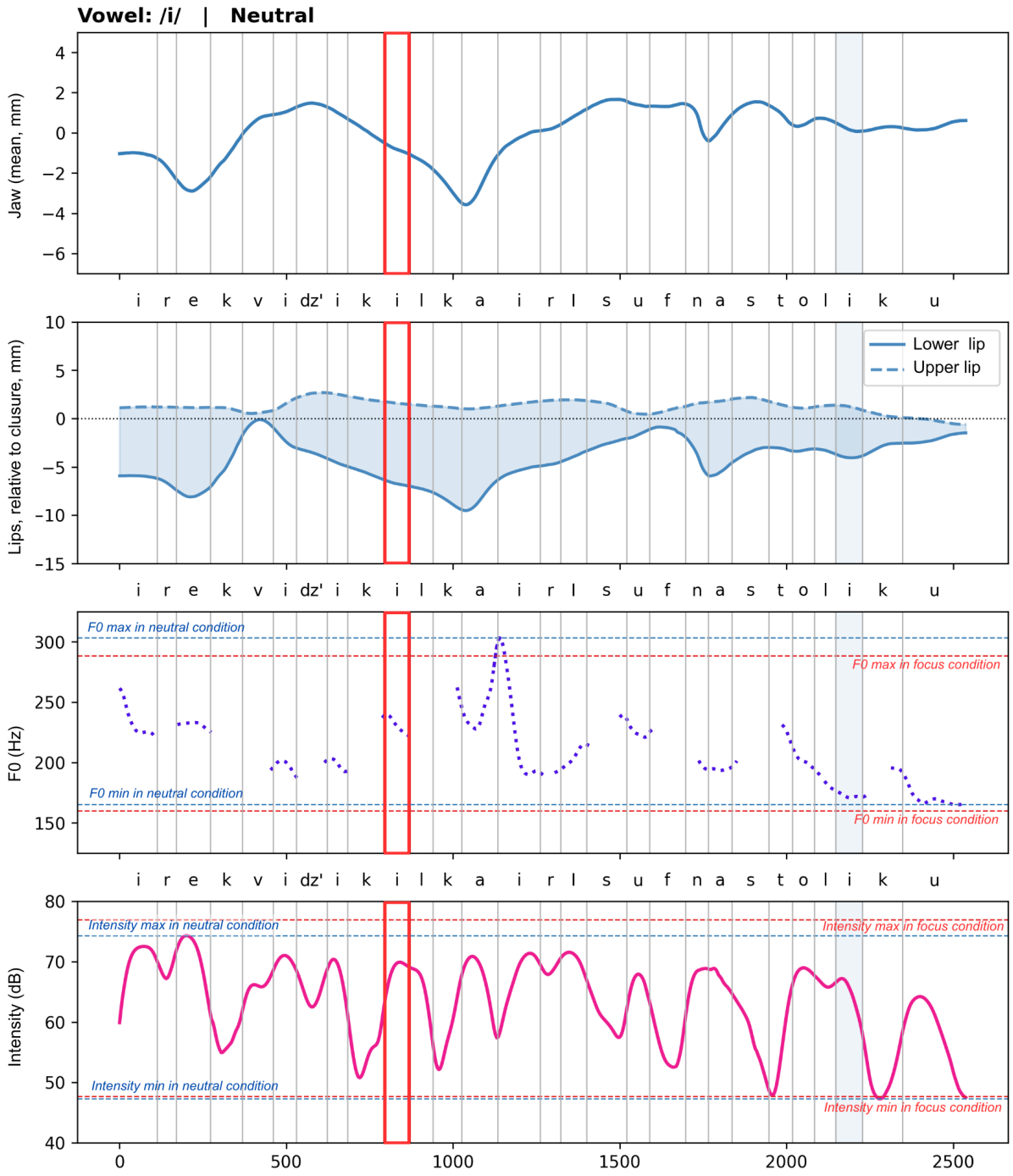

Figure 4.52. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Irek widzi kilka irysów na stoliku* /irek vidˆz'i kilka irIsuf na stoliku/ (Eng. 'Irek sees a few irises on the table'), featuring /i/ as the target vowel. Neutral condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

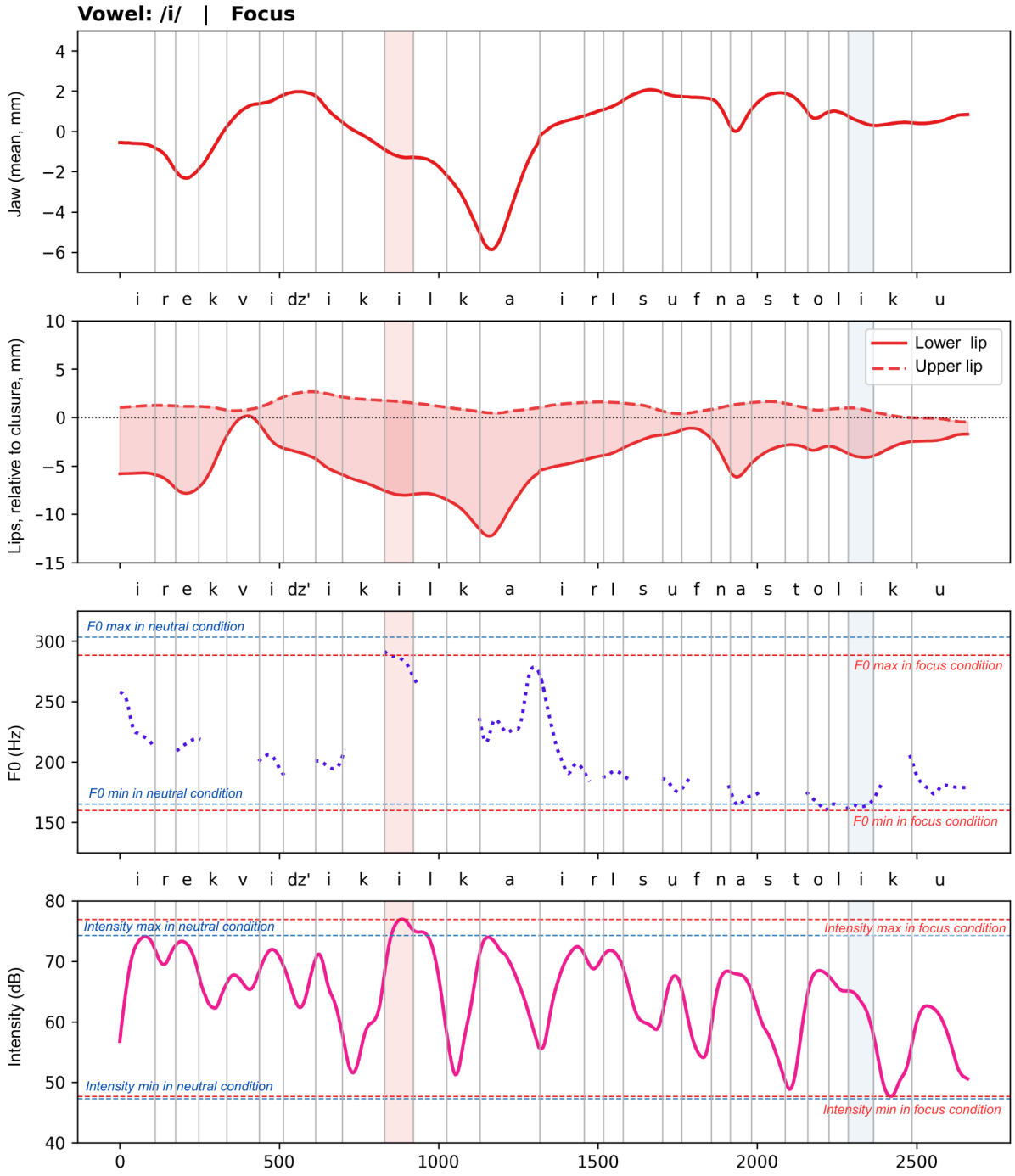

Figure 4.53. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Irek widzi kilka irysów na stoliku* /irek vidˆz'i kilka irIsuf na stoliku/ (Eng. 'Irek sees a few irises on the table'), featuring /i/ as the target vowel. Focus condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

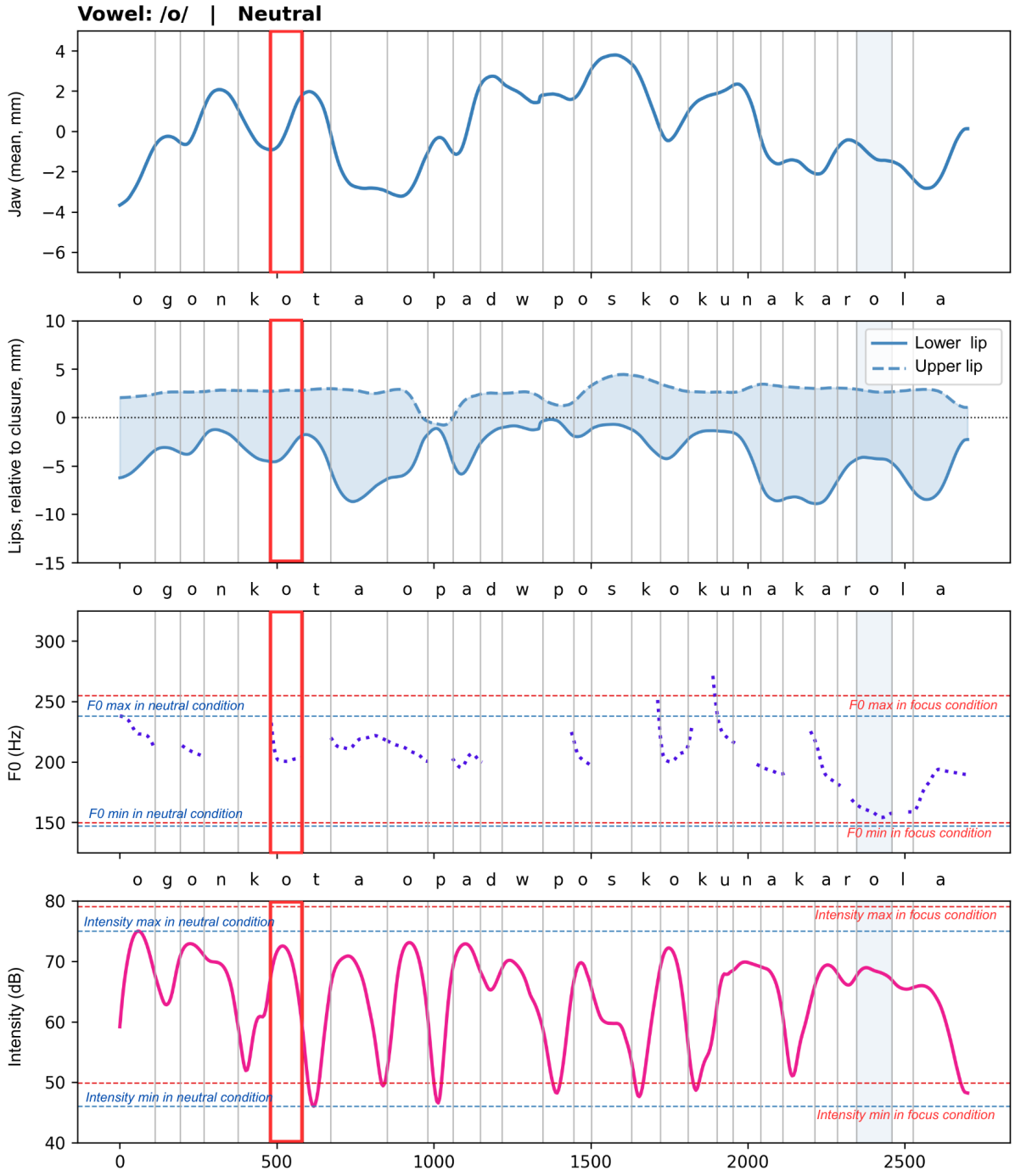

Figure 4.54. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Ogon kota opadł po skoku na Karola* /ogon kota opadw po skoku na karola/ (Eng. 'A cat's tail has dropped after jumping on Karol'), featuring /o/ as the target vowel. Neutral condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

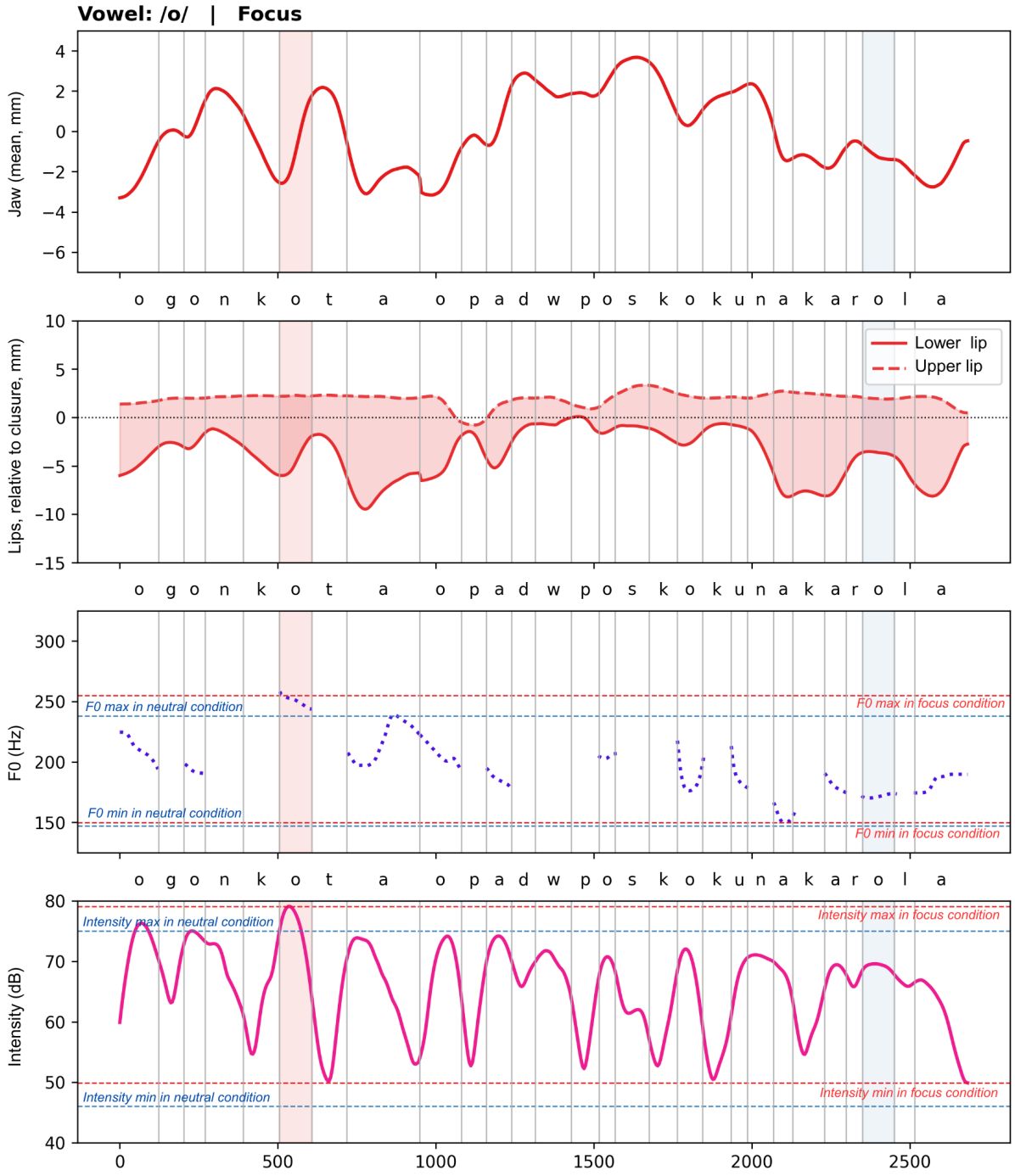

Figure 4.55. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Ogon kota opadł po skoku na Karola* /ogon kota opadw po skoku na karola/ (Eng. 'A cat's tail has dropped after jumping on Karol'), featuring /o/ as the target vowel. Focus condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

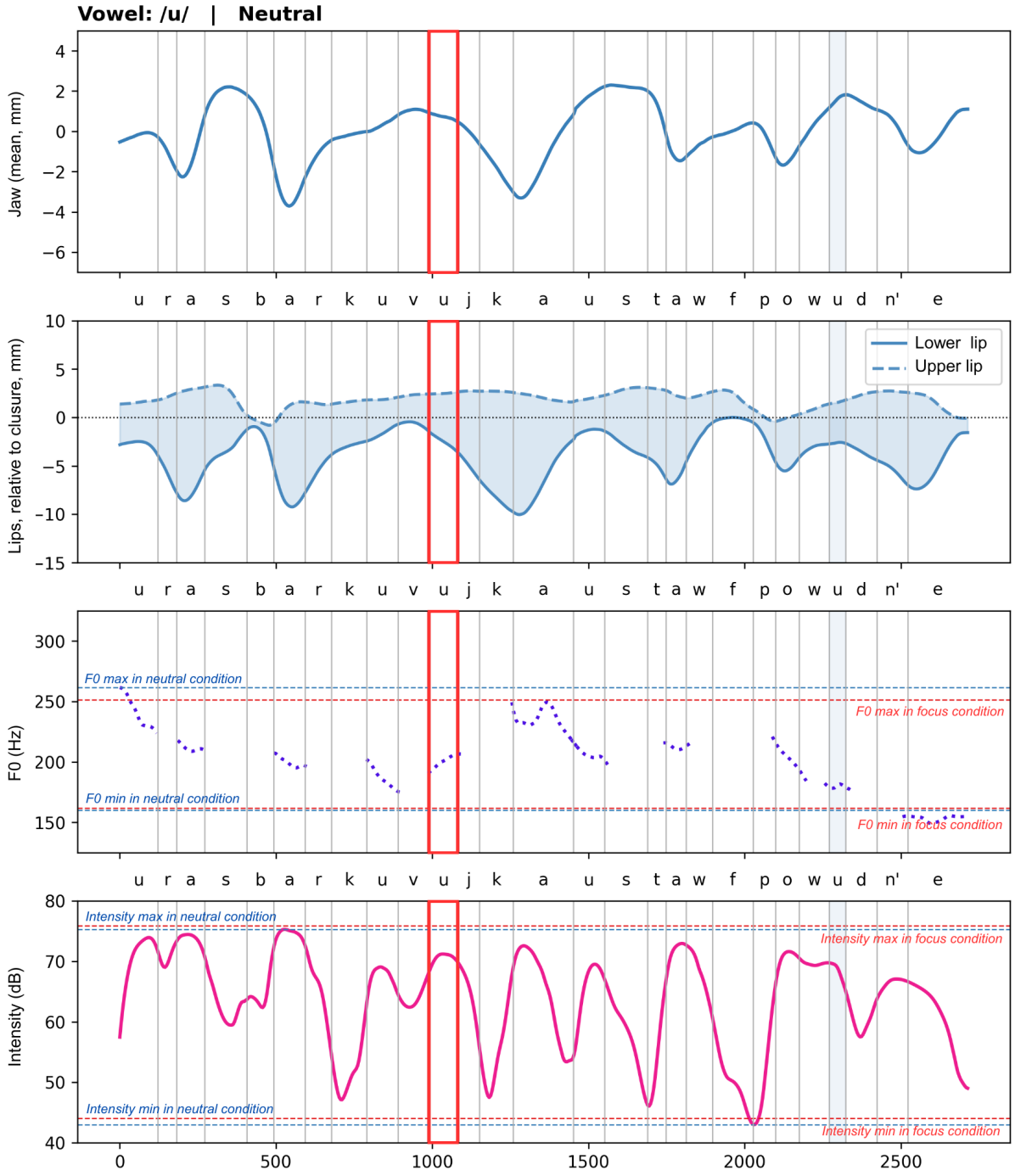

Figure 4.56. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Uraz barku wujka ustał w południe* /uras barku vujka ustaw fpowudn'e/ (Eng. 'Uncle's shoulder injury cleared up at noon.'), featuring /u/ as the target vowel. Neutral condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

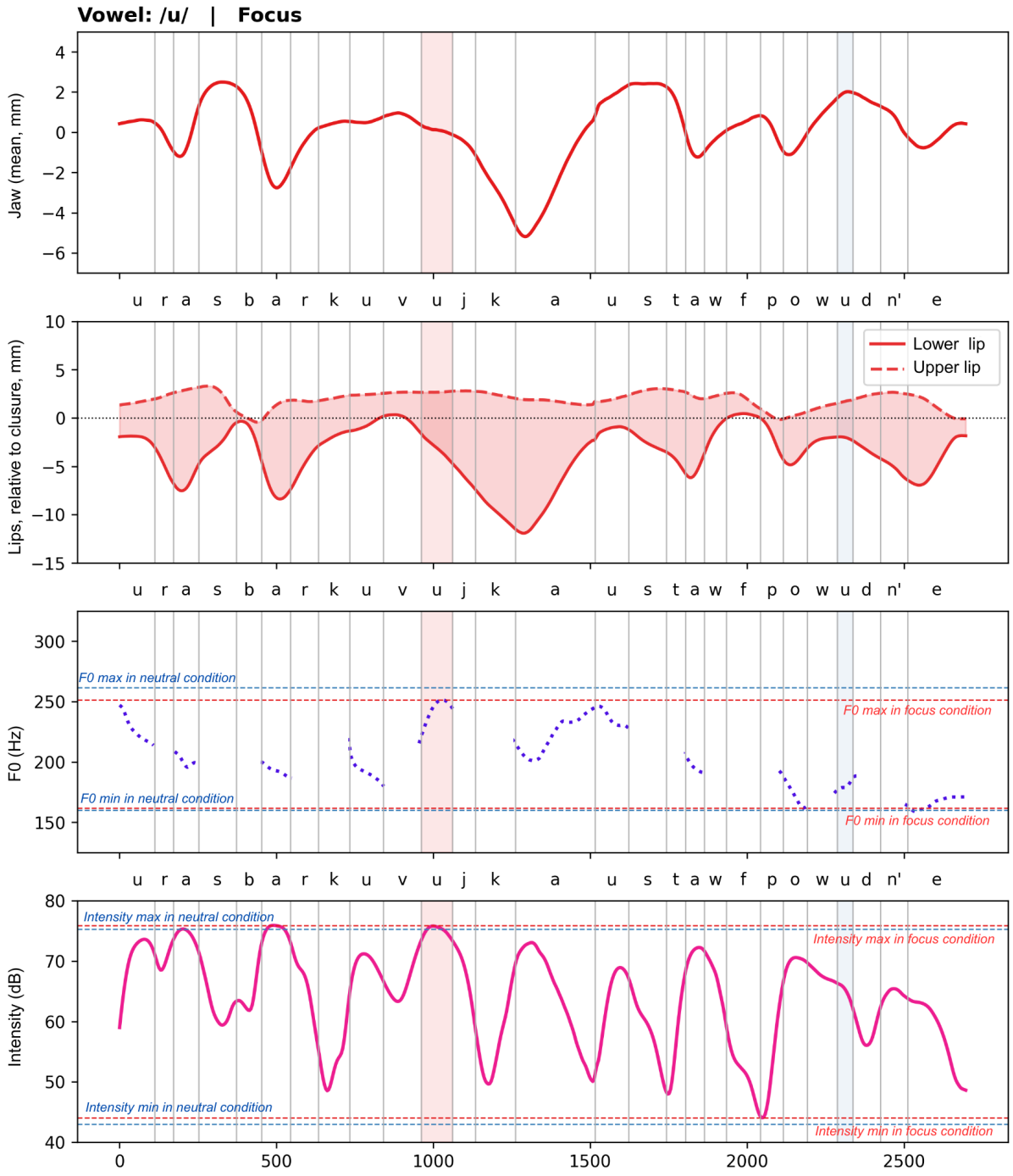

Figure 4.57. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Uraz barku wujka ustał w południe* /uras barku vujka ustaw fpowudn'e/ (Eng. 'Uncle's shoulder injury cleared up at noon.'), featuring /o/ as the target vowel. Focus condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

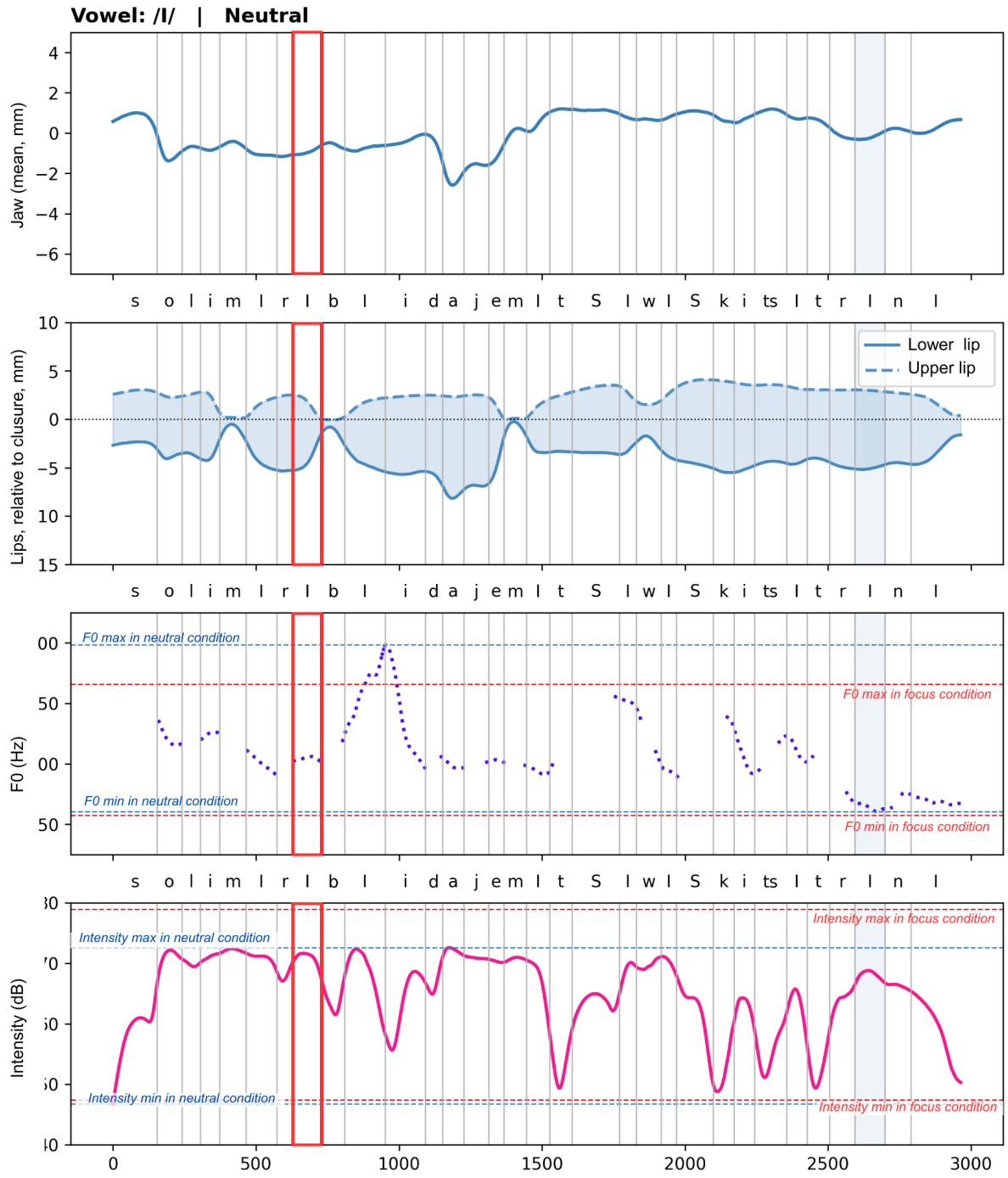

Figure 4.58. Time-aligned articulatory and acoustic trajectories for the realisations of the Polish utterance *Solimy ryby i dajemy trzy łyżki cytryny* /solimI rIbI i dajemI tSI wISki tˆsItrInI/ (Eng. 'Season the fish with salt and add three tablespoons of lemon juice.'), featuring /I/ as the target vowel. Neutral condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

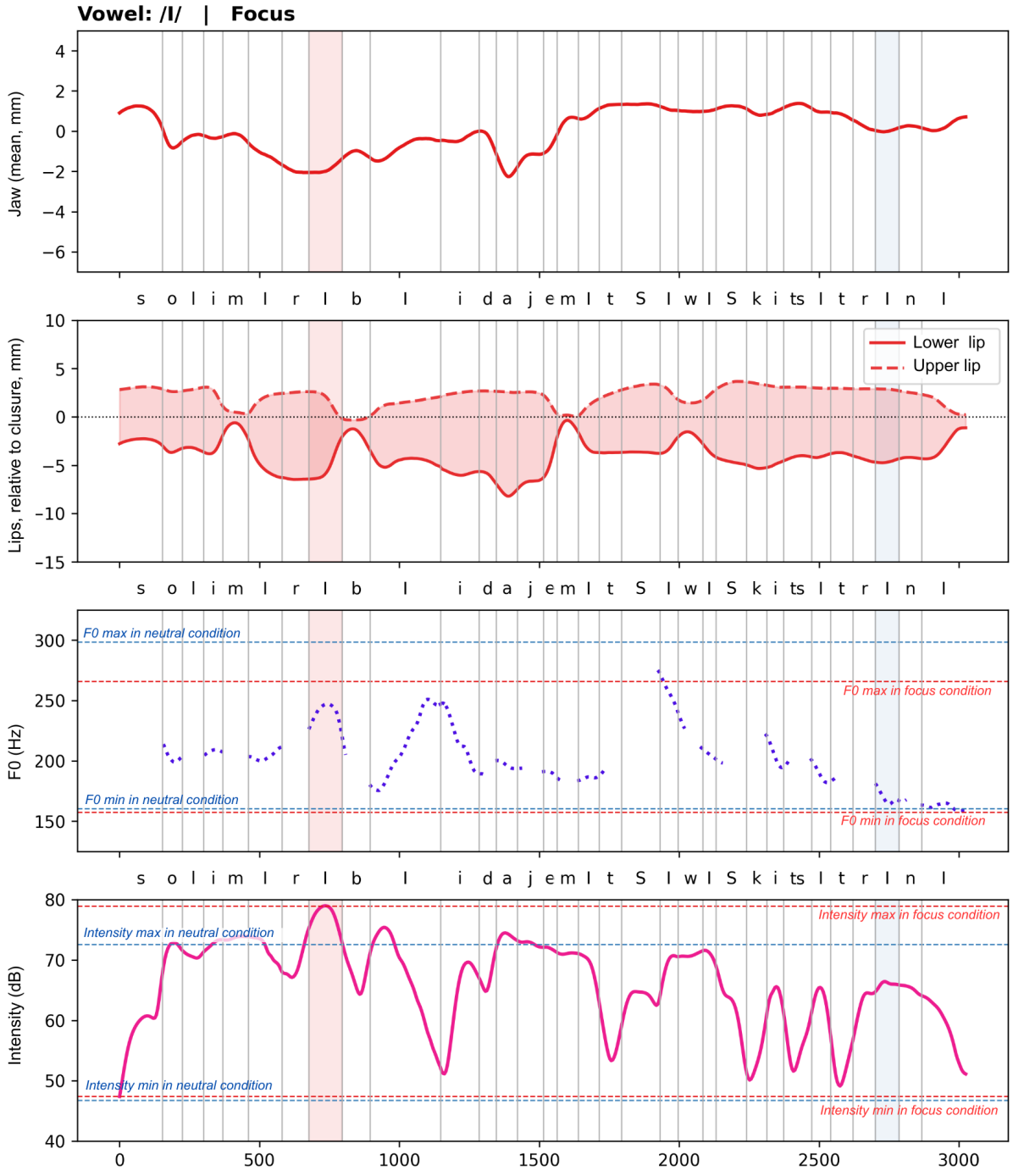

Figure 4.59. Time-aligned articulatory and acoustic trajectories for realisations of the Polish utterance *Solimy ryby i dajemy trzy łyżki cytryny* /solimI rIbI i dajemI tSI wISki tˆsItrInI/ (Eng. 'Season the fish with salt and add three tablespoons of lemon juice.'), featuring /I/ as the target vowel. Focus condition. The panels display jaw displacement, upper and lower lip movements relative to lip closure, F0 contours, and intensity profiles.

phonemic segment is acoustically marked, while the lips introduce their own articulatory adjustments. The lower lip largely follows the jaw but adds separate amplitude or compensatory movements. Here for lips only vertical displacement was studied, it is expected though, that at least in rounded vowels /o/ and /u/ some horizontal adjustments would be present as a result of lip rounding. The upper lip often operates mostly independently, matching the coordinated gesture. In **neutral** conditions, these subsystems show clearer phase shifts. Both articulatory measures responded systematically to prosodic prominence, displaying complementary adjustment profiles. As lip aperture expanded with prominence, jaw displacement increased in downward direction, suggesting a coordinated articulatory mechanism rather than isolated effects. Also, under **focus**, jaw and lips timing converges more, resulting in tighter synchrony between mandibular and labial gestures.

This account could be further elaborated in the light of recent findings by Svensson-Lundmark and Erickson (2024), who emphasise the different role of the articulators. By employing Descriptive Approach to Segmental Articulations (DASA) it was demonstrated that while the lips or the tongue tip synchronize closely with acoustic onsets and offsets of phonemic segments, the jaw functions more independently, marking the oscillatory basis of syllables; the study was conducted for Swedish.

Applying a similar approach to Polish would be a valuable complement to current findings. In particular, a gesture constellation analysis could reveal the distinct contributions of the jaw and lips and compare their coordination in two conditions. While promising, such an analysis extends beyond the scope of the present dissertation and represents a direction for possible further research.

4.8 Hierarchical articulatory gradient of prominence

Building on the observations from the previous section, the data can be interpreted as revealing a clear pattern of articulatory gradient dependent on position: *Other* < *Accent* < *Focus*, observed consistently for both jaw displacement and lip aperture measures.

Jaw displacement increased significantly for *Accent* positions (-0.383 SD) and especially for *Focus* positions (-1.172 SD), $p < .001$; Lip aperture showed analogous increases: $+0.513$ SD for *Accent* and $+1.119$ SD for *Focus*, $p < .001$. These effects are independent of vowel duration, implying that prominence operates via direct articulatory enhancement, not only by just increasing duration.

This hierarchical gradient can be interpreted through several complementary theoretical frameworks. From an articulatory effort perspective, the gradient may suggest that speakers tend to calibrate their 'investment' according to communicative needs, with focus requiring the greatest perceptual salience and thus receiving the strongest articulatory realisation. This would align with the findings on effects of presence of contrastive focus. As shown in the section *Focus influence on global prosodic pattern*, jaw displacement in no-*Focus* positions is reduced. This, however, is not statistically supported for lip aperture, even though such a tendency is visible. The findings also fit with the Hyper- and Hypo-articulation theory (H&H) theory (Lindblom, 1990) which proposes that speakers adjust their articulatory clarity based on the informational

needs of a listener. In the context of contrastive focus, this would mean that speakers tend to hyper-articulate — by increased distinctiveness of articulatory gestures — elements of interest to provide maximum acoustic detail, while reducing articulatory effort on non-*Focus* positions, effectively 'shifting' articulatory resources towards the central element to enhance communicative clarity.

The findings for *Accent* position provide additional support for this resource allocation framework. *Accent* position showed greater articulatory gradient compared to *Other* positions, consistent with Ćwiek and Wagner (2018) findings that Polish acoustic prominence is reliably marked at the phrase-level accent rather than at lexical stress positions. Given that Polish lexical stress lacks distinctive function and serves mainly to delimit word boundaries, speakers might economize their articulatory effort by directing it towards communicatively more significant utterance-level prominence, manifesting as hyperarticulation in *Accent* positions relative to *Other* positions. Polish, with its fairly fixed utterance-level accent, may allocate proportionally more articulatory resources to contrastive focus than languages where lexical stress carries greater informational load.

Beyond the articulatory effort and H&H theory accounts, the articulatory gradient might also be interpreted within metrical stress theory (Hayes, 1995; Liberman & Prince, 1977). The systematic ordering of *Other* < *Accent* < *Focus* reflects the hierarchical organisation of prosodic structure, in which *Accent* positions serve as phrase-level heads, and focus introduces a superordinate level of prominence. Figure 4.60 depicts how successive metrical levels might interact in Polish.

3			|			|	*		|			|						
2			|			|	*		|			|			*		
1	*		|	*		|	*		|	*		|			*		
0	*	*	|	*	*	|	*	*	|	*	*	|	*		*	*		
syllable	u	raz	|	bar	ku	|	vuj	ka	|	u	staw	|	fpo		wu	dn'ie		

Figure 4.60. Example of possible hierarchical stress assignment in Polish. Level 0 — marks syllable positions only, level 1 — lexical stress, level 2 — phrasal final stress / utterance accent, and level 3 — focus stress.

The results may also be interpreted in light of Fujimura's Converter/Distributor model (Fujimura, 2000), which proposes that prominence is implemented through syllable pulses that distribute articulatory energy. The increased jaw displacement and lip aperture observed in focus positions, accompanied by a tendency towards reduced displacement in non-focus segments, are fairly compatible with this account: prominence appears to involve not only local articulatory enhancement but also a redistribution of articulatory resources across the utterance. Such a mechanism seems consistent with the present results, suggesting that vowels outside the focus position exhibit smaller jaw movements when a contrastive focus is present.

In terms of Articulatory Phonology (Browman & Goldstein, 1992), prominence may further be understood as modulation of gestural magnitude. The data for Polish indicate that jaw displacement and lip aperture expand under focus, which can be interpreted as an increase in the amplitude of opening gestures, resulting in greater durational and spatial distinctiveness of vowel articulation.

4.9 Cross-linguistic comparison

The articulatory patterns observed for Polish vowels under **focus** condition align with some cross-linguistic tendencies (see Chapter 1, Section 1.3.3 for details and references) while also exhibiting language-specific properties. Similar to English, increased jaw displacement in prominent positions was accompanied by acoustic reinforcement in F0 and intensity, suggesting that the jaw functions as an articulatory organiser of prominence. In contrast, however, to Mandarin Chinese and French, where jaw displacement is anchored predominantly at phrase edges, Polish prominence appears to operate at the level of contrastive focus irrespective of phrasal position. Compared with Brazilian Portuguese, Polish displays a weaker and largely predictable lexical stress. In both languages declarative utterances are typically associated with a final falling F0 contour, yet the main place of prominence differs: in Polish it is enhanced at the phrasal level, with contrastive focus yielding the strongest articulatory–acoustic convergence, whereas in Brazilian Portuguese prominence is already robustly cued by lexical stress. Both languages thus belong to the head-prominence type, but they diverge in whether prominence is anchored primarily in lexical stress or in higher-level accents.

Therefore, from a cross-linguistic perspective, the findings are consistent with typological observations on prosodic systems (Jun, 2014). In Polish, where lexical stress is mostly fixed and informationally weak, utterance-level accent and, especially, contrastive focus provide the primary cues to prominence. Within Jun's typology, Polish could be classified amongst head-prominence languages with weak lexical stress, a configuration that predicts stronger reliance on phrasal-level marking and, in many cases, a tendency towards more regularised macro-rhythm. With respect to F0, the most frequently realised contour in Polish is flat (F), alongside common occurrences of H_H (cf. Demenko & Oleśkowicz-Popiel, 2016). A tentative classification as medium macro-rhythm therefore seems appropriate: languages that predominantly employ level contours are expected to show weaker macro-rhythm than those characterised by alternating rising or falling patterns typical of the strong group. At the same time, the necessity of using phrasal cues — such as enhanced jaw and lip movements, increased F0, greater intensity, vowel lengthening — to signal focus suggests that the prosodic rhythm of Polish is more structured than would be expected for the weak group.

4.10 Remarks on the difference between utterance-level accent and contrastive focus stress

In the present data, utterance-level accent in Polish appears to be realised primarily through articulatory adjustments, most notably increased jaw displacement and wider lip aperture, accompanied by the vowel lengthening in the *Accent* position. By contrast, contrastive focus emerges as more multi-dimensional: alongside articulatory expansion, consistent acoustic correlates are also observed.

From a functional perspective, these findings resonate with earlier descriptions of Polish lexical stress as primarily delimitative, marking the approach of a phrase boundary or supporting the parsing of lexical units (Jassem, 1962; Ostaszewska & Tambor, 2000; Wierzchowska, 1967).

It may be suggested that this delimitative role extends beyond the lexical level, shaping prominence patterns at the phrasal or utterance level as well. As for the contrastive focus,

Contrastive focus seems to fulfill a different communicative function. Rather than marking boundaries, it highlights new or contrastive information in discourse. Therefore, its realisation appears richer: articulatory expansion is combined with acoustic enhancement, which might increase perceptual salience and help the central element stand out from the backdrop. This pattern can be related to functional accounts of speech production (e.g., Lindblom, 1990), where speakers are assumed to vary the resources they use depending on whether the aim is structural demarcation or informational highlighting. From this perspective, both utterance-level accent and contrastive focus rely on articulatory correlates, but only the latter consistently recruits additional acoustic cues.

The different communicative functions of utterance-level accent and contrastive focus may also help explain the observed difference in rhythmic organisations. Accent seems to serve primarily a rhythmic–structural role, guiding the listener through segmentation and delimitation of the speech stream. By contrast, focus fulfills an informational role, drawing attention to new or contrastive material in discourse.

4.11 Remarks on jaw drift

An additional point concerns the slow drift observable in jaw trajectories (see Chapter 3, Section *3.4.1.5*). Such gradual lowering was not directly addressed in reviewed literature and might bring possible new explanations to already observed phenomena as it coincides with the declination pattern described in F0 contours (cf. Ladd, 2008). Declination is a gradual downward trend of phonetic-acoustic parameters that accompanies the progression of an utterance; it is observable in Polish data as well with descending intensity and F0 contour (see plots in section Jaw and lip movements with corresponding F0 and intensity profiles). Declination might be the answer to why F0 in *Accent* position is lower than in *Other* positions (see Section *4.5.2*). Possibly, jaw drift might be declination's articulatory correlate.

This type of global trend may arise from different sources. One explanation is physiological: as the utterance progresses, jaw posture relaxes because of muscular fatigue. The present dataset is well suited to examine this issue more directly in the future, as it provides sufficient articulatory data collected while working on this dissertation. The naming convention of the recordings, which reflects their sequential order within a session, allows for within-speaker comparisons of jaw detrending across multiple takes.

4.12 Summary of hypotheses

The analyses addressed the hypotheses formulated in Chapter3 (Section 3.2.2). Their outcomes can be summarised as follows:

Effects of *Accent* position

H$_0$: Utterance-level accent does not affect vertical jaw displacement.

H$_1$: Utterance-level accent increases vertical jaw displacement.

The null hypothesis H_0 was rejected. Utterance-level accent significantly affected vertical jaw displacement, with accented vowels showing greater lowering than non-prominent ones.

H$_2$: Utterance-level accent does not affect lip aperture.
H$_3$: Utterance-level accent increases lip aperture.

The null hypothesis H_2 was rejected. Utterance-level accent had a systematic effect on lip aperture.

Effects of *Focus* position

H$_4$: Contrastive focus does not affect vertical jaw displacement.
H$_5$: Contrastive focus increases vertical jaw displacement.

The null hypothesis H_4 was rejected. Contrastive focus significantly affected vertical jaw displacement.

H$_6$: Contrastive focus does not affect lip aperture.
H$_7$: Contrastive focus increases lip aperture.

The null hypothesis H_6 was rejected. Presence of contrastive focus does not systematically alter lip aperture during vowel production.

Global influence of focus condition

H$_8$: The presence of contrastive focus in an utterance has no effect on jaw displacement in non-*Focus* position vowels.
H$_9$: The presence of contrastive focus in an utterance reduces jaw displacement in non-*Focus* position vowels.

The null hypothesis H_8 was rejected. Contrastive focus did affect jaw displacement in non-*Focus* position vowels.

H$_{10}$: The presence of contrastive focus in an utterance has no effect on lip aperture in non-*Focus* position vowels.
H$_{11}$: The presence of contrastive focus in an utterance reduces lip aperture in non-*Focus* position vowels.

The null hypothesis H_{10} was accepted. Focus did not systematically influenced lip aperture outside the focal position.

Duration-controlled effects

H$_{12}$: When vowel duration is controlled for, contrastive focus has no significant effect on jaw displacement.
H$_{13}$: Contrastive focus affects jaw displacement independently of vowel duration.

The null hypothesis H_{12} was rejected. Focus remained a significant predictor of jaw displacement even when vowel duration was controlled.

H$_{14}$: When vowel duration is controlled for, utterance-level accent has no significant effect on

jaw displacement.
H$_{15}$: Utterance-level accent significantly affects jaw displacement, beyond what is explained by vowel duration.

The null hypothesis H$_{14}$ was rejected. Accent continued to significantly influence jaw displacement when controlling for duration.

Chapter 5

Synthesis of findings and research outlook

5.1 Key findings

Bringing the analyses together, several consistent tendencies can now be identified. They are presented below as the central findings of this study.

Hierarchical prominence. The data reveal a gradient: vowels in non-prominent positions are closer to the articulatory baseline, whereas *Accent* position enhances articulation to a moderate degree, and contrastive *Focus* position elicits the strongest adjustments which are present in both.

Articulatory correlation of *Accent* and *Focus* position. Both *Accent* and contrastive *Focus* position rely on articulatory reinforcement. Increased jaw displacement and wider lip aperture were robustly associated with prominent positions, which might confirm that articulatory displacement serves as a stable marker of prominence across speakers and vowel categories.

Acoustic enrichment under contrastive focus. In addition to articulation, contrastive *Focus* position consistently mobilised phonetic-acoustic resources, such as raised F0, increased intensity, and systematic lengthening of the final vowel of the word under contrastive focus. These parameters enhance perceptual salience and new/important information apart from utterance-level accent, which remained primarily articulatory in character.

Global redistribution. Contrastive focus affected not only the target vowels but also the surrounding material, reducing displacement in non-*Focus* positions. Such redistribution points to a global reorganisation of articulatory dynamics in the presence of contrastive focus. This was reflected in articulatory changes within the utterances (diminished jaw displacement and lip aperture, the latter however not being statistically significant), rhythmic organisations of utterances, their greater durational variability with respect to both vowel and consonant segments, and changes of the F0 contour.

Novel findings. Two additional observed tendencies may be of interest for further research. First, detrending revealed traces of slow vertical jaw drift towards the end of an utterance, which bears some resemblance to declination in the F0 contour (cf. Ladd, 1984) and could represent an articulatory correlate of global prosodic trends. Second, subtle protrusion effects were noted in jaw movements, a dimension not often addressed in prominence studies, yet

potentially relevant for a fuller account of multimodal prominence marking.

Beyond these empirical insights, the dissertation also advances methodological practice. By combining rhythm metrics with articulatory data, it establishes a framework for studying prominence in a language with both predictable stress and ambiguous rhythmic classification. The development of analysis and visualisation procedures and techniques for automatic tracking of jaw displacement and lip aperture — together with their integration with acoustic measures — represents a language-independent methodological contribution that might be adapted in future research on other datasets and prosodic phenomena.

5.2 Limitations

Several limitations should be acknowledged. Firstly, the main data were collected in a laboratory setting using electromagnetic articulography. While this ensured high temporal and spatial precision, it may have constrained the natural variability of spontaneous speech prosody.

Second, the analysis revealed differential patterns across vowel categories: low and mid vowels demonstrated the most pronounced prominence effects, whereas high vowels exhibited greater articulatory stability. This phenomenon likely reflects both the biomechanical constraints of the speech production system and the specific distributional characteristics of tokens within the analysed corpus.

Finally, the study focused mainly on articulatory and additionally on phonetic-acoustic dimensions. Other channels of prominence marking, such as visual gestures or aerodynamic cues, and perceptual features were not included, though they may play an important role in natural interaction.

The limitations outlined above suggest several important directions for future research.

5.3 Future directions

The comprehensive dataset and methodological framework established in this study create substantial opportunities for further investigation. Several particularly promising research pathways emerge from the current findings.

The recorded trajectories enable examination of global speech patterns that extend beyond individual segments. Phenomena such as jaw drift throughout utterances may reveal systematic relationships with prosodic declination, offering insights into the broader organisational principles governing connected speech production.

Another promising direction for future work involves analyses based on data obtained from the remaining EMA sensors, already recorded during the present experimental sessions. In particular, tongue sensor data could provide new insights into the realisation of Polish vowels in both the vertical and front–back dimensions under the currently proposed experimental conditions.

The preliminary observations of jaw protrusion effects call for more systematic investigation,

including possible labial shift. Also, study of coordination between the jaw and labial movements employing Descriptive Approach to Segmental Articulations (DASA) methodology (Svensson-Lundmark & Erickson, 2024) would possibly bring new insights on how utterances develop in time and how articulators cooperate in shaping them, especially under varied experimental conditions.

Methodologically, extending this analytical framework to more naturalistic speech contexts represents a crucial next step. Investigating whether the prominence strategies documented here operate consistently across semi-spontaneous conversational speech would significantly enhance the ecological validity of these findings and bridge laboratory observations with real-world communicative behaviour. If not in a more natural setting, collecting acoustic data samples and comparing phonetic-acoustic results between EMA and non-EMA situations would also complement current findings.

To supplement contrastive **focus** condition findings, the same stimuli and instructions could be applied with changing the word in focus. Such modification would help in determining whether observed patterns are indeed *Focus* position specific, or different aspects contribute to the effect described in the findings.

From a comparative linguistic perspective, parallel investigations across related languages — particularly, though not exclusively, within the Slavic family — could determine whether the functional distinction between **accent** and **focus** prominence observed in Polish reflects language-specific characteristics or broader typological universals. Such cross-linguistic validation would strengthen theoretical claims about prominence marking systems.

Perhaps most significantly, future work should pursue multimodal integration by simultaneously analysing articulatory, acoustic, visual, and perceptual prominence channels. This comprehensive approach would advance our understanding of how different communicative modalities coordinate to signal and perceive prominence, ultimately contributing to more complete models of speech production under various conditions.

Funding

This dissertation was supported by two research grants:

- Jaw position as an articulatory correlate of speech rhythm in spoken Polish (*Pozycja żuchwy jako artykulacyjny korelat rytmu w mówionej polszczyźnie*). ID-UB Project Excellence Initiative (PhD Student Minigrant) for the academic year 2022/2023 at Adam Mickiewicz University, Poznań. Grant No. 054/13/SNJL/0002.
- Examination of disordered speech and primary functions using an articulograph CARSTENS AG501 and Acoustic Field Distribution Analyser (*Badanie mowy zaburzonej i funkcji prymarnych za pomocą artykulografu CARSTENS AG501 i analizatora rozkładu pola akustycznego*). The Polish National Science Centre, Grant No. 2021/43/B/HS2/00162.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Katarzyna Klessa and Prof. Anita Lorenc. They have not only supervised my progress on this dissertation but also encouraged me to pursue a new approach to data preparation. My ideas and insights were always met with openness and thoughtful discussion, which fostered an environment of creativity and intellectual growth. Their constructive feedback, patience, and trust have been invaluable, allowing me to grow as a researcher and to shape my own methodological framework.

I would also like to thank the members of the Department of Multimodal Communication at Adam Mickiewicz University in Poznań and the wider academic community, particularly those who took a kind interest in the progress of my work and responded with understanding to my doubts, concerns, and anxieties related to writing. Special thanks to Maciej, Brygida, and Ewa.

There is also a special kind of thanks for my closest companion, family, and friends, who experienced firsthand the daily pressures of my writing and offered patience, encouragement, and support throughout.

It is important to note that my studies in the Doctoral School of Languages and Literatures coincided with an exceptionally challenging period — the COVID-19 pandemic and the resulting isolation, as well as the escalation of the war in Ukraine. The support, kindness, and encouragement I received were of particular significance to me during these difficult circumstances.

Written acknowledgements have never been my strong suit, as I try to express gratitude on an ongoing, day-to-day basis. I therefore hope that my gratitude, though modest in form, is felt as fully as I wish to convey it.

Appendix

A Recording procedure scenario

A.1 Instruction Set I (Neutral Utterances)

On the computer screen, short phrases will appear one by one. Each phrase will be visible for a short while and then disappear. At that moment, please pronounce the phrase aloud naturally, just as you normally speak.

After you do so, a message will appear: *The next text will appear on the screen shortly.* Then the next phrase will follow — and so on until the end.

If at any point you need a break or some rest, please signal this to the researcher. You may pause and rest at any time during the experiment.

A.2 Instruction Set II (With Contrastive Focus)

Short phrases will again appear on the computer screen, one by one. Each phrase will be visible for a short while and then disappear. This time, some words in the phrases will appear in upper case, as in the examples below. This means that we ask you to read the word in a way that emphasises it:

- *Yes, I see that* MAN *in the coat.* → Here you emphasise that it is that particular man, and not someone else, e.g., when searching for him amongst other people.
- *Yes, I* SEE *that man in the coat.* → You emphasise that you actually see him with your own eyes, and not, for example, guess that he is somewhere around or hear him.

At that moment, please pronounce the phrase aloud naturally, emphasising the capitalised word as indicated.

After you do so, a message will appear: *The next text will appear on the screen shortly.* Then the next phrase will follow — and so on until the end.

If at any point you need a break or some rest, please signal this to the researcher. You may pause and rest at any time during the experiment.

B Table with phonemic indices

Vowel	Stimuli	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
/a/	Ala dała Arkowi zapas buraków	a	l	a	d	a	w	a	a	r	k	o	v	i	z	a	p	a	s	b	u	r	a	k	u	f							
/e/	Edek jedzie do Łeby, jeszcze bez adresu	e	d	e	k	j	e	dz'	e	d	o	w	e	b	I	j	e	S	tS	e	b	e	s	a	d	r	e	s	u				
/i/	Irek widzi kilka irysów na stoliku	i	r	e	k	v	i	dz'	i	k	i	l	k	a	i	r	I	s	u	f	n	a	s	t	o	l	i	k	u				
/o/	Ogon kota opadł po skoku na Karola	o	g	o	n	k	o	t	a	o	p	a	d	w	p	o	s	k	o	k	u	n	a	k	a	r	o	l	a				
/u/	Uraz barku wujka ustał w południe	u	r	a	z	b	a	r	k	u	v	u	j	k	a	u	s	t	a	w	f	p	o	w	u	d	n'	e					
/I/	Solimy ryby i dajemy trzy łyżki cytryny	s	o	l	i	m	I	r	I	b	I	i	d	a	j	e	m	I	t	S	I	w	I	S	k	i	ts	I	t	r	l	n	I

Figure 1. Table with phonemic indices showing segment boundaries for target stimuli

C Specification of sensors

Sensor No.	Location	Cable Direction	Calibration Set
16A	Right mastoid process	Cable downward	A
15A	Left mastoid process	Cable downward	A
14A	Nose control	On thumbs	A
1B	Bite plane - left	Cable outside oral cavity	B
2B	Bite plane - right	Cable outside oral cavity	B
3B	Bite plane - central	Cable outside oral cavity	B
13A	Center of upper lip, just above vermilion	Cable upward	A
12A	Center of lower lip, just below vermilion	Cable downward	A
11A	Right corner of mouth	Cable to the right	A
10A	Left corner of mouth	Cable to the left	A
9A	At the border of gum and lower incisors	Cable upward	A
8A	Tongue tip (TT)	Cable forward (not left, not right)	A
7A	Tongue back (TB)	Cable to the left	A
6A	Tongue dorsum (TD)	Cable to the left	A
5A	Tongue front (TF)	Cable to the left	A
4A	Right anterior tongue edge (at TF level)	Cable to the left	A
3A	Left anterior tongue edge (at TF level)	Cable to the left	A
2A	Right posterior tongue edge (at TB level)	Cable to the left	A
1A	* First: Outlines * Then: Left posterior tongue edge (at TB level)	Cable to the left	A

Table 1. Set A as Primary

D Supplementary LMEM Models Results

The following section presents the detailed outputs of linear mixed-effects models (LMEMs) fitted by restricted maximum likelihood (REML) using the `lmerTest` package in R. All models included Subject as a random intercept. Fixed effects comprised VowelPosition (*Accent*, *Focus*, *Other*), Focus (**focus, neutral**), and Vowel categories, with some models additionally including Duration.

D.1 Model A2. Jaw displacement (excl. subject UOKV)

Data: `ema_data_noUOKV`
N = 1479 observations, 5 speakers
REML criterion = 3945

D.1.1 Random effects

- Subject Var = 0.0035 (SD = 0.060)
- Residual Var = 0.836 (SD = 0.915)

D.1.2 Fixed effects

Predictor	Estimate	Std. Error	df	t value	p value
(Intercept)	0.0089	0.0381	5.36	0.234	0.824
VowelPosition: *Accent*	−0.291	0.0674	1473	−4.320	1.7×10^{-5}***
VowelPosition: *Focus*	−1.333	0.0941	1474	−14.168	$< 2 \times 10^{-16}$***

D.2 Model B2. Lip aperture (excl. subject SLDT)

Data: `ema_data_noSLDT`
N = 1434 observations, 5 speakers
REML criterion = 4138

D.2.1 Random effects

- Subject Var = 0.0070 (SD = 0.084)
- Residual Var = 1.040 (SD = 1.020)

D.2.2 Fixed effects

Predictor	Estimate	Std. Error	df	t value	p value
(Intercept)	0.024	0.0486	4.34	0.499	0.642
VowelPosition: *Accent*	0.626	0.0764	1427	8.199	5.4×10^{-16}***
VowelPosition: *Focus*	0.890	0.1075	1428	8.284	2.7×10^{-16}***

E Summary of slopes and intercepts for individual speakers

Subject	Vowel	Focus	N	Slope	Intercept
OKPC	A	Neutral	18	0.578	−0.300
OKPC	A	Focus	16	1.036	−0.173
OKPC	E	Neutral	24	0.500	0.061
OKPC	E	Focus	21	0.541	0.307
OKPC	I	Neutral	15	0.451	0.179
OKPC	I	Focus	15	0.891	0.189
OKPC	O	Neutral	18	0.782	−0.271
OKPC	O	Focus	11	0.999	−0.264
OKPC	U	Neutral	12	0.328	0.449
OKPC	U	Focus	12	0.453	0.586
OKPC	Y	Neutral	36	0.513	0.053
OKPC	Y	Focus	27	0.661	0.057
RHAK	A	Neutral	40	0.565	−0.792
RHAK	A	Focus	19	0.681	−0.479
RHAK	E	Neutral	64	0.215	−0.238
RHAK	E	Focus	32	−0.116	−0.056
RHAK	I	Neutral	35	0.368	0.208
RHAK	I	Focus	20	0.512	0.149
RHAK	O	Neutral	40	0.285	−0.192
RHAK	O	Focus	21	0.647	−0.238
RHAK	U	Neutral	25	0.223	0.723
RHAK	U	Focus	16	0.241	0.840
RHAK	Y	Neutral	71	−0.240	0.108
RHAK	Y	Focus	36	−0.334	0.196
UOKV	A	Neutral	18	0.554	−0.712
UOKV	A	Focus	24	0.204	−0.942
UOKV	E	Neutral	40	0.329	−0.256
UOKV	E	Focus	24	−0.048	−0.082
UOKV	I	Neutral	15	0.428	0.453
UOKV	I	Focus	15	1.065	0.515
UOKV	O	Neutral	14	0.242	−0.198
UOKV	O	Focus	21	−0.161	0.000
UOKV	U	Neutral	12	0.039	0.375
UOKV	U	Focus	8	−0.142	0.275
UOKV	Y	Neutral	36	−0.188	0.060
UOKV	Y	Focus	27	−0.236	0.089

Table 2. Slopes and intercepts for subjects OKPC, RHAK, and UOKV.

Subject	Vowel	Focus	N	Slope	Intercept
VTVK	A	Neutral	21	0.721	−0.330
VTVK	A	Focus	21	0.994	−0.424
VTVK	E	Neutral	22	0.682	−0.180
VTVK	E	Focus	23	0.592	−0.135
VTVK	I	Neutral	14	0.359	0.440
VTVK	I	Focus	15	0.630	0.335
VTVK	O	Neutral	18	0.523	−0.166
VTVK	O	Focus	18	0.731	−0.233
VTVK	U	Neutral	12	0.554	0.368
VTVK	U	Focus	12	0.453	0.380
VTVK	Y	Neutral	27	0.667	−0.065
VTVK	Y	Focus	27	0.746	−0.117
SLDT	A	Neutral	28	0.900	0.117
SLDT	A	Focus	22	0.905	0.017
SLDT	E	Neutral	32	0.524	−0.027
SLDT	E	Focus	28	0.426	0.346
SLDT	I	Neutral	25	0.445	−0.069
SLDT	I	Focus	20	0.801	0.038
SLDT	O	Neutral	24	0.673	0.040
SLDT	O	Focus	24	1.120	−0.072
SLDT	U	Neutral	12	0.416	0.490
SLDT	U	Focus	16	0.428	0.475
SLDT	Y	Neutral	36	0.360	0.137
SLDT	Y	Focus	36	0.462	0.054
RLRG	A	Neutral	20	0.311	−0.864
RLRG	A	Focus	24	0.330	−0.606
RLRG	E	Neutral	24	0.367	−0.126
RLRG	E	Focus	28	0.194	−0.148
RLRG	I	Neutral	15	0.282	0.290
RLRG	I	Focus	15	0.542	−0.041
RLRG	O	Neutral	18	0.426	0.156
RLRG	O	Focus	17	0.595	0.172
RLRG	U	Neutral	12	0.123	0.582
RLRG	U	Focus	9	0.894	0.700
RLRG	Y	Neutral	26	−0.081	0.053
RLRG	Y	Focus	27	−0.244	−0.154

Table 3. Slopes and intercepts for subjects VTVK, SLDT, and RLRG.

List of Figures

List of Tables

Bibliography

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press.

Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107.

Al Bawab, Z., Raj, B., & Stern, R. M. (2008). Analysis-by-synthesis features for speech recognition. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4185–4188.

Articulograph. (n.d.). The AG501. Retrieved September 27, 2025, from https://www.articulograph.de/articulograph-head-menue/the-ag501/

Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66*, 46–63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Audacity Team. (2023). Audacity (Version 3.3.3) [Computer software].

Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction*. Oxford University Press.

Barbosa, P. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, *49*(9), 725–742.

Barbosa, P., & Albano, E. (2004). Brazilian Portuguese. *Journal of the International Phonetic Association*, *34*(2), 227–232.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? *Proceedings of the 15th ICPhS*, 2693–2696.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, *26*(1), 86–125.

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer (Version 6.3.09) [Computer software].

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An overview. *Phonetica*, *49*(3-4), 155–180.

Browman, C. P., & Goldstein, L. M. (1986). Towards an Articulatory Phonology. *Phonology*, *3*, 219–252.

Byrd, D., Browman, C. P., Goldstein, L., & Honorof, D. (1999). Magnetometer and X-ray microbeam comparison. *Proceedings of the 14th International Congress of Phonetic Sciences*, 627–630.

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, *75*, 113–147.

Crosswhite, K. (2003). Spectral tilt as a cue to word stress in Macedonian, Polish, and Bulgarian. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 767–770).

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171.

Cutler, A. (1984). Stress and accent in language production and understanding. In D. Gibbon & H. Richter (Eds.), *Intonation, accent and rhythm: Studies in discourse phonology* (pp. 76–90, Vol. 8). De Gruyter.

Ćwiek, A., & Wagner, P. (2018). The acoustic realization of prosodic prominence in Polish: Word-level stress and phrase-level accent. *Proceedings of Speech Prosody 2018*, 922–926.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences*, *5*, 447–450.

De Leeuw, E., Chang, C. B., & Amengual, M. (2023). Phonetic and phonological L1 attrition and drift in bilingual speech. *The Cambridge Handbook of Bilingual Phonetics and Phonology*.

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(1), 33–40.

Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and speech rhythm* (pp. 231–241). Peter Lang.

Dellwo, V., Wagner, P., Solé, M.-J., Recasens, D., & Romero, J. (2003). Relations between language rhythm and speech rate. *Proceedings of the International Congress of Phonetic Sciences*, 471–474.

Demenko, G., & Oleśkowicz-Popiel, M. (2016). Automatic pitch accent annotation. *Proceedings of the International Conference on Speech Prosody*, *2016*.

Dłuska, M. (1950). *Fonetyka polska*. Państwowe Wydawnictwo Naukowe.

Dłuska, M. (1976). *Prozodia języka polskiego* (2nd ed.). Państwowe Wydawnictwo Naukowe.

Dogil, G. (1980). Autosegmental phonology in contrastive linguistics. In *Theoretical issues in contrastive linguistics* (p. 237, Vol. 12).

Dogil, G. (1999). The phonetic manifestation of word stress in Polish, Lithuanian, Spanish, and German. In H. van der Hulst (Ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 271–311). Mouton de Gruyter.

Domahs, U., Knaus, J., Orzechowska, P., & Wiese, R. (2012). Stress "deafness" in a language with fixed word stress: An ERP study on Polish. *Frontiers in Psychology*, *3*, 439.

Dromey, C., Hunter, E., & Nissen, S. L. (2018). Speech adaptation to kinematic recording sensors: Perceptual and acoustic findings. *Journal of Speech, Language, and Hearing Research*, *61*(3), 593–603.

Erickson, D. (1998). Effects of contrastive emphasis on jaw opening. *Phonetica*, *55*(3), 147–169. https://doi.org/10.1159/000028429

Erickson, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica*, *59*(2-3), 134–149. https://doi.org/10.1159/000066065

Erickson, D. (2004). On phrasal organization and jaw opening. *Proceedings of From Sound to Sense*, 24.

Erickson, D. (2006). An articulatory account of rhythm, prominence and phrasal organization. *Proceedings of Speech Prosody*.

Erickson, D., & Fujimura, O. (1996). Maximum jaw displacement in contrastive emphasis. *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP '96)*, *1*, 141–144.

Erickson, D., Fujimura, O., & Pardo, B. (1998). Articulatory correlates of prosodic control: Emotion and emphasis. *Language and Speech*, *41*(3-4), 399–417.

Erickson, D., Honda, K., & Kawahara, S. (2017). Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis. *Acoustical Science and Technology*, *38*(3), 137–146.

Erickson, D., Iwata, R., & Suemitsu, A. (2016). Jaw displacement and phrasal stress in Mandarin Chinese. *Proceedings of TAL 2016: The Sixth International Symposium on Tonal Aspects of Languages*.

Erickson, D., & Kawahara, S. (2016). Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistics Vanguard*, *2*(1), 20150025. https://doi.org/10.1515/lingvan-2015-0025

Erickson, D., & Niebuhr, O. (2023). Articulation of prosody and rhythm: Some possible applications to language teaching. *Proceedings of the 13th International Conference of Nordic Prosody*, 1–45.

Erickson, D., Rilliard, A., Lundmark, M. S., Silva, A., Couto, L. R., Niebuhr, O., & De Moraes, J. A. (2024). Collecting Mandible Movement in Brazilian Portuguese. *Proceedings of Interspeech 2024*, 3145–3149.

Erickson, D., Suemitsu, A., Shibuya, Y., & Tiede, M. (2012). Metrical structure and production of English rhythm. *Phonetica*, *69*(3), 180–190.

Erickson, D., Villegas, J., Wilson, I., & Iguro, Y. (2015). Spanish articulatory rhythm. *Acoustical Society of Japan Fall Meeting*, 319–322.

Eschenberg, A. (2008). Polish narrow focus constructions. In C. Lee & M. Gordon (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 23–40, Vol. 82). Springer Science & Business Media.

Ferragne, E., & Pellegrino, F. (2004). Rhythm in read British English: Interdialect variability. *Proceedings of Interspeech 2024*, 1573–1576.

Francuzik, K., Karpiński, M., & Kleśta, J. (2002). A preliminary study of the intonational phrase, nuclear melody and pauses in Polish semi-spontaneous narration. *Proceedings of Speech Prosody 2002*.

Francuzik, K., Karpiński, M., Kleśta, J., & Szalkowska, E. (2004). Nuclear melody in polish semi-spontaneous and read speech: Evidence from the Polish Intonational Database Polnt. *Studia Phonetica Posnaniensia*, *7*, 97–128.

Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Proceedings of Speech Prosody*.

Fuchs, S. (2019). Vocal tract variations affect vowel sounds. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0683-6

Fujimura, O. (2000). The C/D model and prosodic control of articulatory behavior. *Phonetica*, *57*(2-4), 128–138.

Gibbon, D. (2003). Computational modelling of rhythm as alternation, iteration and hierarchy. *Proceedings of the 15th International Congress of Phonetic Sciences*.

Gibbon, D. (2013). TGA: A web tool for Time Group Analysis. In D. Hirst & B. Bigi (Eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop* (pp. 66–69).

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. *Seventh European Conference on Speech Communication and Technology*, 95–98.

Giordano, R., & D'Anna, L. (2010). A comparison of rhythm metrics in different speaking styles and in fifteen regional varieties of Italian. *Proceedings of Speech Prosody*.

Gordon, M. (2014). Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels. In H. van der Hulst (Ed.), *Word stress: Theoretical and typological issues* (pp. 83–118). Cambridge University Press.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Papers in laboratory phonology* (pp. 515–546, Vol. 7).

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

Hamlaoui, F., Żygis, M., Engelmann, J., & Wagner, M. (2019). Acoustic correlates of focus marking in Czech and Polish. *Language and Speech*, *62*(2), 358–377.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press.

Hixon, T. J. (1971). An electromagnetic method for transducing jaw movements during speech. *The Journal of the Acoustical Society of America*, *49*(2B), 603–606.

James, L. (1940). *Speech Signals in Telephony*. Sir I. Pitman & Sons.

Jassem, W. (1992). Acoustic-phonetic variability of Polish vowels. *Archives of Acoustics*, *17*(2), 217–233.

Jassem, W. (1962). *Akcent języka polskiego*. Ossolineum.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. https://doi.org/10.1017/S0025100303001191

Jassem, W., & Gibbon, D. (1980). Re-defining English accent and stress. *Journal of the International Phonetic Association*, *10*(1-2), 2–16.

Jassem, W., Hill, D. R., & Witten, I. H. (1984). Isochrony in English Speech: its Statistical Validity and Linguistic Relevance. In D. Gibbon & H. Richter (Eds.), *Intonation, Accent and Rhythm. Studies in Discourse Phonology* (pp. 203–225).

Jassem, W., Morton, J., & Steffen-Batóg, M. (1968). The perception of stress in synthetic speech-like stimuli by Polish listeners. In W. Jassem (Ed.), *Speech analysis and synthesis* (pp. 289–308, Vol. 1). Państwowe Wydawnictwo Naukowe.

Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 520–539).

Karpiński, M., & Kleśta, J. (2001). Intonational database for the Polish language. *Proceedings of Prosody 2000 Workshop*.

Kawahara, S., Erickson, D., Moore, J., Shibuya, Y., & Suemitsu, A. (2014). Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress. *Journal of the Phonetic Society of Japan*, *18*(2), 77–87.

Kawahara, S., Erickson, D., & Suemitsu, A. (2015). Edge prominence and declination in Japanese jaw displacement patterns: A view from the C/D model. *Journal of the Phonetic Society of Japan*, *19*, 33–43.

Klessa, K. (2006). *Analiza iloczasu głoskowego na potrzeby syntezy mowy polskiej* [Doctoral dissertation, Adam Mickiewicz University].

Klessa, K., & Gibbon, D. (2014). Annotation Pro+ TGA: Automation of speech timing analysis. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*, 1499–1505.

Klessa, K., Karpiński, M., & Wagner, A. (2013). Annotation Pro-a new software tool for annotation of linguistic and paralinguistic features. *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence*, 51–54.

Klessa, K., Koržinek, D., Sawicka-Stępińska, B., & Kasperek, H. (2022). Annpro: A desktop module for automatic segmentation and transcription. *Language and Technology Conference*, 65–77.

Kochetov, A. (2020). Research methods in articulatory phonetics II: Studying other gestures and recent trends. *Language and Linguistics Compass*, *14*(6), e12371.

Kohler, K. J. (2008). 'Speech-smile','speech-laugh','laughter'and their sequencing in dialogic interaction. *Phonetica*, *65*(1-2), 1–18.

Kraska-Szlenk, I. (1995). *The phonology of stress in Polish* [Doctoral dissertation, University of Illinois at Urbana-Champaign].

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology*, *1*, 53–74.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*(3), 253–263.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336.

Lin, S. (2021). Observing and measuring speech articulation. *The Cambridge Handbook of Phonetics*.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Kluwer.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608.

Lorenc, A. (2016). *Wymowa normatywna polskich samogłosek nosowych i spółgłoski bocznej*. Dom Wydawniczy Elipsa.

Lorenc, A., Król, D., & Klessa, K. (2018). An acoustic camera approach to studying nasality in speech: The case of Polish nasalized vowels. *The Journal of the Acoustical Society of America*, *144*(6), 3603–3617.

Łukaszewicz, B. (2018). Phonetic evidence for an iterative stress system: The issue of consonantal rhythm. *Phonology*, *35*(1), 115–150.

Łukaszewicz, B., & Rozborski, B. (2008). Korelaty akustyczne akcentu wyrazowego w języku polskim dorosłych i dzieci. *Prace Filologiczne*, *54*, 265–283.

Machač, P., & Skarnitzl, R. (2009). *Principles of Phonetic Segmentation*. Epocha.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*(4), 499–511.

Madelska, L., & Witaszek-Samborska, M. (1991). *Zapis fonetyczny: Zbiór ćwiczeń*. Wydawnictwo Naukowe UAM.

Malisz, Z. (2013). *Speech rhythm variability in Polish and English: A study of interaction between rhythmic levels* [Doctoral dissertation, Adam Mickiewicz University].

Malisz, Z., O'Dell, M., Nieminen, T., & Wagner, P. (2017). Perspectives on speech timing: Coupled oscillator modeling of Polish and Finnish. *Phonetica*, *73*(3-4), 229–255.

Malisz, Z., & Wagner, P. (2012). Acoustic-phonetic realization of Polish syllable prominence: A corpus study. *Speech and Language Technology*, *14/15*, 105–114.

Malisz, Z., & Żygis, M. (2018). Lexical stress in Polish: Evidence from focus and phrase-position differentiated production data. *Proceedings of Speech Prosody 2018*. https://doi.org/10.21437/SpeechProsody.2018-204

McCarthy, J. J. (2003). OT constraints are categorical. *Phonology*, *20*(1), 75–138.

McCarthy, J. J., & Prince, A. (1993). Generalized alignment. In *Yearbook of Morphology 1993* (pp. 79–153). Springer.

Mik, Ł., & Lorenc, A. (2024a). EMAviewer [Computer software to visualise speech organs movement positions] [Available from authors on request].

Mik, Ł., & Lorenc, A. (2024b). PhoneEMAtool [Computer software] [Available from the authors on request. Contact: anita.lorenc@uw.edu.pl].

Milewski, S., & Binkuńska, E. (2024). Struktura fonostatystyczno-fonotaktyczna wybranych polskich słownych ciągów trudnych (lingwołamek). *Prace Językoznawcze*, *26*(4), 143–160.

Mołczanow, J., Łukaszewicz, B., & Łukaszewicz, A. (2018). Rhythmic stress or word-boundary effects? Comparison of primary and secondary stress correlates in segmentally identical word pairs. *Proceedings of the 9th International Conference on Speech Prosody*, 908–912.

Newlin-Łukowicz, L. (2012). Polish stress: Looking for phonetic evidence of a bidirectional system. *Phonology*, *29*(2), 271–329. https://doi.org/10.1017/S0952675712000139

Nieuwenhuis, R., Grotenhuis, M., & Pelzer, B. (2012). Influence. ME: Tools for detecting influential data in mixed effects models. *R Journal*, *4*, 38–47. https://doi.org/10.32614/RJ-2012-011

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130396.

Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*, *34*(2), 139–152.

O'Dell, M. L., & Nieminen, T. (2009). Coupled oscillator model for speech timing: Overview and examples. *Prosody: Proc. 10th conference, Helsinki, Finland*, 179–190.

Oshimat, K., & Gracco, V. L. (1992). Mandibular contributions to speech production. *Proceedings of International Conference on Spoken Language Processing 1992*, 775–778.

Osowicka-Kondratowicz, M. (2021). Wariantywność akcentowa w języku polskim. proparoksytoneza. *Prace Językoznawcze*, *23*(4), 43–60.

Ostaszewska, D., & Tambor, J. (2000). *Fonetyka i fonologia współczesnego języka polskiego*. Państwowe Wydawnictwo Naukowe.

Payne, E. (2021). Comparing and deconstructing speech rhythm across Romance languages. *Manual of Romance Phonetics and Phonology*, 264–298.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430.

Pike, K. L. (1945). *The intonation of american english*. University of Michigan Press.

Pompino-Marschall, B., & Żygis, M. (2003). *Surface palatalization of Polish bilabial stops: Articulation and acoustics*. Universitätsbibliothek Johann Christian Senckenberg.

Posit Software, PBC. (2023). RStudio (Version 2023.06.0) [Integrated development environment].

Python Software Foundation. (2023). Python (Version 3.12.2) [Programming language].

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, 337–342.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.

Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology*, *12*(1), Article 6. https://doi.org/10.5334/labphon.237

Richmond, K. (2002). Estimating articulatory parameters from the acoustic speech signal. *Annexe Thesis Digitisation Project 2017 Block 11*.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In *Linguistic Controversies* (pp. 73–79).

Rogers, J., & Revesz, A. (2019). Experimental and quasi-experimental designs. In *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 133–143). Routledge.

Rojczyk, A. (2019). Quality and duration of unstressed vowels in Polish. *Lingua*, *217*, 80–89. https://doi.org/10.1016/j.lingua.2018.10.012

Rubach, J., & Booij, G. E. (1985). A grid theory of stress in Polish. *Lingua*, *66*(4), 281–320.

Smith, C. L., Erickson, D., & Savariaux, C. (2019). Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers. *Journal of Phonetics*, *77*, 100938.

Sonoda, Y., & Nakakido, K. (1986). Effect of speaking rate on jaw movements in vowel sequences. *Journal of the Acoustical Society of Japan (E)*, *7*(1), 5–12.

Spreafico, L., & Vietti, A. (2022). Techniques and methods for investigating speech articulation: The centrality of instruments. *Laboratory Phonology*, *13*(1), 1–8.

Steffen-Batóg, M. (2000). *Struktura akcentowa języka polskiego*. Wydawnictwo Naukowe PWN.

Stone, M., & Shadle, C. H. (2016). A history of speech production research. *Acoustics Today*, *12*(4), 48–55.

Svensson-Lundmark, M., & Erickson, D. (2024). Segmental and syllabic articulations: A descriptive approach. *Journal of Speech, Language, and Hearing Research*, *67*(10S), 3974–4001.

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, *4*(1), 93–118.

van der Hulst, H. (2002). Stress and accent. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (pp. 246–254, Vol. 4). Nature Publishing Group.

Wagner, A. (2014). Description of Polish speech rhythm using rhythm metrics and the time-delay approach: A comparative study. *Proceedings of the 7th International Conference on Speech Prosody*, 366–370.

Wagner, A. (2017). *Rytm w mowie i języku w ujęciu wielowymiarowym*. Dom Wydawniczy Elipsa.

Wagner, P. (2007). Visualizing levels of rhythmic organization. *Proceedings of the XVIth International Congress of the Phonetic Sciences*.

Wagner, P. (2008). *The Rhythm of Language and Speech: Constraining Factors, Models, Metrics and Applications* [Habilitation thesis, Bielefeld University]. Retrieved April 15, 2022, from https://pub.uni-bielefeld.de/download/1916845/2577895/gesamt.pdf

Wagner, P., & Dellwo, V. (2004). Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. *Proceedings of Speech Prosody*, 227–230.

Warner, N. (2021). Processes in connected speech. In R.-A. Knight & J. Setter (Eds.), *The cambridge handbook of phonetics* (pp. 133–156). Cambridge University Press.

Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, *63*, 38–54.

White, L., & Malisz, Z. (2021). Speech Rhythm and Timing [Original work published 2020]. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody*. Oxford Academic. https://doi.org/10.1093/oxfordhb/9780198832232.013.10

Wierzchowska, B. (1967). *Opis fonetyczny języka polskiego*. Państwowe Wydawnictwo Naukowe.

Wierzchowska, B. (1971). *Wymowa polska*. Państwowe Zakłady Wydawnictw Szkolnych.

Williams, J., Erickson, D., Ozaki, Y., Suemitsu, A., Minematsu, N., & Fujimura, O. (2013). Neutralizing differences in jaw displacement for English vowels. *The Journal of the Acoustical Society of America*, *133*(5_Supplement), 3607–3607.

Yu, J., Gibbon, D., & Klessa, K. (2014). Computational annotation-mining of syllable durations in speech varieties. *Proceedings of the 7th Speech Prosody Conference*, 20–23.