Open Library of Humanities

The encoding of prominence relations in supra-laryngeal articulation across speaking styles

Lena Pagel*, I*f*L Phonetics, University of Cologne, Germany, lena.pagel@uni-koeln.de

Simon Roessig, Department of Language and Linguistic Science, University of York, United Kingdom, mail@simonroessig.de

Doris Mücke, I*f*L Phonetics, University of Cologne, Germany, doris.muecke@uni-koeln.de

*Corresponding author.

Speakers modify their supra-laryngeal articulation on the one hand as a function of prosodic prominence and on the other hand as a function of speaking style. Articulatory movements are spatio-temporally expanded in words that are highlighted due to prosodic strengthening, but also in utterances that are produced in a loud speaking style. What remains unclear, to date, is how these two phenomena interact, or, in other words, how prominence relations are encoded in loud speech. In this paper, we use electromagnetic articulography to investigate lip and tongue kinematics and analyze selected acoustic parameters across several focus conditions in habitual and loud speech. Prosodic prominence relations are studied as the simultaneous strengthening of highlighted words and the attenuation of words in the background. The results show that prominence relations can be encoded multidimensionally in loud speech. Some modifications are even stronger in loud speech than in habitual speech, despite the general articulatory modifications associated with a loud speaking style. In both speaking styles, the attenuation of words in the background tends to be more robust than the strengthening of highlighted words. The paper underlines the flexibility of the multidimensional phonetic space in the encoding of prosodic prominence across varying communicative demands.

1 Background

Speakers structure the information that they convey in an utterance (Vallduví & Engdahl, 1996). Based on their beliefs about which entities are most important for the listener, some parts of an utterance are in *focus* while others are in the *background* (Gussenhoven, 2004; Lambrecht, 1994; Vallduví, 1992). This so-called focus structure of an utterance is prosodically marked in many languages through a range of parameters, including modulations of F0, increased sound pressure, longer segmental durations, as well as more distinct sound productions (Baumann et al., 2007; Breen et al., 2010; Féry & Kügler, 2008; Roessig, Winter, et al., 2022; Xu & Xu, 2005).

The phonetic modifications observed for prosodic focus have similarities to characteristics of a loud speaking style, where these modifications span entire utterances. Since speakers adapt their speech production to communicative demands, they produce loud speech to be more intelligible in adverse listening conditions. The characteristics of a loud speaking style are associated with high vocal effort and include modulations of F0, increased sound pressure, and spatio-temporal changes in supra-laryngeal articulation, which resemble the characteristics of focus marking (Raitio et al., 2013; Tasko & McClean, 2004).

While the two phenomena of focus marking on the one hand and a loud speaking style on the other hand have been widely studied to date, their *interaction* remains unclear. In this paper, we pose the following research question: Can speakers signal focus structure in loud speech, despite the overall modifications associated with this speaking style? In other words, if loud speech "turns up" the speech production system, is there still room left for prosodic focus marking?

To address this question, we conducted an experiment with 20 German speakers, using acoustic recordings and 3D electromagnetic articulography to track the movements of the oral articulators. The speakers were recorded as they interacted with a virtual avatar, producing embedded target words in two utterance positions, two focus conditions and two speaking styles. Acoustic parameters, as well as kinematics of the lips and tongue body, were analyzed in target syllables including two vowel types. Through measurements of duration, F0 and sound pressure level, as well as target positions of the primary articulators, we investigated how focus structure manifests in habitual speech on the one hand and in loud speech on the other. The results reveal acoustic and articulatory signatures of focus structure in both speaking styles. Focus marking emerges as a structural phenomenon requiring a balance of strengthening and weakening strategies affecting the entire phrase: An increase in prosodic prominence in the nuclear condition is accompanied by a simultaneous decrease in prominence in the prenuclear condition. Interestingly, in most analyzed parameters, the acoustic and articulatory differentiation of focus conditions tend to be stronger in a loud speaking style compared to a habitual one. This underscores that the goal of enhanced intelligibility in adverse listening conditions is not constrained by an overall increased production effort.

1.1 Focus structure and prosodic prominence

1.1.1 Focus structure

Focus is a fundamental component of information structure (Féry & Krifka, 2008). Information structure can be defined as the packaging of information to fulfil the communicative requirements of an interlocutor (Chafe, 1976). The focus of an utterance is determined by the communicative context and reflects the speaker's beliefs of which entities are most informative and relevant for the listener (Halliday, 1967; Lambrecht, 1994; Vallduví & Engdahl, 1996). Other entities considered less important in the given situation are out of focus, i.e., in the background of the utterance (Lambrecht, 1994). Typically, the background contains information that has already been established and "anchors the sentence to the previous discourse" (Vallduví & Engdahl, 1996, p. 461).

There are multiple manifestations of focus, differentiated among others by focus breadth and contrastiveness (Gussenhoven, 2004; Ladd, 1980; Repp, 2016). Below are examples illustrating *broad*, *narrow,* and *corrective* focus. Focus is indicated by square brackets with the subscript *F* (Jackendoff, 1974). In Example (1), the entire utterance is in broad focus, because no information is provided by the previous discourse (Ladd, 1980). In Example (2), some information is already supplied by the preceding question. The only new information in the response is *coffee*, which is in focus, while the remaining parts of the utterance are in the background. This condition is termed narrow focus, where only one aspect of an event is highlighted, making the focus narrower compared to broad focus (Ladd, 1980). Example (3) presents a special case of narrow focus, where the focus breadth is identical to Example (2), but here, *coffee* is in corrective focus. It corrects an entity already established in the discourse; in this instance, *coffee* corrects the suggestion *tea* from the preceding question (Gussenhoven, 2004; Repp, 2016; Wagner, 2012).

(1)   Q: What's up?
        A: [Lisa wants to drink coffee.]$_F$

(2)   Q: What does Lisa want to drink?
        A: Lisa wants to drink [coffee.]$_F$

(3)   Q: Does Lisa want to drink tea?
        A: Lisa wants to drink [coffee.]$_F$

What becomes evident from the examples is that the focus structure affects the entire utterance, as prominences are distributed across all entities, with some words in focus and others in the background (Calhoun, 2007). In examples (2) and (3), only the word *coffee* is in focus and is therefore associated with the main prominence of the sentence. However, it can only stand out in comparison to other elements in the utterance that are in the background. In other words, an entity always bears *relative* prominence (Calhoun, 2007; Cangemi & Baumann, 2020; Wagner

& Watson, 2010). Prosodic prominence is commonly defined based on the production and perception of an entity, e.g., as "the strength of a spoken word relative to the words surrounding it" (Cole et al., 2010, p. 425) or as its "acoustic and perceptual salience" (Vogel et al., 2016, p. 123; see Ladd & Arvaniti, 2023, for a review). While parts of an utterance can either be in or out of focus, prosodic prominence is not a binary feature. Instead, an element can be associated with several *degrees* of prominence, which depend, among other factors, on the focus structure of the utterance (Calhoun, 2007). *Coffee* is in focus in all three examples above, but it may be produced with increased prosodic prominence in corrective focus (Example [3]) compared to narrow focus (Example [2]), and in narrow focus compared to broad focus (Example [1]). The effect of focus structure on the entire utterance will be further elaborated in Section 1.1.3.

1.1.2 Correlates of prosodic prominence

Prosodic prominence manifests in a broad array of acoustic correlates (Baumann et al., 2007; Breen et al., 2010; Féry & Kügler, 2008; Roessig, Winter, et al., 2022; Xu & Xu, 2005). In English, German, and other West-Germanic languages, a salient signature of prominence is the placement of a pitch accent on a word. Let's revisit the previously mentioned examples of focus structures. A pitch accent on *coffee* can be expected in all three examples. In broad focus (Example [1]), where the focus spans the entire utterance, the final noun of the sentence (in this case *coffee*) typically serves as the focus exponent for the larger domain. The focus exponent bears the nuclear pitch accent, which is the final and perceptually most salient pitch accent of the intonational phrase (Ladd, 1980). While the placement of a pitch accent already marks an element as prominent, distinctions between degrees of prominence can be conveyed through different types or shapes of accents, as demonstrated, among others, for German and English (Baumann et al., 2007; Pierrehumbert & Hirschberg, 1990; Roessig, Mücke, & Grice, 2019; Roessig, 2021). In German speech production, there are probabilistic preferences for downstepped or falling pitch accents on nuclear-accented words in broad focus as opposed to rising pitch accents in corrective focus, with proportions for narrow focus lying between broad and corrective (Baumann et al., 2007; Grice et al., 2017; Mücke & Grice, 2014). These findings on accent type choice are supported by continuous analyses, which have revealed higher F0 values, greater excursion or tonal onglide, and later peak alignment in corrective compared to narrow focus, and in narrow compared to broad focus (Baumann et al., 2007; Grice et al., 2017; Roessig, Mücke, & Grice, 2019; Roessig, Winter, et al., 2022).

In addition to F0 modulations, prominent entities bearing a pitch accent exhibit various other acoustic modifications compared to less prominent elements. Studies across multiple languages have observed increased sound pressure (Fowler, 1995; Harrington et al., 1995; Roessig, Winter, et al., 2022; Wang, 2020), as well as longer durations of consonants, vowels, syllables, and words (Avesani et al., 2007; Cao & Zheng, 2006; Cho & McQueen, 2005; de Jong, 1995; Edwards et al.,

1991; Hermes et al., 2008; Jabeen & Braun, 2018; Wang, 2020) in prominent words compared to less prominent or non-prominent words. Again, fine-grained acoustic differences can be observed between degrees of prominence, i.e., between words that are nuclear-accented but occur in different focus conditions. For instance, speakers produce sounds with longer acoustic durations in corrective focus compared to narrow focus and/or in narrow focus compared to broad focus (Baumann et al., 2007; Kügler, 2008; Mücke & Grice, 2014; Roessig, Winter, et al., 2022). Similarly, studies have reported higher sound pressure in accented words in corrective focus compared to narrow focus, and in narrow focus compared to broad focus (Breen et al., 2010; Roessig, Winter, et al., 2022).

Prominence also affects the supra-laryngeal articulation of an entity, a phenomenon known as prosodic strengthening (Cho, 2004). Numerous kinematic experiments have been conducted to investigate prominence-induced prosodic strengthening in the articulation of the jaw, lips, and tongue. Evidence for larger (as well as mostly longer and faster) vocalic jaw and lip opening under increased prominence has been gathered for English (Cho, 2005, 2006; de Jong, 1991, 1995; Fowler, 1995; Harrington et al., 1995, 2000; Kim & Cho, 2011), German (Hermes et al., 2008; Mücke & Grice, 2014), and Italian (Avesani et al., 2007), among other languages. Fine-grained analyses of prosodic strengthening at the segment level have revealed greater jaw lowering in vowels compared to greater raising in consonants (Beckman et al., 1992; Macchi, 1985). Additionally, the lips are more protruded in rounded vowels when strengthened (de Jong, 1991, 1995; Kent & Netsell, 1971). While the elicitation methods vary among the studies mentioned above, the majority do not explicitly investigate differences between several focus conditions or degrees of prominence. However, more recent evidence suggests that differences occur not only between highly diverging focus conditions (e.g., background vs. corrective focus). Instead, lip kinematics are also spatially and temporally modified to some extent in a step-wise manner between background, broad, narrow, and corrective focus (Hermes et al., 2008; Krivokapić et al., 2017; Mücke & Grice, 2014; Roessig, 2021; Roessig & Mücke, 2019).

In addition to jaw and lip kinematics, prosodic strengthening is also evident in vocalic tongue body movements. Studies indicate that under prominence, the tongue is lowered in /ɑ, ʌ, ɐ, o/ (Cho, 2002, 2005; de Jong, 1991; Katsika, 2018), while it is raised in /i/ (Katsika, 2018; Kent & Netsell, 1971; Shin et al., 2015). Additionally, there is evidence of tongue body retraction in /o, ʊ/ (de Jong, 1991) and fronting in /i, ɪ/ (Cho, 2005; Kent & Netsell, 1971; Kim & Cho, 2011). These findings suggest that prominence is correlated with the tongue body moving closer towards its "presumed articulatory target" (Kent & Netsell, 1971, p. 42), although strategies may vary among speakers (Harrington et al., 2000). These enhancements tend to occur in contrasts that are relevant within the respective language's sound system (Cho & McQueen, 2005). Once again, fine-grained modulations of tongue body kinematics can be observed between several focus

structures or degrees of prominence, i.e., between background, broad, narrow, and corrective focus (Roessig, 2021; Roessig, Mücke, & Pagel, 2019; Roessig & Mücke, 2019).

Findings on prosodic strengthening have been explained by two predominant accounts. The increased jaw and lip opening in vowels can be interpreted as a strategy of *sonority expansion*. This assumes that the greater opening of the vocal tract allows for more energy to radiate from the mouth, thereby enhancing sonority contrasts between consonants and vowels (Beckman et al., 1992). The prominence-induced modifications of the tongue body, on the other hand, can be viewed as a strategy of *localized hyperarticulation*. This assumes that sounds are produced with increased distinctness under prominence (de Jong, 1995). These two strategies may serve different functions: The sonority expansion primarily contributes to an enhancement of syntagmatic contrasts, while the localized hyperarticulation mainly enhances paradigmatic contrasts (Harrington et al., 2000; Mücke, 2018). Since the two strategies predominantly recruit different articulators, they can mostly go hand in hand. However, there are contexts where they appear to conflict, such as in the tongue body in the high front vowel /i/. In such cases, there is evidence that speakers tend to hyperarticulate those features of the sound that are simultaneously compatible with a sonority expansion (Cho, 2005). For example, in /i/, the tongue body may be fronted but not necessarily raised under prominence, as raising it could interfere with sonority expansion (Cho, 2005).

1.1.3 Correlates of focus structure in the pre-nuclear domain

Up to this point, we have primarily concentrated on the correlates of *increased* prosodic prominence of a word, particularly a focused and pitch-accented element in comparison to an element in the background. However, as previously mentioned, focus structure impacts the entire utterance rather than solely the most prominent entity. Most of the existing literature on this subject is concerned with the post-nuclear region of the utterance, which is the portion of an utterance following the word with the nuclear pitch accent. In this region, post-nuclear (or post-focal) deaccentuation, or compression, is observed in many languages (Xu et al., 2012). Conversely, the pre-nuclear region, which is the portion preceding the nuclear-accented word, has received less attention thus far. Prosody in this domain is often regarded as following a default pattern that does not contribute significantly to the meaning of an utterance (Büring, 2007).

However, some studies suggest that the importance of this portion of the utterance has been underestimated. They indicate that a word in a pre-nuclear utterance position is realized differently when it is in broad focus compared to when it is in the background before the focus; in other words, when it is focal compared to when it is pre-focal. For instance, Baumann et al. (2007) observe a higher proportion of pre-nuclear accents in German when the following nuclear-accented word is in broad focus than when it is in corrective focus, i.e., when the pre-nuclear

word is focal than when it is pre-focal. Additionally, research on German and French reveals that words in the pre-nuclear region are produced with lower F0 contours when the nuclear-accented word is in corrective focus compared to when it is in broad focus (Dohen & Lœvenbruck, 2004; Roessig, 2024). Similar evidence exists for the pre-nuclear domain in German and Mandarin, showing differences when the nuclear-accented word is in narrow focus compared to when it is in broad focus (Féry & Kügler, 2008; Roessig, 2024; Yang & Chen, 2020).

Moreover, pre-nuclear compression is not only evident in F0 contours but also in word durations and other phonetic characteristics. Words preceding a nuclear-accented word in corrective focus can exhibit shorter durations compared to those preceding a word in broad focus in German (Kügler, 2008; Roessig, 2024) and American English (Erickson & Lehiste, 1995). In a separate study on American English, segment durations are examined in a trisyllabic word preceding either a word in corrective focus or in the background (Cho et al., 2013). Shorter vowel durations can be observed in the penultimate and ultimate syllable when the word precedes a correctively focused word, a phenomenon termed "preaccentual shortening" by the authors (Cho et al., 2013, p. 388). Similarly, shorter durations and lower intensities are reported for words preceding a nuclear-accented word in narrow focus compared to broad focus in Mandarin (Yang & Chen, 2020). A study on Bulgarian measures the phonetic strength of pre-nuclear pitch accents and observes weaker accents when they precede corrective or narrow focus than when they precede broad focus (Andreeva et al., 2017). These weaker pre-nuclear accents are characterized, among others, by lower mean F0, shorter durations, and lower intensities. Similar results are reported for the pre-focal domain in various Arabic varieties (Alzaidi et al., 2023; Alzamil & Hellmuth, 2021; Chahal & Hellmuth, 2014).

A recent study on Korean expands upon the existing body of evidence regarding the realization of focal and non-focal words in different utterance positions by incorporating results from an articulatory investigation (Im et al., 2023). The authors observe that lip opening and closing movements are spatio-temporally reduced in pre- and post-focal words compared to the focal words they surround. Crucially, this reduction tends to be more pronounced in pre-focal words than in post-focal words, highlighting once again that focus realization affects not only the highlighted word but the entire utterance in intricate ways.

Upon initial examination, the observed differences described above might seem attributable not to the distinction between various focus conditions, but rather to the differentiation between focal and pre-focal. However, further evidence indicates that words are also realized differently across different focus conditions when they are all pre-focal. For instance, in German, Baumann et al. (2007) observe fewer pre-nuclear accents on pre-focal words when they occur before corrective focus compared to when they occur before narrow focus. Similarly, evidence suggests that pre-nuclear pitch accents are shallower before corrective focus than before narrow focus in

Bulgarian (Andreeva et al., 2017). In a recent analysis, Roessig (2024) demonstrates that pre-focal words are produced with lower or compressed F0 and shorter syllable durations before corrective focus compared to before narrow focus in German.

Taken together, the findings suggest that when the nuclear-accented word is produced with increased prosodic prominence, words in the pre-nuclear domain are simultaneously produced with decreased prosodic prominence. This indicates an inverse relationship between the nuclear and pre-nuclear domains, wherein *prominence relations* between different parts of an utterance are realized in speech production. This pattern is observed both between focal and pre-focal words preceding nuclear-accented words, as well as between pre-focal words preceding nuclear-accented words in different focus conditions. To our knowledge, the first study to include a kinematic investigation of supra-laryngeal articulation in the pre-nuclear part of an utterance has recently been conducted for Korean (Im et al., 2023). Korean, however, is a language without lexical stress and is considered an edge-prominence language, where prominence is primarily encoded through phrasing (Jun, 2006). Therefore, it seems beneficial to expand analyses of the pre-nuclear domain of an utterance to other languages, such as German, which is a head-prominence language.

1.2 Loud speech

Modulations associated with loud speech have been termed "global" variations, not only because they span long stretches of speech but also because they affect several speech subsystems: respiration, phonation, and supra-laryngeal articulation (Dromey & Ramig, 1998). In the existing literature, loud speech is distinguished from other effortful speaking styles, such as Lombard speech (speech produced in background noise) and clear speech (e.g., speech directed at a hearing-impaired or non-native listener). While all these speaking styles share numerous phonetic correlates, they also exhibit some striking differences, both in their elicitation methods and their phonetic realization (Bond & Moore, 1990; Darling & Huber, 2011; Godoy et al., 2014; Huber & Chandrasekaran, 2006; Mefferd, 2017; Tjaden, Richards, et al., 2013). In this paper, our primary focus is loud speech, but we also include some evidence for Lombard speech when appropriate, as these two speaking styles show strong similarities and have been believed to arise from the same production mechanism (Bond & Moore, 1990). Therefore, supplementary results from Lombard speech can offer valuable insights but should be interpreted with caution in the context of loud speech without background noise. Additionally, selected evidence from clear speech is presented. It is important to note that the following literature review includes studies on pathological speech; however, only results for the typical control group are reported here.

In acoustic analyses, the most striking correlate of loud speech is an increase in sound pressure. This higher sound pressure in loud speech is consistently reported across multiple languages,

including English (Bond & Moore, 1990; Darling & Huber, 2011; Dromey & Ramig, 1998; Garnier et al., 2022; Huber & Chandrasekaran, 2006; Kleinow et al., 2001; Mefferd, 2017, 2019; Mefferd & Green, 2010; Tasko & McClean, 2004; Wohlert & Hammen, 2000), German (Geumann, 2001a, 2001b; Koenig & Fuchs, 2019; Xue et al., 2021), French (Liénard & Di Benedetto, 1999), Swedish (Traunmüller & Eriksson, 2000), and Finnish (Raitio et al., 2013). A further acoustic correlate of loud speech is a higher mean F0 (Bond & Moore, 1990; Dromey & Ramig, 1998; Garnier et al., 2022; Geumann, 2001a, 2001b; Liénard & Di Benedetto, 1999; Raitio et al., 2013; Traunmüller & Eriksson, 2000; Xue et al., 2021), which is closely linked to an increased sound pressure level, likely due to physiological and structural reasons (Alku et al., 2002; Gramming et al., 1988). Additionally, studies report longer vowel durations and either unchanged or shorter consonant durations (Bonnot & Chevrie-Muller, 1991; Geumann, 2001a; Schulman, 1989; Traunmüller & Eriksson, 2000), as well as an expanded acoustic vowel space (Mefferd, 2017; Mefferd & Green, 2010; Tjaden, Lam, et al., 2013, 2013; Tjaden & Wilding, 2004).

A number of studies have investigated supra-laryngeal articulation in loud speech, using various kinematic tracking techniques. These studies demonstrate that vowels are produced with larger and often faster jaw and/or lip opening movements in this speaking style (Darling & Huber, 2011; Dromey & Ramig, 1998; Garnier et al., 2022; Geumann, 2001a; Huber & Chandrasekaran, 2006; Mefferd, 2019; Mefferd & Dietrich, 2020; Schulman, 1989; Tasko & McClean, 2004; Xue et al., 2021). At the same time, some studies indicate that these spatial enhancements of jaw and lip gestures occur to a greater extent in vowels and not to the same extent in consonants, thus leading to enhanced contrasts between vowels and consonants (Geumann, 2001a; Schulman, 1989). Studies on tongue body kinematics have observed larger and often also faster and longer lingual gestures in loud speech compared to habitual speech, either across tongue body gestures overall (Tasko & McClean, 2004) or, more specifically, between /i/ and /a/ (Mefferd, 2017, 2019; Mefferd & Dietrich, 2020; Mefferd & Green, 2010). Additionally, an expanded kinematic vowel space is found across the entire sentence (Whitfield et al., 2018). In a study on shouted sustained vowels, an overall lower tongue position is observed across a wide range of vowels (Xue et al., 2021). Supplementary evidence can be drawn from studies examining Lombard speech, where individuals produce speech in background noise. As previously mentioned, it is important to interpret these results with caution when considering their applicability to loud speech without background noise. Articulatory analyses show that Lombard speech is associated with greater tongue displacements (Nicolaidis, 2012; Šimko et al., 2016) and increased contrasts between close and open vowels (Garnier et al., 2018) in the vertical (high–low) movement dimension. In the horizontal (front–back) movement dimension, an experiment on Lombard speech indicates a general lingual fronting in vowels alongside an increased articulatory contrast between front and back vowels (Garnier et al., 2018).

1.3 Prosodic prominence across speaking styles

It becomes evident that the two phenomena of loud speech and the prosodic marking of focus structure have been extensively studied to date. However, to our knowledge, there is currently no report on how focus structure is realized *in* loud speech, examining the interaction between the two phenomena. While there is a small number of studies providing indirect evidence for the interaction of prosodic prominence and effortful speaking styles, they do not investigate loud speech but rather other effortful speaking styles, such as Lombard speech or clear speech. These analyses suggest that when transforming habitual speech into effortful speech, not all elements in an utterance undergo the same transformation. Instead, there appear to be interactions with prosodic structure.

For instance, Vainio et al. (2012) investigate focus marking in Finnish Lombard speech. They discover that despite the expansion of F0 mean and range in this speaking style, the focus-related intonation contours produced in Lombard speech still resemble those that are typical for focus marking in habitual speech. Another study has shown that when transforming habitual speech into Lombard speech in English, F0 mean and movement are increased across the entire utterance, but this increase is not equally strong for all words. Instead, F0 changes are more pronounced in accented words and less so in unaccented words, resulting in greater F0 differences between accented and unaccented words in Lombard speech (Rivers & Rastatter, 1985).

Another piece of evidence comes from a study on clear speech directed at a non-native listener (Cho et al., 2011). Clear speech is characterized by a slow speaking rate (usually slower than in loud speech), increased sound pressure (typically not as high as in loud speech), as well as a hyperarticulation of vowels and consonants (Baker & Bradlow, 2009; Mefferd, 2017; Smiljanić & Bradlow, 2005; Tjaden, Lam, et al., 2013; Tjaden & Martel-Sauvageau, 2017). While findings on clear speech cannot be assumed to directly apply to loud speech, they are worth mentioning in this context because clear speech, like loud speech, is a speaking style that requires high vocal effort, and there are similarities between the two speaking styles. In their study on prosodic prominence in Korean clear speech, Cho et al. (2011) find prominence-induced temporal expansions and increased sound pressure in vowels, similar to habitual speech. The authors conclude that "the clear speech effects are conditioned by prosodic factors in a way that segments positioned in prosodically strong locations are weighted more" (Cho et al., 2011, p. 355).

Taken together, these acoustic analyses suggest that effects of prosodic structure may also manifest in effortful speaking styles such as Lombard and clear speech. However, it remains unclear how these results relate to deliberately loud speech. Loud speech is associated with an even greater increase in sound pressure than Lombard and clear speech, which may require even more production effort and could potentially impede the encoding of prosodic prominence. Furthermore, to our knowledge, existing studies are limited to acoustic analyses and evidence on supra-laryngeal articulatory correlates of prosodic structure in any effortful speaking style is still lacking to date.

1.4 The present study

The literature overview highlights the overlap in phonetic correlates between prosodic prominence (related to focus structure) and loud speech. Modifications such as longer vowel durations, F0 modulations, increased sound pressure, as well as larger jaw, lip, and tongue body gestures are associated with both phenomena. However, the specific interaction between focus structure and loud speech remains largely unexplored. The crucial question arises: Can speakers still signal prominence relations in a speaking style that is produced with increased vocal effort, or are such fine modulations overridden by the overall aim to speak loudly? Alternatively, could it be the case that in a loud speaking style, where the speech production system is "scaled up," the differences between prominence degrees are realized to an even greater extent? While there is sparse acoustic evidence for an interaction between prosodic prominence and other effortful speaking styles (i.e., Lombard and clear speech), to our knowledge, this interaction has not been investigated for loud speech, nor has it been analyzed articulatorily. Addressing these questions could significantly contribute to our understanding of how prosodic prominence and loud speech intersect, shedding light on the underlying mechanisms of speech production under increased vocal effort.

The present study adopts an exploratory approach to examine the acoustic and supra-laryngeal articulatory signatures of focus structure in both habitual and loud speech. We investigate acoustic durations, F0, and sound pressure level, in addition to lip and tongue body kinematics, in words that occur in different focus structures. In doing so, we compare words associated with different degrees of prominence, acknowledging that prominence is not a binary feature, and fine-grained differences can surface in the realization of words. An innovative aspect of our study is the extension of the analysis to the pre-nuclear domain (where words may be produced with decreased prominence), reflecting our understanding that focus structure affects not just the most prominent word. Rather, it impacts the whole utterance and distributes relative degrees of prominence across all its elements. With this comprehensive approach, we aim to elucidate how speakers navigate the expression of focus structure within various speech production contexts.

2 Methods

2.1 Research questions and predictions

We address two research questions:

A) What are the acoustic and articulatory correlates of a loud speaking style compared to a habitual speaking style?

B) What are the acoustic and articulatory correlates of focus structure in habitual and loud speaking styles?

Based on previous evidence, we expect to find the following results:

For A) Vowel durations, F0, and sound pressure level, as well as lip aperture will be increased, and the tongue body target will be lowered in loud speech compared to habitual speech.

For B) Similar focus-related modifications are anticipated in habitual and loud speech, though they may be stronger in habitual speech. Vowel durations, F0, sound pressure level, and lip aperture will be increased, and tongue body targets will be hyperarticulated when associated with high prosodic prominence. Vowel durations, F0, SPL, and lip aperture will be decreased, and tongue body targets will be hypoarticulated when associated with low prosodic prominence.

2.2 Participants

For this experiment, 20 German speakers were recorded, ranging in age from 22 to 27 years old (mean: 24.7, SD: 1.3). The group comprised of 10 female and 10 male participants. All participants were native German speakers who were raised monolingually in Germany. None reported any diagnosed voice or speaking disorders, hearing impairments, or reading and writing disorders. Additionally, none of the participants were experts in phonetics, phonology, or speech analysis. This study was approved by the Local Ethics Committee of the University of Cologne. Each participant gave written informed consent before study participation. The research was conducted in accordance with the Declaration of Helsinki.

2.3 Experimental procedure

The kinematic signal was recorded using 3D electromagnetic articulography (Carstens AG 501) at a sampling rate of 1250 Hz, which was subsequently downsampled to 250 Hz for analysis. Sensors were placed on the chin, lower and upper lips, tongue tip, tongue blade, and tongue dorsum/body. For the present paper, only the sensors on the lips and the tongue body will be analyzed. Additional reference sensors were placed on the bridge of the nose and behind both ears. Simultaneously, speakers were recorded using an AKG C544 headset microphone connected to a PC via a Tascam US-4x4 audio interface, recording at 48 kHz and 16-bit resolution. The microphone was placed at a distance of 7 cm from the speaker's mouth, angled between 45–90° at the side of the mouth, following the guidelines by Patel et al. (2018). Speakers were seated underneath the articulograph in front of a screen at approximately 3 meters and were asked to follow the instructions displayed on the screen. The experiment was written as a browser-based application using HTML/CSS and JavaScript.

We devised the experiment to ensure controlled utterances while fostering a natural communication atmosphere. Participants engaged in a game-like task set in a soccer stadium, where they interacted with a virtual avatar. In the task scenario, participants attended a soccer match with their friend Marie, represented by the virtual avatar. It was explained that Marie had forgotten her glasses at home, impairing her ability to see the game. Therefore, participants were tasked with answering Marie's questions and describing the match to her. **Figure 1** shows a screenshot of one experimental trial. Each trial began with an animation depicting a scene on the field, where one soccer player passed the ball to another player. Subsequently, Marie's image appeared in the bottom corner of the screen, accompanied by an auditory presentation of her question over loudspeakers. The question was also displayed visually in a speech bubble. Marie's questions, which served as trigger questions for the participant, were pre-recorded by a female native speaker of German, who was a trained phonetician and familiar with study's objectives. She was instructed to produce the questions without any distinct focus structure marking and produced them in two speaking styles, generating two sets of questions: one in habitual speech and the other in loud speech. Following the presentation of the trigger question in the experiment, participants responded based on the written sentence displayed on the screen, which was the target utterance.
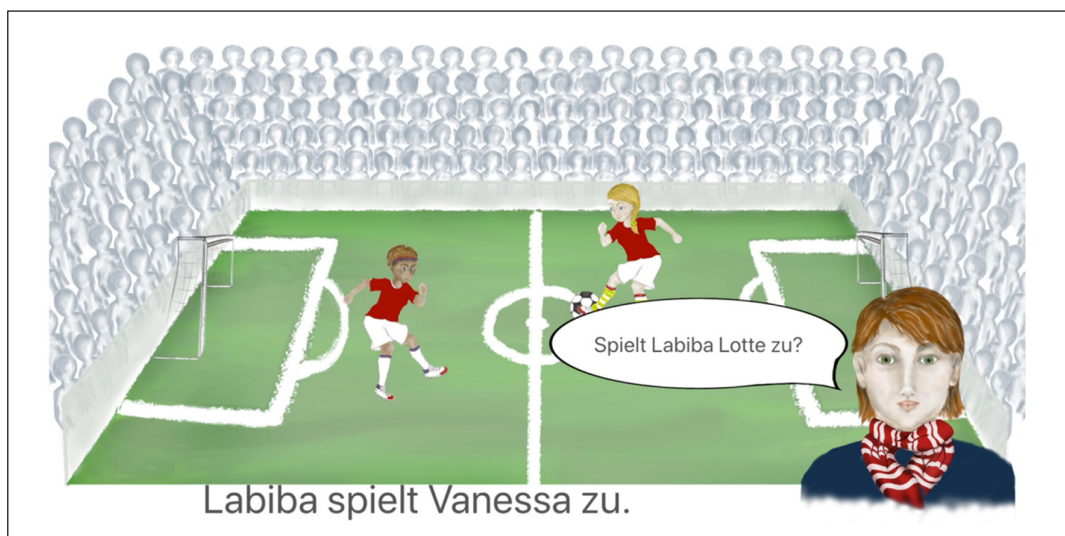

Figure 1: Example of the experiment screen at the end of a trial. In this trial, *Labiba* passes the ball to *Vanessa*. The virtual avatar asks, *Spielt Labiba Lotte zu?* 'Does Labiba pass the ball to Lotte?', and the participant responds, *Labiba spielt Vanessa zu*. 'Labiba passes the ball to Vanessa.', where Vanessa is in corrective focus.

After familiarizing themselves with the target words and completing a training phase of six trials, participants proceeded to the main experiment, which comprised two blocks. Each block consisted of 48 target utterances plus six filler trials. In the first block, participants produced utterances in habitual speech without any specific instructions regarding speaking style. They were unaware that they would later be instructed to produce loud speech. The pre-recorded avatar's questions were from the set of habitual speech. For the second block, participants were told that during the halftime break, the stadium became very noisy, requiring them to speak loudly for Marie to understand them. No actual background noise was played during the experiment (except during the filler trials).

Previous studies have shown that if background noise is played in an experiment, the volume and specific type of noise (e.g., broadband vs. babble noise) can influence the characteristics of the produced Lombard speech (Garnier et al., 2010; Lu & Cooke, 2008). We wanted to rule out this possible confounding factor and opted to elicit deliberately loud speech instead of Lombard speech to enhance the generalizability of our results. The mention of background noise in the instructions solely aimed to make the task more relatable for participants. To further encourage participants to speak loudly, the experimenter communicated with them in a loud speaking style, and the pre-recorded avatar's questions were from the set of loud speech productions, additionally presented at an increased volume. Furthermore, a brief acclimatization phase was included at the beginning of the loud speech block. The combination of these measures aimed to alleviate participants' inhibitions, enabling them to adapt to the game-like task and increase their sound pressure despite the laboratory setting and absence of actual background noise.

Each of the 20 speakers produced 96 target utterances, resulting in a total of 1,920 recorded target utterances. The trial order was pseudo-randomized for each speaker, where repetitions of critical items in two successive utterances were avoided. Twelve utterance productions were excluded from the analysis due to technical problems or unnoticed speech disfluencies, leaving 1,908 utterances available for analysis.

2.4 Speech material and conditions

Each utterance contained one of six trisyllabic target words, which were German-sounding pseudo words: *Nimami* (/ni'ma:mi/), *Sibabi* (/zi'ba:bi/), *Libami* (/li'ba:mi/), *Nabima* (/na'bi:ma/), *Samima* (/za'mi:ma/), and *Labiba* (/la'bi:ba/). The target syllable was the lexically stressed penultimate syllable, which invariably comprised the bilabial consonant /b/ or /m/ followed by the target vowel /a:/ or /i:/. The trisyllabic structure ensured a controlled vocalic context for the target syllable: The target vowel /a/ was consistently flanked by the vowel /i/, and the target vowel /i/ was consistently flanked by the vowel /a/. Furthermore, the consonantal context was controlled so that the target syllable was always succeeded by a bilabial consonant.

The target words were introduced to the participants as the names of female soccer players on the field. The carrier sentences, within which the target words were embedded, followed the structure: subject + verb stem + object + verb affix, as illustrated in examples (4) and (5). The target words were positioned either in the subject position (referred to as utterance-*initial* position), as demonstrated in Example (4), or in the object position (referred to as utterance-*medial* position), as demonstrated in Example (5). In the respective other position, a filler name was inserted, drawn from the set of the four German names *Vanessa* (/vaˈnɛsa/), *Rebecca* (/ʁəˈbɛka/), *Annette* (/ʔaˈnɛtə/), and *Carlotta* (/kaˈlɔta/). The verb stem and affix options were either *spielt … zu* 'passes the ball to…' or *passt … zu* 'passes the ball to …'.

(4)   *Nabima*   *spielt*   *Vanessa*   *zu.*
        subject   predicate   object   predicate
                  (verb stem)   (verb affix)
        'Nabima passes the ball to Vanessa.'

(5)   *Carlotta* *spielt*   *Nabima*   *zu.*
        subject   predicate   object   predicate
                   (verb stem)   (verb affix)
        'Carlotta passes the ball to Nabima.'

The utterances were produced in two focus conditions, referred to as [broad, broad] and [background, corrective]. These two focus conditions were elicited through a question-answer paradigm, where two types of trigger questions (posed by the virtual avatar) prompted two focus types in the answers (given by the participant). **Table 1** demonstrates exemplary question-answer pairs for the two focus conditions and the two utterance positions. In the [broad, broad] condition, both the subject (initial position) and the object (medial position) were in broad focus. The elicitation question for this focus condition was *Was passiert gerade?* 'What is going on?' In the [background, corrective] condition, the subject (initial position) was in the background, while the object (medial position) was in corrective focus. The elicitation question for this focus condition was *Passt/Spielt X Y zu?* 'Does X pass the ball to Y?', where X represented the subject of the answer utterance, and Y was a competitor name introduced in the trigger question. The competitor names included the female German names *Lotte* (/ˈlɔtə/) and *Holly* (/ˈhɔli/) for target words with the vowel /a/, and *Jutta* (/ˈjʊta/) and *Ulla* (/ˈʔʊla/) for target words with the vowel /i/.

Based on the combination of utterance position and focus condition, the target word could occur either in broad focus or in the background when elicited in the initial position, as opposed to broad focus or corrective focus when elicited in the medial position. This enabled a

comprehensive analysis of focus marking: By comparing the [background, corrective] condition with the [broad, broad] condition, we could simultaneously explore the nuclear and pre-nuclear domains. This allowed us to analyze the entity transitioning from broad to corrective focus and the one transitioning from broad focus to the background.

position	focus	question-answer pair
initial	[broad, broad]	Q: *Was passiert gerade?* What is going on? A: *[<u>Nabima</u> spielt Vanessa zu.]*$_F$ [<u>Nabima</u> passes the ball to Vanessa.]$_F$
	[background, corrective]	Q: *Spielt Nabima Holly zu?* Does Nabima pass the ball to Holly? A: <u>*Nabima*</u> *spielt [Rebecca]*$_F$ *zu.* <u>Nabima</u> passes the ball to [Rebecca.]$_F$
medial	[broad, broad]	Q: *Was passiert gerade?* What is going on? A: *[Carlotta spielt <u>Nabima</u> zu.]*$_F$ [Carlotta passes the ball to <u>Nabima.</u>]$_F$
	[background, corrective]	Q: *Spielt Annette Lotte* zu? Does Annette pass the ball to Lotte? A: *Annette spielt [<u>Nabima</u>]*$_F$ *zu.* Annette passes the ball to [<u>Nabima.</u>]$_F$

Table 1: Exemplary question-answer pairs for the target word *Nabima* in two utterance positions and two focus types. Focus is marked by square brackets and the subscript letter F; the target word is underlined.

2.5 Data processing and measurements

Annotations for words and segments were forced-aligned to the acoustic data using the Montreal Forced Aligner (McAuliffe et al., 2017) in Praat (Boersma & Weenink, 2022). A pre-trained acoustic model based on the German Prosodylab dictionary was employed for this purpose. After the automatic segmentation, manual correction was applied to the first four segments of the target word where necessary. All acoustic and articulatory measurements that are described in the following correspond to the vowel of the target vowel (i.e., the vowel of the penultimate syllable of the target word, which was either /a/ or i/).

The acoustic analyses involved measurements of vowel duration, F0, and sound pressure level (SPL). All processing steps for the acoustic data were carried out in R (R Core Team, 2020), predominantly using the packages PraatR (Albin, 2014) and rPraat (Bořil & Skarnitzl, 2016). The code is accessible in the online repository (cf. Data accessibility statement). Duration was determined as the time in milliseconds between the beginning and end of the target vowel, as annotated in Praat. F0 was assessed as the maximum F0 during the target vowel, expressed in semitones relative to a speaker-specific baseline (the 5th percentile of F0 values during the entirety of utterances produced by a speaker). The Praat command "Get maximum..." for Pitch objects was applied, with floor and ceiling values adjusted based on speaker gender and speaking style. To enable SPL analyses independent of recording gain, a reference tone was played during the recording at the beginning of each block (the habitual and the loud block) at a constant volume and distance from the microphone. SPL was computed as the RMS amplitude of the target vowel relative to the RMS amplitude of the reference tone for the respective speaker and recording block. The RMS amplitude was extracted using the Praat command "Get root mean square..." for the respective time window, and SPL was calculated using the formula:

$$SPL_{target} = 20 \cdot \log_{10} \frac{RMS_{target}}{RMS_{reference}}$$

The articulatory data of the lips and tongue body underwent several processing steps. Initially, using the NormPos program provided by Carstens Medizinelektronik GmbH, the data were corrected for head movement and transposed for each speaker into a head-based coordinate system based on the biteplate recording. Subsequently, the kinematic signal from the electromagnetic articulograph was processed using the ema2wav converter (Buech et al., 2022) to transform it into multi-channel wav files, which were then displayable in Praat, and CSV files containing kinematic data, which were used for the analyses described below. The signal was smoothed using a rolling mean with a window width of 20 samples.

For the articulatory analyses, we investigated the lips (labial system) and tongue body (lingual system) during the target vowel. We analyzed three articulatory parameters, as depicted in **Figure 2**: lip aperture, vertical tongue body, and horizontal tongue body. For the labial system, we calculated the Euclidean distance between the upper and lower lip, considering both vertical (y-axis, high–low dimension) and horizontal (x-axis, front–back dimension) movements. Regarding the lingual system, we examined tongue body kinematics in vertical and horizontal movement dimensions separately. For both articulator systems, we detected the *target* of the vowel gesture, employing a semi-automatic algorithm in R (cf. Data accessibility statement for the code). **Figure 3** illustrates this procedure for an exemplary target word with the target vowel /i/. For the lips, the algorithm detected the target time as the first zero crossing in the

velocity curve of the lip distance following the peak velocity of the gesture. For the tongue body, the target was detected as the time of the first minimum in the tangential velocity curve following the tangential peak velocity. With this procedure, a target time was identified in all three articulatory parameters (lip aperture, vertical tongue body, and horizontal tongue body) for each target vowel production.

All automatically detected target times underwent manual verification and correction, if necessary. Corrections were made when trajectories exhibited irregularities, such as multiple velocity peaks, leading to premature "target" detection before the vowel onset. In 29 out of 1,908 utterances, erroneous detections were manually corrected to the visually identified accurate target position by the first author. Subsequently, the corresponding articulator positions were extracted at these points in time (cf. Data accessibility statement for the code). For the lips, this represented the maximum opening during the vowel. For the tongue, this corresponded to a low and central/retracted position in /a/ and a high and fronted position in /i/. It is important to note that the target measure employed here differs from the measure of displacement, which is commonly used in articulatory analyses as the spatial distance between the onset and target position of an articulatory gesture. We opted for the target measure, as it solely considers the endpoint of the articulatory movement, which is associated with the maximum position of the articulator in the oral tract to produce the target vowel. Other than displacement, the target measure does not depend on the onset position of the articulator, which is associated with the preceding vowel instead of the target vowel.
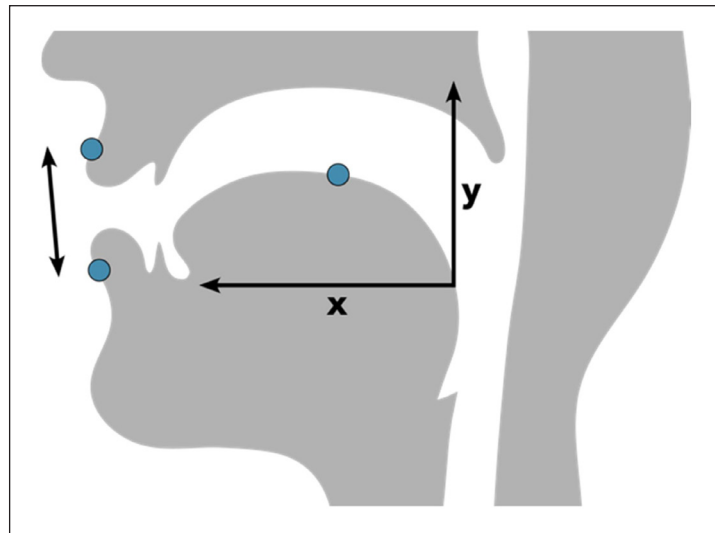

Figure 2: Schema of articulatory parameters. Lip aperture is calculated as the Euclidean distance between the two lip sensors. The tongue body is analyzed on the vertical y-axis (high–low) and the horizontal x-axis (front–back).
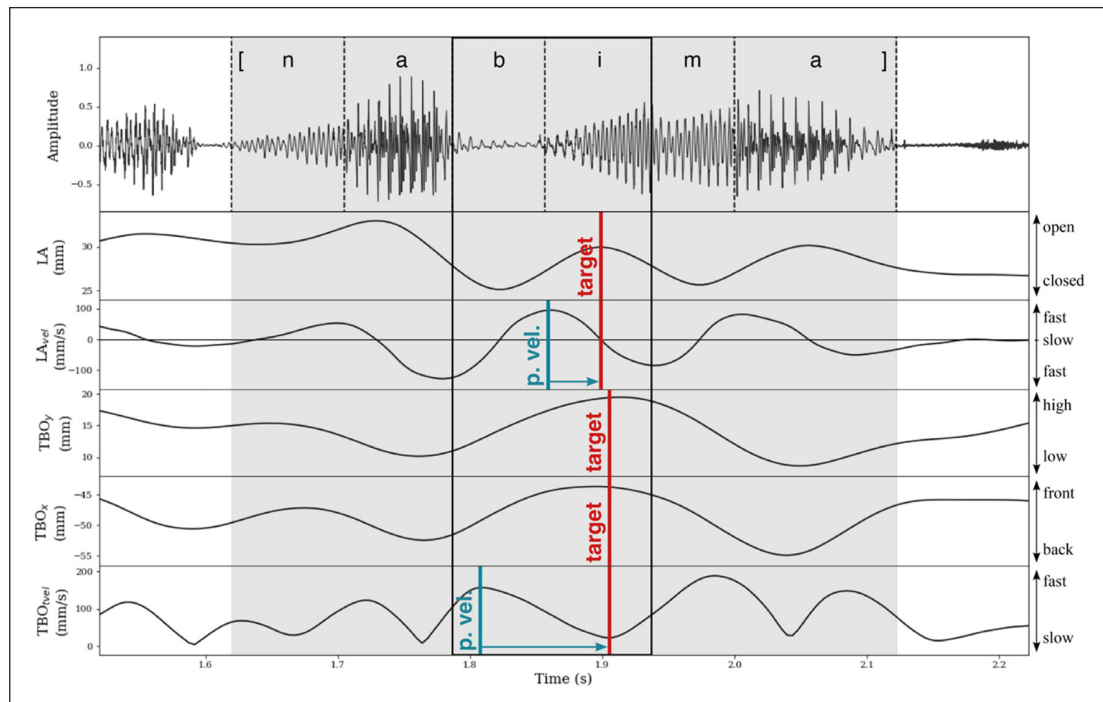
Figure 3: Illustration of algorithm used to find articulatory targets. From top to bottom: Waveform with acoustic annotation, lip aperture (LA), velocity of lip aperture (LA$_{vel}$), vertical position of tongue body (TBO$_y$), horizontal position of tongue body (TBO$_x$), and tangential velocity of tongue body (TBO$_{vel}$). Blue vertical lines indicate the time of the peak velocity in the velocity curves. Red vertical lines indicate the time of the target.

2.6 Bayesian modeling

The acoustic and articulatory results were modeled using Bayesian hierarchical linear models with the brms package (Bürkner, 2017) in R (R Core Team, 2020). Throughout the analysis, the tidyverse package was used for data processing (Wickham et al., 2019), and the plots were created with ggplot2 (Wickham, 2016), tidybayes (Kay, 2023) and patchwork (Petersen, 2020). The complete modeling code can be accessed in the provided online repository (cf. Data accessibility statement). Each acoustic parameter (duration, maximum F0, and mean SPL), and articulatory parameter (lip aperture target, vertical tongue body target, and horizontal tongue body target) was modeled as a function of the factors focus ([broad, broad] vs. [background, corrective]) and speaking style (habitual vs. loud), as well as their two-way interaction. Random intercepts were included for speaker and word, and random slopes for the effects of focus and speaking style by speaker. The reference levels for the predictors were set to '[broad, broad]' for focus and 'habitual' for speaking style. Note that response variable values (i.e., acoustic or articulatory parameter values) went into the analysis without additional standardization (e.g., z-score) to retain as much natural variation in the data as possible.

We used weakly informative priors for the regression coefficients, with normal distributions centered around zero with a standard deviation of 10. The brms package's default priors were employed for all other parameters. Each model was sampled using four Markov Chain Monte Carlo (MCMC) chains with 9,000 iterations, including 4,500 warmup iterations. This process yielded 18,000 posterior samples per model for inference. We conducted thorough checks for convergence issues, which revealed no divergent transitions or $\hat{R}$ values above 1.01. Additionally, visual inspection of the predictive posterior checks showed no irregularities.

The results were modeled for each acoustic and articulatory parameter (i.e., three acoustic and three articulatory parameters). Models were fitted separately for each utterance position (i.e., medial and initial). Additionally, for the articulatory parameters, separate models were fitted for each target vowel (i.e., /a/ and /i/). This approach was adopted because the effects of focus and speaking style on the articulatory target were anticipated to vary depending on vowel quality and utterance position. For example, a shift in focus from [broad, broad] to [background, corrective] might result in a more extreme target in the medial position but a less extreme target in the initial position. Also, a more extreme target may be associated with a lower tongue position in the vowel /a/ but a higher tongue position in the vowel /i/. Conversely, for acoustic parameters, no relevant differences were anticipated between the two vowels, so their data were merged for modelling purposes. As a result, we compiled six models for acoustic parameters (3 parameters $\times$ 2 utterance positions) and 12 models for articulatory parameters (3 parameters $\times$ 2 utterance positions $\times$ 2 target vowels), all following the same basic structure. The subsequent summary presents the models that were fitted for each acoustic or articulatory parameter, along with the number of observations in the data that were included in the respective model:

For acoustic parameters:
- Medial: Target words in medial utterance position (954 observations)
- Initial: Target words in medial utterance position (954 observations)

For articulatory parameters:
- Medial /a/: Target words with target vowel /a/, in medial utterance position (476 observations)
- Medial /i/: Target words with target vowel /i/, in medial utterance position (478 observations)
- Initial /a/: Target words with target vowel /a/, in initial utterance position (477 observations)
- Initial /i/: Target words with target vowel /i/, in initial utterance position (477 observations)

The 2 $\times$ 2 factorial design with the two categorical predictors focus and style is depicted in **Figure 4**. To address the two research questions, three comparisons between cells in the design

matrix were conducted, as indicated by the blue arrows in the figure. To address research question A), general differences between speaking style were explored by comparing cells I and III in the focus condition [broad, broad]. This comparison aimed to establish a foundational understanding of production differences between habitual and loud speech in the present data set (cf. Section 3.1). To address research question B), differences between focus types were examined: initially for habitual speech by comparing cells I and II, followed by an examination of loud speech by comparing cells III and IV (cf. Section 3.2).
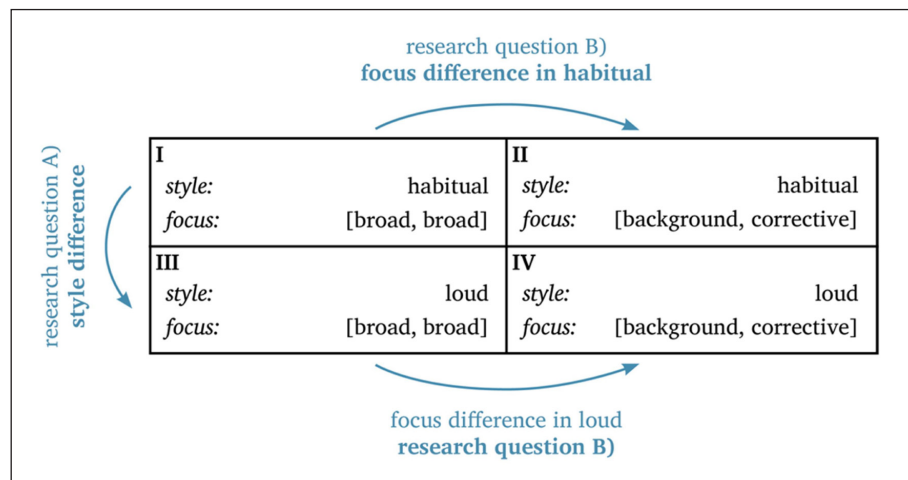
Figure 4: Schematic factorial design matrix with factors focus (horizontal) and speaking style (vertical).

The results for the three acoustic parameters (duration, F0, and SPL) and the three articulatory parameters (lip aperture target, vertical tongue body target, and horizontal tongue body target) are consistently presented in terms of the estimated mean difference of the respective cell comparison ($\hat{\beta}$), a posterior distribution and 95% credible interval (*CI*) of this difference, along with the probability that the difference is above or below zero, $Pr(\hat{\beta} > 0)$ or $Pr(\hat{\beta} < 0)$. We consider there to be reliable evidence for a hypothesis if the probability $Pr(\hat{\beta} > 0)$ or $Pr(\hat{\beta} < 0)$ is close to one, if nearly all posteriors for the difference $\hat{\beta}$ are distributed on one side of zero, and if zero is not within the 95% credible interval of $\hat{\beta}$. Given our decision not to adhere to a strict threshold for interpreting effect sizes, the results are evaluated in a graded manner.

3 Results

In a first step, loud speech is compared to habitual speech to address research question A) (cf. Section 3.1). In a second step, differences between focus structures are examined in the two speaking styles to tackle research question B) (cf. Section 3.2). The results are reported for the Bayesian models of acoustic duration, maximum F0, and relative sound pressure level (SPL), as

well as articulatory targets of lip aperture, vertical tongue body, and horizontal tongue body target, each during the target vowels /a/ or /i/. The included plots consistently illustrate the results for the analyzed parameter with estimated conditional means and their 95% credible intervals in the left panel, and the posterior distribution of the difference $\hat{\beta}$ between these estimates, including a mean (represented as a black dot) and a quantile-based 95% credible interval (represented as a black horizontal line), as well as the probability of $\hat{\beta}$ being above or below zero in the right panel.

3.1 Correlates of loud speech

This section aims to provide answers to research question A). To capture general differences between speaking styles without the influence of focus structure, the parameters are compared between speaking styles in the baseline condition [broad, broad] only. First, results for the acoustic parameters of duration, maximum F0, and SPL during the target vowel are presented. We then turn to the articulatory results for the lip aperture target (associated with the position of maximum opening during the target vowel) and the tongue body target (corresponding to a low and central-retracted position in the vowel /a/ and a high and fronted position in the vowel /i/).

3.1.1 Acoustic parameters

Figure 5 illustrates the estimated conditional means of vowel duration (on the left side) and the posterior distribution of the difference between these estimates, including a mean and 95% credible interval (on the right side). In both utterance positions, the estimates of duration are higher in loud speech compared to habitual speech, indicating a positive difference $\hat{\beta}$ between them (medial: $\hat{\beta} = 34.38$; initial: $\hat{\beta} = 25.80$). This suggests that the target vowel is produced with a longer duration in a loud style than in a habitual one, regardless of utterance position. The posterior distribution for $\hat{\beta}$ demonstrates that the entire posterior mass falls onto the positive side of zero in both utterance positions, and no 95% credible intervals cross zero (medial: lower CI = 26.93, upper CI = 40.78; initial: lower CI = 18.63, upper CI = 32.17). The probability of the difference being above zero ($Pr(\hat{\beta} > 0)$) is 1.00 in both cases. In summary, the difference $\hat{\beta}$, its posterior distribution, and credible intervals provide robust evidence that vowel duration is increased in loud speech compared to habitual speech in both utterance positions.

The model results for maximum F0 during the vowel are displayed in **Figure 6**. The estimates reveal clear differences $\hat{\beta}$ between speaking styles (medial: $\hat{\beta} = 12.21$; initial: $\hat{\beta} = 11.35$), indicating higher F0 maxima in loud speech compared to habitual speech. The posterior distributions illustrate the robustness of these differences, since the entire mass of the distribution falls onto the positive side of zero and the 95% credible intervals do not cross zero (medial: lower CI = 10.87, upper CI = 13.53; initial: lower CI = 10.04, upper CI = 12.66). The probability that the difference $\hat{\beta}$ is above zero ($Pr(\hat{\beta} > 0)$) is 1.00 in both utterance positions.
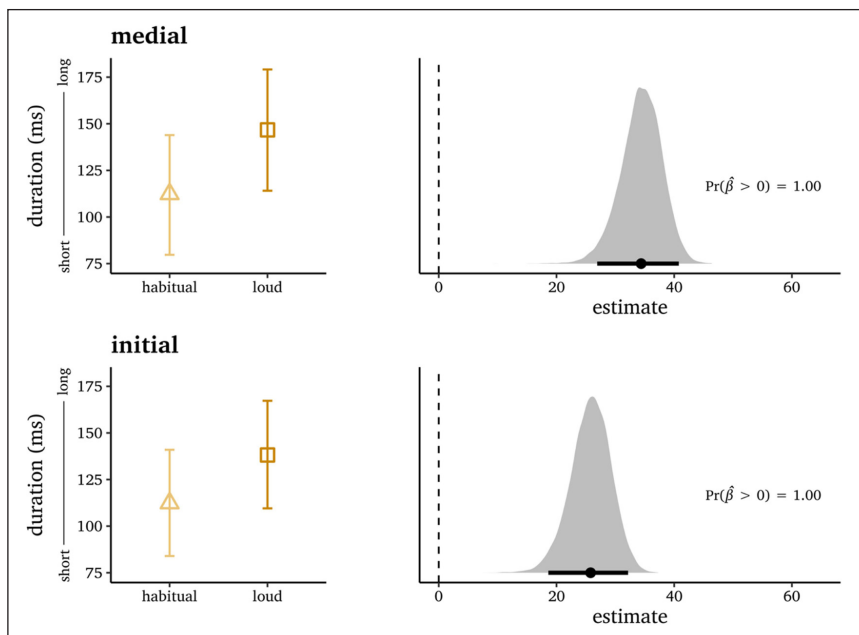
Figure 5: Correlates of speaking style in **vowel duration**. Left: Estimated conditional means for habitual and loud speech in [broad, broad]. Right: Posterior distributions of the difference between these estimates.
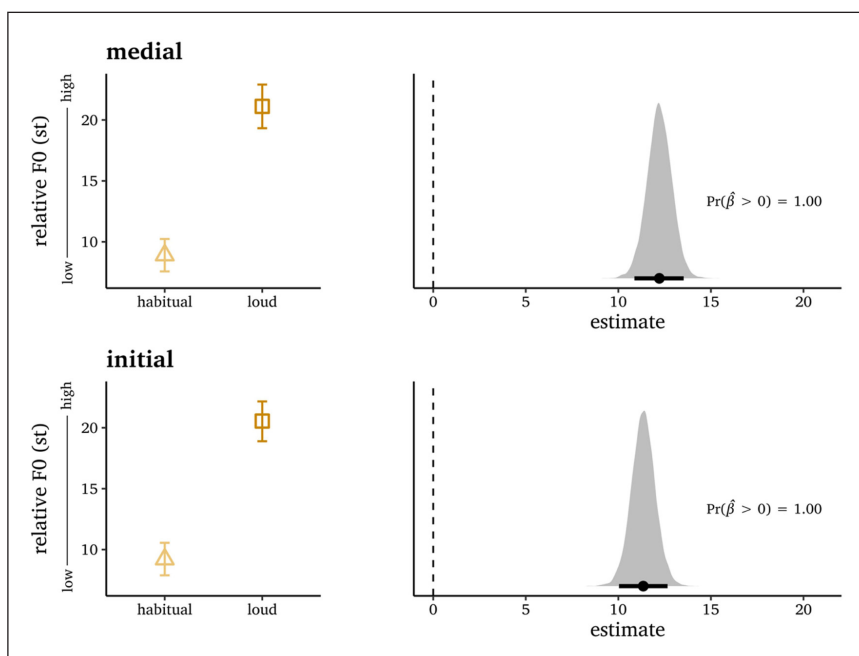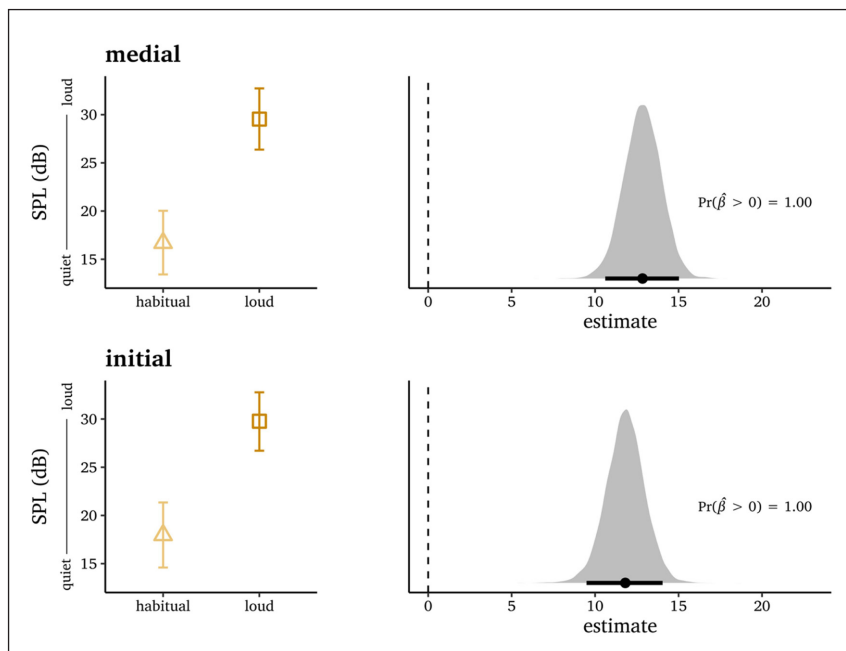

Figure 6: Correlates of speaking style in **maximum F0 of vowel**. Left: Estimated conditional means for habitual and loud speech in [broad, broad]. Right: Posterior distributions of the difference between these estimates.

Figure 7 illustrates the results for the mean SPL of the vowel relative to the reference tone. For this parameter, too, the estimates differ vastly between speaking styles, with the difference $\hat{\beta}$ (medial: $\hat{\beta} = 12.84$; initial: $\hat{\beta} = 11.81$) indicating a higher SPL in loud speech compared to habitual speech in both utterance positions. The entire mass of the posterior distribution of $\hat{\beta}$ falls onto the positive side of zero and the 95% credible interval does not cross zero in either utterance position (medial: lower CI = 10.61, upper CI = 15.02; initial: lower CI = 9.48, upper CI = 14.05). The posterior probability of $Pr(\hat{\beta} > 0) = 1.00$ demonstrates a robust effect of speaking style for both utterance positions.
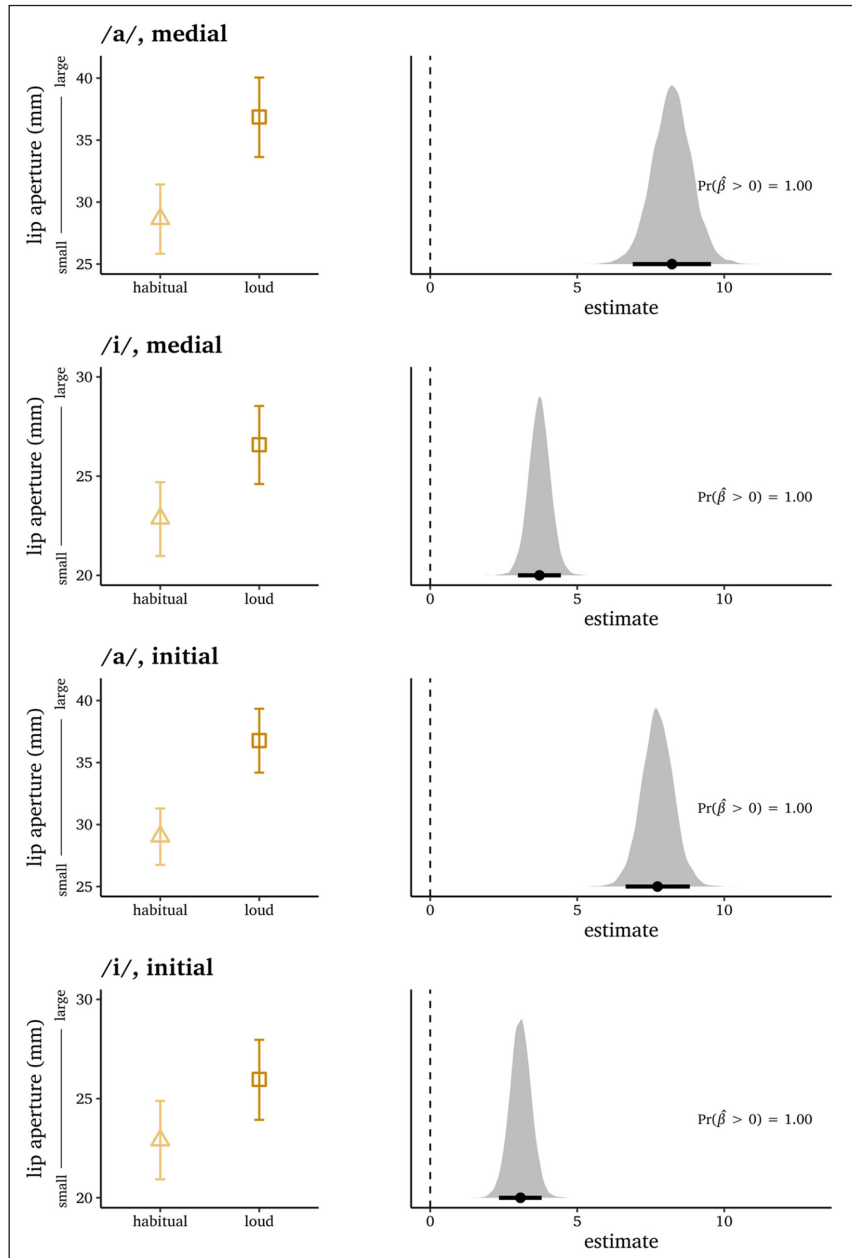

Figure 7: Correlates of speaking style in **mean SPL of vowel**. Left: Estimated conditional means for habitual and loud speech in [broad, broad]. Right: Posterior distributions of the difference between these estimates.

3.1.2 Articulatory parameters

Now we turn to the articulatory results of lip aperture, as depicted in **Figure 8**. In all four models, i.e., in both vowels and both positions, the estimates of the lip aperture target are higher in loud speech than in habitual speech, corresponding to a positive difference $\hat{\beta}$ between them (medial /a/: $\hat{\beta} = 8.22$; medial /i/: $\hat{\beta} = 3.72$; initial /a/: $\hat{\beta} = 7.73$; initial /i/: $\hat{\beta} = 3.07$). This indicates that the lip aperture during vowel production is greater in a loud than in a habitual speaking style in both utterance positions and vowels. It can be observed that this increase in lip aperture in loud speech is greater in /a/ than in /i/, as indicated by higher $\hat{\beta}$ values in this vowel.

The posterior distribution for $\hat{\beta}$ shows that the entire posterior mass falls onto the positive side of zero in all four models, and no 95% credible intervals cross zero (medial /a/: lower CI = 6.89,

upper CI = 9.55; medial /i/: lower CI = 2.98, upper CI = 4.44; initial /a/: lower CI = 6.65, upper CI = 8.83; initial /i/: lower CI = 2.34, upper CI = 3.79). The probability of the difference being above zero ($Pr(\hat{\beta} > 0)$) is 1.00 in all cases. Taken together, the difference $\hat{\beta}$, its posterior distribution, and credible intervals can be interpreted as robust evidence that lip aperture is increased in loud speech compared to habitual speech across all vowels and utterance positions.

Figure 8: Correlates of speaking style in **lip aperture targets**. Left: Estimated conditional means for habitual and loud speech in [broad, broad]. Right: Posterior distributions of the difference between these estimates.

Figure 9 illustrates the results for the correlates of loud speech in tongue body targets. Once again, the results are displayed in terms of estimated conditional means (on the left) and posterior distributions of the difference between these estimated means (on the right). The estimates on the left side can be interpreted similarly to a vowel chart, where high values on the y-axis correspond to high tongue positions and values on the right of the x-axis correspond to retracted tongue body positions (cf. **Figure 2** for an illustration). When abbreviating the two movement dimensions, y will denote vertical (high–low) movements, and x will denote horizontal (front–back) movements. It is important to note that the x-axis is flipped in the estimate plots to match the visual display of the vowel chart, meaning that higher values correspond to front tongue positions when inspecting the raw x-numbers.

In all four models, the estimated means indicate that the overall tongue position is lower (i.e., lower y-values) and fronted (i.e., higher x-values) in loud speech compared to habitual speech (medial /a/: $\hat{\beta}_y = -2.58$, $\hat{\beta}_x = 1.14$; medial /i/: $\hat{\beta}_y = -1.10$, $\hat{\beta}_x = 1.03$; initial /a/: $\hat{\beta}_y = -2.57$, $\hat{\beta}_x = 0.61$; initial /i/: $\hat{\beta}_y = -1.12$, $\hat{\beta}_x = 0.80$). Although this is consistent across both utterance positions and vowels, there are some differences between vowels: In /a/, there is a greater lowering of the tongue associated with loud speech than in /i/, as indicated by more extreme negative $\hat{\beta}_y$ values.

When examining the posterior distributions of $\hat{\beta}$ for the vertical movement dimension, it is observed that the entire posterior mass falls onto the negative side of zero and the 95% credible interval is outside of zero in all four cases, i.e., in /a/ and /i/ in both utterance positions (medial /a/: lower $CI_y = -3.62$, upper $CI_y = -1.55$; medial /i/: lower $CI_y = -1.53$, upper $CI_y = -0.66$; initial /a/: lower $CI_y = -3.53$, upper $CI_y = -1.64$; initial /i/: lower $CI_y = -1.60$, upper $CI_y = -0.63$). The probability of the estimate being below zero, $Pr(\hat{\beta}_y < 0)$, is 1.00 in all cases. For the horizontal movement dimension, the posterior mass falls onto the positive side of zero and the 95% credible interval is outside of zero in three out of four cases, i.e., in medial /a/, medial /i/ and initial /i/ (medial /a/: lower $CI_x = 0.07$, upper $CI_x = 2.23$; medial /i/: lower $CI_x = 0.26$, upper $CI_x = 1.82$; initial /a/: lower $CI_x = -0.45$, upper $CI_x = 1.66$; initial /i/: lower $CI_x = 0.09$, upper $CI_x = 1.49$). The probability of the estimate being above zero is very high in these three cases (medial /a/: $Pr(\hat{\beta}_x > 0) = 0.98$, medial /i/: $Pr(\hat{\beta}_x > 0) = 0.99$, initial /i/: $Pr(\hat{\beta}_x > 0) = 0.99$) but indicates only a trend in the case of initial /a/ $(Pr(\hat{\beta}_x > 0) = 0.88)$. All things considered, there is compelling evidence for the tongue body to be lower in /a/ and /i/ in loud speech than in habitual speech in both utterance positions. Additionally, in loud speech, there is a robust effect of fronting in /i/ in both utterance positions and in /a/ in the medial position, as well as a trend for fronting in /a/ in the initial position.
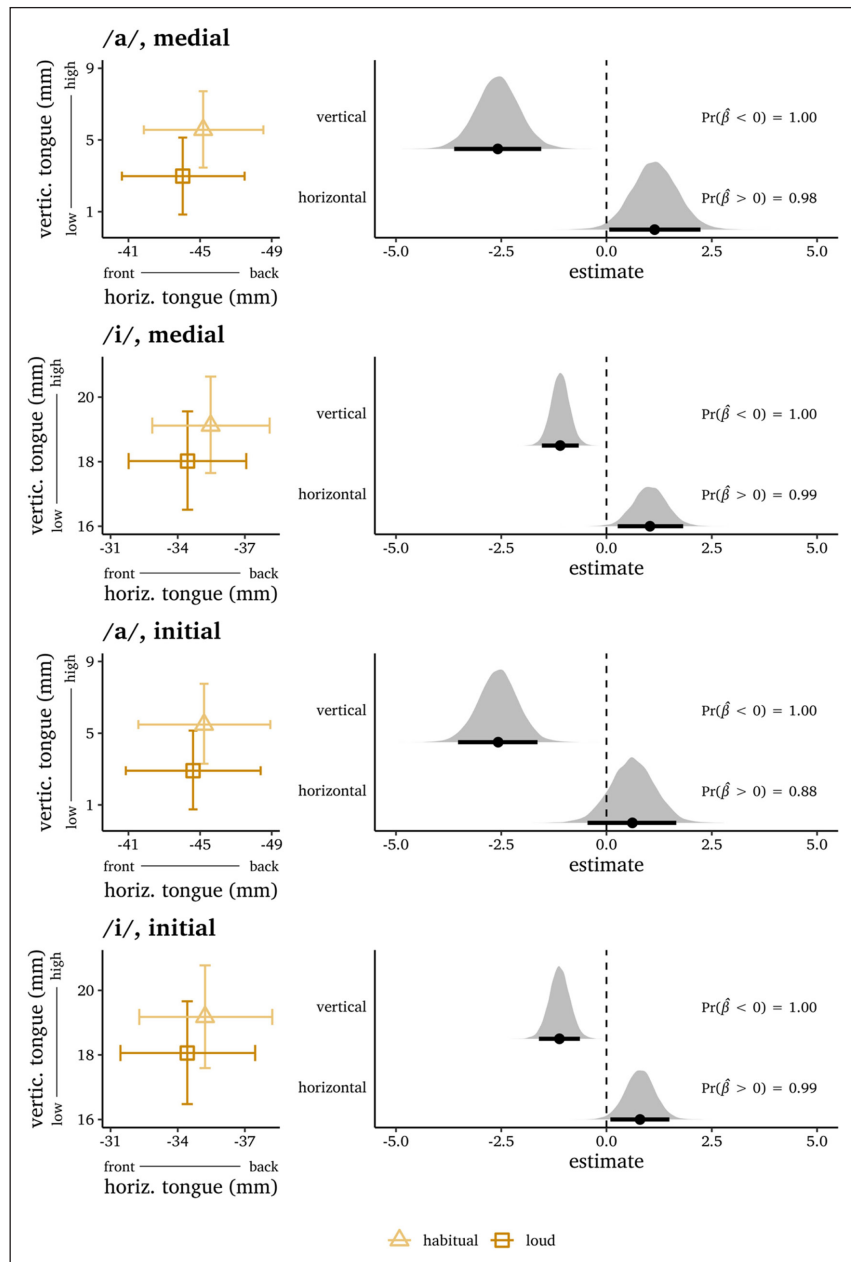
Figure 9: Correlates of speaking style in **tongue body targets**. Left: Estimated conditional means for habitual and loud speech in [broad, broad]. Right: Posterior distributions of the difference between these estimates.

3.2 Correlates of focus in habitual and loud speech

This step of the analysis addresses research question B). We present results for the acoustic parameters of duration, F0, and SPL, as well as for the articulatory targets of the lip aperture, vertical, and horizontal tongue body movements, and compare them between focus conditions in both habitual and loud speech. Within each speaking style, differences are investigated between the condition [broad, broad] (or abbreviated to [br, br]) and the condition [background, corrective] (or abbreviated to [ba, co]). The analysis includes both utterance positions, capturing a change

from broad to corrective focus in the medial position and from broad focus to the background in the initial position. First, acoustic results are presented for vowel duration, maximum F0, and mean SPL. Next, we turn to articulatory results for the lip aperture target (which represents the position of maximum opening during the target vowel) and the tongue body target (which corresponds to a low and central/retracted position in the vowel /a/ and a high and fronted position in the vowel /i/).

3.2.1 Acoustic parameters

Results for the vowel duration are displayed in **Figure 10**. In the medial utterance position, the estimated means show only subtle differences between focus conditions in habitual speech but more robust differences in loud speech (medial habitual: $\hat{\beta} = 1.40$; medial loud: $\hat{\beta} = 2.67$). These observations are supported by the posterior distributions, credible intervals, and posterior probabilities (medial habitual: lower CI = –1.25, upper CI = 4.02; medial loud: lower CI = –0.03, upper CI = 5.35). They indicate that vowel durations in the medial position are longer in corrective focus than in broad focus in loud speech ($Pr(\hat{\beta} > 0) = 0.97$), but there is only a trend for longer durations in habitual speech ($Pr(\hat{\beta} > 0) = 0.85$).

In the initial position, we observe that the estimated means differ markedly between focus conditions in both speaking styles (initial habitual: $\hat{\beta} = -8.53$; initial loud: $\hat{\beta} = -11.52$). The posterior distributions, credible intervals, and posterior probabilities emphasize the robustness of the focus effect in both speaking styles (initial habitual: lower CI = –13.48, upper CI = –3.49; initial loud: lower CI = –16.55, upper CI = –6.37). Consequently, we find that vowel durations are reliably shorter in the background than in broad focus in both habitual and loud speech (habitual: $Pr(\hat{\beta} < 0) = 1.00$; loud: $Pr(\hat{\beta} < 0) = 1.00$).
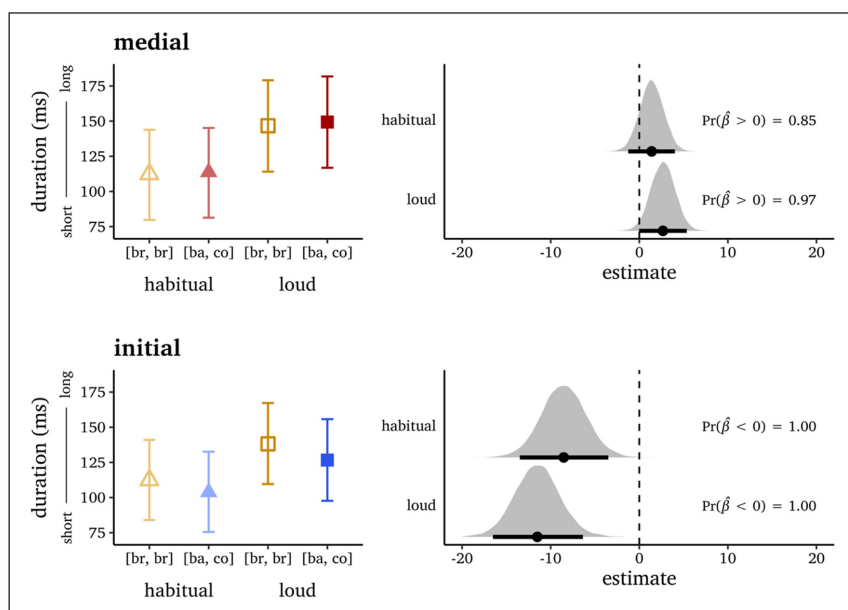

Figure 10: Correlates of focus in **vowel duration**. Left: Estimated conditional means for both focus conditions in habitual and loud speech. Right: Posterior distributions of the difference between these estimates.

Figure 11 presents model results for the maximum F0 of the target vowel. In the medial position, between-focus differences in the estimated means are more pronounced in habitual than in loud speech (medial habitual: $\hat{\beta}$ = 0.70; medial loud: $\hat{\beta}$ = 0.19). The posterior distributions and credible intervals confirm that while the differences in F0 between broad and corrective focus are robust in habitual speech, they are less reliable in loud speech (medial habitual: lower CI = 0.26, upper CI = 1.13; medial loud: lower CI = –0.23, upper CI = 0.63). We find that with a posterior probability of $Pr(\hat{\beta} > 0)$ = 1.00, vowels are robustly produced with a higher maximum F0 in corrective than in broad focus in habitual speech, whereas there is only a subtle trend into the same direction in loud speech ($Pr(\hat{\beta} > 0)$ = 0.81).

In the initial position, marked between-focus differences can be observed in both speaking styles, indicating that maximum F0 is lower in the background than in broad focus (initial habitual: $\hat{\beta}$ = –0.62; initial loud: $\hat{\beta}$ = –0.58). Since the entire mass of the posterior distribution falls onto the negative side of zero, and neither 95% credible interval crosses zero (initial habitual: lower CI = –0.99, upper CI = –0.26; initial loud: lower CI = –0.95, upper CI = –0.21), we find that maximum F0 is reliably decreased in the background compared to broad focus in both habitual and loud speech (habitual: $Pr(\hat{\beta} < 0)$ = 1.00; loud: $Pr(\hat{\beta} < 0)$ = 1.00).
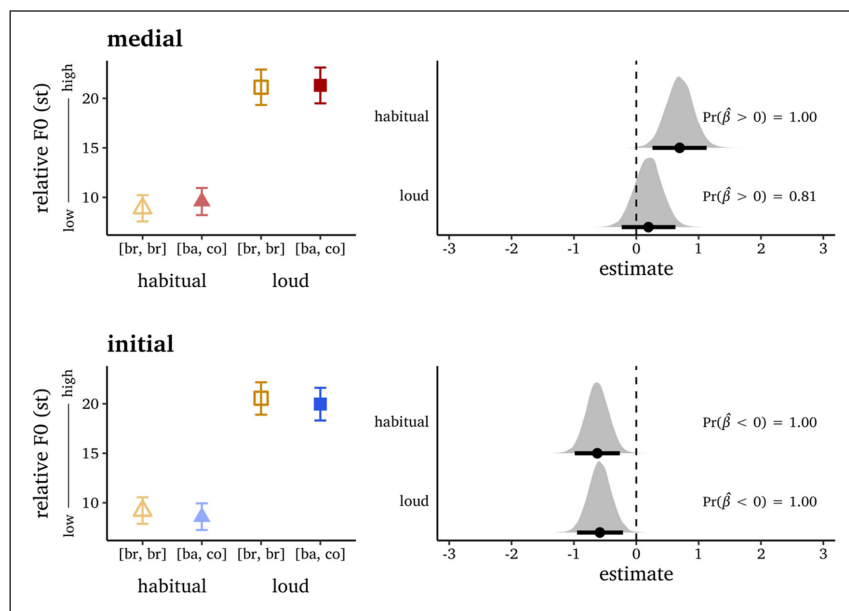

Figure 11: Correlates of focus in **maximum F0 of vowel**. Left: Estimated conditional means for both focus conditions in habitual and loud speech. Right: Posterior distributions of the difference between these estimates.

Results for the third and last acoustic parameter, mean SPL, are presented in **Figure 12**. In the medial position, estimated means are higher in corrective than in broad focus in habitual and loud speech (medial habitual: $\hat{\beta} = 0.51$; medial loud: $\hat{\beta} = 0.50$). In the posterior distributions and credible intervals, we find robust evidence for between-focus differences in both speaking styles (medial habitual: lower CI = 0.02, upper CI = 0.99; medial loud: lower CI = 0.01, upper CI = 0.98). The posterior probability that the target vowel is produced with a higher SPL in corrective focus than in broad focus is $Pr(\hat{\beta} > 0) = 0.98$ in both speaking styles.

In the initial position, the estimated means are only slightly lower in the background than in broad focus in both habitual and loud speaking styles (initial habitual: $\hat{\beta} = -0.19$; initial loud: $\hat{\beta} = -0.28$). The posterior distributions, 95% credible intervals, and posterior probabilities show that these between-focus differences of decreased SPL are only trends and not robust in either habitual or loud speech, but the trend is more pronounced in loud speech (initial habitual: lower CI = -0.62, upper CI = 0.25, $Pr(\hat{\beta} < 0) = 0.80$; initial loud: lower CI = -0.71, upper CI = 0.16, $Pr(\hat{\beta} < 0) = 0.89$).
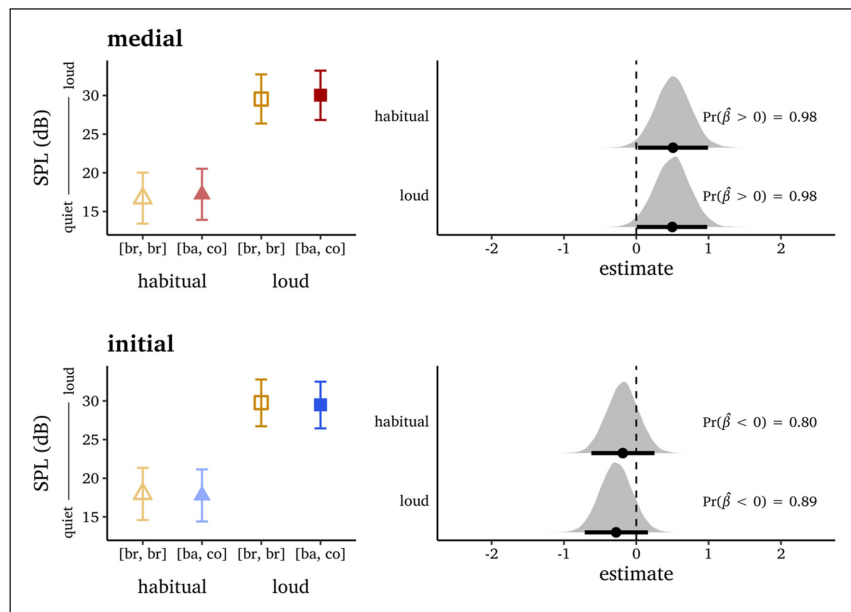

Figure 12: Correlates of focus in **mean SPL of vowel**. Left: Estimated conditional means for both focus conditions in habitual and loud speech. Right: Posterior distributions of the difference between these estimates.

3.2.2 Articulatory parameters

We now delve into the articulatory results, starting with the lip aperture (cf. **Figure 13**). For the medial position in habitual speech, the between-focus differences as shown by the differences between estimated means are rather small (medial /a/: $\hat{\beta}$ = 0.08; medial /i/: $\hat{\beta}$ = 0.24). However, upon inspecting the posterior distributions, credible intervals, and posterior probabilities, a clear trend emerges for a greater lip aperture in corrective focus in /i/ but not in /a/ (medial /a/: lower CI = –0.46, upper CI = 0.62, $Pr(\hat{\beta} > 0)$ = 0.62; medial /i/: lower CI = –0.09, upper CI = 0.56, $Pr(\hat{\beta} > 0)$ = 0.93). Conversely, for the medial position in the loud speaking style, the differences between estimated means indicate a consistent increase in lip aperture in corrective focus compared to broad focus (medial /a/: $\hat{\beta}$ = 0.52; medial /i/: $\hat{\beta}$ = 0.45). Examining the posterior distributions of $\hat{\beta}$ for loud speech, it is apparent that most of the posterior mass falls onto the positive side of zero, with the 95% credible interval being either entirely or nearly outside of zero for both vowels (medial /a/: lower CI = –0.02, upper CI = 1.08; medial /i/: lower CI = 0.13, upper CI = 0.77). The robustness of this observation is also reflected in the high probability of the posterior being above zero (medial /a/: $Pr(\hat{\beta} > 0)$ = 0.97; medial /i/: $Pr(\hat{\beta} > 0)$ = 1.00).

For the initial position, we explore differences in lip aperture when the word is in the background as opposed to broad focus. In habitual speech, the between-focus differences in the initial position suggest that lip aperture is smaller in the background (initial /a/: $\hat{\beta}$ = –0.81; initial /i/: $\hat{\beta}$ = –0.43). The posterior distributions, credible intervals, and posterior probabilities provide robust evidence for a smaller lip aperture in the background in habitual speech (initial /a/: lower CI = –1.56, upper CI = –0.06, $Pr(\hat{\beta} < 0)$ = 0.98; initial /i/: lower CI = –0.96, upper CI = 0.10, $Pr(\hat{\beta} < 0)$ = 0.95). For loud speech, the between-focus differences in the initial position go into the same direction and are even more pronounced than in habitual speech, again revealing a smaller lip aperture in the background than in broad focus (initial /a/: $\hat{\beta}$ = –1.24, lower CI = –2.00, upper CI = –0.50, $Pr(\hat{\beta} < 0)$ = 1.00; initial /i/: $\hat{\beta}$ = –1.20, lower CI = –1.74, upper CI = –0.66, $Pr(\hat{\beta} < 0)$ = 1.00).

Taking both utterance positions into account, the evidence suggests that lip aperture is mostly increased when a word is in corrective focus and reduced when it is in the background, compared to broad focus, respectively. While there is only weak evidence for habitual speech in some cases (i.e., only a trend in medial /i/ and no modification in medial /a/), the differences are larger and the evidence is robust in all cases in loud speech.
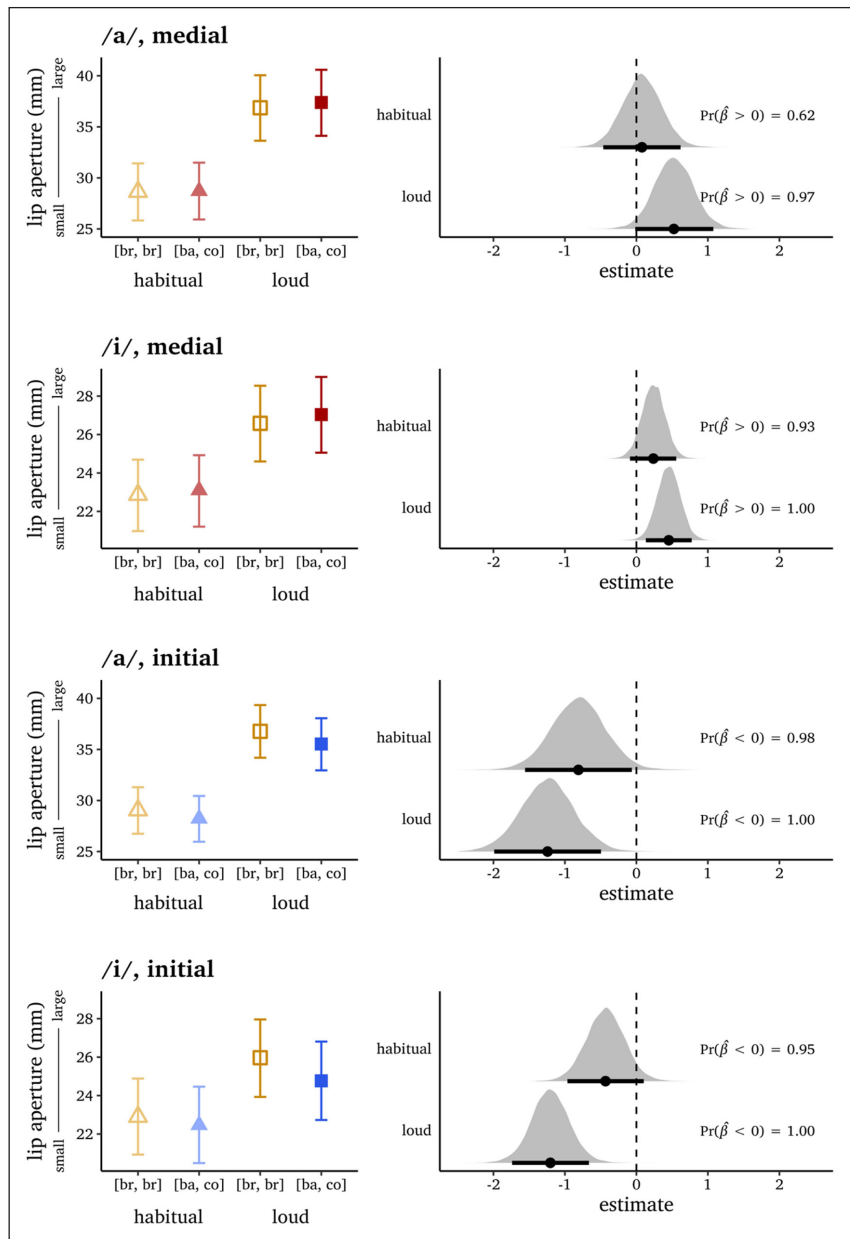
Figure 13: Correlates of focus in **lip aperture targets**. Left: Estimated conditional means for both focus conditions in habitual and loud speech. Right: Posterior distributions of the difference between these estimates.

Results for the tongue body targets are depicted in **Figure 14**. Once more, the estimated means in the left panel can be interpreted akin to vowel charts, where high values on the y-axis correspond to high tongue positions, and values on the right of the x-axis correspond to retracted

tongue positions (cf. **Figure 2** for an illustration). In the medial position, i.e., when comparing corrective to broad focus, habitual speech exhibits no clear modifications between focus types in either vowel or movement dimension (medial /a/: $\hat{\beta}_y$ = 0.10, $\hat{\beta}_x$ = –0.10; medial /i/: $\hat{\beta}_y$ = 0.01, $\hat{\beta}_x$ = 0.09). Posterior distributions, credible intervals, and posterior probabilities affirm that the tongue body target does not change systematically in the medial position in this speaking style (medial /a/: lower CI_y = –0.28, upper CI_y = 0.48, $Pr(\hat{\beta}_y < 0)$ = 0.30, lower CI_x = –0.50, upper CI_x = 0.30, $Pr(\hat{\beta}_x < 0)$ = 0.69; medial /i/: lower CI_y = –0.21, upper CI_y = 0.23, $Pr(\hat{\beta}_y > 0)$ = 0.54, lower CI_x = –0.35, upper CI_x = 0.51, $Pr(\hat{\beta}_x > 0)$ = 0.65).

For the medial position in loud speech, clear differences between focus conditions are observed only in the vertical dimension in /a/, whereas the horizontal dimension in /a/ and both dimensions in /i/ exhibit only minor differences (medial /a/: $\hat{\beta}_y$ = –0.45, $\hat{\beta}_x$ = –0.19; medial /i/: $\hat{\beta}_y$ = 0.03, $\hat{\beta}_x$ = 0.25). Regarding /a/, the posterior distributions, credible intervals, and posterior probabilities provide compelling evidence for a lower tongue body target in corrective focus compared to broad focus in the vertical dimension. For the horizontal dimension, they reveal only a marginal trend toward tongue retraction in corrective focus compared to broad focus (medial /a/: lower CI_y = –0.83, upper CI_y = –0.07, $Pr(\hat{\beta}_y < 0)$ = 0.99), lower CI_x = –0.59, upper CI_x = 0.22, $Pr(\hat{\beta}_x < 0)$ = 0.82). For /i/, posterior distributions, credible intervals, and posterior probabilities emphasize the absence of a systematic between-focus difference in the vertical dimension and show only a trend towards a fronted tongue in the horizontal dimension in corrective focus (medial /i/: lower CI_y = –0.19, upper CI_y = 0.26, $Pr(\hat{\beta}_y > 0)$ = 0.62, lower CI_x = –0.18, upper CI_x = 0.69, $Pr(\hat{\beta}_x > 0)$ = 0.88).

In the initial position in habitual speech, notable modifications can be observed in the background compared to broad focus, namely a higher and fronted tongue in /a/ and a lowered and retracted tongue in /i/ (initial /a/: $\hat{\beta}_y$ = 0.26, $\hat{\beta}_x$ = 0.46; initial /i/: $\hat{\beta}_y$ = –0.26, $\hat{\beta}_x$ = –0.32). The posterior distributions, credible intervals, and posterior probabilities indicate systematic patterns for all these between-focus differences, except for the higher tongue in /a/, where they reveal only a trend (initial /a/: lower CI_y = –0.28, upper CI_y = 0.81, $Pr(\hat{\beta}_y > 0)$ = 0.84, lower CI_x = –0.04, upper CI_x = 0.98, $Pr(\hat{\beta}_x > 0)$ = 0.96; initial /i/: lower CI_y = –0.60, upper CI_y = 0.08, $Pr(\hat{\beta}_y < 0)$ = 0.93, lower CI_x = –0.74, upper CI_x = 0.09, $Pr(\hat{\beta}_x < 0)$ = 0.94).

For the initial position in loud speech, we see a higher tongue in /a/ and a lowered and retracted tongue in /i/ in the background (initial /a/: $\hat{\beta}_y$ = 0.63, $\hat{\beta}_x$ = 0.08; initial /i/: $\hat{\beta}_y$ = –0.37, $\hat{\beta}_x$ = –0.36). The evidence supporting these modifications in the posterior distributions, credible intervals, and posterior probabilities is compelling, whereas there is no clear pattern in initial /a/ in the horizontal dimension (initial /a/: lower CI_y = 0.10, upper CI_y = 1.17, $Pr(\hat{\beta}_y > 0)$ = 0.99, lower CI_x = –0.43, upper CI_x = 0.59, $Pr(\hat{\beta}_x > 0)$ = 0.62; initial /i/: lower CI_y = –0.72, upper CI_y = –0.02, $Pr(\hat{\beta}_y < 0)$ = 0.98, lower CI_x = –0.78, upper CI_x = 0.05, $Pr(\hat{\beta}_x < 0)$ = 0.96).

Taken together, the analysis reveals that in habitual speech, the tongue body target is not systematically modified between corrective focus and broad focus. In loud speech, however, the tongue is lowered and slightly retracted in /a/ in corrective focus, as well as slightly fronted in /i/. In the background compared to broad focus, habitual speech shows a fronted and slightly higher tongue body target in /a/, as well as a lowered and retracted tongue body target in /i/. In loud speech, the target is higher in /a/ and lowered and retracted in /i/ in the background compared to broad focus.
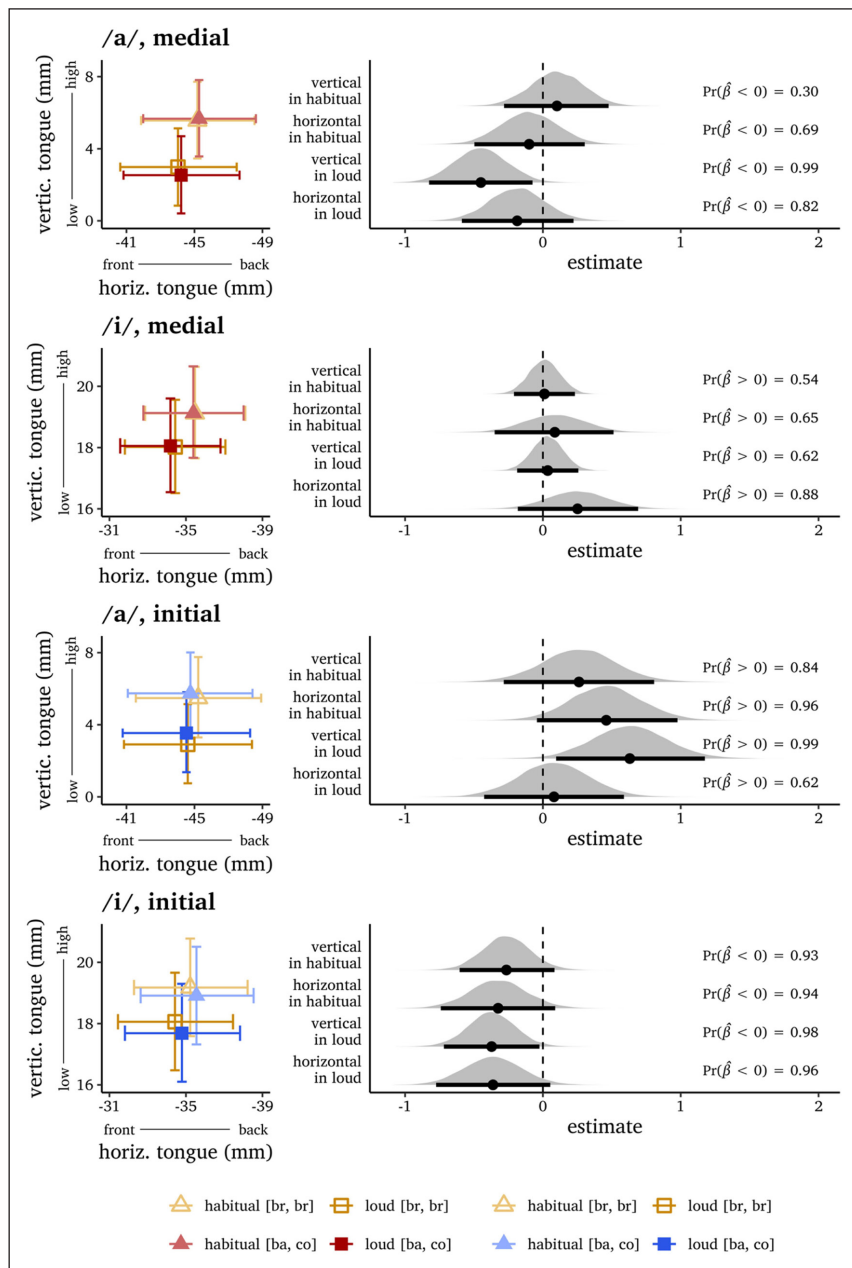

Figure 14: Correlates of focus in **tongue body targets**. Left: Estimated conditional means for both focus conditions in habitual and loud speech. Right: Posterior distributions of the difference between these estimates.

4 Discussion

We carried out a study on acoustic parameters and articulatory kinematics of loud speech and of focus marking in habitual and loud speech. A summary of the results is provided in **Table 2** for research question A), the differences between speaking styles, and in **Table 3** for research question B), the differences between focus conditions.

In **Table 2**, Brackets indicate when only a trend could be observed. The results indicate the characteristics of loud speech compared to the baseline habitual speech (e.g., "greater" means that the given parameter is greater in loud than in habitual speech).

parameter	medial		initial	
	/a/	**/i/**	**/a/**	**/i/**
duration	longer		longer	
max. F0	higher		higher	
mean SPL	louder		louder	
lip aperture	greater	greater	greater	greater
vertical tongue body	lower	lower	lower	lower
horizontal tongue body	fronted	fronted	(fronted)	fronted

Table 2: Summary of results on differences **between speaking styles**.

In **Table 3**, a dash indicates no robust results, and brackets indicate when only a trend could be observed. In the medial position, the results indicate the characteristics of corrective focus compared to the baseline broad focus (e.g., "greater" means that the given parameter is greater in corrective than in broad focus); in the initial position, the results indicate the characteristics of background compared to the baseline broad focus (e.g., "smaller" means that the given parameter is smaller in the background than in broad focus).

parameter	habitual				loud			
	medial corr. vs. broad		**initial backgr. vs. broad**		**medial corr. vs. broad**		**initial backgr. vs. broad**	
	/a/	**/i/**	**/a/**	**/i/**	**/a/**	**/i/**	**/a/**	**/i/**
duration	(longer)		shorter		longer		shorter	
max. F0	higher		lower		(higher)		lower	
mean SPL	louder		(quieter)		louder		(quieter)	
lip aperture	–	(greater)	smaller	smaller	greater	greater	smaller	smaller
vertical tongue body	–	–	(higher)	lower	lower	–	higher	lower
horiz. tongue body	–	–	fronted	retracted	(retracted)	(fronted)	–	retracted

Table 3: Summary of results on differences **between focus conditions** in both speaking styles.

First, the results reveal overall differences between speaking styles regarding acoustic and supra-laryngeal articulatory parameters of target vowels. Loud speech is systematically produced with longer acoustic durations, a higher maximum F0, higher mean sound pressure level (SPL), a greater lip aperture, and a tongue body that is lowered and mostly fronted in both investigated vowels. Second, concerning correlates of focus structure, the results show that in habitual speech, target vowels are produced with a higher maximum F0 and mean SPL, and tend to have longer durations when in corrective as opposed to broad focus. When they are produced in the background as opposed to broad focus, vowels in habitual speech exhibit shorter durations and a lower F0, and tend to have a lower SPL. In loud speech, when vowels occur in corrective focus as opposed to broad focus, they are longer, louder, and tend to have a higher F0. When they occur in the background as opposed to broad focus, they are shorter, have a lower F0, and tend to be produced with a lower SPL. In terms of articulatory kinematics, there is a trend for lip aperture in habitual speech to be greater in corrective than in broad focus only in /i/. Importantly, lip aperture in habitual speech is systematically smaller in the background than in broad focus in both vowels. In loud speech, there is compelling evidence for an enhanced lip aperture in corrective focus compared to broad focus, as well as a reduced aperture in the background compared to broad focus in both vowels respectively, and the observed between-focus differences are overall greater than in habitual speech. Concerning the tongue body target, the data for habitual speech do not show systematic modifications connected to corrective focus compared to broad focus. However, in the background compared to broad focus, the target in habitual speech is fronted and tends to be higher in /a/, and it is lowered and retracted in /i/. In loud speech, the tongue target is lower and tends to be retracted in /a/ and fronted in /i/ in corrective focus compared to broad focus. Additionally, in the background compared to broad focus, it is higher in /a/ and lowered and retracted in /i/. In many instances, between-focus differences in the tongue target tend to be larger, and the evidence is generally stronger in loud speech than in habitual speech.

The primary objective of the study was to investigate focus marking in two speaking styles. The inquiry centered on whether the marking of prominence relations could be maintained in loud speech, a speaking style that is produced with high vocal effort. The findings indicate that not only can correlates of focus be observed in loud speech, but in certain parameters, they appear to be even more pronounced than in habitual speech. In the following section, the results are discussed in greater depth in relation to the two research questions.

4.1 Correlates of loud speech

The study contributes new findings on articulatory correlates of a loud speaking style (irrespective of focus, for now), which provide answers to research question A). In terms of acoustic parameters, our results reveal longer vowel durations, a higher maximum F0, and higher SPL in loud speech compared to habitual speech. These findings suggest that participants in our study successfully

produced deliberately loud speech, despite the laboratory setting. Our results on German add to the existing body of literature on loud speech across various languages, which consistently report increased vowel durations (e.g., Bonnot & Chevrie-Muller, 1991; Geumann, 2001b; Tjaden, Richards, et al., 2013), higher F0 (e.g., Dromey & Ramig, 1998; Geumann, 2001b; Liénard & Di Benedetto, 1999; Raitio et al., 2013), and elevated SPL (e.g., Koenig & Fuchs, 2019; Mefferd & Green, 2010; Tjaden, Richards, et al., 2013). While the increased SPL is inherent to loud speech, the higher F0 may partly be attributed to the tight interconnection between F0 and SPL in physiological and structural ways (Alku et al., 2002; Gramming et al., 1988). The longer vowel durations may further contribute to enhancing the SPL by prolonging the duration of highly sonorant sounds, where the vocal tract is opened wide.

Turning to the articulatory parameters, we find that the lip aperture is greater in loud speech compared to habitual speech, evident in both high /i/ and low /a/ vowels. The findings are in line with previous studies on loud speech (Darling & Huber, 2011; Dromey & Ramig, 1998; Garnier et al., 2022; Huber & Chandrasekaran, 2006; Schulman, 1989). A larger lip aperture contributes to a more open vocal tract, enhancing the sonority of loud vowels and increasing the SPL, which is crucial for a loud speaking style. It has also been proposed that an increased lip opening is correlated not only with increased overall SPL but also with heightened spectral energy in specific frequency bands (Garnier et al., 2006), potentially aiding in differentiating speech from background noise (Garnier & Henrich, 2014). Additionally, as lip movements are typically visible to listeners in face-to-face conversations, the greater lip opening may serve as a listener-driven strategy to enhance intelligibility in the presence of auditory disturbances by exploiting the visual channel (Fitzpatrick et al., 2015; Garnier et al., 2018).

In addition to modifications of the labial system, loud speech is produced with an overall lowered and fronted tongue body in the present data. Interestingly, the lingual system is modified in the same direction in /a/ and /i/, despite their different phonological targets. This suggests that the modifications are not specific to individual vowels in terms of feature enhancement but might instead be generally associated with the production of loud speech – at least for non-back vowels.

The finding of a lower tongue body is in line with an existing study on shouted vowels in German (Xue et al., 2021). Moreover, indirect evidence for a tongue lowering in loud speech is provided by acoustic analyses showing higher first formants in this speaking style (Geumann, 2001b; Koenig & Fuchs, 2019; Tjaden & Wilding, 2004; Xue et al., 2021). A lower tongue position contributes to a greater opening of the vocal tract, thereby potentially contributing to an enhanced SPL. Additionally, the lowered tongue may be necessary for preserving the perceived vowel quality: It has been suggested that the perception of vowels is connected to the relation between F0 and the first formant (Syrdal & Gopal, 1986). Therefore, if F0 is raised, as typically the case in loud speaking styles, the F1 may need to be raised as well to maintain vowel quality

(Geumann, 2001b; Šimko et al., 2016). Another factor could be that the same degree of vowel constriction that is produced in habitual speech, such as in /i/, may result in friction when intraoral pressure is increased in loud speech, hence necessitating a lower tongue position than in habitual speech (Schulman, 1989). Furthermore, if both vowels are lowered in the vowel space in loud speech, the overall perceptual distance between them can be preserved.

The finding of a fronted tongue in loud speech presents challenges in interpretation, as it has not been explicitly described by previous studies on this speaking style, which primarily investigated *contrasts* between vowel types rather than vowel-specific tongue targets. It should be noted that the larger differences reported between /a/ and /i/ in these previous studies (Mefferd, 2017, 2019; Mefferd & Dietrich, 2020; Mefferd & Green, 2010) do not necessarily contradict our finding of a general fronting of vowels. Such increased contrast is possible despite an overall fronting of all vowels, although this aspect has not yet been explored in the present data. Some indirect evidence supporting the fronting of /a/ and /i/ in loud speech can be found in second formant analyses, showing a (trend towards) a higher F2 (Bond et al., 1989; Bond & Moore, 1990; Geumann, 2001a; Huber & Chandrasekaran, 2006; Lu & Cooke, 2008; Tjaden & Wilding, 2004). Additionally, an articulatory study on Lombard speech has reported a general fronting of vowels (Garnier et al., 2018). However, this is not consistent across studies, with some indicating a lower F2 or showing varied and non-significant results for Lombard or loud speech (Bond et al., 1989; Huber et al., 1999; Koenig & Fuchs, 2019). Several hypotheses could explain the fronted tongue position in loud speech. It could be the case that fronting helps enhance the perceptual quality and salience of loud vowels by raising the F2. Alternatively, it may simply be a concomitant of the simultaneous tongue lowering, given the strong physiological link between the two movement dimensions of one articulator. For a more precise interpretation of the fronted tongue position in loud speech, further investigations with a wider set of vowels would be necessary.

4.2 Correlates of focus in acoustic parameters

To answer research question B), we investigated focus structure in both habitual and loud speech. In the acoustic domain, we examined duration, maximum F0, and mean SPL of the target vowel. Our findings for habitual speech are in line with previous research, revealing a higher F0 (e.g., Baumann et al., 2007; Grice et al., 2017; Roessig, Winter, et al., 2022) and higher SPL (e.g., Fowler, 1995; Roessig, Winter, et al., 2022; Wang, 2020), along with a trend for longer durations (e.g., Cao & Zheng, 2006; Cho & McQueen, 2005; de Jong, 1995) in corrective focus compared to broad focus. These parameter modulations can be associated with an increase in prosodic prominence. For focus differences in loud speech, we observe longer durations, a higher SPL, and a trend for higher F0 in corrective focus compared to broad focus, once again suggesting an association with increased prosodic prominence. Remarkably, correlates of increased prosodic prominence are very similar in both speaking styles, despite the overall increase in duration,

F0, and SPL found in loud speech (cf. Section 4.1). This is, to our knowledge, novel evidence, as prior studies have not yet explored focus marking in loud speech (but cf. Cho et al., 2011; Rivers & Rastatter, 1985; Vainio et al., 2012, for insights into interactions between prosodic structure and speaking style).

Another notable contribution of our study is the incorporation of a second utterance position in the analysis, wherein background is compared to broad focus, to investigate prominence *relations*. In this condition, we observe shorter acoustic durations, a lower F0, and a trend for a lower SPL in the background, both in habitual and loud speech. These parameter modulations can be associated with a decrease in prosodic prominence in both speaking styles and they complement the modulations associated with increased prominence in the other utterance position. This underscores that the *attenuation* of words, simultaneously to the strengthening of others, can contribute to the differential production of focus structures. The observation that the modifications associated with decreased prominence are similar in both speaking styles further emphasizes that focus encoding remains feasible in loud speech despite the overarching modifications characteristic of this speaking style.

Interestingly, we observe that in both utterance positions, the effects of focus on the acoustic parameters of duration and SPL tend to be more pronounced in loud speech compared to habitual speech (as indicated by the modeled posterior distributions and posterior probabilities). However, there is an exception for F0, where an opposite pattern can be observed, with weaker effects in loud speech compared to habitual speech. It is possible that prominence-related modulations of F0 are more limited in loud speech due to a physiological relation between F0 and intensity. For a more detailed account on intonational focus marking across speaking styles, we refer readers to Roessig et al. (2022), where a comprehensive analysis of F0 for a subset of the data is presented. It is demonstrated that the variability in pitch accent realization is decreased in loud speech, resulting in a reduced intonational differentiation between broad and corrective focus, and potential physiological factors underlying these findings are discussed.

To sum up, we solidify and extend the existing studies on acoustic correlates of prosodic prominence by adding evidence for German, investigating simultaneous correlates of increased and decreased prominence, and comparing them across speaking styles.

4.3 Correlates of focus in articulatory parameters

Now, we turn to the articulatory results. The analysis of lip aperture yields somewhat unexpected results for habitual speech. While a smaller lip aperture can be observed in the background compared to broad focus in both vowels, there is only a trend for greater lip aperture in words in corrective focus than in broad focus for /i/ and no clear trend at all for /a/. These findings contrast with the majority of existing studies which have found greater lip opening associated with prosodic prominence, which is connected, for example, to corrective focus (Cho, 2005;

Harrington et al., 1995; Mücke & Grice, 2014; Roessig, 2021; Roessig & Mücke, 2019). This greater aperture has been interpreted in previous research as a means of enhancing vowel sonority, enabling a greater amount of acoustic energy to radiate from the mouth (Beckman et al., 1992).

The divergent findings of the present study could potentially be attributed to the experimental design. The focus conditions compared in the medial utterance position, broad and corrective focus, represent two cases where the target word is in focus and receives the nuclear pitch accent. Consequently, these words exhibit only slightly different degrees of prominence. In contrast, previous experiments often compared words in the background with those in corrective focus, representing two highly divergent poles on a scale of prosodic prominence. It could be hypothesized that the differences between the two conditions in the present study are too subtle to elicit clear articulatory modifications in all cases. However, in other studies on German, such modifications still surface between these focus conditions (Mücke & Grice, 2014; Roessig, 2021), leaving the lack of reliable effects in habitual speech in the present study an open question.

Nonetheless, an important insight from our data is that despite the partial subtlety of articulatory differences between nuclear-accented words across focus conditions in habitual speech, marked distinctions still surface when considering the entire utterance. In fact, they can be found in the initial position. The observed smaller lip aperture in the background compared to broad focus can be related to a sonority reduction in the initial position, wherein the acoustic energy of a vowel is reduced. As discussed previously for the acoustic domain, this emphasizes the relevance of the attenuation of words in encoding focus structures – here, the attenuation appears to be even more consistent than the corresponding strengthening in the medial position.

In contrast to habitual speech, our findings for loud speech reveal a dual pattern of articulatory adjustments in both vowels: a reduced lip aperture in the background compared to broad focus, and simultaneously, an enhanced aperture in corrective focus compared to broad focus. This suggests that speakers make use of the wider range of prosodic options on the articulatory tier in loud speech: They do not only (or mainly) attenuate words in the pre-nuclear domain but also strengthen the nuclear-accented word. Notably, this pattern is consistent in both open /a/ and close /i/ vowels. In loud speech, we thus find between-focus modifications in all four examined cases, encompassing both vowels and both utterance positions. Furthermore, all differences are more pronounced and supported by more robust evidence in loud speech compared to habitual speech. This suggests that the loud speaking style does not impede the signaling of focus structure in the labial system compared to a habitual speaking style; rather, it amplifies it. Despite the overall increase in lip aperture associated with a loud speaking style, speakers can still open their lips even further in words with higher prosodic prominence, and conversely, reduce lip aperture in words with lower prominence. This results in greater between-focus differences than those observed in habitual speech.

Now, we discuss the findings regarding the tongue body. In habitual speech, the lingual data reveal no systematic modifications in corrective focus compared to broad focus. As it was the case for habitual speech in the labial system, this absence of systematic lingual modifications was rather surprising, given that previous studies have identified prominence-related differences, such as a lower target in /a, ɑ/ (Cho, 2002; de Jong, 1991; Macchi, 1985) and fronting and/or raising in /i/ (Cho, 2002; Harrington et al., 2000; Katsika, 2018; Kent & Netsell, 1971; Kim & Cho, 2011; Roessig, 2021; Roessig & Mücke, 2019). However, once again, this may be attributed to the subtle differences between focus structures elicited in the present experiment, and robust between-focus differences might emerge in more divergent focus conditions.

On the contrary, for the initial utterance position, there are some clear tongue body differences in the background compared to broad focus in habitual speech. The lowered and retracted tongue body target in /i/ corresponds to a hypoarticulation, or centralization, of the high front vowel, reflecting a decrease in prominence. This implies that the tongue body target produced in the background does not reach the same high front position as in broad focus. This hypoarticulation can be understood as the opposite of the localized hyperarticulation associated with increased prominence (de Jong, 1995), and we might call it *localized hypoarticulation* for structural reasons. In /a/, the trend towards a higher tongue target can also be interpreted as a hypoarticulation of the low vowel. The case is less clear for the observed fronted target in this vowel, as /a/ is a central vowel in German and phonologically neither a front nor a back vowel. In our study, /a/ was always flanked by the front vowel /i/. In this context, a retraction of the tongue in /a/ would correspond to an enhanced contrast between /a/ and /i/, whereas a fronting would correspond to a weakened contrast. Thus, the observed fronting in /a/ in the background condition does not directly equate to a hypoarticulation strategy but is still in accordance with a weakening of contrasts.

In loud speech, unlike habitual speech, we observe clear tongue body modulations in corrective compared to broad focus. In corrective focus, the observed lower (and slightly retracted) target in /a/, as well as the slightly fronted target in /i/, can all be related to the strategy of localized hyperarticulation, enhancing the respective vowel's phonological place features on the paradigmatic axis. It is plausible that the target in /i/ is fronted but not additionally higher because this would conflict with the low tongue body required for an open vocal tract in loud speech. This may indicate that prominence-induced modulations occur preferably in those articulatory dimensions where no conflicts arise with physiological or phonological requirements or sonority expansion (Cho, 2005). In the background, the higher tongue target in /a/, as well as the lowered and retracted target in /i/, are all clear cases of localized hypoarticulations, weakening their characteristic place features. Overall, the between-focus differences are mostly greater, and the evidence is more robust in loud speech than in habitual speech.

4.4 Summary: Correlates of focus across speaking styles

Taking into consideration both the results for the acoustic domain (duration, F0, and SPL) and the articulatory domain (labial and the lingual system), the present study provides evidence that a loud speaking style does not impede the marking of focus structure compared to a habitual speaking style. In fact, quite the opposite is observed in many of the analyzed parameters. In the acoustic domain, although loud speech is produced with altered characteristics, prominence-induced modifications can still be observed in this speaking style, similar to habitual speech, with effects on duration and SPL tending to be even greater than in habitual speech. Similarly, in the articulatory domain, while loud speech seems to require a different overall configuration of the lips and tongue body, this does not prevent the articulators from exhibiting clear differences between focus conditions. These differences are mostly even greater than in habitual speech. This is evident, firstly, in cases where there are no differences in habitual speech, but differences emerge in loud speech (primarily in corrective compared to broad focus), and secondly, in cases where differences are present in both speaking styles, but they are stronger in loud speech.

Our current findings suggest that the effects of focus structure are not impeded by the overarching goal of increasing the sound pressure level. Rather, these effects appear to be cumulative. While speaking loudly may serve various purposes, it is probable that, in our scenario, the primary objective of a loud speaking style is to enhance intelligibility in adverse listening conditions. This objective does not seem to be achieved by simply "scaling up" the vocal effort of all words in an utterance uniformly. Instead, focus structure, which shall facilitate speech comprehension and processing for the listener, appears to be effectively conveyed in this speaking style as well. It is conceivable that the overall heightened production effort in loud speech extends beyond the goal of speaking loudly but encompasses other aspects of speech production, such as focus marking.

However, it should be noted that focus marking may not be uniformly achieved across all parameters. As the analysis of F0 shows, the effect sizes of focus-related modifications of this parameter are smaller in loud speech compared to habitual speech. This implies that some speech production subsystems may retain their flexibility across speaking styles, whereas others (such as intonation) appear to be more constrained. One could speculate that this reduced modulation strength in F0 is partially compensated for by the higher modulation strengths observed in other systems. Further, it could be hypothesized that in loud speech, the multiple parameters of focus encoding are weighted differently in the multidimensional phonetic space.

On a side note, a multimodal analysis of a subset of the data, as presented in Pagel et al. (2023), further corroborates the present findings of focus-related effects in loud speech. We discovered that differences between broad and corrective focus are also encoded in the kinematics of speech-accompanying head movements. While this can be observed in both habitual and loud speech,

the differences in terms of movement displacement and velocity are more pronounced in loud than in habitual speech, despite head movements being generally enhanced in loud speech. This underlines that focus can be encoded in loud speech in a range of dimensions and modalities.

Another important finding that emerges from the results on acoustic parameters, as well as the labial and lingual systems, is that focus structure indeed influences the production of the entire utterance. This has been pointed out by previous studies for intonational and temporal characteristics (Baumann et al., 2007; Dohen & Lœvenbruck, 2004; Féry & Kügler, 2008; Kügler, 2008; Roessig, 2024; Yang & Chen, 2020). The present study adds supporting evidence on supra-laryngeal kinematics, which, to our knowledge, have only been investigated by Im et al. (2023) thus far. While most existing studies report results for words immediately preceding a nuclear-accented word, we extend these findings to words occurring earlier in the utterance. Both the acoustic and articulatory data indicate that words in corrective focus are produced with greater prominence, while words in the background are produced with reduced prominence, compared to broad focus respectively. Thus, we observe not only an articulatory strengthening of the word that receives more prosodic prominence (in the nuclear position), but also a simultaneous weakening, or attenuation, of other words in the utterance (in the pre-nuclear position). The two inverse mechanisms (strengthening and weakening, or for the articulation: localized hyperarticulation and localized hypoarticulation) may both be present in some cases, whereas only one can be observed in others. In our results on lip and tongue kinematics, the attenuation in the pre-nuclear domain appears to be even more robust than the strengthening in the nuclear domain, especially in habitual speech. This is particularly interesting because the pre-nuclear domain of an utterance has been previously described as following a meaningless default pattern (Büring, 2007). Our results contradict this view and underscore the importance of the pre-nuclear domain for signaling focus structure. In line with previous interpretations (e.g., Cho et al., 2013; Erickson & Lehiste, 1995; Im et al., 2023; Roessig, 2024), we propose that the attenuation of pre-nuclear elements may serve as a "boost" for the saliency of the following prominent word in the nuclear position. As suggested by Im et al. (2023) this may simultaneously be listener-oriented (by increasing perceptual saliency) and speaker-oriented (by enhancing motor efficiency).

Overall, the present data emphasize the flexibility of the speech system in the encoding of focus structure. There are numerous potential parameters through which it can be realized, and all of them can but do not necessarily have to be modified simultaneously. This surfaces in complementary processes of strengthening and weakening across different positions of a word within the utterance and, on the other hand, in various acoustic and articulatory parameters. Crucially, this flexibility of the speech system in the encoding of prominence relations in a multidimensional phonetic space is not confined to a habitual speaking style but is maintained, and in some respects even enhanced, in loud speech.

4.5 Limitations and future directions

The present study exclusively focuses on selected acoustic parameters (duration, F0, and SPL) and supra-laryngeal articulation (lips and tongue body), while disregarding other parameters, such as additional articulators or non-verbal cues like speech-accompanying gestures, which may play an important role in the encoding of prosodic structure. A full account on focus marking across speaking styles should encompass a broader range of parameters to fully elucidate the underlying mechanisms at play. In the present study, F0 is modulated as a function of focus type to a smaller extent in loud than in habitual speech. Based on this observation, it is interesting to investigate potential underlying reasons for the differential behavior of this parameter and examine the modulation patterns of further parameters across speaking styles. Moreover, considering speaker-specific differences in addition to aggregate data across speakers could offer a deeper understanding of potential individual strategies. Another limitation of the experimental design could be the selection of two focus conditions in the medial position (broad and corrective focus) that elicit relatively similar degrees of prominence. This might have resulted in less pronounced results compared to focus conditions eliciting highly divergent degrees of prominence. However, we view the choice of less divergent focus conditions as a strength of the study, demonstrating that even subtle differences in prominence degrees can manifest in clear acoustic and articulatory correlates in certain cases.

It should also be noted that we did not decouple the tongue position from the jaw movement, as proposed and implemented in certain previous papers (de Jong, 1991; Geumann, 2001a; Henriques & Van Lieshout, 2013; Mefferd, 2017; Westbury et al., 2002). Therefore, it cannot be ruled out that the lowered tongue body targets found in the data might simply be a concomitant of a lowered jaw in loud speech. Nevertheless, previous research has indicated that the tongue can move relatively independently, if necessary. For example, it can still produce close constrictions despite a lowered jaw (Geumann, 2001a; Macchi, 1985). This suggests that the tongue body can compensate for jaw movement and is passively lowered with the jaw only when the resulting lingual configuration is intended. Nonetheless, future work could move beyond the analysis of composite lingual movements by further investigating the intrinsic tongue position decoupled from the jaw to compare the two measures.

Additionally, our study could be complemented by a perception experiment. Thus far, we have investigated speech production, and can only speculate on how the observed strategies of focus marking may shape the perception of an utterance by listeners. A controlled perception study could shed light on the detectability and relative importance of the focus-related modifications in supra-laryngeal articulation.

5 Conclusion

An articulatory study with 20 German speakers was conducted to investigate the realization of focus structure in acoustic duration, F0, and sound pressure level (SPL), as well as lip and tongue kinematics in habitual and loud speaking styles. The findings reveal that not only can acoustic and articulatory correlates of focus structure be found in loud speech, but the focus-related modifications are partially even stronger compared to habitual speech. In adverse listening conditions, a loud speaking style aims for maximum intelligibility, which appears to include a clear encoding of prosodic structure. Specifically, we generally observe longer vowel durations, higher maximum F0, increased SPL, greater lip aperture, and hyperarticulated tongue body targets associated with increased prosodic prominence, as well as shorter durations, lower maximum F0, decreased SPL, smaller lip aperture, and hypoarticulated tongue body targets associated with decreased prosodic prominence. Notably, in habitual speech, the articulatory modifications associated with decreased prominence are more pronounced than those associated with increased prominence. This highlights the importance of the entire utterance, including the pre-nuclear domain, in focus realization. Conclusively, the study emphasizes the flexibility of the speech production system in the encoding of focus structure across different speaking styles in a multidimensional phonetic space.

Data accessibility statement

All materials, including data and code, are accessible as an online repository via the following link: https://osf.io/svdq3/ (DOI: 10.17605/OSF.IO/SVDQ3).

Competing interests

The authors have no competing interests to declare.

Authors' contributions

All authors contributed to the conceptualization of the study. All authors contributed to the experiment design. Lena Pagel conducted the experiment and data processing. Lena Pagel and Simon Roessig contributed to the analysis. Lena Pagel drafted the manuscript. All authors contributed to manuscript revision. All authors read and approved the submitted version.

References

Albin, A. (2014). An architecture for controlling the phonetics software "Praat" with the R programming language. *Journal of the Acoustical Society of America, 135*(4), 2198.

Alku, P., Vintturi, J., & Vilkman, E. (2002). Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation. *Speech Communication, 38*(3–4), 321–334. https://doi.org/10.1016/S0167-6393(01)00072-3

Alzaidi, M. S. A., Xu, Y., Xu, A., & Szreder, M. (2023). Analysis and computational modelling of Emirati Arabic intonation – A preliminary study. *Journal of Phonetics*, *98*, 101236. https://doi.org/10.1016/j.wocn.2023.101236

Alzamil, A., & Hellmuth, S. (2021). The prosodic realisation of focus in Saudi Arabic dialects in comparative perspective. *Proceedings of the 1st International Conference on Tone and Intonation (TAI), 6–9 December 2021, Sonderburg, Denmark*, 195–199. https://doi.org/10.21437/TAI.2021-40

Andreeva, B., Barry, W. J., & Koreman, J. (2017). Local and global cues in the prosodic realization of broad and narrow focus in Bulgarian. *Phonetica*, *73*(3–4), 256–278. https://doi.org/10.1159/000448044

Avesani, C., Vayra, M., & Zmarich, C. (2007). On the articulatory bases of prominence in Italian. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS), 6–10 August, Saarbrücken, Germany*, 981–984.

Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, *52*(4), 391–413. https://doi.org/10.1177/0023830909336575

Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & R. D. Ladd (Eds.), *Gesture, Segment, Prosody* (pp. 68–89). Cambridge University Press. https://doi.org/10.1017/cbo9780511519918.004

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS), 6–10 August, Saarbrücken, Germany*, 1029–1032.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* [Computer software]. https://praat.org

Bond, Z. S., & Moore, T. J. (1990). A note on loud and lombard speech. *Proceedings of the First International Conference on Spoken Language Processing, 18–22 November, Kobe, Japan*, 969–972. https://doi.org/10.21437/ICSLP.1990-255

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *Journal of the Acoustical Society of America*, *85*(2), 907–912. https://doi.org/10.1121/1.397563

Bonnot, J.-F. P., & Chevrie-Muller, C. (1991). Some effects of shouted and whispered conditions on temporal organization. *Journal of Phonetics*, *19*, 473–483. https://doi.org/10.1016/s0095-4470(19)30339-0

Bořil, T., & Skarnitzl, R. (2016). Tools rPraat and mPraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech and Dialogue* (pp. 367–374). Springer International Publishing.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Buech, P., Roessig, S., Pagel, L., Mücke, D., & Hermes, A. (2022). ema2wav: Doing articulation by Praat. *Proceedings of INTERSPEECH, 18–22 September, Incheon, Korea*. https://doi.org/10.21437/Interspeech.2022-10813

Büring, D. (2007). Semantics, intonation, and information structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp. 445–474). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199247455.013.0015

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Calhoun, S. (2007). *Information structure and the prosodic structure of English: A probabilistic relationship* [Doctoral dissertation, University of Edinburgh].

Cangemi, F., & Baumann, S. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics*, *81*, 100993, 1–6. https://doi.org/10.1016/j.wocn.2020.100993

Cao, J., & Zheng, Y. (2006). Articulatory strengthening and prosodic hierarchy. *Proceedings of Speech Prosody, 2–5 May, Dresden, Germany*, 289–292. https://doi.org/10.21437/speechprosody.2006-71

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 27–55). Academic Press.

Chahal, D., & Hellmuth, S. (2014). The intonation of Lebanese and Egyptian Arabic. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 365–404). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.003.0013

Cho, T. (2002). *The effects of prosody on articulation in English*. Routledge.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*, 141–176. https://doi.org/10.1016/S0095-4470(03)00043-3

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *The Journal of the Acoustical Society of America*, *117*(6), 3867–3878. https://doi.org/10.1121/1.1861893

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 519–548). Mouton De Gruyter. https://doi.org/10.1515/9783110197211.3.519

Cho, T., Kim, J., & Kim, S. (2013). Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *The Journal of the Acoustical Society of America*, *133*(5), 384–390. https://doi.org/10.1121/1.4800179

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*, 344–361. https://doi.org/10.1016/j.wocn.2011.02.005

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*, 121–157. https://doi.org/10.1016/j.wocn.2005.01.001

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2), 425–452. https://doi.org/10.1515/labphon.2010.022

Darling, M., & Huber, J. E. (2011). Changes to articulatory kinematics in response to loudness cues in individuals with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 54*, 1247–1259. https://doi.org/10.1044/1092-4388(2011/10-0024)a

de Jong, K. (1991). *The oral articulation of English stress accent* [Dissertation]. Ohio State University.

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America, 97*(1), 491–504. https://doi.org/10.1121/1.412275

Dohen, M., & Lœvenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French. *Proceedings of INTERSPEECH, 4–8 October, Jeju, Korea*. Interspeech 2004. https://doi.org/10.21437/Interspeech.2004-296

Dromey, C., & Ramig, L. O. (1998). Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation. *Journal of Speech, Language, and Hearing Research, 41*, 1003–1018. https://doi.org/10.1044/jslhr.4105.1003

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America, 89*(1), 369–382.

Erickson, D., & Lehiste, I. (1995). Contrastive emphasis in elicited dialogue: Durational compensation. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS), 14–19 August, Stockholm, Sweden* (pp. 352–355).

Féry, C., & Krifka, M. (2008). Information structure: Notional distinctions, ways of expression. In P. van Sterkenburg (Ed.), *Unity and diversity of languages* (pp. 123–135). John Benjamins. https://doi.org/10.1075/z.141.13kri

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*(4), 680–703. https://doi.org/10.1016/j.wocn.2008.05.001

Fitzpatrick, M., Kim, J., & Davis, C. (2015). The effect of seeing the interlocutor on auditory and visual speech production in noise. *Speech Communication, 74*, 37–51. https://doi.org/10.1016/j.specom.2015.08.001

Fowler, C. A. (1995). Acoustic and kinematic correlates of contrastive stress accent in spoken English. In F. Bell-Bert & L. J. Raphael (Eds.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris* (pp. 355–372). AIP Press.

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Lœvenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. *Proceedings of the INTERSPEECH – ICSLP, 17–21 September, Pittsburgh, USA*. https://doi.org/10.21437/Interspeech.2006-323

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech and Language, 28*, 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research, 53*, 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Garnier, M., Ménard, L., & Alexandre, B. (2018). Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? *The Journal of the Acoustical Society of America, 144,* 1059–1074. https://doi.org/10.1121/1.5051321

Garnier, M., Smith, J., & Wolfe, J. (2022). Lip hyper-articulation in loud voice: Effect on resonance-harmonic proximity. *The Journal of the Acoustical Society of America, 152*(6), 3695–3705. https://doi.org/10.1121/10.0016595

Geumann, A. (2001a). Invariance and variability in articulation and acoustics of natural perturbed speech. In P. Hoole (Ed.), *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* (Vol. 38, pp. 265–393). Institut für Phonetik und Sprachliche Kommunikation, Ludwig-Maximilans-Universität München.

Geumann, A. (2001b). Vocal intensity: Acoustic and articulatory correlates. In B. Maassen, W. Hulstijn, R. D. Kent, & P. H. H. M. Van Lieshout (Eds.), *Speech Motor Control in Normal and Disordered Speech. Proceedings of the 4th International Speech Motor Conference* (pp. 70–73). Uitgeverij Vantilt.

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. *Computer Speech & Language, 28,* 629–647. https://doi.org/10.1016/j.csl.2013.09.007

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice, 2*(2), 118–126. https://doi.org/10.1016/S0892-1997(88)80067-5

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics, 64,* 90–107. https://doi.org/10.1016/j.wocn.2017.03.003

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

Halliday, M. A. K. (1967). *Intonation and grammar in British English.* De Gruyter. https://doi.org/10.1515/9783111357447

Harrington, J., Fletcher, J., & Beckman, M. E. (2000). Manner and place conflicts in the articulation of accent in Australian English. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 40–51). Cambridge University Press.

Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics, 23,* 305–322. https://doi.org/10.1016/S0095-4470(95)80163-4

Henriques, R. N., & Van Lieshout, P. (2013). A comparison of methods for decoupling tongue and lower lip from jaw movements in 3D articulography. *Journal of Speech, Language, and Hearing Research, 56*(5), 1503–1516. https://doi.org/10.1044/1092-4388(2013/12-0016

Hermes, A., Becker, J., Mücke, D., Baumann, S., & Grice, M. (2008). Articulatory gestures and focus marking in German. *Proceedings of Speech Prosody, 6–9 May, Campinas, Brazil,* 457–460.

Huber, J. E., & Chandrasekaran, B. (2006). Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults. *Journal of Speech, Language, and Hearing Research, 49*(6). https://doi.org/10.1044/1092-4388(2006/098)

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America, 106*(3), 1532–1542. https://doi.org/10.1121/1.427150

Im, S., Kim, S., & Cho, T. (2023). Some asymmetrical pre- versus post-focal effects on articulatory realization of prominence distribution in Korean. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS), 7–11 August, Prague, Czech Republic.*

Jabeen, F., & Braun, B. (2018). Production and perception of prosodic cues in narrow & corrective focus in Urdu/Hindi. *Proceedings of the 9th International Conference on Speech Prosody, 13–16 June, Poznan, Poland*, 30–34. https://doi.org/10.21437/SpeechProsody.2018-6

Jackendoff, R. (1974). *Semantic interpretation in generative grammar*. MIT Press.

Jun, S.-A. (2006). Intonational phonology of Seoul Korean revisited. In T. J. Vance & K. Jones (Eds.), *Japanese-Korean Linguistics* (Vol. 14, pp. 15–26).

Katsika, A. (2018). The kinematic profile of prominence in Greek. *Proceedings of Speech Prosody, 13–16 June, Poznań, Poland*, 764–768. https://doi.org/10.21437/SpeechProsody.2018-155

Kay, M. (2023). *tidybayes: Tidy data and geoms for Bayesian models* (v3.0.4) [Computer software]. https://doi.org/10.5281/zenodo.1308151

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica, 24*, 23–44. https://doi.org/10.1159/000259350

Kim, S., & Cho, T. (2011). Articulatory Manifestation of Prosodic Strengthening in English /i/ and /ɪ/. *Journal of the Korean Society of Speech Sciences, 3*(4), 13–21.

Kleinow, J., Smith, A., & Ramig, L. O. (2001). Speech motor stability in IPD: Effects of rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 44,* 1041–1051. https://doi.org/10.1044/1092-4388(2001/082)

Koenig, L. L., & Fuchs, S. (2019). Vowel formants in normal and loud speech. *Journal of Speech, Language, and Hearing Research, 62*, 1–18. https://doi.org/10.1044/2018_JSLHR-S-18-0043

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory Phonology, 8*(1), 1–26. https://doi.org/10.5334/labphon.75

Kügler, F. (2008). The role of duration as a phonetic correlate of focus. *Proceedings of Speech Prosody, 6–9 May, Campinas, Brazil,* 591–594. https://doi.org/10.21437/SpeechProsody.2008-134

Ladd, R. D. (1980). *The structure of intonational meaning: Evidence from English*. Indiana University Press.

Ladd, R. D., & Arvaniti, A. (2023). Prosodic prominence across languages. *Annual Review of Linguistics, 9*(1), 171–193. https://doi.org/10.1146/annurev-linguistics-031120-101954

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Laurinavichyute, A., & Von Der Malsburg, T. (2022). Semantic attraction in sentence comprehension. *Cognitive Science*, *46*(2), e13086. https://doi.org/10.1111/cogs.13086

Liénard, J.-S., & Di Benedetto, M.-G. (1999). Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America, 106*, 411–422. https://doi.org/10.1121/1.428140

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124*, 3261–3275. https://doi.org/10.1121/1.2990705

Macchi, M. (1985). *Segmental and supersegmental features and lip and jaw articulators* [Doctoral dissertation, New York University].

McAuliffe, M., Socolof, M., Mihuc, S., & Wagner, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of INTERSPEECH, 20–24 August, Stockholm, Sweden*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Mefferd, A. S. (2017). Tongue- and jaw-specific contributions to acoustic vowel contrast changes in the diphthong /ai/ in response to slow, loud, and clear speech. *Journal of Speech, Language, and Hearing Research, 60,* 3144–3158. https://doi.org/10.1044/2017_JSLHR-S-17-0114

Mefferd, A. S. (2019). Effects of speaking rate, loudness, and clarity modifications on kinematic endpoint variability. *Clinical Linguistics & Phonetics*, *33*(6), 570–585. https://doi.org/10.1080/02699206.2019.1566401

Mefferd, A. S., & Dietrich, M. S. (2020). Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers with dysarthria due to Amyotrophic Lateral Sclerosis: Effects of loud, clear, and slow speech. *Journal of Speech, Language, and Hearing Research, 63*(8), 2625–2636. https://doi.org/10.1044/2020_JSLHR-19-00309

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research, 53*(5). https://doi.org/10.1044/1092-4388(2010/09-0083)

Mücke, D. (2018). *Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie*. Language Science Press. https://doi.org/10.5281/zenodo.1188764

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation—Is it mediated by accentuation? *Journal of Phonetics, 44*, 47–61. https://doi.org/10.1016/j.wocn.2014.02.003

Nicolaidis, K. (2012). Consonant production in Greek Lombard speech: An electropalatographic study. *Rivista Di Linguistica, 24*(1), 65–101.

Pagel, L., Sóskuthy, M., Roessig, S., & Mücke, D. (2023). A kinematic analysis of visual prosody: Head movements in habitual and loud speech. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS), 7–11 August, Prague, Czech Republic*, 4130–4134. https://doi.org/10.5281/zenodo.10299230

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliynski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, *27*, 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Petersen, T. L. (2020). *patchwork: The composer of plots* (1.1.1) [Computer software]. https://CRAN.R-project.org/package=patchwork

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271–311). MIT Press. https://doi.org/10.7551/mitpress/3839.003.0016

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software].

Raitio, T., Suni, A., Pohjalainen, J., Airaksinen, M., Vainio, M., & Alku, P. (2013). Analysis and synthesis of shouted speech. *Proceedings of INTERSPEECH, 25–29 August, Lyon, France*, 1544–1548. https://doi.org/10.21437/Interspeech.2013-391

Repp, S. (2016). Contrast: Dissecting an elusive information-structural notion and its role in grammar. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 270–289). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642670.013.006

Rivers, C., & Rastatter, M. P. (1985). The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults. *Journal of Auditory Research*, *25*, 37–45.

Roessig, S. (2021). *Categoriality and continuity in prosodic prominence*. Language Science Press. https://doi.org/10.5281/zenodo.4121875

Roessig, S. (2024). The inverse relation of pre-nuclear and nuclear prominences in German. *Laboratory Phonology*, *15*(1). https://doi.org/10.16995/labphon.9993

Roessig, S., & Mücke, D. (2019). Modeling dimensions of prosodic prominence. *Frontiers in Communication*, *4*(44), 1–19. https://doi.org/10.3389/fcomm.2019.00044

Roessig, S., Mücke, D., & Grice, M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLoS ONE*, *14*(5), e0216859. https://doi.org/10.1371/journal.pone.0216859

Roessig, S., Mücke, D., & Pagel, L. (2019). Dimensions of prosodic prominence in an attractor model. *Proceedings of INTERSPEECH, 15–19 September, Graz, Austria*, 2533–2537. https://doi.org/10.21437/Interspeech.2019-2227

Roessig, S., Pagel, L., & Mücke, D. (2022). Speaking loudly reduces flexibility and variability in the prosodic marking of focus types. *Proceedings of Speech Prosody, 23–26 May, Lisbon, Portugal*. https://doi.org/10.21437/SpeechProsody.2022-102

Roessig, S., Winter, B., & Mücke, D. (2022). Tracing the phonetic space of prosodic focus marking. *Frontiers in Artificial Intelligence*, *5*, 1–24. https://doi.org/10.3389/frai.2022.842546

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America*, *85*(1). https://doi.org/10.1121/1.397737

Shin, S., Kim, S., & Cho, T. (2015). What is special about prosodic strengthening in Korean: Evidence in lingual movement in V# V and V# CV. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), 10–14 August, Glasgow, UK.*

Šimko, J., Beňuš, Š., & Vainio, M. (2016). Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *The Journal of the Acoustical Society of America, 139,* 151–162. https://doi.org/10.1121/1.4939495

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America, 118,* 1677–1688. https://doi.org/10.1121/1.2000788

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America, 79*(4), 1086–1100. https://doi.org/10.1121/1.393381

Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research, 47,* 85–100. https://doi.org/10.1044/1092-4388(2004/008)

Tjaden, K., Lam, J., & Wilding, G. (2013). Vowel acoustics in Parkinson's disease and Multiple Sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research, 56,* 1485–1502. https://doi.org/10.1044/1092-4388(2013/12-0259)

Tjaden, K., & Martel-Sauvageau, V. (2017). Consonant acoustics in Parkinson's disease and Multiple Sclerosis: Comparison of clear and loud speaking conditions. *American Journal of Speech-Language Pathology, 26,* 569–582. https://doi.org/10.1044/2017_AJSLP-16-0090

Tjaden, K., Richards, E., Kuo, C., Wilding, G., & Sussman, J. (2013). Acoustic and perceptual consequences of clear and loud speech. *Folia Phoniatrica et Logopaedica, 65,* 214–220. https://doi.org/10.1159/000355867

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783. https://doi.org/10.1044/1092-4388(2004/058)

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America, 107,* 3438–3451. https://doi.org/10.1121/1.429414

Vainio, M., Aalto, D., Suni, A., Arnhold, A., Raitio, T., Seijo, H., Järvikivi, J., & Alku, P. (2012). Effect of noise type and level on focus related fundamental frequency changes. *Proceedings of INTERSPEECH, 9–13 September, Portland, USA,* 671–674. https://doi.org/10.21437/Interspeech.2012-206

Vallduví, E. (1992). *The informational component.* Garland.

Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics, 34*(3), 459–519. https://doi.org/10.1515/ling.1996.34.3.459

Vogel, I., Athanasopoulou, A., & Pincus, N. (2016). Prominence, contrast, and the functional load hypothesis: An acoustic investigation. In J. Heinz, R. Goedemans, & H. Van Der Hulst (Eds.),

Pagel et al: The encoding of prominence relations in supra-laryngeal articulation across speaking styles
55

Dimensions of Phonological Stress (1st ed., pp. 123–167). Cambridge University Press. https://doi.org/10.1017/9781316212745.006

Wagner, M. (2012). Focus and givenness: A unified approach. In I. Kučerová & A. Neeleman (Eds.), *Contrasts and positions in information structure* (pp. 102–148). Cambridge University Press. https://doi.org/10.1017/CBO9780511740084.007

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7), 905–945. https://doi.org/10.1080/01690961003589492

Wang, M. (2020). The acoustic effect of speaking rate, focus and prosodic position on syllables in Chinese. *Journal of Chinese Linguistics*, *48*(1), 174–205. https://doi.org/10.1353/jcl.2020.0004

Westbury, J. R., Lindstrom, M. J., & McClean, M. D. (2002). Tongues and lips without jaws. *Journal of Speech, Language, and Hearing Research*, *45*(4), 651–662. https://doi.org/10.1044/1092-4388(2002/052)

Whitfield, J. A., Dromey, C., & Palmer, P. (2018). Examining acoustic and kinematic measures of articulatory working space: Effects of speech intensity. *Journal of Speech, Language, and Hearing Research*, *61*(5), 1104–1117. https://doi.org/10.1044/2018_JSLHR-S-17-0388

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V. …Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wohlert, A. B., & Hammen, V. L. (2000). Lip muscle activity related to speech rate and loudness. *Journal of Speech, Language, and Hearing Research*, *43*, 1229–1239. https://doi.org/10.1044/jslhr.4305.1229.

Xu, Y., Chen, S.-W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, *29*(1), 131–147. https://doi.org/10.1515/tlr-2012-0006

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

Xue, Y., Marxen, M., Akagi, M., & Birkholz, P. (2021). Acoustic and articulatory analysis and synthesis of shouted vowels. *Computer Speech & Language*, *66*, 101156. https://doi.org/10.1016/j.csl.2020.101156

Yang, Y., & Chen, S. (2020). Revisiting focus production in Mandarin Chinese: Some preliminary findings. *Proceedings of Speech Prosody, 25–28 May, Tokyo, Japan*, 260–264. https://doi.org/10.21437/SpeechProsody.2020-53