Journal of Phonetics 111 (2025) 101426

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Advancements of phonetics in the 21st century: Phonetic universals, language variation, and phonetic grammar ☆

Taehong Cho *

HIPCS, Hanyang University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 7 July 2024
Received in revised form 25 May 2025
Accepted 28 May 2025
Available online xxxx

Keywords:
Phonetic universals
Language variation
Extended Model of Phonetic Grammar
Segmental-level and utterance-level phonetic universals
Phonetics-prosody interface
Speaker control
Automatic phonetic processes

ABSTRACT

This review, part of the journal's special collection on *Advancements* of *Phonetics in the 21st Century*, examines the interplay between phonetic universals and language variation at both segmental and utterance levels. It traces the physiological and biomechanical foundations of phonetic universals established by 20th-century research while focusing on cross-linguistic variation explored predominantly in 21st-century research. Segmental phonetic universals include *the role of the syllable* in organizing segments and gestures, *intrinsic vowel duration* influenced by vowel height, *extrinsic vowel duration* due to coda voicing, *intrinsic and co-intrinsic f0 variation* affected by vowel height and onset consonant characteristics, respectively, and *place effects on closure duration and VOT.* While segmental universals stem from distinct mechanical bases, utterance-level universals emerge from respiratory and articulatory resets at utterance onset, shaping the entire speech production system—a perspective substantiated here based primarily on 21st-century phonetic research. These resets structure prosodic organization, leading to weakening effects at the right edge (e.g., *f0 declination, articulatory declination, phrase-final lengthening*) and strengthening effects at the left edge (e.g., *domain-initial strengthening*) and occasionally at the right edge as well (e.g., *phrase-final strengthening*) when sufficient time permits. Extensive evidence demonstrates that phonetic universals are further shaped by language-specific factors and the interaction between *system-oriented* and *output-oriented* constraints. This diversity calls for detailed phonetic descriptions tailored to each language, with phonetic grammar, as proposed here, fine-tuning phonetic realization accordingly. Research in the 21st century has also illuminated that segmental and utterance-level universals, traditionally regarded as distinct, are deeply interconnected, if not inseparable. *The Extended Model of Phonetic Grammar* is introduced as a framework for mediating this relationship within the phonetics-prosody interface as well as interactions with other higher-order linguistic structures. Furthermore, language variation within phonetic universals suggests that many phonetic processes, once considered automatic, are actively controlled by speakers, reflecting the unique evolutionary pathways of different languages.

1. Introduction

A widely embraced premise that aids in understanding the phonetic characteristics of the world languages is that humans possess a speech production system characterized by unique anatomical traits, particularly when viewed from a broad perspective on the origins of speech. The shared anatomical structure distinguishing humans from other primates suggests an inherent physiological capacity to articulate specific sounds common across languages. Setting aside the nuanced complexities this may introduce with respect to theories of language evolution, this view underscores our species' shared phonetic capabilities and encourages deeper reflection on the implications of these common traits for understanding the universality among the diverse sound systems of world languages. Thus, in exploring phonetic universals, phoneticians often interpret recurring patterns observed across multiple languages as outcomes of natural speech production processes, rooted in the physiological and biomechanical traits shared by speakers of the world languages. Keating (1985, p. 126) referred to this process as "the physical operation of the speaking device." This speaking device enables the production of

☆ This article is part of a special issue entitled: *Advancements of Phonetics in the 21st Century* published in Journal of Phonetics.

* Corresponding author.
E-mail address: tcho@hanyang.ac.kr

https://doi.org/10.1016/j.wocn.2025.101426

phonetic forms that can be remarkably similar across languages, a phenomenon that has long intrigued phoneticians investigating phonetic universals to uncover the underlying bases of these cross-linguistically applicable phonetic patterns (cf. Keating, 1985; Maddieson, 1997; Cho & Ladefoged, 1999).

As much as speakers of world languages share commonalities, they also differ in numerous ways. One obvious fact is that no two speakers have the exact same anatomical geometry to produce the physically exact same sound. This phenomenon leads to what is termed “organic” variation, as discussed by Laver (1980, 1994) and Beck (1997). ***Organic variation*** refers to the variability in sounds caused by differences in the anatomical structures of speakers' speech organs, even within the same speech community. This kind of variation likely arises purely automatically, as speakers have no control over their anatomical structures, and thus falls clearly outside the realm of “linguistic phonetics” (cf. Ladefoged, 1971). However, speakers may also differ—across languages or speech communities—in the way they use their speech organs (such as the jaw, lips, tongue, velum, and vocal folds), to produce sounds that may be labeled similarly (e.g., a voiceless aspirated bilabial stop) across languages. These differences contribute to the linguistic diversity in the sound systems of world languages. Thus, a cross-linguistic diversity may result not only from using different sound inventories and phonological processes across languages, but it may also arise as speakers utilize the available speech organs in subtly different ways to produce sounds that render the particular language or dialect unique from others. As a consequence, many phonetic universals, assumed to originate from the physical operation of the speech organs that produce cross-linguistically similar sounds, may vary in the exact way they are produced from language to language or within different varieties of the same language. These nuances of fine phonetic detail shape the language's unique sound system, thus providing the motivation behind many studies aimed at understanding ***language variation within a phonetic universal.***

The objective of this article, contributing to the special issue of the journal themed ***Advancements of Phonetics in the 21st Century,*** is to delve into the realms of the phonetic universals and language variation. This exploration will draw on insights garnered from phonetic research conducted from the late 20th century into the 21st century. Clearly, the vast range of phonetic literature related to this ambitiously broad topic cannot be fully encapsulated within a single article. Therefore, the focus will be on the most significant advancements, particularly those that align with my expertise and understanding of the field over the past few decades. Accordingly, this review does not aim to be exhaustive but rather to provide a selective perspective, highlighting key developments in phonetic universals and language variation in the 21st century. Since many of these advancements are rooted in foundational research conducted in the 20th century, a substantial discussion of earlier work is necessary to provide an appropriately contextualized and comprehensive background, reinforcing the continuity between past and present research on the topic.

To structure this discussion, the remainder of Section 1 examines theoretical considerations related to phonetic universals, language variation, and phonetic arbitrariness, while also introducing the role of phonetic grammar in capturing variation within phonetic universals. Section 2 focuses on segmental phonetic universals and language variation, beginning with the role of syllable structure and exploring the fundamental relationship between phonetics and phonology—i.e., how symbolic representations are phonetically implemented within and across languages. Section 3 extends this discussion to the utterance level, an area that has historically received less attention but has been more fully substantiated in 21st-century research. This section examines phonetic universals and language variation within the phonetics-prosody interface, investigating how utterance-level phonetic patterns are shaped by prosodic structure. These sections primarily highlight findings from the 21st century while tracing developments from the 20th century, but where appropriate, they are framed within the extended view of phonetic grammar. Additionally, I will discuss how phonetic universals and language variation emerge from the interaction between system-oriented and output-oriented constraints. Through these perspectives, this review aims to broaden our understanding of the world's sound systems, adopting a more ecologically grounded perspective. Section 4 provides a brief summary and explores an extended phonetic grammar that integrates the phonetics-prosody interface, along with potential interactions with other higher-order linguistic structures, into the framework of phonetic universals and language variation. Finally, Section 5 concludes the review along with outlining some potential future research directions that may address unresolved issues remaining in the 21st century.

1.1. Theoretical premise: Phonetic arbitrariness, phonetic grammar and phonetic rules

Language variation within phonetic universals reflects both phonetic arbitrariness in selecting modal values for phonetic features and the extent to which a language adheres to general phonetic principles that shape sound systems over time. Cho and Ladefoged (1999) examined VOT variation across 18 languages, later expanded in Ladefoged and Cho (2001), demonstrating that while most languages follow the general phonetic tendency of longer VOT for more posterior places of articulation, deviations exist. Languages arbitrarily select modal phonetic values (i.e., typical or representative values for a given category in a language) for voiceless aspirated and unaspirated stops, making a value typical in one language anomalous in another. This ***phonetic arbitrariness*** contributes to language-specific variation that cannot be fully explained by biomechanical constraints or general phonetic principles such as low-cost articulatory strategies, contrast maximization, and the quantal nature of sound (e.g., Lindblom, 1986, 1990; Stevens, 1989; Docherty, 1992).

Later in the 21st century, Cho, Whalen, and Docherty (2019) expanded upon earlier research by examining 19 additional languages, reaffirming the extensive cross-linguistic variation in VOT patterns. Some languages, such as Burushaski, exhibit a highly polarized three-way contrast (voiced, voiceless unaspirated, voiceless aspirated), with VOT values ranging from −131 ms to 91 ms. However, not all languages display such extreme polarization. Others with a three-way contrast, such as Thai, Vietnamese, and Khmer, show more moderate values for their voiced stops (−60 to −74 ms). Furthermore, languages such as Russian, Turkish, Pashto, and Wakhi, which maintain a two-way contrast between voiced and voiceless unaspirated stops, differ in their modal VOT values. Turkish, with a relatively long VOT for voiceless unaspirated stops (41 ms), might be expected to exhibit a moderate negative VOT for its voiced counterpart, yet it shows a rather extreme −77 ms. Pashto and Wakhi push this boundary further, with values of −128 ms and −139 ms, respectively. Given the articulatory effort required for such extreme values (Ohala, 1997), this variation likely reflects a language-specific balance between system-oriented production efficiency and output-oriented contrast maximization (cf. Lindblom, 1986)—i.e., some languages prioritize ease of articulation and physiological constraints, while others focus on enhancing phonological contrasts for perceptual distinctiveness.

These observations have further theoretical implications. When the phonological component of a given language feeds the phonetic component with a sound specified by phonetic features such as {voiceless unaspirated} or {voiceless aspirated} (curly brackets are used to refer to phonetic features, following Keating, 1984, to be differentiated with phonological features) for motor execution, the "speaking device" must be informed how to implement the assigned phonetic features in a way that conforms to the modal phonetic value determined by the language's phonetic arbitrariness. Cho and Ladefoged (1999) suggest that this phonetic arbitrariness must be internalized in the grammar of the language. Since this grammar governs the phonetic implementation of phonetic features necessary for expressing phonological contrast, it is often referred to as *phonetic grammar,* a concept proposed by Keating (1985, 1990) but used by Cho and Ladefoged with a broader application that encompasses phonetic arbitrariness. Thus, ***phonetic grammar*** describes the part of the phonetic component that regulates language-specific phonetic rules, mediating between the phonological input and the phonetic output by guiding motor execution on how to implement the features provided by phonology. Crucially, acquiring this phonetic grammar, alongside the language's phonological grammar, is essential to becoming a native speaker within a speech community, as both contribute to the proper phonetic realization of phonological contrasts. As such, phonetic grammar enriches phonological representations with fine-grained phonetic detail prior to motor execution, ensuring that speech outputs align with the pronunciation norms characteristic of each language or linguistic community.

Ladefoged and Cho (2001) expanded on the concept of Articulatory VOT by reinterpreting it in terms of articulatory timing between oral and laryngeal events—a perspective that shares conceptual ground with the intergestural timing notions of Articulatory Phonology (Browman & Goldstein, 1990, 1992). Unlike conventional acoustically defined VOT—measured from stop release to the onset of voicing—***Articulatory VOT*** refers to the timing between the release gesture and the start of vocal fold vibration gesture. This model, building on Keating (1985, 1990) and Cohn (1993), posits that phonology selects a modal VOT value for each stop category (e.g., voiceless unaspirated, voiceless aspirated), while language-specific rules regulate intergestural timing to define Articulatory VOT. A key advantage of Articulatory VOT is its direct control within articulatory coordination, making it a manipulable phonetic parameter governed by a language's phonetic grammar. These timing targets feed into automatic phonetic implementation, constrained by physical laws. Fig. 1 illustrates the multistage pathway from phonology to speech, where phonetic grammar, once informed by phonology, interacts with physical constraints. The right panels depict VOT production stages, from selecting a modal value to applying language-specific phonetic rules and executing speech movements. This framework highlights how phonetic variation emerges within universal principles, often incorporating idiosyncratic phonetic arbitrariness shaped by language-specific rules.

The concept of phonetic grammar is useful because it extends beyond merely regulating the phonetic implementation of segmental features in a language-specific manner. It functions as a hub, integrating input from phonology and other higher-order linguistic structures to refine phonetic details before they are executed in speech production. While languages vary in the phonetic realization of features as influenced by higher-order linguistic structures, one such structure explored in a growing body of 21st-century phonetic literature is prosodic structure (see Fletcher, 2010; Cho, 2016, for related reviews). ***Prosodic structure***, central to this discussion, plays a crucial role in shaping phonetic realization—an interaction commonly referred to as the ***phonetics-prosody interface*** (e.g., Cho & Keating, 2009; Byrd & Choi, 2010; Cho, 2011, 2016, 2022; Mücke, Grice, & Cho, 2014; Silpachai, 2024). Here, the ***phonetics-prosody interface*** refers to the interplay between the low-level phonetic processes and the higher-order prosodic structures that encompass prosodic boundaries and prominence distributions. In the theoretical framework proposed here, the language-specific manipulation of the phonetics-prosody interface—such as the realization of the {voiceless unaspirated} feature based on both modal VOT values and prosodic position—must be internalized within phonetic grammar. This extends phonetic grammar beyond a simple phonology-to-phonetics mapping, incorporating the prosodic conditioning unique to each language. These language-specific phonetic rules, governed by a language's phonetic grammar, operate at a later production stage—following phonological encoding but preceding motor execution—marking the transition from controlled phonetic adjustments to automatic phonetic processes.

Without such a mechanism regulating phonetic output in reference to various factors, language-specific variation can only be described rather than understood in a principled way. A purely descriptive approach risks reducing variation to an arbitrary collection of language-specific quirks rather than recognizing it as a structured system governed by identifiable principles. Phonetic grammar provides a necessary framework to explain how phonetic realization is not only shaped by phonology but also fine-tuned by higher-order linguistic structures such as prosodic structure, as well as by articulatory constraints and perceptual considerations. Without this intermediary system, it becomes difficult to account for why similar phonological categories exhibit different phonetic realizations across languages or why the same phonetic contrast may be implemented differently depending on prosodic position or possibly other linguistic structures. Moreover, phonetic grammar is not just an abstract theoretical construct; it is an essential component of language acquisition. To become fully competent members of their speech community, children must

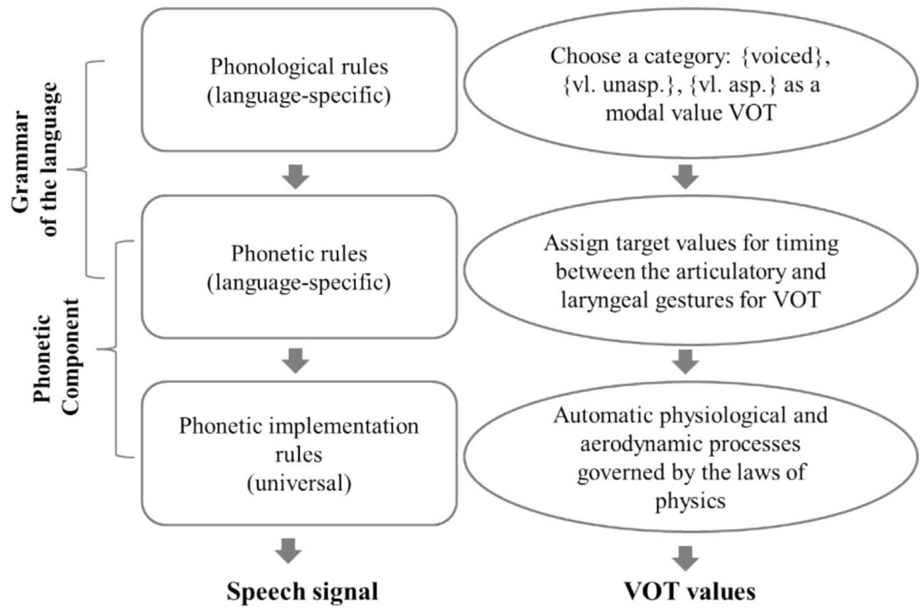

Fig. 1. Multiple processes from phonology to speech signal. The diagram drawn here by the author is based on models proposed and discussed by Keating (1985, 1990), Cohn (1993), and Cho and Ladefoged (1999). In this model, the language-specific phonetic rules operate within the phonetic component, part of which falls under the domain of the linguistic grammar of the language, and part of which is governed by the laws of physics. In this model, {}, as in {vl. unasp.}, denotes a phonetic feature that guides the application of language-specific phonetic rules in shaping the actual phonetic realization.

acquire not only the phonological system of their language but also the phonetic grammar that governs the fine details of phonetic implementation. This includes internalizing language-specific phonetic rules that determine how phonological contrasts are realized in different contexts, how prosodic structure influences articulation, and how speech patterns align with both system-oriented and output-oriented constraints. The fact that speakers within a linguistic community, despite physiological differences, exhibit remarkably consistent phonetic patterns that systematically differ from those of other language communities suggests that phonetic grammar is learned and shared rather than purely innate. Thus, phonetic grammar serves as the crucial link between innate articulatory mechanisms and the socially embedded linguistic norms that define a speech community.

In this theoretical context, it is important to clarify the terms *system-oriented* and *output-oriented*, which have been widely discussed in relation to phonetic universals and language variation (e.g., Lindblom, 1986, 1990; Kingston & Diehl, 1994; Ohala, 1997; Maddieson, 1997; Flemming, 2004; Keyser & Stevens, 2006). ***System-oriented*** refers to speech production processes fundamentally shaped by the physiological and biomechanical constraints of the human speech system. These constraints arise from the anatomical structure and motor control of the vocal tract, influencing articulatory patterns based on principles such as *ease of articulation* and *effort minimization*. Since these tendencies naturally emerge from the physical properties of speech organs, system-oriented processes are often considered ***speaker-oriented,*** as they do not impose substantial cognitive or articulatory effort on the speaker beyond the inherent physiological and biomechanical constraints of speech production and as such can be seen to benefit the speaker.

By contrast, ***output-oriented*** refers to the ways in which speakers actively modulate or control their speech beyond purely physiological constraints, often in response to perceptual or communicative demands or the speech norms expected within the speech community to which the speaker belongs. These modifications often serve to enhance phonological contrasts and improve intelligibility for listeners. Because such adjustments are driven by the ecological demands of speech communication—such as perceptual distinctiveness, lexical retrieval, and listener adaptation from which listeners benefit—output-oriented processes can be considered ***listener-oriented.*** This listener-oriented perspective assumes that certain phonetic universals are shaped by ***ecological constraints,*** reflecting how languages optimize phonetic patterns to maximize contrast and minimize ambiguity in communication. This notion can also extend to phonetic arbitrariness shared within a speech community, shaping language-specific phonetic norms and variations.

In the theoretical perspective proposed here, languages vary in how they balance physiological (*system-oriented*) constraints and ecological (*output-oriented*) demands in the phonetic implementation of features. This balance, reflecting Lindblom's (1986) *tug-of-war* between articulatory ease and communicative clarity, is fine-tuned and systematically regulated by a language's phonetic grammar. In other words, phonetic grammar governs language-specific phonetic realization by integrating system-oriented articulatory tendencies with output-oriented adjustments, ensuring that phonetic implementation aligns with both speaker-internal constraints and communicative pressures. However, as seen in the vast cross-linguistic variation in VOT distribution, some languages prioritize system-oriented constraints, while others emphasize output-oriented demands, reflecting the language-specific operation of phonetic grammar. Crucially, this regulation is shaped by higher-order linguistic structures, particularly prosodic structure, which will be a central focus in the later part of this review.

2. Segmental-level phonetic universals and language variation

In the previous section, I examined language variation within phonetic universals, emphasizing phonetic arbitrariness governed by language-specific rules within phonetic grammar. Building on this, the current section explores phonetic universals, often regarded as natural outcomes of speech production processes, alongside language variation. While the focus is on 21st-century advancements rooted in 20th-century research, I also examine, where appropriate, how these universals manifest across languages and how divergence arises, reflecting a balance between system-oriented and output-oriented constraints shaped by each language's phonetic grammar.

2.1. Syllables as basic units of segmental realization

The syllable's universality is widely recognized, as it serves as a fundamental organizational unit in linguistic systems, structuring phonetic features, gestures, and segments (Clements & Keyser, 1983; Goldsmith, 1990; Browman & Goldstein, 1992; Gafos & Goldstein, 2012). Developmental research shows that infants acquire syllabic structures rapidly, suggesting that the basic CV syllable is an innate feature of the human vocal apparatus, facilitating alternating mouth movements (Oller, 1980, 2000; Stark, 1980; 1981; Vihman et al., 1985). This early acquisition underscores the syllable's central role in language development, independent of the language being learned (Stark, 1980; Jusczyk, 1997; Kuhl, 2004; Werker & Tees, 1984; Nazzi et al., 2000). The universality of syllables aligns with evolutionary perspectives (Lieberman, 1984; MacNeilage, 1998; MacNeilage & Davis, 2000). While theories on syllable evolution vary, a theory-neutral view suggests that modifications to the human vocal tract over millions of years enabled a versatile syllable production system. MacNeilage's (1998) *Frame Content Theory* posits that rhythmic jaw oscillations provide a structural *frame* for speech, with tongue and lip movements articulating consonants and vowels within this framework. This theory links syllabic structure to non-speech functions like mastication, which likely preceded more complex vocal capabilities. Supporting evidence from infant babbling research indicates that early speech development relies on these rhythmic jaw movements (MacNeilage, 1998; MacNeilage & Davis, 2000), although alternative views have been proposed (see Whalen et al., 2011, for a critical response to this theory).

Despite the intuitive appeal of the syllable as a fundamental speech unit, its physiological role in speech organization was not firmly established in phonetic theory until the late 20th century. Krakow (1999), building on Browman & Goldstein's (1995) work, provided a comprehensive review and empirical support for the syllable as a physiological unit, organized by articulatory patterns and timing. This perspective defines *physiological* in terms of articulatory coordination, with evidence showing systematic timing differences based on syllable structure, particularly in CV versus VC contexts. These articulatory timing patterns, which vary by syllable position, highlight fundamental aspects of syllable organization across languages, suggesting a universal basis for syllable structure.

At the turn of the 21st century, Goldstein and colleagues further developed the concept of syllable organization within the framework of ***Articulatory Phonology*** (Browman & Goldstein, 2000; Goldstein & Fowler, 2003; Nam & Saltzman, 2003; Goldstein et al., 2006; Gafos and Goldstein, 2012). This approach defines articulatory gestures as both units of action and phonological contrast, with phonological distinctions directly expressed through the temporal and spatial coordination of gestures. The refined model (Goldstein et al., 2006) posits that inter-articulatory timing is dictated by planning oscillators, which coordinate gestures based on a ***coupling graph*** encoded in the lexicon. Gestures exhibit two primary coupling modes specified in the coupling graph: ***in-phase*** (synchronous timing, typically for onset consonants and vowels) and ***anti-phase*** (sequential timing, typically for vowels and coda consonants). This coordination mirrors human bimanual movements, suggesting that these phase relationships are deeply rooted in motor control mechanisms. By framing Krakow's (1999) observations of syllable-based timing differences in terms of *in-phase* and *anti-phase* coupling, this model provides a physiologically grounded rationale for syllabic organization across languages. The gestural coordination principles suggest a universal basis for syllable structure, while also allowing for cross-linguistic variation in intergestural timing (see, Gafos and Goldstein, 2012, for further elaboration).

The intergestural timing and coupling of articulatory gestures within syllable organization are well exemplified by vowel nasalization. In coda nasal contexts, the velar lowering gesture is anti-phase coupled with the vowel, meaning the velum's lowering onset aligns with the vowel's articulatory target. Since the acoustic vowel duration extends beyond this target, acoustic vowel nasalization begins before the nasal consonant's full oral constriction. This reflects a systematic coarticulation of velum lowering with the vowel. Such anticipatory vowel nasalization may appear cross-linguistically similar due to system-driven (anti-phase coupling) constraints associated with VC coordination, but variations also suggest that its implementation is shaped in a language-specific manner, regulating the extent and magnitude of coarticulatory effects (Farnetani & Recasens, 2010; Solé, 2007; Zellou, 2022).

Solé (2007) investigated vowel nasalization in VN contexts across Spanish, French, Japanese, Italian, and Swedish, finding that vowels followed by a tautosyllabic nasal exhibited minimal coarticulatory nasalization, constrained by the time required for velum lowering—likely a physiologically driven process reflecting anti-phase coupling. In contrast, American English showed extensive nasalization, suggesting a controlled, speaker-modulated process rather than a purely automatic one. Solé analyzed oral and nasalized vowel portions in Spanish and American English at different speech rates. In English, nasalization varied proportionally with vowel duration, while in Spanish, it remained stable, supporting the idea that English nasalization is programmed and under speaker control, whereas Spanish exhibits a low-level coarticulatory process.

Building on this, Zellou (2022) examined vowel nasalization in 60 American English speakers, identifying three distinct patterns: (1) *phonetic-mechanical*, where nasalization remained stable across speech rates (similar to Spanish in Solé's study); (2) *phonologized,* where nasalization increased proportionally with vowel duration (as observed by Solé in American English); and (3) *phonetic-coda enhancement*, where nasalization increased at slower speech rates, suggesting a hyperarticulatory effect linked to clear speech or prosodic strengthening (cf. Lindblom, 1990; de Jong, 1995, 2004; Cho et al., 2017). As such, the study revealed significant speaker variation, reinforcing the idea that coarticulatory vowel nasalization is not purely automatic but subject to individual differences and potential phonologization. This aligns with Keating's (1985) claim that any controllable phonetic parameter is, in principle, available for speaker manipulation. As Zellou argues, individual variation may reflect different phonological grammars—akin to phonetic grammar—within a speech community (Beddor, 2009, 2023; Yu & Zellou, 2019), contributing to synchronic variation and potentially driving sound change (Lindblom et al., 1995; Ohala, 1993; Beddor, 2023).

Nasalization in NV contexts differs from VN contexts, with left-to-right carryover nasalization generally less extensive than right-to-left anticipatory nasalization (Cohn, 1993; Farnetani & Recasens, 2010). This asymmetry may stem from syllable structure, where in-phase coupling between the onset velar and vocalic gestures synchronizes their onsets, unlike the anti-phase coupling in codas. This synchronization reduces the opportunity for velar lowering to influence the acoustic vowel, resulting in weaker nasalization and a shorter nasal murmur. Suppressed nasality in NV contexts may also relate to the tighter oral constriction of syllable onsets (Krakow, 1999) or domain-initial strengthening (Cho et al., 2017), which could reinforce velar closure. Cross-linguistic studies confirm this asymmetrical pattern related to syllable structure, with weaker nasalization in NV than VN contexts observed in American English (Cho & Keating, 2001; Cho et al., 2017), Australian English (Joo et al., 2019), French (Fougeron, 2001), Korean (Jang et al., 2018), and Mandarin Chinese (Li et al., 2020). Given the overall lower nasality in onsets, cross-linguistic variation in NV contexts is relatively minor. However, Korean exhibits a striking language-specific effect, with onset nasality exceptionally reduced, sometimes approaching denasalization (Yoo & Nolan, 2020; Lee et al., 2023). This robust nasal reduction may stem from strong domain-initial strengthening (Keating et al., 2003), illustrating another case of how a universal phonetic pattern shaped by syllable structure can be further refined in a language-specific manner.

These examples illustrate just a few of the potential phonetic universals and language-specific variations shaped by syllable structure, which serves as a fundamental unit in speech production. Other effects include timing relationships among features, gestures, and segments within a syllable (see Gafos & Goldstein, 2012, for a related review). These may encompass *closed syllable shortening* (e.g., Maddieson, 1985), *closed syllable laxing* versus *open syllable tensing* (Storme, 2019*), polysyllabic shortening* (Lehiste, 1972), and *various trading relationships* that preserve syllable-level timing among articulatory gestures, segments, or phonetic events. Such interactions likely further shape phonetic universals and cross-linguistic variation in segmental realization, underscoring the pivotal role of syllable structure in speech patterns. Further research in the 21st century is needed to refine our understanding of the interplay between syllable structure and phonetic realization within the broader framework of phonetic universals and language variation. In particular, future studies should explore how syllable structure operates across languages with different rhythmic typologies (e.g., so-called stress-timed vs. syllable-timed), where syllables may function in distinct ways.

2.2. Intrinsic vowel duration

Intrinsic vowel duration is one of the most frequently cited phonetic universals, referring to the tendency for low vowels to be longer than high vowels (Lindblom, 1967; Lehiste, 1970; Lisker, 1974; Keating, 1985; Maddieson, 1997). (Note that, following Keating (1985) and Maddieson (1997), the term *intrinsic* is used for variation in vowel duration arising from a segment's inherent phonetic properties, whereas *extrinsic* refers to durational changes influenced by external factors.) This is often attributed to jaw mechanics: transitioning to and from a low vowel requires greater jaw movement, inherently lengthening its duration (Lindblom, 1967, 1990). However, speakers may partially compensate for this effect to minimize excessive durational contrasts (Lindblom, 1967, 1990). For instance, in Swedish /ɪ'bɑːbɪ/, the jaw begins to open for /ɑː/ before the release of /b/, yet speakers control the timing to prevent premature opening. Similarly, when closing the lips for the second /b/, they do so before the jaw fully closes, shortening /ɑː/ to maintain temporal balance with /biː/ in the /ɪ'biːbɪ/ contexts. This partial compensation reflects a trade-off between mechanical constraints and the need for durational stability, aligning with Maddieson's (1997) notion of phonetic universals balancing physiological and ecological factors. Further evidence suggests that intrinsic vowel duration may remain under speaker control (Westbury & Keating, 1980; Keating, 1985; Maddieson, 1997; Cho, 2015). An Electromyographic (EMG) study by Westbury and Keating (1980) examined the anterior belly of the digastric (ABD), a primary jaw-lowering muscle. If vowel duration differences were purely mechanical, ABD activity would remain constant across vowel heights. Instead, lower vowels showed prolonged and intensified EMG activity, indicating active speaker control. These findings support Keating's (1985) argument that vowel duration is a ***controllable parameter***, making it available for language-specific manipulation.

Solé and Ohala (2010) responded to this view by investigating whether speakers actively manipulate height-related vowel duration to enhance vowel height contrast by analyzing American English, Catalan, and Japanese. Using acoustic data across varying speech rates, they found that in English and Catalan, speakers systematically adjusted vowel durations: low vowels became even longer relative to high vowels at slower speech rates. This adjustment, they argued, helps maintain stable durational ratios across rates, reflecting an output-oriented constraint. A perceptual test with Catalan listeners confirmed that vowel duration cues vowel height, with shorter durations increasing the likelihood of perceiving a higher vowel. However, Japanese exhibited a different pattern. While phonologically long-short vowel contrasts in Japanese showed similar speech rate adjustments to English and Catalan, the height-related durational differences between high and low vowels remained unchanged. This suggests that Japanese primarily relies on a biomechanical basis for intrinsic vowel duration, possibly linked to jaw movement, while controlling duration only when it serves phonological contrast.

Wilson and Chodroff (2017) expanded on the controlling aspect of intrinsic vowel duration in American English by examining a large speech dataset: in isolated speech with 24 speakers and in connected speech with 391 speakers. They demonstrated that although the raw data plotted across speakers seemed to show substantial variation in vowel-height related durational differences, when the data were normalized across speakers (by "uniform translation," in their terms), patterns of intrinsic vowel duration are highly correlated across speakers. They concluded that while intrinsic vowel durations may be mechanically motivated, the remarkable uniformity among American English speakers suggests that this low-level effect is nonetheless controlled by speakers (see Chodroff & Wilson, 2017, for a related discussion on VOT).

Toivonen et al. (2015) examined intrinsic vowel duration from a cross-linguistic perspective, analyzing the relationship between first formant (F1), which reflects tongue height, and vowel duration in American English and Swedish. Their approach differed from previous studies by considering F1 variation within the same vowel. While a general positive correlation between F1 and vowel duration emerged when all vowels were analyzed collectively, this relationship disappeared within individual vowels, weakening the mechanistic explanation. This finding aligns with Lisker (1974), who argued against a strictly jaw-driven account, demonstrating that steady-state formants, rather than CV transitions, primarily account for height-related durational effects—i.e., if vowel height differences were purely mechanical, CV formant transitions would be more pronounced. Building on this reasoning and their own findings, Toivonen et al. proposed that vowel height-related duration differences should be specified in phonology or at least in *higher-level phonetics*, which can be interpreted as phonetic grammar. From the current perspective, these differences are not directly encoded in phonology but are mediated by phonetic grammar, which operates in reference to phonological structure.

These studies, taken together, highlight how phonetic grammar may encode intrinsic vowel duration differently across languages. While English and Catalan regulate it in an output-oriented manner, actively manipulating duration presumably for perceptual stability, Japanese appears to rely more on system-driven biomechanical constraints, showing less speaker control. Moreover, despite speaker variation, uniformity within a linguistic community observed across multiple speakers of English suggests that phonetic grammar internalizes these patterns, mediating between phonological structure and language-specific phonetic realization, thereby fostering stability within speaker variation.

2.3. Extrinsic vowel duration due to coda voicing

Another widely discussed phonetic universal involves ***extrinsic vowel duration*** variation based on the voicing of the following consonant, with vowels typically longer before voiced obstruents (Belasco, 1953; Peterson & Lehiste, 1960; Halle & Stevens, 1967; Chen, 1970; Lisker, 1974; Maddieson & Gandour, 1977; Keating, 1985; Maddieson, 1997, among others). For instance, in English, the vowel in *cab* is longer than in *cap*, a pattern observed across various languages. The term *extrinsic* reflects that this variation originates outside the vowel. As with many phonetic universals, the underlying mechanism remains debated, potentially involving biomechanical and auditory-perceptual factors (e.g., Maddieson, 1997). Some key proposed mechanisms, though controversial, can be summarized as follows.

2.3.1. Possible physiological and biomechanical bases of extrinsic vowel duration

Belasco (1953) proposed that voiceless consonants require greater articulatory force than voiced ones, leading to vowel shortening as energy shifts toward their production. Conversely, Halle and Stevens (1967) and Chomsky and Halle (1968) suggested that vowels lengthen before voiced consonants to allow time for laryngeal adjustments needed to sustain vocal fold vibration during the upcoming constriction. These adjustments, such as larynx lowering and glottal loosening, help counteract the rising intraoral pressure and require additional time. Chen (1970) attributed this vowel duration difference to varying rates of vowel-to-consonant closure transitions. He argued that the open glottis of voiceless stops creates greater aerodynamic resistance, necessitating increased articulatory effort and faster closure to counteract it. However, Maddieson (1997) challenged this view, noting that closure onset for voiceless stops often occurs earlier than for voiced stops, suggesting timing differences rather than movement velocity as the key factor. Additionally, Chen's account fails to explain durational variations in consonant clusters (e.g., *build*-*built*, *send*-*sent*), where preceding liquids and nasals also show length differences despite already having formed their closure. Empirical data further challenge Chen's hypothesis. Since articulatory velocity is linked to articulatory force (Kuehn & Moll, 1976) and biomechanical effort (Nelson, 1983), increased movement velocity has been suggested as a factor underlying extrinsic vowel duration effects. If vowel shortening before voiceless stops resulted from increased articulatory force, as Belasco (1953) also proposed, a corresponding increase in peak velocity would be expected. However, Son et al. (2011) found that in Korean VCV sequences, vowels were longer before lenis (voiced intervocalically) stops than before fortis or aspirated stops. Yet, lip-closing movements for voiceless stops were neither shorter in duration nor faster in velocity, contradicting the notion that greater articulatory force drives vowel shortening.

2.3.2. Some typologies of the coda voicing effects on vowel duration

Given this uncertainty, pinpointing the precise physiological underpinnings of extrinsic vowel duration remains a formidable challenge—one that may never be fully resolved. This difficulty arises from the possibility that even if a phonetic universal has a physiological basis, languages have diverged along distinct evolutionary paths. Over generations, these developmental trajectories may have shaped unique articulatory strategies, leading to language-specific phonetic norms. As Maddieson (1997) suggests, a phonetic pattern observed in one context (e.g., *sat* or *sad*) may be adapted to another (e.g., *sent* or *send*), reinforcing these norms within a language's phonetic grammar. Such adaptations, deeply rooted in a language's linguistic heritage, cannot be fully explained by mechanistic accounts alone. Against this backdrop, languages can be broadly classified, as discussed in Cho (2015), into three types based on how consonantal voicing influences the duration of preceding vowels, reflecting cross-linguistic variation in the evolution of phonetic grammar.

One type includes ***phonological encoding languages***, such as English and German, which encode coda voicing contrast primarily through vowel duration (de Jong, 2004; de Jong & Zawaydeh, 2002). This effect often exceeds what is expected from purely mechanical factors, as seen in English (de Jong, 2004; Choi et al., 2016), where durational differences are exaggerated in contrast-enhancing contexts, particularly in stressed syllables with phrase-level pitch accents. In these languages, mechanically driven variation transitions into controlled, phonologized patterns shaped by output-oriented constraints, resulting in a phonetic rule governed by the interaction between phonetic grammar and phonology. German aligns with English in this regard, as vowels are systematically longer before phonologically voiced consonants in intervocalic contexts (Braunschweiler, 1997; Piroth & Janker, 2004), despite the coda itself carrying minimal phonetic voicing cues. German also exhibits utterance-final neutralization of coda voicing, though often incompletely, as evidenced by residual vowel duration differences (Kleber, John, & Harrington, 2010; Roettger et al., 2014). Similarly, Dutch shows intervocalic vowel duration effects linked to coda voicing, though these weaken in final position while still reflecting incomplete neutralization (Warner et al., 2004, 2006). Further research should investigate whether German and Dutch, like English, enhance coda voicing contrast under focus, reinforcing their classification as phonological encoding languages.

Another type includes ***phonetically defaulting languages***, such as Catalan and Arabic, among others, where coda voicing is related to vowel duration but only in a physiologically preferred manner (de Jong & Zawaydeh, 2002; Solé, 2007). It is noteworthy that Arabic, as demonstrated by de Jong and Zawaydeh (2002), incorporates phonological vowel-length (quantity) contrasts, potentially reserving vowel duration for the quantity contrast, yet still exhibits the coda voicing effect on vowel duration. Unlike English, Arabic does not further enhance vowel duration for phonological purposes of the coda voicing contrast, even when the coda voicing contrast is focused. Thus, it appears that the phonetic grammar of the phonetically defaulting languages has evolved in such a way that the phonetic implementation of the coda voicing simply follows a system-oriented constraint.

The third type includes ***phonetically idiosyncratic languages*** like Polish and Czech, where coda voicing has no effect on vowel duration (Keating, 1985). Despite Polish maintaining a clear voicing contrast intervocalically, vowel duration remains unaffected by the following consonant. Similarly, Czech, another West Slavic language, shows no such effect, likely due to its phonological vowel-length contrast, which reserves duration for signaling phonemic distinctions. This suggests that an output-oriented constraint suppresses the coda voicing effect, outweighing system-oriented constraints. Polish, however, lacks phonological vowel-length contrast yet still diverges from the expected phonetic universal. This suggests that its phonetic grammar has evolved along a less natural pathway, foregoing a low-cost, physiologically driven implementation. Whether this results from unknown influences on phonetic grammar or simply reflects phonetic arbitrariness remains an open question, warranting further research in the 21st century.

Several 21st-century studies have further examined the expansion of extrinsic vowel duration in English. Physiological accounts suggest that this effect is localized to the latter part of the vowel, aligning with the closing gesture for the coda consonant. Moreton (2004) found that in diphthongs (/ai ɔi ei aʊ/), offglides—not nuclei—become more peripheral before voiceless obstruents, indicating that the voicing effect is concentrated in the vowel's final portion. Although Moreton did not focus on duration, he argued that hyperarticulated offglides reflect increased articulatory force for voiceless consonants, consistent with Belasco's (1953) hypothesis. Other studies, however, report more global effects. Pycha and Dahan (2016) found that coda voicing influences both the nucleus and offglide timing in /aɪ/, regulating the entire vowel gesture. Choi et al. (2016) extended this finding, showing that not only vowel duration but also VOT in the preceding onset consonant is longer before voiced stops in both L1 English speakers and L2 Korean learners. Since VOT is inherently voiceless, its expansion cannot be attributed to the closing gesture of the coda or perceptual enhancement of the voiced consonant. However, when considering the intergestural timing between consonantal and vocalic gestures within the framework of Articulatory Phonology (Browman & Goldstein, 1992), the vocalic gesture may often begin unfolding well before the consonantal release, thereby overlapping with the period traditionally labeled as VOT. This overlap can make VOT temporally coextensive with an earlier phase of the vowel gesture (see Cho et al., 2014, for a related discussion). If the coda voicing effect were strictly localized to the coda, it would not be expected to influence the earlier vocalic gesture where VOT occurs. Similarly, Hawkins and Nguyen (2004) found that coda voicing affects the temporal realization of onset /l/, suggesting that voicing impacts the temporal coordination of the entire syllable, at least in syllables containing /l/.

These findings suggest that coda voicing influences the temporal organization of the entire vocalic gesture rather than being restricted to the coda-adjacent portion. This global timing adjustment appears to be a language-specific effect rather than a mere biomechanical consequence, at least in English, where vowel duration serves as a primary cue for coda voicing contrast. However, it remains unclear whether such global modifications are unique to phonological encoding languages like English or if they extend to other languages. Further research is needed to determine whether this pattern reflects a broader cross-linguistic mechanism of vowel duration adjustment or if it emerges specifically in systems where vowel duration is phonologized.

2.3.3. Output-oriented perspectives of extrinsic vowel duration due to coda voicing

The hypothesized global articulatory timing effect aligns with an auditory-perceptual account, suggesting that an entire vowel may lengthen to enhance the perception of a voiced coda, as explored by Raphael (2005). This perspective fits within an output-oriented speech production model: if vowel lengthening is perceptually motivated, there is no reason to confine it to the latter part of the vowel. However, this explanation does not account for the impact of coda voicing on onset VOT, as VOT neither relates to coda closure formation nor directly aids voicing perception. The challenge, then, is determining whether the coda voicing effect stems from auditory perception, biomechanical constraints, or both. Given the stability of this phonetic universal, physiological explanations cannot be dismissed, but perceptual mechanisms must also be considered, reflecting a balance between system-oriented efficiency and contrast maximization.

Javkin (1976), cited in Maddieson (1997), proposed that continued voicing in a consonant creates an auditory illusion, making the preceding vowel seem longer—an effect that, once internalized, could lead to actual lengthening. Kluender et al. (1988) offered a related hypothesis, noting that voiced stops generally have shorter closures (Lehiste, 1970; Ohala, 1983). To enhance contrast, speakers may lengthen vowels before voiced stops, making the already short closure perceptually even shorter. If this adjustment becomes encoded in articulation, the extended VOT as a "voiceless" segment of the vowel can be explained. However, Fowler (1992) found that vowels are perceived as longer before consonants with longer closures, contradicting Kluender et al.'s claim. In response, Maddieson (1997) proposed a speech production-based alternative, arguing that vowel and closure duration are inversely related due to timing constraints that maintain syllable-level timing balance. If this timing pattern is articulated, the elongated VOT may emerge from coordinated vowel-coda consonant timing, emphasizing articulatory organization over purely auditory effects. In short, Kluender et al.'s (1988) auditory-based account prioritizes perceptual distinctiveness between voiced and voiceless consonants, while Maddieson's (1997) speech production mechanism highlights timing constraints at the syllabic level. Regardless of its physiological or auditory basis, such timing differences influence perception (see Raphael, 2005).

In the 21st century, Sanker (2019) explored this perceptual impact through experiments with American English listeners judging vowel duration in Hindi speech. Listeners perceived vowels as longer before voiceless stops, suggesting perceptual compensation linked to English's phonologized coda voicing effect. That is, given the expectation that vowels shorten before voiceless stops, a vowel of equal duration sounds longer in that context. However, when the voiced coda was produced with breathy phonation, this compensation disappeared, as the breathiness itself contributed to a longer perceived vowel—indicating a low-level auditory rather than phonological effect. Interestingly, when the coda and VC formant transitions were removed, the effect reversed: vowels from voiced contexts were perceived as longer than those from voiceless ones. Sanker attributed this to intrinsic vowel cues such as first formant (F1) variations, influencing perception independent of compensation effects (cf. Moreton, 2004).

Insights from an auditory-perceptual perspective suggest that consonant voicing effects on preceding vowel duration cannot be attributed solely to physiological and biomechanical constraints—auditory-perceptual factors must also be involved. It is likely that once a physiologically grounded surface form emerges from system-oriented constraints within a language, it undergoes further modulation through the establishment of an auditory-perceptual target. Once internalized within phonetic grammar, this target may evolve in a language-specific manner. Future research in the 21st century should continue investigating this issue across a broader range of languages to gain deeper insights into how different linguistic systems negotiate system-oriented constraints with output-oriented adaptations in shaping coda-voicing-related extrinsic vowel duration.

2.4. Intrinsic f0 variation by vowel height (If0)

The height of vowels not only results in intrinsic vowel duration differences, as discussed earlier, but also causes variations in the fundamental frequency (f0) intrinsically associated with the vowel, a phenomenon commonly observed across languages. For example, high vowels such as /i/ and /u/ exhibit higher f0s compared to non-high vowels like /æ/ and /ɑ/ (e.g., Ohala, 1983; Whalen & Levitt, 1995; Whalen, Gick, Kumada, & Honda, 1998; Honda, 2004; Hoole, 2006; see Hoole & Honda, 2011, for a review). As is the case for other putative phonetic universals, the physiological underpinnings of this intrinsic f0 variation have also been long debated, potentially involving aerodynamical factors, laryngeal muscle activities (especially cricothyroid activity), as well as auditory-perceptual factors (see Maddieson, 1997, for related discussion).

2.4.1. Possible physiological and biomechanical bases of If0

The *tongue pull* hypothesis, first proposed by Ladefoged (1964), suggests that high vowel articulation exerts an upward pull on the larynx, increasing vocal fold tension and raising f0. Honda (1983) and later Hoole and Honda (2011) expanded this explanation, emphasizing the role of the posterior genioglossus muscle. Its contraction elevates the tongue, pulling the hyoid bone forward and tilting the thyroid cartilage, thereby increasing vocal fold tension and raising f0. Ohala (1978), building on Ohala and Eukel (1976), linked this effect to vertical tension in the vocal folds, primarily controlled by the cricothyroid (CT) muscle, which tilts the thyroid cartilage forward to raise pitch and relaxes to lower it. Since then, research into CT activity's role in high vowel f0 raising has continued. A long-standing debate, reviewed in Hoole and Honda (2011), concerns whether intrinsic f0 (If0) results from a passive mechanical process or active control. The mechanical view argues that f0 raising occurs automatically due to the tongue pull, supported by cross-linguistic data (Whalen & Levitt, 1995) and physiological evidence (Whalen et al., 1998). Most recently, Chen, Whalen, and Tiede (2021) further explored this perspective in 44 American English speakers using the X-ray microbeam database (Westbury, 1994) and the Haskins rate comparison database (Tiede et al., 2017). They examined the roles of the tongue and jaw, finding both correlated with f0, though the jaw had a greater impact on If0 than the tongue. Based on these findings, Chen et al. (2021) proposed a dual-mechanism hypothesis: in non-low vowels (including high vowels), f0 is primarily raised by the tongue-pull mechanism (linked to CT activity), while in non-high vowels (including low vowels), f0 is primarily lowered by a jaw-push mechanism, where jaw lowering pushes back the hyoid bone and thyroid cartilage, slackening the vocal folds (Erickson, Honda, & Kawahara, 2017). These studies, as argued by the authors, reinforce the mechanical basis of If0 rather than an actively controlled process.

2.4.2. Output-oriented perspectives of If0

Intrinsic f0 (If0) may involve active control rather than being purely mechanical. Electromyography (EMG) data show that CT muscle activity, which is controllable, correlates with vowel height (Dyhr, 1990; Honda & Fujimura, 1991). Because the posterior genioglossus influences the position of the hyoid bone, which in turn affects CT activity, speakers may modulate If_0 through intentional articulatory adjustments. Hoole and Honda (2011) suggest that such control could lead to phonologization over time. This hypothesis has shaped auditory-perceptual accounts, proposing that when two closely spaced spectral components (e.g., F1 and f0 within ~3.5 Bark) are perceived together, f0 perception shifts upward (Stevens, 1997; cf. Maddieson, 1997). Because high vowels have low F1 values close to f0, this integration may enhance vowel height perception. Speakers may replicate this effect by deliberately raising f0, actively controlling If0 to reinforce vowel height contrast (Diehl & Kluender, 1989; Hoemeke & Diehl, 1994). This aligns with phonetic knowledge (Kingston & Diehl, 1994) and featural enhancement (Keyser & Stevens, 2006), where phonetic adjustments optimize perception. As such, languages may actively adjust If0 for phonological purposes or due to phonetic arbitrariness. This aligns with Hoole and Honda's (2011) hybrid account, which posits that articulatory contingency drives If0 variation, though speakers may amplify these effects as part of linguistic behavior. Once again, biomechanical (system-oriented) constraints interact with linguistic (output-oriented) factors, including auditory-perceptual mechanisms, illustrating how languages adapt vocal mechanics within their sound systems. Potential influences include f0′s functional load in tone languages and its interaction with vowel inventory size.

At the start of the 21st century, Connell (2002) examined If0 in four African tone languages, each with four speakers, varying in tonal inventory size: Ibibio (two tones: High-Low), Kunama (three: High-Mid-Low), and Dschang and Mambila (four: High-UpperMid-LowerMid-Low). The study investigated whether If0 effects were present and how the functional load of f0 in tonal contrasts interacts with biomechanical constraints, given that strong If0 effects could interfere with tonal distinctions. Connell found If0 effects in Ibibio, Kunama, and Dschang, consistent with expected phonetic universals, though significantly weaker than those reported in Whalen and Levitt's (1995) cross-linguistic study. This suggests that tonal languages may suppress If0 to preserve tonal contrasts, unlike non-tonal languages. However, Mambila, despite having four tones like Dschang, exhibited no significant If0 variation, highlighting phonetic arbitrariness—where Mambila suppresses If0 more than Dschang, prioritizing output-oriented constraints in tone realization. Connell proposed that Mambila's suppression of If0 may be linked to its phonetic structure: it allows more bitonal patterns, creating a denser tonal space than Dschang. Some contours combine similar tone levels (e.g., UpperMid and High), reducing f0 distinctions. This structure may further suppress If0 to prevent excessive f0 perturbation. Connell's findings again underscore the interaction between biomechanical processes and phonological environments, showing that If0 control is language-specific, shaped not only by tonal inventory size but also by how tones are phonetically implemented in the language.

Van Hoof and Verhoeven (2011) investigated the interaction between If0 and the size of a language's vowel inventory, hypothesizing that If0 control may be utilized to enhance vowel height distinctions. The rationale behind the hypothesis is that If0 enhancement would be more pronounced in languages with larger vowel inventories because the crowded acoustic vowel space could lead to perceptual ambiguities between vowels. Therefore, If0 correlated with vowel height could help reduce these ambiguities by enhancing the height-related distinction among vowels. To test this, they compared If0 effects in Dutch, which has a 12-vowel system, to Arabic, which has a smaller 3-vowel system (although it makes additional distinctions by quantity). Their findings revealed a more significant If0 effect in Dutch than in Arabic, suggesting that If0 adjustments are tailored to the phonological demands of the language. Furthermore, they observed that Arabic speakers transferred their specific If0 usage into their second language (L2) English. This can be interpreted as implying that language-specific If0 control is deeply embedded in the phonetic grammar of the first language and may be carried over into L2 production. However, the phonetic traits observed in L2 could also result from simple low-level phonetic transfer, such as motor habits, necessitating further investigation into the relationship between L1 and L2 production at detailed phonetic levels.

2.5. Co-intrinsic f0 variation by consonant (Cf0)

Another extensively studied phonetic universal in the 21st century is Co-Intrinsic (consonant-dependent) f0 variation. The term Co-Intrinsic f0, often abbreviated as Cf0, has been widely used since its early adoption by Di Cristo and Hirst (1986). This recurrent pattern, first observed in the mid-20th century (House & Fairbanks, 1953; Peterson & Lehiste, 1961), refers to the tendency of vowel f0 to vary depending on the voicing or laryngeal configuration of the preceding obstruent. Specifically, vowels exhibit a higher initial f0 when preceded by a voiceless consonant rather than a voiced one, a phenomenon documented across languages (Di Cristo & Hirst, 1986; Kingston & Diehl, 1994; Cho, Jun, & Ladefoged, 2002; Francis et al., 2006; Kingston, 2007; Hanson, 2009; Chen, 2011; Dmitrieva et al., 2015; Kirby & Ladd, 2016; Kirby, 2018; Gao & Arai, 2019; Xu & Xu, 2021, among others). While this type of f0 variation is extrinsic in that it arises from a consonantal, thus external influence on the vowel, the term Co-Intrinsic highlights its inherent link to the consonant's phonetic properties. This f0 variation is generally attributed to physiological and biomechanical processes associated with onset obstruent voicing. Since the effect is typically localized to the vowel onset, transiently deviating from the macro-level f0 contour, it is also referred to as *f0 perturbation* (Hombert, 1978; Hombert, Ohala, & Ewan, 1979), *pitch skip* (Haggard, Ambler, & Callow, 1970; Francis et al., 2006; Hanson, 2009), or *micro f0* (Kohler, 1990; Jun 1996). Various theoretical accounts have been proposed to explain this phonetic universal, which will be discussed below.

2.5.1. Possible physiological and biomechanical bases of Cf0

One possible mechanism for Cf0 is aerodynamic. During voiced stop closure, rising oral pressure reduces transglottal airflow, weakening voicing and risking devoicing unless compensatory adjustments occur. While passive expansion of the supralaryngeal cavity helps (Ohala & Riordan, 1979), active strategies like laryngeal lowering or cheek expansion further reduce intraoral pressure, sustaining vocal fold vibration (Ohala, 1997). Since laryngeal lowering lowers f0 (Ewan & Krones, 1974; Riordan, 1980), this compensatory action may inadvertently affect pitch, resembling deliberate modulation (Hombert, 1978; Maddieson, 1997) and potentially becoming a controllable parameter in Cf0 modulation. Another explanation involves vocal fold tension. Halle and Stevens (1971) proposed that voiceless obstruents increase vocal fold tension, raising f0, while voiced obstruents relax the folds, lowering f0. The extent to which this variation is automatic or speaker-controlled remains debated. Some suggest that Cf0 is regulated by cricothyroid (CT) muscle activity (Sonesson, 1982), as CT contraction tilts the thyroid cartilage forward, stretching the vocal folds and raising f0. Löfqvist et al. (1989) found higher CT muscle activity during voiceless obstruents in English and Dutch speakers, supporting active control. However, they also observed that CT activity peaks during consonant closure, before vowel onset, where Cf0 effects emerge. This timing mismatch suggests Cf0 is not directly controlled but rather a biomechanical byproduct of vocal fold tension during voiceless stops.

Hoole and Honda (2011), in an EMG study of three German speakers, provided further evidence linking CT activity to voicing. They found that increased CT activity closely follows the timing of glottal opening and closing gestures, supporting Löfqvist et al. (1989). However, they also observed cases where Cf0 occurred without significant CT activity, suggesting that Cf0 effects may extend beyond what would be predicted by EMG data alone. Interestingly, they also found instances of robust CT activity during the vowel itself, potentially enhancing Cf0 to accentuate consonantal differences. This effect, unrelated to consonantal voicing, suggests an active modulation of Cf0. Based on these findings, Hoole and Honda concluded that "[t]he driving force comes from the articulatory contingency, but once established, speakers can deliberately emphasize its effects" (p.165). This supports the idea that Cf0 may shift from a system-oriented constraint to an output-oriented feature as a language evolves, with German possibly internalizing Cf0 within its phonetic grammar.

2.5.2. Output-oriented perspectives of Cf0

The variability in Cf0 led Kingston and Diehl (1994) to argue that its correlation with voicing arises from speakers' phonetic knowledge—specifically, that adjusting Cf0 enhances phonological voicing contrasts, driven by output-oriented constraints. Their view is based on the assumption that voiced sounds inherently concentrate energy at lower frequencies more than voiceless sounds. This acoustic pattern can be further reinforced by maintaining a lower f0 following a voiced consonant, sharpening the perceptual distinction between voiced and voiceless phonemes. Kingston and Diehl's argument underscores the primacy of auditory representation in speech processing, paralleling Kluender et al.'s (1988) analysis of If0. This perspective has influenced many 21st-century studies exploring how Cf0 contributes to voicing contrasts, particularly in relation to VOT. Several of these cases are reviewed below.

Kirby (2018) examined Cf0's role in voicing contrasts across three Southeast Asian languages—Central Thai, Northern Vietnamese (tonal), and Khmer (non-tonal)—each with a three-way stop contrast: (pre)voiced, voiceless unaspirated, and voiceless aspirated. The results showed that while VOT consistently distinguished stops across all languages, Cf0 varied: Khmer (non-tonal) showed the strongest Cf0 effect, whereas Thai and Vietnamese (tonal) exhibited weaker Cf0, likely to avoid tonal interference. Similarly, Ladd and Schmid (2018) investigated Swiss German, where fortis and lenis stops contrast mainly in closure duration rather than VOT. They found that fortis stops had a consistently higher Cf0, persisting into the vowel and exceeding what would be expected from a purely mechanical effect. Moreover, an emerging aspirated stop category with even higher Cf0 suggests an ongoing sound change. Al-Tamimi and Khattab (2018) explored Lebanese Arabic, analyzing Cf0 in singleton and geminate stops. While VOT distinguished voiced-voiceless pairs, Cf0 primarily helped differentiate geminates, suggesting context-dependent control to enhance the four-way (Voicing $\times$ Quantity) contrast. Drawing from these cross-linguistic patterns—similar VOT distinctions and systematic Cf0 variation, as Kirby (2018) and Ladd & Schmid (2018) argued, Cf0 complements VOT, underscoring the need to refine phonetic and phonological models of voicing contrasts (see Cho et al., 2019; Kirby & Ladd, 2016, for related discussions on *true voicing* languages like French and Italian).

A more pronounced transformation observed in the 21st century regarding Cf0 is its role in tonogenetic sound change, as seen in languages like Afrikaans (Coetzee et al., 2018; see Kingston, 2011) and Seoul Korean (Kang, 2014; Bang et al., 2018; Choi et al., 2020; see Cho, 2022). In Afrikaans, Coetzee et al. (2018) found that while speakers varied in prevoicing production, they consistently lowered f0 after voiced stops, creating a robust f0 contrast between voiced and voiceless stops. Listeners also relied on f0 as a cue, with younger speakers favoring it over prevoicing, suggesting an ongoing shift in which f0 is replacing prevoicing as the primary cue for voicing contrast. In Seoul Korean, the traditional VOT distinction between lenis and aspirated stops has weakened, while f0 has emerged as a key cue, especially in phrase-initial positions. Lenis stops induce lower f0 and aspirated stops higher f0, despite both being phonetically voiceless. This shift—VOT decreasing and f0 increasing—has been interpreted as an ongoing tonogenetic change (Silva, 2006; Oh, 2011; Kang, 2014; Bang et al., 2018).

Alternatively, Choi et al. (2020) proposed a *prosodic account*, arguing that Korean's f0 shift is not true tonogenesis but a prosodic restructuring. In Korean intonational phonology, phrase-initial lenis stops receive a low tone, while aspirated and fortis stops receive a high tone (Jun 1998, 2000). This phonologizes Cf0 into phrase-level tonal contrast, distinct from the developed phonemic tone found in languages such as Afrikaans or in tone languages like Kammu and Cham (Kingston, 2011). In phrase-medial positions, the lenis-aspirated contrast remains clearly expressed via VOT and phonetic voicing, with f0 playing only a perturbation role. Thus, Seoul Korean does not exhibit transphonologization (Hyman, 2008; Bang et al., 2018), as it has not incorporated new tones into its phonemic system. Instead, post-lexical tones have taken over the phonological contrast between lenis and aspirated stops in phrase-initial positions, rendering VOT redundant. This shift, driven by effort minimization (Flemming, 1995; Boersma, 1998), has reorganized cue primacy, favoring post-lexical tones over VOT. The prosodic account thus highlights a unique case where a physiologically driven phonetic process (Cf0) has reshaped tonal organization, modifying cue weighting in an intonation-dependent manner.

All these observations, within the framework of output-oriented perspectives on Cf0, are consistent with other proposed phonetic universals. While Cf0 reflects universal phonetic patterns, speakers modulate it to conform to language-specific phonologies and prosodic structures, supporting the idea that Cf0 is internalized within a language's phonetic grammar, shaped by higher-order influences such as lexical tonal contrast and post-lexical intonational structure.

2.6. Consonant place effects: Closure duration and VOT

As the last case of a segmental-level phonetic universal to be discussed in this article, let us consider the recurring effects of place of articulation, focusing on how these effects manifest in closure duration and VOT in stop production. Numerous studies over the past two centuries have indeed shown that the place of articulation of consonants significantly influences both the timing of closures and the onset of voicing.

2.6.1. Place effect on stop closure and VOT

Place of articulation systematically affects closure duration and VOT across languages. Stop closure duration generally follows the pattern bilabials > alveolars > velars, though the alveolar-velar relationship varies (Weismer, 1980; Stathopoulos & Weismer, 1983; Docherty, 1992; Maddieson, 1999). A key biomechanical factor is intraoral pressure dynamics: in velar stops, the smaller oral cavity leads to faster pressure buildup and quicker release. However, Maddieson (1997) suggested that the compressibility of the tongue against the soft palate in velars may prolong closure, whereas the tongue tip's rapid movement in alveolar (coronal) allows for a shorter closure.

VOT follows a cross-linguistic trend: velars > alveolars > bilabials (Fischer-Jørgensen, 1954; Lisker & Abramson, 1964; Cho & Ladefoged, 1999). This is influenced by multiple biomechanical factors, as discussed in Cho & Ladefoged (1999). Here again, the smaller cavity behind velar constrictions causes rapid intraoral pressure buildup, delaying vocal fold vibration. Additionally, the larger air mass in front of velar constrictions slows pressure release, delaying transglottal pressure drop. The slower movement of the tongue dorsum may further prolong pressure release, while extended contact between the tongue dorsum and the soft palate may create a more gradual release due to the Bernoulli effect, in which rapid airflow through a narrowing constriction temporarily pulls articulators together. This, along with a slower decrease in intraoral pressure post-release, can prolong VOT and occasionally result in double releases.

2.6.2. Relationship between closure duration and VOT, and uniformity constraint on VOT

Research in the 1980s and 1990s examined the inverse relationship between closure duration and VOT, investigating whether a fixed glottal abduction duration creates temporal invariance. Studies in American English and French suggest that the total voiceless interval (closure + VOT) remains relatively constant across places of articulation (Weismer, 1980; Laeufer, 1996). This pattern aligns with the idea that glottal abduction overlaps with closure and VOT (Maddieson, 1997). Weismer (1980) explains that the glottal abduction gesture begins with oral constriction, while voicing starts only after glottal adduction, when the vocal folds close sufficiently for vibration. If abduction duration is fixed, longer closures (e.g., bilabials) leave less time for aspiration, shortening VOT, while shorter closures (e.g., velars) allow more time, extending VOT. This principle also accounts for shorter VOT in /s/-stop clusters in American English (Browman & Goldstein, 1986; Goldstein, 1992; Cho et al., 2014). Since glottal abduction likely begins with /s/, by the time the stop is released, the abduction phase is nearly complete, leaving less time for aspiration and reducing VOT.

Some studies further explored the relationship between closure duration and VOT. Smiljanic and Bradlow (2008) compared English, an "aspirating" language, and Croatian, a "true voicing" language, and found that English voiceless stops had increased VOT in clear speech compared to conversational speech, whereas Croatian showed the opposite trend. However, when VOT was measured as a percentage of total stop duration, it remained stable across speaking styles and languages. This relational invariance reflects language-specific pronunciation norms that maintain consistent temporal patterns for stop voicing contrasts. This finding does not imply a simple inverse relationship between closure duration and VOT from a purely mechanistic perspective. Studies on British English (Docherty, 1992) and French (Abdelli-Beruh, 2009) found no clear inverse correlation between the two measures. Docherty (1992) argued that language-specific phonetic rules introduce micro-variability, refining stop production timing in British English, and Abdelli-Beruh (2009) reached similar conclusions for French. These findings reinforce the view that while physiological mechanisms, such as glottal abduction timing, shape cross-linguistic patterns, the coordination of closure duration and VOT exhibits systematic language-specific variation, likely internalized within each language's phonetic grammar.

3. Utterance-level phonetic universals and language variation

Speech production is orchestrated through three essential components—respiration, phonation, and (supralaryngeal) articulation—each harmoniously contributing to the refined articulation of the speech signal. Among these, it appears that respiration has received less scholarly attention in linguistic phonetics, presumably because phonetic theories predominantly define linguistic contrasts through features related to phonation and articulation. Nonetheless, it is essential to recognize that phonation and articulation are inextricably dependent on the process of respiration. This is because the respiratory process regulates the egressive pulmonic airstream in response to motor commands likely issued by the central nervous system, aimed at achieving specific phonatory and articulatory targets.

Ladefoged (1967), in *Three Areas of Experimental Phonetics*, provided key insights into the respiratory mechanisms of speech production, drawing on an EMG study from the 1950s in Edinburgh, later refined at UCLA (Ladefoged & Loeb, 2002). His findings highlight how respiratory muscles regulate subglottal pressure and lung volume during speech. Elastic recoil forces initiate exhalation after inspiration, but external intercostal muscles and the diaphragm actively modulate airflow to prevent rapid lung volume loss, ensuring smooth speech. As lung volume decreases, passive recoil becomes insufficient, requiring activation of internal intercostals, rectus abdominis, and internal obliques to maintain pulmonic egressive airflow. This supports the hypothesis that the central nervous system integrates passive elasticity and active muscle control to regulate subglottal pressure, achieving motor equivalence and acoustic stability across varying lung volumes (Ladefoged & Loeb, 2002, p.59). Ladefoged (1967) also proposed that phonetic stress is implemented as a respiratory gesture, with respiratory muscles modulating air volume and subglottal pressure to produce stressed articulation. His analysis, based on full-length utterances, emphasized phrase-level stress (pitch accents), reinforcing that phonetic prominence—particularly nuclear pitch accents in English—requires active respiratory control to align pulmonic airflow with articulatory demands. More broadly, his work underscores the close interaction between respiration, laryngeal function, and vocal tract dynamics in shaping speech, reinforcing the central role of respiration in linguistic phonetics, as extensively reviewed in Fuchs and Rochet-Capellan (2021).

The discussion thus far has emphasized the controlled aspect of respiratory processes in shaping the aerodynamic conditions necessary for speech production at the utterance level. However, as Ladefoged explained, these processes operate within the physiological constraints of the speech production system as a whole. The interplay between respiratory dynamics and articulation unfolds within these confines, shaping the phonetic structure of utterances. Consequently, cross-linguistic similarities in speech must arise, manifesting as phonetic universals at the utterance level. These universals stem from respiratory constraints, which fundamentally influence phonation and articulation across all three phases of speech production.

A fundamental concept linked to respiratory processes across all human languages is the ***breath group***, defined by Lieberman (1966) as "a prosodic pattern that delimits the boundaries of unemphatic, declarative sentences in normal speech" (p. 27). While theoretical refinements exist, it is broadly understood as closely tied to subglottal air pressure, regulated by respiratory muscles to sustain an utterance within a single expiratory phase (e.g., Wang et al., 2010; cf. Rochet-Capellan & Fuchs, 2013). Each breath group typically aligns with a unit of speech produced on a single breath cycle, often forming an utterance or phrase. As phonation continues, increased muscular effort is needed to compress the rib cage and maintain airflow, leading to a gradual decrease in subglottal pressure and f0 declination (Ladefoged, 1967). In Lieberman's model, this occurs in an *unmarked* breath group with automatic f0 fall, contrasting with a *marked* breath group, such as an interrogative, where f0 rises due to active control.

Beyond its physiological basis, the breath group could be seen as a potential foundation for prosodic phrasing, possibly influencing frameworks in prosodic and intonational phonology (e.g., Selkirk, 1984, 1986; Nespor & Vogel, 1986; Beckman & Pierrehumbert, 1986; see Shattuck-Hufnagel & Turk, 1996; Beckman, 1996; Ladd, 2008). Many utterance-level phonetic universals arise from respiratory constraints interacting with phonation and articulation, as well as from prosodic strengthening and weakening patterns. The breath group can thus be interpreted as a prosodic constituent, such as an Intonational Phrase. At its onset, a ***respiratory reset*** due to inspiration increases subglottal pressure and may enhance laryngeal and supralaryngeal articulatory force. As speech progresses, these forces decline, weakening the utterance-final position. This asymmetry—a stronger left edge and a weaker right edge—underlies f0 declination, articulatory declination, phrase-final lengthening, and domain-initial strengthening, topics explored in the following subsections.

3.1. f0 declination

F0 declination refers to the gradual decrease in f0 over the course of an utterance (e.g., Cohen & 't Hart, 1967; Gelfer, Harris, & Baer, 1987; Maeda, 1976; Nooteboom, 1997; 'tHart, Collier, & Cohen, 1990). Observed cross-linguistically (see Fuchs et al., 2015), it occurs continuously when phonologically driven intonational contours are factored out. However, a low boundary tone in a language's intonational grammar may reinforce the downdrift pattern, even though Lieberman (1966) classified it as *unmarked*. F0 declination is widely regarded as a universal property of intonation (Ladd, 2008), evident in the gradual decline of pitch peaks and troughs within phonologically defined landmarks, such as L*, L + H*, and H* in English or rising pitch accents in Japanese (Beckman & Pierrehumbert, 1986). The effect is particularly pronounced in longer utterances and is often reset at phrase or sentence boundaries (*declination reset*), where the starting f0 of a new phrase is raised before declining again. While there is debate over whether f0 declination is actively controlled (see Fuchs & Rochet-Capellan, 2021), it is generally linked to subglottal pressure decline corresponding with decreasing lung volume (e.g., Gelfer et al., 1987; Maddieson, 1997). This physiological basis likely contributes to its cross-linguistic prevalence, reinforcing its status as a putative phonetic universal.

Various studies in the 21st century further reinforce that f0 declination is not solely driven by respiratory physiology (e.g., Watson et al., 2003; Fuchs et al., 2015; Fuchs & Rochet-Capellan, 2021). Fuchs et al. (2015) examined this relationship in German speakers, testing whether f0 declination correlates with rib cage movement (linked to lung volume and subglottal pressure). If f0 declination were purely biomechanical, its slope would vary with utterance length (syllable count) and the presence of voiceless obstruents (e.g., /f/), which require greater airflow than sonorants (e.g., /m/) (Gelfer et al., 1987). Specifically, more voiceless obstruents should lead to steeper f0 declination and greater rib cage compression. However, their results showed that while f0 declination depended on syllable count, rib cage movement did not. Moreover, f0 declination remained unaffected by voiceless obstruent count, though rib cage compression was influenced by it. This suggests that f0 declination and rib cage movement are not directly linked, contradicting a purely biomechanical explanation. Instead, Fuchs et al. concluded that f0 declination is actively modulated, reflecting anticipatory mechanisms in speech planning (cf. Fuchs, Petrone, Krivokapić, & Hoole, 2013; see Scholz & Chen, 2014; Kim & Tilsen, 2024, for discussions on planning and f0 scaling). Additionally, f0 declination was steeper in read than in spontaneous speech, suggesting a possible adjustment to output-oriented constraints, likely enhancing listener comprehension in specific contexts. This parallels various segmental-level phonetic universals: while f0 declination may have physiological underpinnings, it is further modulated by speakers, interacting with higher-order linguistic and communicative factors, which must then be internalized in the phonetic grammar of a language.

3.2. Articulatory declination and phrase-final lengthening, weakening or strengthening

3.2.1. Articulatory declination

Just as f0 declines over the course of an utterance, articulatory movements also tend to decrease in magnitude (e.g., Vayra & Fowler, 1992; Krakow et al., 1985; Vatikiotis-Bateson & Fowler, 1988; Krakow, 1993; Krakow, Bell-Berti, & Wang, 1995). This ***articulatory declination*** has been observed in velic positions (Krakow, 1993; Krakow et al., 1995), tongue position for open vowels (Vayra & Fowler, 1992), jaw movements (Vatikiotis-Bateson & Fowler, 1988), and the lower lip (Krakow et al., 1991). Analogous to f0 declination, this pattern likely reflects general speech production mechanisms. As the energy from the initial expiratory phase is expended, articulatory force diminishes, contributing to both articulatory and f0 declination.

Articulatory declination has received less attention than f0 declination in 21st-century phonetic research, likely because f0 declination is closely linked to intonational grammar, whereas articulatory declination appears less directly tied to phonological contrasts. While articulatory declination should theoretically follow a linear pattern, empirical data often do not reflect this clearly (Keating et al., 2003). Instead, articulation tends to weaken mid-phrase rather than decline consistently throughout. Identifying causal links between physiological mechanisms and surface articulatory patterns is thus more complex than for laryngeal-respiratory behaviors (Krakow et al., 1995). Language-specific factors, such as stress, further complicate its study (Krakow, 1993; Krakow et al., 1995). Given the complexity of articulatory declination and its interaction with language-specific factors, further studies are particularly welcome in the years to come in the 21st century to better understand the mechanisms underlying this phenomenon.

Despite these complexities, articulatory gestures generally start with greater force at the beginning of a breath group, aligning with respiratory reset, much like the higher initial f0. However, this process is not always linear, and some languages exhibit phrase-final articulatory strengthening rather than weakening. While speakers tend to initiate utterances with strong articulatory force, language-specific factors—including segmental inventories, tonal systems, and prominence marking—can modify articulation throughout the utterance, counteracting purely mechanistic declination. These patterns will be explored further in the sections on phrase-final lengthening, phrase-final strengthening, and domain-initial articulatory strengthening.

3.2.2. Preboundary lengthening

Phrase-final lengthening, or ***preboundary lengthening*** (henceforth PBL), refers to the elongation of segments or articulatory gestures near the end of a prosodic constituent, typically before a boundary (e.g., Edwards et al., 1991; Wightman et al., 1992; Gussenhoven & Rietveld, 1992; Berkovits, 1993; Byrd, 2000; Byrd & Saltzman, 2003; Cho, 2006; Turk & Shattuck-Hufnagel, 2007; Paschen et al., 2022). This phenomenon is widely attested across languages (Fletcher, 2010; Cho, 2016; Paschen et al., 2022), with segments showing greater duration in IP-final positions than in non-IP-final or IP-medial positions. The causal mechanisms behind preboundary lengthening remain complex, though biomechanical constraints offer one explanation. Lindblom (1968) proposed that it results from a natural deceleration of articulatory movements toward the end of an utterance, akin to a car slowing before a stoplight. This frames final lengthening as *supralaryngeal declination in the temporal domain*. However, this deceleration is not uniform across an utterance but rather localized near the boundary, with the degree of lengthening increasing as the segment nears the prosodic edge (Byrd & Saltzman, 2003; Byrd & Krivokapić, 2021). Speakers thus appear to dynamically adjust articulatory timing, ensuring smooth phrase-final transitions rather than a continuous slowing from utterance onset.

Lindblom (1968) further discussed that PBL helps distribute energy across syllables, ensuring that each is produced with constant energy. As subglottal pressure and intensity naturally decline toward the end of an utterance (Öhman, 1967), lengthening compensates for this loss, maintaining the syllable's acoustic prominence. This also explains why phrase-final lengthening does not always involve spatial expansion (Edwards et al., 1991; Beckman et al., 1992), though languages differ in how they balance weakening vs. strengthening at phrase boundaries. Beyond energy distribution, other factors may contribute to PBL. One possibility is respiratory regulation—speakers may slow down near the end of an utterance to manage breath intake for the next phrase (Lieberman, 1967). Another explanation relates to speech planning: PBL may provide extra time for organizing the upcoming phrase (Cooper & Paccia-Cooper, 1980). However, White et al. (2020) newly observed that speakers still lengthen phrase-finally even when they are not planning the next utterance, suggesting that PBL may also function to signal turn-taking by regulating breath in preparation for upcoming speech events. Moreover, as pauses also serve a planning function (Grosjean & Collins, 1979; Krivokapić, 2007), it remains unclear how final lengthening alone interacts with cognitive planning. Some studies suggest a trading relationship between the two—longer PBL often correlates with shorter pauses, though some recent studies indicate that this varies cross-linguistically (e.g., a stronger correlation in Mandarin than in English, Wang et al., 2019; but see Kentner et al., 2023, for German). Further research is certainly needed to clarify the role of pauses in relation to PBL, particularly in how different languages balance final lengthening and pausing in prosodic structuring and speech planning.

PBL also interacts with phrase-final pitch movements. Many languages exhibit delayed pitch peaks at prosodic boundaries, requiring extra time for tone realization (Gussenhoven, 2016). More complex tones tend to occur phrase-finally, where sufficient duration allows for their full articulation (Zhang, 2004). Supporting this, Li, Kim, & Cho (2023) found that in Mandarin Chinese, PBL is more extended for Tone 3 (low-dipping) and Tone 4 (falling) than for Tone 1 (high level). However, rather than PBL causing tonal complexity, researchers suggest that tonal distribution is constrained by existing durational limits, highlighting co-occurrence rather than causation (Zhang, 2004; DiCanio et al., 2021).

Significant advancements in 21st-century research on PBL have emerged from a perceptual perspective, emphasizing its critical role in lexical segmentation. Prosodic boundaries provide crucial cues that help listeners parse continuous speech into discrete words. While traditionally viewed as a phrasal-level phenomenon, phrasing directly influences word recognition by shaping listeners' expectations about word boundaries in conjunction with computed prosodic boundaries. Numerous studies have indeed demonstrated that prosodic junctures—marked by preboundary lengthening as well as other boundary-related cues such as pitch resets, and pauses—not only facilitate segmentation but also reduce lexical competition, thereby enhancing spoken word recognition (e.g., Cutler, 2012; McQueen & Dilley, 2020; Steffman et al., 2022). PBL, in particular, serves as a key cue for prosodic boundaries, reinforcing lexical segmentation and speech comprehension (Cutler, 2012; McQueen & Dilley, 2020; White et al., 2020; Warner, 2023). While it remains still unclear whether perceptual advantages actively shape PBL, speakers may modulate its extent based on contextual demands (Cutler, 2012; White et al., 2020). This modulation likely reflects speakers' phonetic knowledge (Kingston & Diehl, 1994), allowing them to fine-tune temporal adjustments to optimize listener comprehension. Listeners, in turn, process speech by computing prosodic structure, with PBL serving as one of many cues (see Cho et al., 2007; McQueen & Dilley, 2020; Steffman, et al., 2022, for discussions on prosodic structure in spoken word recognition).

3.2.2.1. Cross-linguistic variation of prebounndary lengthening. A substantial body of 21st-century research on PBL has further clarified language variation within the putative universal PBL effect. While PBL may initially appear to be a physiologically driven phenomenon, it is highly language-specific, interacting with various linguistic factors such as lexical stress (e.g., English: Turk & Shattuck-Hufnagel, 2007; Cho, Kim, & Kim, 2013; Kim, Jang, & Cho, 2017; Greek: Katsika, 2016), mora structure (e.g., Japanese: Shepherd, 2008; Seo, Kim, Kubozono, & Cho, 2019), vowel quantity (e.g., Finnish: Nakai et al., 2009; other languages: Paschen et al., 2022), and tonal complexity (e.g., Mandarin Chinese: Li et al., 2020; Yoloxóchitl Mixtec: DiCanio et al., 2021). The core premise that emerges from these studies is that the fine phonetic details of PBL should be included in linguistic descriptions of the phonetics-prosody interface (cf. Section 1.1) as part of the phonetic grammar of a language. Below, I will briefly outline some of the notable developments on this aspect in the 21st century, focusing on examples that illustrate how PBL interacts with the phonological and prosodic structures across different languages.

PBL interacts with lexical stress placement in head-prominence languages such as English and Greek. Studies by Turk and Shattuck-Hufnagel (2007) in English and Katsika (2016) in Greek show that while phrase-final lengthening is robust in the rime of the final syllable, non-final stressed syllables (penultimate or antepenultimate) may also attract boundary-related lengthening. In English, this attraction can bypass an intervening unstressed syllable, as seen in words like *Michigan*, indicating multiple targets for PBL (Turk & Shattuck-Hufnagel, 2007), where both the final syllable and a non-final stressed syllable (if pitch-accented) serve as lengthening targets. However, Cho et al. (2013) observed cases where lengthening extended to the antepenultimate syllable even when it was unstressed (*banana*), suggesting a leftward gradient effect independent of stress and complicating the identification of PBL's phonetic target.

In Greek, Katsika (2016) found intricate interactions between stress and phrase-boundary lengthening (PBL). Using electromagnetic articulography (EMA), Katsika examined how PBL interacts with lexical stress and phrasal accent, showing that, unlike the English case reported in Turk and Shattuck-Hufnagel (2007) where multiple PBL targets are observed, Greek exhibits a gradient leftward spread of lengthening from the final syllable toward non-final stressed syllables without directly targeting them. Notably, this attraction of PBL toward non-initial stressed syllables occurs independently of phrase accent, indicating that PBL operates with direct reference to lexical stress rather than being mediated by phrase-level prominence. This contrasts with English, where PBL attraction to a stressed syllable appears to occur when it receives a phrasal accent. These findings highlight the complexity of the phonetics-prosody interface and the need for language-specific models to account for PBL-stress interactions. Katsika (2016) proposes a dynamical systems framework in which stress- and boundary-related gestures are coupled, explaining the gradient nature of PBL in Greek. For further discussion, see Katsika (2016), Byrd & Krivokapić (2021), and Iskarous & Pouplier (2022).

PBL also interacts with moraic structure, particularly in Japanese. Shepherd (2008) proposed that PBL is confined to the final mora rather than the final vowel, indicating a restrictive domain. However, Seo et al. (2019) found a more complex pattern, showing that final syllable duration in CV words (one mora, V) was comparable to CVN words (two moras, V and N). This suggests that PBL is governed by syllable structure (final rime) rather than strictly by moraic structure (final mora). Their study further revealed that PBL spreads leftward, affecting all segments of disyllabic words except the first onset, with lengthening extending to the first vowel across all word types (CV.CV, CV.CVN, CVN.CV, CVN.CVN), regardless of mora count. Moreover, words with N.C clusters (CVN.CV, CVN.CVN) exhibited *non-final boosting*, where the presence of a medial mora led to pronounced lengthening in the non-final rime. This suggests a complex interaction between PBL and moraic structure. PBL in Japanese is also influenced by lexical pitch accent. When the initial syllable of a disyllabic word carried a pitch accent, the rime of the final (non-pitch-accented) syllable showed suppressed PBL. This contrasts with English and Greek, where PBL expands in stressed, non-final syllables while the final syllable undergoes robust lengthening (Turk & Shattuck-Hufnagel, 2007; Katsika, 2016). In Japanese, this suppression appears to prevent excessive prominence on the final syllable, preserving the relative prominence of the pitch-accented initial syllable.

A similar interaction between PBL and prosodic contrast is observed in Finnish, where vowel quantity plays a role in maintaining syntagmatic distinctions (Nakai et al., 2009). In Northern Finnish, which exhibits phonological length contrasts aligned with moraic structure, PBL extends to non-final stressed syllables, similar to Greek. However, this effect is constrained by vowel quantity: PBL of a short vowel is suppressed to preserve its contrast with long vowels, and PBL of a long vowel is restricted when preceded by another long vowel. Nakai et al. attributed this to ***syntagmatic constraints***—maintaining contrast between adjacent vowels. A phrase-final long vowel following a short vowel can be freely lengthened to enhance contrast, but when both vowels are long, excessive final lengthening would compromise the phonological length distinction between them Similarly, in Japanese, the suppression of PBL in the final, non-pitch-accented rime may serve to maintain temporal *syntagmatic* contrast with the initial, pitch-accented rime. The results of these two studies suggest that while different languages manage prominence and boundary effects differently, they may be guided by a common principle of maintaining syntagmatic contrast in organizing temporal structure along with PBL. Here again, such syntagmatically-driven temporal realization may also be specified in the phonetic grammar of these languages, in relation to their phonological quantity system or pitch accent system, respectively.

Paschen et al. (2022) conducted a large-scale study on PBL, analyzing final lengthening and vowel duration across 25 languages using spontaneous speech data from the DoReCo corpus (Seifart, Paschen, & Stave, 2022). Among their findings, a key observation emerged from languages with vowel length contrast that allow both short and long vowels in final and non-final positions. These languages exhibited three distinct patterns. The first involved strong lengthening in both final and penultimate positions (e.g., Fanbyak, Svan), with PBL strongest in the final syllable and significantly reduced in the penultimate. The second pattern showed lengthening only in the penultimate position, with no PBL in the final position (e.g., Beja, Bora, Evenki, Resígaro). In this group, Bora and Resígaro exhibited final devoicing, which may explain the absence of observable PBL. In Beja and Evenki, the suppression of final PBL may prevent excessive lengthening, preserving vowel length contrast. The third pattern involved complex interactions between vowel length and position (e.g., Arapaho, Baïnounk Gubëeher, Dolgan, Movima). In these languages, short vowels exhibited progressive lengthening from the penultimate to the final position, whereas long vowels lengthened only in the penultimate position. The suppression of PBL on final long vowels likely prevents excessive lengthening, maintaining syntagmatic contrast—a pattern similar to that observed in Northern Finnish. Based on their observations of PBL in 25 languages, Paschen et al. suggested that PBL is a prosodic phenomenon “deeply entrenched in the phonology of individual languages,” and its manifestation reflects the need to preserve phonological contrasts. Although the actual vocabularies used may differ, this suggestion aligns with the view proposed in the present article: that PBL observed across languages as a phonetic universal is indeed internalized within the phonetic grammar of a given language and must be refined with reference to its phonological system as well as the prosodic structure of the language, engendering language variation within a phonetic universal.

3.2.3. Preboundary (phrase-final) strengthening

Thus far, I have discussed preboundary lengthening with the assumption that its physiological origin is related in some ways to articulatory relaxation and declination towards the end of an utterance. This process may also coincide with a single respiratory cycle, culminating in a single breath group that is likely to form an intonational phrase (IP). In line with this assumption, the phrase-final position has often been understood as a locus for phonological neutralization due to phonetic weakening or loss of important cues to phonological contrasts (e.g., Beckman, 1998; Myers, 2007; Hyman, 2009). In fact, a kinematic study on English by Beckman et al. (1992) showed that PBL, despite its temporal expansion, is not accompanied by spatial expansion, whereas stress-related temporal expansion is. Nevertheless, just as we have seen ample evidence that the PBL process is systematically modulated by speakers of a given language in relation to various factors, studies in the 21st century have also indicated that the phrase-final articulation may be controlled, often countering the presumably default weakening process at the end of a phrase. Articulatory studies on some languages have indeed shown ***preboundary strengthening*** rather than weakening at the end of a phrase (in English, Cho, 2004, 2005; in French, Tabain, 2003; Tabain & Perrier, 2005, 2007; in Korean, Kim, Kim, & Cho, 2024; in Mandarin Chinese, Li, Kim, & Cho, 2023). The 21st century has thus witnessed the phrase-final position serving as a locus for both articulatory weakening and strengthening, demonstrating *dual positional processes*. In a similar vein, it is worth noting that in Scottish English, an utterance-final position also acts as a locus for both weakening and strengthening of /r/ due to socio-indexical factors, as demonstrated in working-class versus middle-class speech, respectively (Lawson & Stuart-Smith, 2021; cf. Kendall, Pharao, Stuart-Smith, & Vaughn, 2023). This dual positional effect aligns with Barnes’ (2002) view that the phrase-final position serves as a locus for phonologically licensing both strengthening and weakening processes. For the remainder of this section, I will further discuss preboundary strengthening effects documented in 21st-century studies across various languages, highlighting how these effects interact with phonetic and prosodic structures specific to each language.

Prosodic boundary strengthening is often described as a process that enhances the syntagmatic contrast between the phrase-final vowel and the upcoming consonant at the start of a new prosodic domain (e.g., Fougeron & Keating, 1997). This effect is evident in vowel articulation, where English /ɑ/ and /i/ become more peripheral phrase-finally than phrase-medially, with /i/ showing tongue raising and /a/ showing tongue lowering and backing (Cho, 2004); and in French, where phrase-final /i/ and /u/ exhibit increased tongue fronting and backing, respectively (Tabain & Perrier, 2005, 2007). In French, this strengthening effect aligns with its head/edge-prominence system, where prominence is marked at the right edge of the phrase (Jun, 2014). However, its presence in English suggests an intrinsic articulatory effect linked to phrase-final position, possibly due to greater temporal availability for articulatory expansion, independent of language-specific prominence system.

Studies in the 21st century further reveal that preboundary strengthening patterns, when present, vary across languages depending on their prosodic typology (cf. Jun, 2014). In English, a head-prominence language where a stressed syllable with a pitch accent functions as the head of the phrase, preboundary strengthening differs from prominence-related strengthening in its kinematic properties. While prominence-induced strengthening involves hyperarticulation across all dimensions—resulting in larger, longer, and faster movements (Fowler, 1995; de Jong, 1995; Cho, 2005)—preboundary strengthening typically leads to larger and longer but not necessarily faster articulation (Cho, 2005). In contrast, Korean, an edge-prominence language, marks prominence and prosodic junctures through phrasing, exhibiting larger, longer, and faster articulation phrase-finally, paralleling prominence effects in English (Kim et al., 2024; cf. Jang, 2023; Jang and Katsika, 2024). Similarly, Li et al. (2023) found that Mandarin Chinese exhibits phrase-final articulatory strengthening with spatio-temporal expansion, resembling prominence-induced strengthening in English. In Mandarin, the phrase-final position enhances tonal contrast, much like in Yoloxóchitl Mixtec, where tonal distinctions are reinforced phrase-finally (DiCanio et al., 2021).

All these observations suggest that languages may choose phrase-final positions where ample timing is available as a potential physiological underpinning for fully realizing the articulatory targets, as found in French, Korean, and Mandarin Chinese, all of which do not use lexical stress as the locus for phrase-level prominence. In such cases, preboundary lengthening—possibly rooted in a low-level phonetic universal of articulatory declination—appears to serve as a hyperarticulated locus, counteracting the default weakening at the right edge of prosodic constituents. This phenomenon can be interpreted as an outcome of speaker control, likely facilitated by the additional time available for its realization.

Before concluding this section, it is worth briefly addressing boundary-related temporal variations within the task dynamic model and Articulatory Phonology (Browman & Goldstein, 1992; Goldstein, Byrd, & Saltzman, 2006). This topic is extensively covered by Iskarous & Pouplier (2022) in the special issue to which this article belongs, so I will highlight only the key concept of the π-*gesture* theory. Developed by Byrd and colleagues (e.g., Byrd & Saltzman, 2003; Byrd, Krivokapić, & Lee, 2006; see Byrd & Krivokapić, 2021, for a review), this theory proposes that boundary-related temporal variations arise not from changes in the dynamical parameters of vocal-tract constriction gestures but from modulation by a prosodic gesture, the π-*gesture*. Anchored at a prosodic boundary, the π-*gesture* slows down articulatory gestures, with its effect strongest at the boundary and tapering off with distance. Its influence is assumed to scale with boundary strength, peaking at the IP boundary. This approach provides a cross-linguistically applicable explanation for prosodic juncture-related temporal and spatial variations (Byrd & Saltzman, 2003), offering a dynamical underpinning for phonetic universals of boundary-related lengthening effects. However, it remains difficult to fully account for language-specific preboundary patterns solely through the π-*gesture* (see Cho, 2016, and Iskarous & Pouplier, 2022, for discussion; but also see Katsika, 2016, for an attempt to integrate the π-*gesture* with a stress-related μ-*gesture*). While a π-*gesture*-like prosodic gesture may universally govern articulation at prosodic boundaries, its specific effects are likely shaped by the phonetic grammar of each language. The language variation discussed in this section represents just a fraction of the boundary-related phenomena that may continue to emerge as research expands across languages in the 21st century.

3.3. Domain-Initial Strengthening (DIS)

Domain-initial strengthening (henceforth **DIS**) generally refers to articulatory strengthening that is frequently observed when a segment (typically a consonant) is produced at the beginning of a larger prosodic constituent or phrase-initially than at the beginning of a smaller prosodic constituent or phrase-medially. While the exact patterns of articulatory strengthening differ from language to language, as with other putative phonetic universals, the strengthening effect has been observed across languages, suggesting it as a possible phonetic universal: in English (Fougeron & Keating, 1997; Cho & Keating, 2009); in French (Fougeron, 2001; Georgeton & Fougeron, 2014); in Korean (Cho & Keating, 2001; Cho & Jun 2000); in Japanese (Onaka, 2003; Onaka, Watson, Palethorpe, & Harrington, 2003); in Taiwanese (Hayashi, Hsu, & Keating, 1999); in German (Kuzla, Cho, & Ernestus, 2007; Kuzla & Ernestus, 2011; Bombien, Mooshammer, & Hoole, 2013); in Dutch (Cho et al., 2007); in Thai (Silpachai, 2024); in Spanish (Napoleão de Souza, 2023); in Portuguese (Napoleão de Souza, 2023). (See Keating et al. (2003) for cross-linguistic comparisons, and Cho (2016) for a review.).

The concept of DIS is the inverse of articulatory declination, yet they are interrelated. DIS emphasizes that articulatory energy or force is at its peak at the beginning of an utterance, leading to articulatory strengthening at the left edge of a prosodic constituent, typically larger than the word level such as an Intonational Phrase or an intermediate level of phrase. Thus, DIS can be seen as a flip side effect of articulatory declination (Krakow et al., 1995), indicating that the two phenomena arising at the opposite edges of a prosodic constituent may stem from speech production mechanisms operating on similar utterance-level biomechanical bases. For example, the fact that the velum position becomes lower, and thus weakened, towards the right edge of an utterance means that the utterance starts with an elevated velum position at its left edge. Studies on DIS in the 21st century have indeed consistently reported such an effect: the nasal murmur for a word-initial nasal consonant is significantly reduced phrase-initially compared to phrase-medially (Fougeron, 2001; Cho & Keating, 2001, 2009; Cho, Kim, & Kim, 2017; Jang, Kim, & Cho, 2018). Such elevated velum position phrase-initially is considered as domain-initial strengthening of oral articulation. For a possible physiological underpinning of DIS, Fougeron (1999) posited that it is attributable to "articulatory force" (cf. Straka, 1963). This force is defined as "the amount of energy necessary for all muscular effort involved in the production of a consonant" (Delattre, 1940, translated). From this perspective, the spatio-temporal expansion observed for a domain-initial consonant is considered a manifestation of domain-initial *articulatory* strengthening (Fougeron & Keating, 1997). Such articulatory force, possibly along with aerodynamic force due to a reset of the respiratory cycle, affects the supralaryngeal articulation. This leads to an elevated velum, thereby reducing both the duration and the acoustic nasal energy of the nasal murmur, while the oral constriction itself for the nasal is spatio-temporally strengthened (Cho & Keating, 2001, 2009).

Another consideration for the basis of DIS is the sufficient time required to achieve target articulation. DIS, particularly for oral articulations, is highly correlated with temporal expansion. Cho and Keating (2001) proposed an undershoot-based account in the sense of Lindblom (1963) and Moon and Lindblom (1994), suggesting that the physiological basis lies in the ample time available to fully reach the articulatory target as also discussed with regard to preboundary articulation. This perspective underscores the importance of temporal factors in ensuring precise and effective speech production. On a related point, however, this account does not explain the often-observed language-specific articulatory weakening rather than strengthening towards the end, where no spatial expansion is typically observed, despite the ample time provided by temporal expansion. In such cases, intrinsic articulatory weakening at the right edge reflects the absence of active strengthening despite the availability of sufficient time, highlighting the default side of the dual positional function at the right edge.

DIS effects, like other putative phonetic universals, exhibit language-specific variation, interacting with higher-order linguistic factors. Cho and McQueen (2005) demonstrated this in Dutch, where the voiceless stop /t/ shows a shorter VOT in domain-initial positions compared to domain-medial ones, contrasting with English, where /t/ lengthens domain-initially due to DIS of the glottal abduction gesture. They argue that despite both languages sharing the phonological feature [-voice], their voiceless stops differ in phonetic specification: Dutch /t/ is {−spread glottis}, while English /t/ is {+spread glottis}. This suggests that DIS operates within language-specific phonetic constraints, shaping phonetic realization in accordance with a language's phonological system and becoming internalized in phonetic grammar. A recent study on Thai (Silpachai, 2024) further supports this claim. Thai's three-way stop contrast /b p pʰ/ shows DIS effects aligned with each stop's phonetic features: /b/ exhibits more negative VOT, /pʰ/ more positive VOT, while /p/ remains largely unchanged. Similarly, Cho and Jun (2000) found that in Korean, DIS effects reflect phonetic feature distinctions. The fortis stop, specified with {constricted glottis}, shows no VOT change, while the aspirated stop, marked by {spread glottis}, exhibits longer VOT domain-initially. The lenis stop, assumed to be underspecified for laryngeal features (e.g., Cho et al., 2002), still undergoes default strengthening, producing a longer VOT domain-initially. (But recall that the phonetic (VOT) distance between lenis and aspirated stops, once distinct in early 21st-century Korean (Cho & Jun, 2000), has since diminished significantly as discussed above in Section 2.5.2).

These studies on DIS in Dutch, Thai, and Korean support the view that DIS operates in relation to a language's phonetic features, though the interaction is complex and language-specific. Differences in prosodic structure, phonetic feature realization, and potential phonetic arbitrariness further complicate the picture, requiring further investigation. Specifically, it remains unclear whether DIS-driven enhancement in a given language primarily serves to increase paradigmatic contrast among phonemes (Cho & Jun, 2000; Cho & McQueen, 2005; Silpachai, 2024), similar to prominence-induced enhancement (e.g., de Jong, 1995), or to enhance syntagmatic contrast between consonants and adjacent vowels for juncture demarcation (Fougeron & Keating, 1997; Fougeron, 1999, 2001; Cho, 2016), or both. Emerging evidence suggests that while DIS is closely tied to a language's phonological system, its implementation varies across languages. In the following subsections, I will explore these language-specific nuances through illustrative cases from 21st-century studies.

3.3.1. Domain-initial strengthening across different prosodic typologies

Keating et al. (2003) examined DIS in four languages and found language-specific effects potentially linked to each language's prominence system. Similar to Korean's robust phrase-final strengthening (Section 3.2.2), they observed that Korean exhibits stronger DIS than English, French, and Taiwanese. As an edge-prominence language without lexical stress, Korean does not need to reserve the initial syllable for stress marking, allowing it to employ DIS more extensively than English, which emphasizes the head rather than the edge for prominence at the phrase level, or French, which emphasizes prominence at the right edge (cf. Jun, 2014). Barnes (2002) further suggests that DIS interacts with how initial syllables signal stress. While English reserves vowel duration for stress or pitch accent marking, Turkish—where stress is typically final (Gordon, 1998; Inkelas & Orgun, 2003)—and Korean, which lacks lexical stress, exhibit DIS effects on both consonants and vowels. This, as noted by Barnes (2002), also explains why some scholars have misidentified initial syllables in Turkish as stressed, as they perceived DIS effects when hearing words in isolation, where the word-initial position coincides with the IP-initial position. French shows DIS effects on vowels (Georgeton & Fougeron, 2014) and consonants (Fougeron, 2001), aligning with Barnes' view that French vowels are not primarily reserved for marking lexical stress, particularly word-initially. Korean DIS data also reveal effects on vowels, influencing vowel duration (Cho & Keating, 2001) and lip opening movements (Cho, Son, & Kim, 2016), with effects extending even to the second syllable (Cho, Lee, & Kim, 2011)—none of which are observed in English. These findings demonstrate that DIS manifests differently across languages, shaped by prosodic typologies.

DIS also influences tonal realization. Since f0 is regulated by subglottal pressure, which is modulated through DIS, domain-initial strengthening naturally interacts with tone. The phrase-initial f0 reset, marked by a raised pitch register and expanded f0 range (Ladd, 2008), can be viewed as an f0-related DIS effect. However, an across-the-board higher f0 at the onset could obscure tonal distinctions in tone languages. Research on Thai (Silpachai, 2024) and Taiwanese Min (Pan, 2009) shows that tone languages systematically modulate DIS effects on f0 to preserve tonal contrasts. In Thai, Silpachai (2024) found that the maximum f0 during the vowel was lower in IP-initial than in phrase-medial positions, regardless of tone type, diverging from the typical initial f0 reset. This suppression was more pronounced for low tones than for falling and mid tones, with the falling tone affected more than the mid tone. Silpachai suggested that these tonal DIS effects enhance contrast at the IP level, particularly by making low tones more distinct.

3.3.2. Domain-initial strengthening of laryngeal articulation

DIS effects also manifest as laryngeal strengthening, particularly through enhanced glottal abduction. The lengthening of VOT in phrase-initial positions, especially for aspirated stops, is often attributed to increased glottal abduction gestures (Pierrehumbert & Talkin, 1992; Cho & Keating, 2001; Cho, 2016). This correlation is well-documented in transillumination and acoustic studies of American English, where a greater glottal opening corresponds to longer VOT (Cooper, 1991). The strengthening of the glottal abduction gesture is likely driven by both laryngeal and supralaryngeal articulatory forces, aligning with prosodic resets and increased respiratory effort, which raises subglottal pressure. However, language-specific constraints, as observed in Dutch, Thai, and Korean (see the discussion beore Section 3.3.1), may modulate this effect, highlighting the need for further investigation.

Glottalization of word-initial vowels at larger prosodic boundaries, such as IP or utterance-initial positions, represents another form of laryngeal strengthening (e.g., Pierrehumbert & Talkin, 1992; Dilley et al., 1996; Gordon, 1998; Redi & Shattuck-Hufnagel, 2001; Fougeron, 2001; Pompino-Marschall & Żygis, 2010; Mitterer et al., 2019, 2020; Shin et al., 2023; Hwang et al., 2023). Unlike aspiration, which results from glottal abduction, glottalization arises from reinforced glottal adduction, often involving a "braced" rather than fully "pressed" glottal configuration (Pierrehumbert & Talkin, 1992; Ladefoged & Maddieson, 1996). This "braced" setting, common in natural speech, produces irregularities in glottal excitation pulses rather than a full glottal stop, suggesting controlled vocal fold activity. Esling et al. (2019) describe glottalization as involving vocal fold adduction (linked to a "pressed" configuration) or ventricular fold incursion (associated with a "braced" configuration), accounting for variations in glottalization strength (Davidson, 2021). Regardless of configuration, increased subglottal pressure and articulatory force likely reinforce glottalization at prosodic boundaries. This supports the view that domain-initial glottalization is an utterance-level phonetic universal, sharing key mechanisms—such as subglottal pressure and laryngeal articulation—with other phonetic universals discussed thus far.

Such glottalization as a DIS effect has often been considered another articulatory signature of prosodic organization employed by languages around the world (Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001; Fougeron, 2001; Mitterer et al., 2019; 2020; Shin et al., 2023; Hwang et al., 2023). This again highlights the possibility that the low-level process of glottalization is modulated in reference to prosodic structure, though language-specific aspects are pronounced. For example, two recent studies on this aspect—one on American English (Shin et al., 2023) and one on Korean (Hwang et al., 2023)—demonstrate such language variation. In resolving syntactically ambiguous coordinate structures ("A and B or C", which can be parsed either as "[A and B] or [C]" or "[A] and [B or C]"), speakers of both languages tend to glottalize the word-initial vowels along with phrasings where all the nouns and conjunctions started with vowels. Comparing the results of the two studies indicates that American English speakers exhibit much stronger and more frequent glottalization phrase-initially compared to Korean speakers. This language variation may be due to languages employing non-contrastive glottalization differently, demonstrating some phonetic arbitrariness. However, it may not be purely arbitrary, but may also be linked to the distinct prosodic typologies of the two languages, with English being a head-prominence language and Korean an edge-prominence language. This would be especially true when the initial vowels in English are produced with pitch accent, another factor that induces glottalization (e.g., Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001; Garellek, 2014, 2022; Steffman, 2023), while Korean has no such additional factor involved. In particular, a possible explanation may lie in Korean's intonational phonology, where phrase-initial vowels are assigned a Low tone (Jun, 1998, 2000), whose phonetic implementation differs from that of other languages. Unlike English pitch-accented L* or Mandarin's low-dipping Tone 3, Korean's initial Low tone is not produced with creaky phonation, which could otherwise enhance glottalization. In fact, the lenis stop, which is known to be produced with a somewhat breathy voice quality, is characterized by this breathiness phrase-initially when an L tone is assigned. While this interpretation remains preliminary, the cross-linguistic differences suggest language-specific modulation of glottalization—English influenced by pitch accent, Korean by tonal restrictions at the phrase level.

Finally, it is important to note several other key aspects of glottalization. First, glottalization occurs not only phrase-initially but also phrase-finally (Dilley et al., 1996; Davidson, 2021), likely due to a braced laryngeal configuration affecting articulation around prosodic junctures. However, phrase-final glottalization, or "phrase-final creak," differs from DIS-induced glottalization, as it is linked to articulatory weakening and f0 declination rather than laryngeal reinforcement (Garellek & Keating, 2015; González et al., 2022; Ogden, 2001; White et al., 2020), while still serving as an "end-of-sentence" marker (Davidson, 2021). Second, Garellek (2014) suggests that phrase-initial glottalization, especially in English, is not solely due to DIS but is often linked to pitch accent, emphasizing the role of prominence. But again, Shin et al. (2023) confirm that phrase-initial glottalization also occurs in non-pitch-accented words, as observed in English (Pierrehumbert & Talkin, 1992) and German (Pompino-Marschall & Żygis, 2010; cf. Kohler, 1994), reinforcing its connection to prosodic positioning. Third, glottalization occurs even in Maltese, where the glottal stop is phonemic (Mitterer et al., 2019, 2020). Despite the potential for phonological confusion, Maltese speakers still employ glottalization for prosodic boundary marking. Perceptual studies (Mitterer et al., 2019, 2020) show that glottalization aids spoken word recognition and helps resolve syntactic ambiguity by aligning syntactic and prosodic boundaries, indicating its cross-linguistic role as a prosodic marker, regardless of whether the language uses glottalization contrastively or not.

3.3.3. More on output-oriented perspectives of domain-initial strengthening

DIS, like other phonetic universals, may be shaped by or linked to output-oriented factors. Research in the 21st century suggests that listeners benefit from DIS in speech perception (Cho et al., 2007; White et al., 2020; McQueen & Dilley, 2020). White et al. (2020) found that speakers of English, Hungarian, and Italian use consonantal strengthening as a cue for lexical segmentation in artificial language learning but do not rely on the increased vowel duration that follows. They argue that consonantal DIS provides a more universally accessible cue than preboundary lengthening (PBL). Similarly, Cho et al. (2007) showed that onset consonantal strengthening facilitates lexical segmentation in English (see McQueen & Dilley, 2020, for a related discussion). Further, Kim and Cho (2013) and Mitterer et al. (2016) found that listeners perceive the same VOT duration for voiceless aspirated stops in English as shorter when preceded by an IP boundary compared to a smaller prosodic boundary. These findings suggest that DIS reinforces syntagmatic CV and VC contrasts across prosodic junctures, aiding lexical segmentation much like PBL. However, while output-oriented constraints may shape speech for the listener's benefit, they do not solely drive these patterns. Assuming DIS originates from physiological and biomechanical foundations within utterance-level phonetic universals, its effects likely evolved across languages in response to system-oriented and output-oriented constraints, giving rise to language-specific patterns in perception.

4. Summary and discussion

In this review article, I have explored the intricate interplay between phonetic universals and language variation, contributing to the journal's special collection on *Advancements of Phonetics in the 21st Century*. In particular, I have endeavored to offer balanced perspectives on phonetic universals and language variation across both segmental and utterance-level phenomena. Utterance-level phenomena, as can be inferred from Maddieson (1997), have often been relegated to a supplementary role in the literature. Thus, by giving comparable emphasis to both dimensions and underscoring the significance of utterance-level phenomena closely related to the phonetics-prosody interface, this approach seeks to provide a more integrated perspective on the topic, which may help enhance our appreciation of the human language system in a more holistic manner.

In Section 2, I explored putative phonetic universals at the segmental level, beginning with *the role of syllables* in organizing features, gestures, and segments. I then examined various segmental phonetic universals, including *intrinsic vowel duration* influenced by vowel height, *extrinsic vowel duration* affected by coda voicing, *intrinsic f0 variation (If0)* linked to vowel height, *co-intrinsic f0 variation (Cf0) driven* by preceding consonant voicing, and *place effects on closure duration and VOT.* All these phenomena seem to have emerged as phonetic universals inherent in the physiological and biomechanical underpinnings stemming from anatomical structures shared by humans. These universal patterns are imposed on the segmental production of speech sounds, reflecting consistent operational mechanisms of the "speaking device," which elucidates common speech production mechanisms across languages. Crucially, however, languages also exhibit unique variations in the phonetic implementation of these putative phonetic universals, leading to diverse aspects of language variation. These language variations may arise from the distinct sound systems of individual languages, among other factors. For example, tonal languages, which rely on f0 variations for phonological distinctions, may shape f0-related phonetic universals differently from non-tonal languages. Languages that use quantity for phonological contrast may accommodate duration-related phonetic universals differently from those that do not use quantity. Some languages may emphasize the auditory-perceptual outcomes of each phonetic universal, prioritizing output-oriented constraints for listeners over system-oriented constraints. Remarkably, some languages may not follow these patterns at all, demonstrating phonetic arbitrariness that underscores language-specific idiosyncrasies. All in all, these language variations within phonetic universals highlight unique aspects of each language's sound system and imply its different evolutionary pathways in accommodating putative phonetic universals inherent in human physiological and biomechanical traits. The discussed insights thus illuminate how phonetic processes that may have originally arisen as *automatic* have long been *controlled* differently by speakers of different languages, which highlights linguistic diversity—although there may still be some automatic low-level mechanisms underlying these language universals.

In Section 3, I continued the discussion of language variation within phonetic universals by examining utterance-level phenomena. Foundational insights from Ladefoged's work in the 20th century (1967) underscore the role of the respiratory system in speech production, emphasizing the interaction among respiratory mechanisms, phonation, and articulation in shaping utterance-level dynamics across languages. The concept of the breath group, introduced by Lieberman (1966), further illustrates how respiratory control influences speech production in relation to prosodic phrasing, thereby reflecting underlying prosodic structure. This presumed reset of the entire production system may give rise to utterance-level complementarities at the left and right edges of prosodic constituents—a phenomenon frequently observed across languages and documented in a body of subsequent studies, including recent 21st-century work (see Section 3). These include *f0 declination*, *articulatory declination*, and *phrase-final (preboundary) lengthening (PBL)*, which are understood as articulatory weakening processes toward the end of an utterance, although some, particularly f0 declination, may occur continuously throughout a phrase. But they also encompass possible strengthening processes such as *domain-initial strengthening (DIS)* and *phrase-final (preboundary) strengthening*. Unlike segmental-level universals, each with distinct physiological bases, the foundations of these utterance-level universals appear to share common denominators, stemming from interrelated aspects of respiratory and articulatory "resets." These resets typically align with utterance onset and can be considered as systematic recalibrations of the speech production system, involving reinforced articulatory force and energy. They are assumed to impact respiration, phonation, and articulation simultaneously, contributing to the recurrent patterns observed throughout an utterance across languages.

Yet, as with segmental-level phonetic universals, we have seen ample evidence that the actual implementation of these phenomena can be regulated in a language-specific manner. For example, the manifestation of f0 declination or articulatory declination can vary depending on a language's use of tones, intonational phonology, and prosodic conventions. Similarly, the extent and manner of domain-initial strengthening and preboundary lengthening appear to be shaped by these language-specific constraints. More specifically, many of the putative utterance-level phonetic universals can vary depending on prosodic typologies: whether the language is a head-prominence language, where vowels may be reserved for prominence marking and therefore boundary marking is less vowel-dependent, or whether it is an edge-prominence language, where prominence is marked by highlighting both edges. In addition, boundary-related strengthening may interact with the segmental inventory of the language from the viewpoint of paradigmatic contrast: It may be suppressed if the outcome becomes detrimental to phonological contrast or augmented if there is a need to enhance phonological contrast. Furthermore, if the language employs quantity contrast, it may consider syntagmatic contrasts in relative terms, such that, for example, a boundary-related lengthening effect on a vowel may be suppressed if the outcome undermines its quantity contrast syntagmatically with adjacent vowels. These interactions between phonetic universals and language-specific constraints give rise to the observed variations of the utterance-level phonetic universals, leading to a rich diversity of linguistic systems.

Importantly, the phonetic diversity observed at both the segmental and utterance levels across languages also means that these patterns can no longer be solely accounted for by physiological and biomechanical bases. As language-specific patterns deviate further from these phonetic universals, the original driving force of these universals becomes elusive, necessitating nuanced phonetic descriptions tailored to each language.

4.1. Integrating phonetic universals and language variation through Phonetic Grammar: An extended model

From a broader theoretical perspective, the present review has also sought to advance our knowledge beyond understanding the physiological and biomechanical traits of putative phonetic universals and cross-linguistic variation. It aims to make strides toward applying these insights in a theoretically informed way, providing a perspective closer to understanding language variation within phonetic universals as they occur in real-world scenarios. In these contexts, the dynamic and functional aspects of language—such as the interplay between system-oriented and output-oriented constraints, or between the principles of effort minimization and contrast maximization—indeed manifest themselves in intricate phonetic variations and fine phonetic details. These details are not merely non-contrastive low-level phonetic variations or random noise; they contain systematic linguistic information that reflects the intricate subtleties and nuances of the linguistic message intended by interlocutors. The converging evidence discussed in the present review indicates that such information-rich phonetic granularity is language-specific and shared by speakers within the same linguistic community. Understanding this phonetic fine-tuning cannot be achieved by simply examining the mapping between the abstract, symbolic representation of sounds governed by phonology and the low-level phonetic implementation that follows the laws of physics. An intermediary component of the spoken language system is needed to mediate these two stages, ensuring that the actual phonetic realization of the symbolic phonological representation is fine-tuned to reflect all these language-specific constraints and interactions before the final motor commands are issued. Such a component is the *phonetic grammar*, as elaborated in the introduction (Section 1.1) and continuously discussed throughout this review, and in this section, I elaborate on an extended model of phonetic grammar.

The phonetic grammar, as originally conceptualized by Keating (1985, 1990), is a key grammatical component of the linguistic structure. It refines phonological representations with fine-grained phonetic detail before motor commands are issued, ensuring that the surface outputs are adherent to pronunciation *norms* unique to each language or each linguistic community. As discussed in Section 1.1, the notion of phonetic grammar has then been extended by Cho and Ladefoged (1999) to account for phonetic arbitrariness that cannot be captured by phonetic principles and constraints imposed on speech output, and by Cho (2015, 2016) to capture the fine phonetic detail that arises from the phonetics-prosody interface, that is, the speaker's modulation of phonetic realization in reference to prosodic structure in speech production. The language variations within phonetic universals discussed in this review further support the concept of language-specific modulation of putative phonetic universals within the extended framework of phonetic grammar. This modulation, being influenced by various phonological, prosodic, and auditory-perceptual factors, is assumed to be achieved through the phonetic grammar of individual languages. This also implies that each phonetic grammar has evolved through different pathways, shaping the present-day phonetic granularity of phonetic universals as they currently exist in a language-specific manner.

To account for the wide range of language variation within phonetic universals, ***the Extended Model of Phonetic Grammar***, briefly alluded to in Section 1.1, can be fully defined as follows: It not only specifies which phonetic features a language chooses to express phonological contrast and how these selected features are phonetically implemented—often resulting in idiosyncrasies (phonetic arbitrariness) that do not align with established principles and constraints—but it also shapes phonetic content in reference to various higher-order linguistic (and possibly extra-linguistic) structures, governing the interface between phonetics and prosodic structure as well as other linguistic systems.

With this extended model of phonetic grammar in mind, let us recall that in Section 3, utterance-level phonetic universals were taken to be closely related to the formation of breath groups, which reflect the prosodic structure of a given utterance. The observed cross-linguistic similarities (and differences), however, are not exclusively pertinent to utterance-level phenomena *per se.* This is because utterance-level phonetic universals concern how the segmental and suprasegmental details of *individual sounds* may vary over the course of an utterance due to physiological and biomechanical constraints that are imposed at the utterance level of speech production. Therefore, systematic segmental variations arising from interactions between phonetics and prosodic structure can be seen as related to utterance-level phonetic universals. In accounting for such interactions, I have elaborated, whenever necessary, in Section 3 on the rationale behind the proposal that language-specific details of utterance-level phonetic universals must be specified within the phonetic grammar, which may fine-tune phonetic realization with reference to prosodic structure.

To further illustrate the intricate relationship between segmental-level and utterance-level phonetic universals, consider the example of co-intrinsic f0 (Cf0) perturbation due to the voicing of the preceding onset, superimposed on the segmental realization of vowels, as discussed in Section 2.5. The universality of Cf0 may be rooted in either aerodynamic factors intricately linked to the respiratory system, vocal fold tension associated with laryngeal strengthening, or both. Interestingly, the respiratory system and laryngeal strengthening, which are precisely the mechanisms arising from utterance-level phonetic universals discussed in Section 3, play a critical role in Cf0 perturbation. That is, the respiratory and articulatory resets at the beginning of an utterance are expected to influence the underlying mechanisms of Cf0. Indeed, previous studies have indicated that Cf0 impacts are more robust at the phrase-initial position compared to phrase-medial positions, suggesting modulation of Cf0 by prosodic structure associated with the articulatory force stemming from utterance-level phonetic universals. This case is crystalized in the case of Korean where segmentally driven Cf0 (*micro* f0) is directly linked to use of *macro* f0 in intonational phonology, shaping utterance-level f0 realization. This implies that understanding phonetic universals necessitates a combined approach that integrates both segmental and utterance-level phenomena. It is, in fact, conceivable that they are inseparable after all, as segmental phonetic universals must be produced within an utterance governed by utterance-level phonetic universals.

To further substantiate this proposal, it is important to note that prosodic structure (or prosody) is generally assumed to be an integral component of spoken language and is considered a grammatical entity in its own right, interacting with various other linguistic and extra-linguistic structures (Beckman, 1996; Shattuck-Hufnagel and Turk, 1996; Fletcher, 2010; Cho, 2016, 2022). The notion of the phonetics-prosody interface, which implies that phonetic realization is fine-tuned with reference to prosodic structure, then raises the important question of how such fine-tuning is actually implemented within the complex relationship between phonetic grammar and the phonetics-prosody interface. Specifically, the complexity arises because an essential role of prosodic structure is to provide a frame for articulation, serving as one of the pivotal linguistic structural components underlying the phonetic encoding process in speech planning (e.g., Keating, 2006; Keating & Shattuck-Hufnagel, 2002). If this frame for articulation includes detailed specifications of phonetic realization in a fine-grained way with reference to prosodic structure (along with specifications of phrasing and prominence distribution), the role of phonetic grammar in regulating fine phonetic detail would seem redundant. On the other hand, if we assume that prosodic structure is still an abstract representation that must be planned in advance to provide a rather coarsely defined frame for articulation (e.g., Keating, 2006; see Cho, 2016, 2022 for related discussion), there must be a mechanism for phonetic realization to reference this prosodic structure. Such a mechanism may be what Keating (2006) refers to as the “phonetic encoding of prosodic structure” (see Cho, 2022, for further discussion). This mediating role between phonetics and prosodic structure is fulfilled by the phonetic grammar. The Extended Model of Phonetic Grammar, offers the potential to encompass the fine-tuning of phonetic realization in reference to various higher-order linguistic systems. These interfaces can be labeled as, but are not limited to, the phonetics-phonology interface, the phonetics-morphology interface, the phonetics-syntax interface, the phonetics-semantics interface, and the phonetics-prosody interface. Within these possibilities, a more parsimonious theory would suggest that phonetic grammar serves as the central *hub* where all factors, including those from prosodic structure, that influence phonetic realization are collectively taken into account, ultimately determining the final phonetic shape of an utterance— this being the crux of the Extended Model of Phonetic Grammar proposed here.

This, however, does not mean that phonetic realization is not fine-tuned with reference to prosodic structure, as we have ample evidence to the contrary. It means that the prosodic structure created online during speech planning does not encompass all the fine phonetic details needed to determine the final surface forms. Rather, it is the phonetic grammar that governs the actual phonetic encoding of prosodic structure, fine-tuning the phonetic realization accordingly at the last stage before motor commands are issued. Nonetheless, the role of prosodic structure extends beyond this rather simplified description. Independent evidence from the literature indicates that prosodic structure may also interact with various other linguistic systems, as reflected in the labels of the prosody-syntax interface, the prosody-semantics interface, and the prosody-discourse interface, among others. Thus, the frame for articulation provided by prosodic structure is already information-rich, encapsulating these interactions cohesively. This frame is therefore more than just coarsely defined; it reflects influences from other linguistic systems, as well as the interplay between segmental-level and utterance-level phonetic universals. But its output, still as a *frame*, is fed into the phonetic grammar, where the final fine-tuning of phonetic content occurs.

5. Conclusion and future directions

This review has meticulously explored the intricate interplay between phonetic universals and language variation, emphasizing the importance of integrating both segmental and utterance-level phenomena through the extended notion of phonetic grammar, now labeled as *The Extended Model of Phonetic Grammar.* Highlighting key advancements in phonetic research from the early 21st century to the present, while tracing back to foundational studies of the 20th century, this work underscores the role of language-specific modulation in shaping the phonetic realization of universal patterns. By discussing how phonetic grammar mediates the fine-tuning of segmental phonetic details in reference to prosodic structure whose effects are assumed to be driven by utterance-level phonetic universals, this review offers a new perspective on the relationship between segmental and utterance-level phonetic universals. Traditionally considered separately, these universals are shown to be intricately related, or possibly inseparable, while their phonetic implementations vary across different languages. It is proposed that phonetic grammar fulfills the role of fine-tuning segmental phonetic universals in relation to utterance-level phonetic universals within the framework of the phonetics-prosody interface. This process refines the phonetic output in a language-specific manner, engendering language variation within phonetic universals. Furthermore, these insights contribute to our understanding of language evolution, illustrating how phonetic processes, once considered automatic, are controlled by speakers in ways that reflect the unique evolutionary pathways of different languages. The speaker control may take into account both system-oriented and output-oriented constraints in interaction with various other language-specific factors. Through these insights, we can appreciate the dynamic and functional aspects of language, recognizing that phonetic processes are intricately shaped by speakers of different languages. This adaptation, which varies across languages, contributes to the rich diversity of linguistic systems within phonetic universals observed today.

Future Directions. Looking ahead, a key challenge in 21st-century phonetic research is the need for systematic cross-linguistic studies to investigate how languages with different prosodic typologies (e.g., stress-timed vs. syllable-timed, head-prominence vs. edge-prominence) modulate utterance-level phonetic universals. As research continues to accumulate data and refine our understanding, patterns once thought to be idiosyncratic or purely arbitrary may either be confirmed as language-specific exceptions or revealed as systematic processes governed by linguistic principles that have remained unrecognized due to limited data. Expanding the empirical foundation is essential for advancing our understanding of language variation, particularly in how phonetic grammar processes input from yet-unidentified linguistic and cognitive factors that ultimately shape phonetic realization. This pursuit holds the potential to uncover hidden regularities in phonetic patterns, bringing us closer to a comprehensive model of phonetic universals and their language-specific adaptations.

Another critical challenge lies in understanding how language variation within phonetic universals, governed by phonetic grammar, interacts with interspeaker variation within a linguistic community. While this review has only briefly addressed this issue (cf. Sections 2.1 and 2.2), studies suggest that even within a single language, speakers may differ in how they balance system-oriented and output-oriented constraints—a dynamic that may contribute to phonological change. Despite such individual variation, languages often exhibit a remarkable degree of phonetic uniformity. One possible explanation is that modal phonetic values are encoded relationally within the phonetic grammar, allowing structured variation to persist while maintaining language-wide consistency. Yet, research on interspeaker variation in phonetic universals and language variation remains severely limited. Future studies must explore how such variation—particularly when it reflects systematic fine phonetic detail—can be integrated into the framework of phonetic grammar while maintaining community-wide pronunciation norms. Addressing these issues will be essential for achieving a more comprehensive understanding of phonetic systems, further illuminating the intricate interplay between universality, variation, and speaker-specific phonetic tuning.

Declaration of competing interest

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I would like to express my gratitude to the handling editor, Natasha Warner, for her invaluable guidance and encouragement throughout the revision process. I am also appreciative of the reviewers' constructive and insightful feedback, which has greatly enhanced the manuscript. Additionally, I extend my thanks to Sahyang Kim and other members of HIPCS for their thoughtful input and comments. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A2A03084924). This work was also supported by the research fund of Hanyang University (HY-202400000003125).

References

Abdelli-Beruh, N. (2009). Influence of place of articulation on some acoustic correlates of the stop voicing contrast in Parisian French. *Journal of Phonetics, 37*, 66–78.

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics, 71*, 306–325.

Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics, 66*, 120–144.

Barnes, J. A. (2002). *Positional neutralization: A phonologization approach to typological patterns.* Berkeley: Doctoral dissertation).University of California.

Beck, J. M. (1997). Organic variation of the vocal apparatus. *The Handbook of Phonetic Sciences*, 256–297.

Beckman, J. N. (1998). *Positional Faithfulness (Doctoral dissertation)*. University of Massachusetts Amherst..

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11* (1–2), 17–68. https://doi.org/10.1080/016909696387213.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty & B. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 68–86). Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255–309.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*(4), 785–821.

Beddor, P. S. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change. *Journal of Phonetics, 98*, 101228. https://doi.org/10.1016/j.wocn.2023.101228.

Belasco, S. (1953). The influence of articulation of consonants on vowel duration. *Journal of the Acoustical Society of America, 25*, 1015–1016.

Berkovits, R. (1993). Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics, 21*, 479–489. https://doi.org/10.1016/S0095-4470(19)30231-1.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives (Doctoral dissertation)*. University of Amsterdam.

Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in wordinitial clusters of German. *Journal of Phonetics, 41*(6), 546–561.

Braunschweiler, N. (1997). Integrated Cues of Voicing and Vowel Length in German: A Production Study. *Language and Speech, 40*(4), 353–376. https://doi.org/10.1177/002383099704000403.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.
Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 341–376). Cambridge University Press.
Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3–4), 155–180.
Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 175–193). MIT Press.
Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée, 5*, 25–34.
Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica, 57*, 3–16. https://doi.org/10.1159/000028456.
Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 31–59). Mouton de Gruyter.
Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics, 7*(1), 31–53.
Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America, 120*, 1589–1599. https://doi.org/10.1121/1.2217135.
Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary adjacent lengthening. *Journal of Phonetics, 31*, 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2.
Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22*, 129–159.
Chen, W., Whalen, D., & Tiede, M. K. (2021). A dual mechanism for intrinsic f0. *Journal of Phonetics, 87*, 101063. https://doi.org/10.1016/j.wocn.2021.101063.
Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics, 39*, 612–625.
Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics, 32*, 141–176.
Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /A, i/ in English. *Journal of the Acoustical Society of America, 33*, 121–157.
Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory phonology, 8*, 519–548.
Cho, T. (2011). Laboratory phonology. *The continuum companion to phonology*, 343–368.
Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 505–529). Wiley-Blackwell.
Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass, 10*(3), 120–141.
Cho, T. (2022). Linguistic functions of prosody and its phonetic encoding with special reference to Korean. In K. Horie, K. Akita, Y. Kubota, D. Y. Oshima and A. Utsugi (Eds.), *Japanese/Korean Linguistics*, 29, 1–24. CLSI Publications (Stanford University).
Cho, T., & Jun, S.-A. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society, 36*(CLS 36), 31–44.
Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*(2), 193–228.
Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics, 37*(4), 466–485. https://doi.org/10.1016/j.wocn.2009.08.001.
Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190. https://doi.org/10.1006/jpho.2001.0131.
Cho, T., Kim, J., & Kim, S. (2013). Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *Journal of the Acoustical Society of America, 133*(5), EL384–EL390.
Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics, 64*, 71–89. https://doi.org/10.1016/j.wocn.2016.12.003.
Cho, T., Lee, Y., & Kim, S. (2011). Communicatively-driven versus prosodically-driven hyper-articulation in Korean. *Journal of Phonetics, 39*(3), 344–361. https://doi.org/10.1016/j.wocn.2011.02.005.
Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics, 46*, 128–146. https://doi.org/10.1016/j.wocn.2014.06.003.
Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*, 207–229.
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*(2), 121–157.
Cho, T., McQueen, J., & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35*, 210–243.
Cho, T., Son, M., & Kim, S. (2016). Articulatory reflexes of the three-way contrast in labial stops and kinematic evidence for domain-initial strengthening in Korean. *Journal of the International Phonetic Association*. https://doi.org/10.1017/S0025100315000481.
Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics, 72*, 52–65. https://doi.org/10.1016/j.wocn.2018.11.002.
Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47.
Choi, J., Kim, S., & Cho, T. (2016). Phonetic encoding of coda voicing contrast under different focus conditions in L1 vs. L2 English. *Frontiers in Psychology, 7*, 624. https://doi.org/10.3389/fpsyg.2016.00624.
Choi, J., Kim, S., & Cho, T. (2020). An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLOS ONE, 15*(10), e0240682.
Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row. (Reprinted, 1991, Boston: The MIT Press.)
Clements, G. N., & Keyser, S. J. (1983). *CV Phonology: A Generative Theory of the Syllable*. MIT Press.
Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics, 66*, 185–216.
Cohen, A. R., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua, 19*, 177–192.
Cohn, A. C. (1993). Nasalisation in English: Phonology or phonetics. *Phonology, 10*(1), 43–81. https://doi.org/10.1017/S0952675700001731.
Connell, B. (2002). Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics, 30*, 101–129.
Cooper, A. (1991). *Glottal Gestures and Aspiration in English (Doctoral dissertation)*. Yale University.
Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.
Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. Cambridge, MA: The MIT Press.
Davidson, L. (2021). Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of Phonetics, 88*, 101075. https://doi.org/10.1016/j.wocn.2021.101075.
de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 97*(1), 491–504.
de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32*(4), 493–516.
de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*(1), 53–75.
Delattre, P. (1940). La force d'articulation consonantique en français. *The French Review, 14*, 220–232.
Di Cristo, A., & Hirst, D. J. (1986). Modelling French micromelody: Analysis and synthesis. *Phonetica, 43*(1–3), 11–30.
DiCanio, C., Benn, J., & García, R. C. (2021). Disentangling the Effects of Position and UtteranceLevel Declination on the Production of Complex Tones in Yoloxóchitl Mixtec. *Language and Speech, 64*(3), 515–557. https://doi.org/10.1177/0023830920939132.
Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 121–144.
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*(4), 423–444. https://doi.org/10.1006/jpho.1996.0023.
Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49*, 77–95.
Docherty, G. (1992). *The timing of voicing in British English obstruents*. Foris Publications.
Dyhr, N. (1990). The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica, 47*(3–4), 141–154.
Edwards, J. E., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America, 89*, 369–382.
Erickson, D., Honda, K., & Kawahara, S. (2017). Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis. *Acoustical Science and Technology, 38*(3), 137–146.
Esling, J., Moisik, S., Benner, A., & Crevier-Buchman, L. (2019). *Voice quality: The laryngeal articulator model*. Cambridge University Press.
Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics, 2*(4), 327–335.
Farnetani, E., & Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 316–352). Wiley-Blackwell. https://doi.org/10.1002/9781444317251.ch9.
Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Le Maître Phonétique, 32*, 42–59.
Flemming, E. (1995). Auditory representations in phonology (Doctoral dissertation). UCLA.
Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 232–276). Cambridge University Press.
Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd edition, pp. 523–602). Oxford: Wiley-Blackwell.
Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics, 97*, 1–74.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics, 29*, 109–135.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*, 3728–3740.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics, 20*(1), 143–165.

Fowler, C. A. (1995). Acoustic and kinematic correlates of contrastive stress accent in spoken English. In F. Bell-Berti & J. J. Raphael (Eds.), *Producing Speech: Contemporary Issues for Katherine Safford Harris* (pp. 355–373). AIP Publishing.

Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *Journal of the Acoustical Society of America, 120*(4), 2884–2895.

Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics, 41*(1), 29–47.

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U. D., & Koenig, L. L. (2015). Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics, 52*, 35–45. https://doi.org/10.1016/j.wocn.2015.04.002.

Fuchs, S., & Rochet-Capellan, A. (2021). The Respiratory Foundations of Spoken Language. *Annual Review of Linguistics, 7*(1), 13–30. https://doi.org/10.1146/annurev-linguistics-031720-103907.

Gafos, A., & Goldstein, L. (2012). Articulatory representation and phonological organization. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 220–231). Oxford University Press.

Gao, J., & Arai, T. (2019). Plosive (de-) voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics, 77*, 100932.

Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics, 45*, 106–113. https://doi.org/10.1016/j.wocn.2014.04.001.

Garellek, M. (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics, 94*, 101155. https://doi.org/10.1016/j.wocn.2022.101155.

Garellek, M., & Keating, P. A. (2015). Phrase-final creak: Articulation, acoustics, and distribution. *Paper presented at the 89th Annual Meeting of the Linguistic Society of America, Portland*.

Gelfer, C. E., Harris, K. S., & Baer, T. (1987). Controlled variables in sentence intonation. In T. Baer, C. Sasaki, & K. S. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 422–435). College Hill Press.

Georgeton, L., & Fougeron, C. (2014). Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation. *Journal of Phonetics, 44*, 83–95. https://doi.org/10.1016/j.wocn.2014.02.006.

Goldsmith, J. A. (1990). *Autosegmental and Metrical Phonology.* Basil Blackwell.

Goldstein, L. (1992). Comments on chapters 3 and 4. In G. Docherty & D. R. Ladd (Eds.), *Laboratory phonology, Vol. 2: Gesture, segment, prosody*, 120–124. Cambridge: Cambridge University Press.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–208). Mouton de Gruyter.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 215–249). Cambridge University Press.

González, C., Weissglass, C., & Bates, D. (2022). Creaky voice and prosodic boundaries in Spanish: An acoustic study. *Studies in Hispanic and Lusophone Linguistics, 15*(1), 33–65.

Gordon, M. (1998). The phonetics and phonology of non-modal vowels: A crosslinguistic perspective. *Berkeley Linguistics Society, 24*, 93–105.

Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica, 36*(2), 98–114.

Gussenhoven, C. (2016). Analysis of Intonation: The Case of MAE_ToBI. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 7*(1), 10. https://doi.org/10.5334/labphon.30.

Gussenhoven, G., & Rietveld, A. C. M. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics, 20*, 283–303. https://doi.org/10.1016/S0095-4470(19)30636-9.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America, 47*(2B), 613–617.

Halle, M., & Stevens, K. N. (1967). Remarks on Analysis by Synthesis and Distinctive Features. In W. Wathen-Dunn (Ed.), *Models for the Perception of Speech and Visual Form* (pp. 88–102). MIT Press.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly progress report, 101*, 198–213.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America, 125*, 425–441.

Hawkins, S., & Nguyen, N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics, 32*(2), 199–231.

Hayashi, W., Hsu, C.-S., & Keating, P. A. (1999). Domain-initial strengthening in Taiwanese: A follow-up study. *UCLA Working Papers in Phonetics, 97*, 152–156.

Hoemeke, K. A., & Diehl, R. L. (1994). Perception of vowel height: The role of F 1–F distance. *The Journal of the Acoustical Society of America, 96*(2), 661–674.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (Ed.), *Tone: a linguistic survey* (pp. 77–111). Academic Press.

Hombert, J.-M., Ohala, J. J., & Ewan, W. (1979). Phonetic explanation for the development of tones. *Language, 55*, 37–58.

Honda, K. (1983). Relationship between pitch control and vowel articulation. *Haskins Laboratories Status Report on Speech Research, SR, 73*, 269–282.

Honda, K. (2004). Physiological factors causing tonal characteristics of speech: From global to local prosody. In *Speech Prosody 2004*. International Conference, 739–744.

Honda, K., & Fujimura, O. (1991). Intrinsic vowel F0 and Phrase final F0 lowering: Phonological vs. Biological Explanations. In J. Gauffin & B. Hammerberg (Eds.), *Vocal Fold Physiology* (pp. 149–157). Singular Publishing Group.

Hoole, P. (2006). Experimental studies of laryngeal articulation. *Habilitationschrift*. Germany: University of Munich.

Hoole, P., & Honda, K. (2011). *Automaticity vs. feature-enhancement in the control of segmental F0. In Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories*. John Benjamins Publishing Company.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustic characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105–113.

Hwang, S., Kim, S., & Cho, T. (2023). Differential effects of prosodic boundary on glottalization of word-initial vowels in Korean. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp.1474-1478). Guarant International.

Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review, 25*, 83–137.

Hyman, L. M. (2009). Penultimate lengthening in Bantu: Analysis and spread. *UC Berkeley PhonLab Annual Report, 5*, 195–209.

Inkelas, S., & Orgun, C. O. (2003). Turkish stress: A review. *Phonology, 20*(1), 139–161.

Iskarous, K., & Pouplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics, 95*, 101195.

Jang, J. (2023). *Functions of prosodic boundaries in an edge-prominence language: Marking focus and phrasing in Seoul Korean (Doctoral dissertation)*. Santa Barbara: University of California.

Jang, J., & Katsika, A. (2024). Phrase boundaries lacking word prosody: An articulatory investigation of Seoul Korean. *Journal of the Acoustical Society of America, 155*(5), 3521–3536. https://doi.org/10.1121/10.0026081.

Jang, J., Kim, S., & Cho, T. (2018). Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical Society of America, 144*, EL33–EL39. https://doi.org/10.1121/1.5044641.

Javkin, H. R. (1976). The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Berkeley Phonology Laboratory, 1*, 78–92.

Joo, H., Jang, J., Kim, S., Cho, T., & Cutler, A. (2019). Prosodic structural effects on coarticulatory vowel nasalization in Australian English in comparison to American English. In S Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), Proceedings of the 19th International Congress of Phonetic Sciences (pp. 835–839). Australasian Speech Science and Technology Association.

Jun, S.-A. (1996). Influence of microprosody on macroprosody: A case of phrase initial strengthening. *UCLA Working Papers in Phonetics, 92*, 97–116.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15* (2), 189–226.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions. *Speech Sciences, 7*(1), 143–170.

Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520–540). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:Oso/9780199567300.001.0001.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. MIT Press.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76–90.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics, 55*, 149–181.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286–319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in Honor of Peter Ladefoged* (pp. 115–132). Orlando FL: Academic Press.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics, 18*, 321–334.

Keating, P. A. (2006). Phonetic Encoding of Prosodic Structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, Phonetic Processes, and Techniques* (pp. 167–186). Psychology Press, New York and Hove: Macquarie Monographs in Cognitive Science.

Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology 6: Phonetic interpretations* (pp. 145–163). Cambridge University Press.

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics, 101*, 112–156.

Kendall, T., Pharao, N., Stuart-Smith, J., & Vaughn, C. (2023). Advancements of phonetics in the 21st century: Theoretical issues in sociophonetics. *Journal of Phonetics, 98*, 101226. https://doi.org/10.1016/j.wocn.2023.101226.

Kentner, G., Franz, I., Knoop, C. A., & Menninghaus, W. (2023). The final lengthening of pre-boundary syllables turns into final shortening as boundary strength levels increase. *Journal of Phonetics, 97*, 101225.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language, 82*(1), 33–63.

Kim, J. J., Kim, S., & Cho, T. (2024). Preboundary lengthening and articulatory strengthening in Korean as an edge-prominence language. *Laboratory Phonology, 15*(1), 1–36.
Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America, 134*(1), EL19-EL25.
Kim, S., Jang, J., & Cho, T. (2017). Articulatory characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English. *The Journal of the Acoustical Society of America, 142*(4), EL362–EL368.
Kim, S. E., & Tilsen, S. (2024). Planning for the future and reacting to the present: Proactive and reactive F0 adjustments in speech. *Journal of Phonetics, 104*, 101322.
Kingston, J. (2007). Segmental influences on f0: Automatic or controlled? In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes, volume 2: Experimental studies in word and sentence prosody* (pp. 171–201). Berlin: Mouton de Gruyter.
Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.). *Blackwell companion to phonology* (Vol. 4, pp. 2304–2334). Oxford, UK: Blackwell Publishing.
Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*, 419–454.
Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics, 71*, 326–354.
Kirby, J., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel f0: Evidence from "true voicing" languages. *The Journal of the Acoustical Society of America, 140*, 2400–2411.
Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics, 38*(2), 185–196.
Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16* (2), 153–169.
Kohler, K. J. (1990). Macro and micro F0 in the synthesis of intonation. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology Volume 1: Between the Grammar and Physics of Speech* (pp. 115–138). Cambridge University Press.
Kohler, K. J. (1994). Glottal stops and glottalization in German. *Phonetica, 51*(1–3), 38–51. https://doi.org/10.1159/000261957.
Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In *Nasals, nasalization, and the velum* (pp. 87–116). Academic Press.
Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics, 27*(1), 23–54.
Krakow, R. A., Bell-Berti, F., & Wang, Q. E. (1985). Supralaryngeal declination: Evidence from the temporal coordination of tongue gestures. *Phonetica, 42*(3), 163–174. https://doi.org/10.1159/000261726.
Krakow, R. A., Bell-Berti, F., & Wang, Q. E. (1991). Variability in the timing of articulatory movements: Evidence for supra-segmental influences. *Journal of Phonetics, 19*(1), 67–75. https://doi.org/10.1016/S0095-4470(19)30306-6.
Krakow, R. A., Bell-Berti, F., & Wang, Q. E. (1995). Supralaryngeal declination: Evidence from the velum. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues. For Katherine Safford Harris* (pp. 333–353). New York, NY: AIP Press.
Krivokapić, J. (2007). The planning, production and perception of prosodic structure (Doctoral dissertation).University of Southern California.
Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics, 4*(4), 303–320.
Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533.
Kuzla, C., Cho, T., & Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics, 35*, 301–320.
Kuzla, C., & Ernestus, M. (2011). Prosodic conditioning of phonetic detail in German plosives. *Journal of Phonetics, 39*(2), 143–155. https://doi.org/10.1016/j.wocn.2011.01.001.
Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press.
Ladd, D. R., & Schmid, S. (2018). Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics, 71*, 229–248.
Ladefoged, P. (1964). *A phonetic study of West African Languages*. Cambridge University Press.
Ladefoged, P. (1967). *Three areas of experimental phonetics*. Oxford University Press.
Ladefoged, P. (1971). Phonological features and their phonetic correlates. *UCLA Working Papers in Phonetics, 21*, 3–12.
Ladefoged, P., & Cho, T. (2001). Linking linguistic contrasts to reality: The case of VOT. In N. Gronnum & J. Rischel (Eds.), *Travaux Du Cercle Linguistique De Copenhague, vol. XXXI. (To Honour Eli Fischer-Forgensen.)* (pp. 212–223). Copenhagen: C.A. Reitzel.
Ladefoged, P., & Loeb, G. (2002). Preliminary studies on respiratory activity in speech. *UCLA Working Papers in Phonetics, 101*, 50–60.
Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell Publishers.
Laeufer, C. (1996). The acquisition of a complex phonological contrast: Voice timing patterns of English initial stops by native French speakers. *Phonetica, 53*(1–2), 86–110.
Laver, J. (1980). The phonetic description of voice quality. *Cambridge Studies in Linguistics London, 31*, 1–186.
Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.
Lawson, E., & Stuart-Smith, J. (2021). Lenition and fortition of /r/ in utterance-final position: An ultrasound tongue imaging study of lingual gesture timing in spontaneous speech. *Journal of Phonetics, 86*, 101053.
Lee, J., Cho, T., & Kim, S. (2023). Gender-related variation of nasality and sound change of denasalization driven by prosodic boundaries in Seoul Korean. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1395–1399). Guarant International.
Lehiste, I. (1970). *Suprasegmentals*. MIT Press.
Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America, 51*(6B), 2018–2024.
Li, H., Kim, S., & Cho, T. (2020). Prosodic structurally conditioned variation of coarticulatory vowel nasalization in Mandarin Chinese: Its language specificity and cross-linguistic generalizability. *The Journal of the Acoustical Society of America, 148*(3). https://doi.org/10.1121/10.0001743.
Li, H., Kim, S., & Cho, T. (2023). Preboundary lengthening and its kinematic characteristics in Mandarin Chinese in interaction with focus and lexical tone. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2149–2153). Guarant International.
Lieberman, P. (1966). *Intonation, perception, and language (Doctoral dissertation)*. Massachusetts Institute of Technology.
Lieberman, P. (1967). *Intonation, perception, and language*. MIT Research Monograph.
Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.
Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*(11), 1773–1781.
Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory Quarterly Progress Status Report, 4*, 1–29.
Lindblom, B. (1968). Temporal organization of syllable production. *Speech Transmission Laboratory Quarterly Progress Status Report, 9*, 1–5.
Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 13–44). Academic Press.
Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. *Speech Production and Speech Modeling, 55*, 403–439.
Lindblom, B., Guion, S., Hura, S., Moon, S. J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica, 7*, 5–36.
Lisker, L. (1974). On "explaining" vowel duration variation. In *winter meeting of the Linguistic Society of America* (Vol. 28, p. 225).
Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.
Löfqvist, A., Baer, T., McGarr, N., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*, 1314–1321.
MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*(4), 499–546.
MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science, 288*(5465), 527–531.
Maddieson, I. (1985). Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 203–221). Academic Press.
Maddieson, I. (1997). Phonetic universals. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 619–639). Oxford: Blackwell.
Maddieson, I. (1999). In search of universals. In *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 3, pp. 2521–2528).
Maddieson, I., & Gandour, J. (1977). Vowel length before aspirated consonants. *Indian linguistics, 38*(1), 6–11.
Maeda, S. (1976). *A Characterization of American English Intonation (Doctoral dissertation)*. Massachusetts Institute of Technology.
McQueen, J. M., & Dilley, L. C. (2020). Prosody and spoken-word recognition. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 509–521). Oxford University Press.
Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics, 54*, 68–79. https://doi.org/10.1016/j.wocn.2015.09.002.
Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language, 108*(104034), 1–19. https://doi.org/10.1016/j.jml.2019.104034.
Mitterer, H., Kim, S., & Cho, T. (2020). The role of segmental information in syntactic processing through the syntax-prosody interface. *Language and Speech*. https://doi.org/10.1177/0023830920974401.
Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*, 40–55.
Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics, 32*(1), 1–33. https://doi.org/10.1016/S0095-4470(03)00004-4.
Mücke, D., Grice, M., & Cho, T. (2014). More than a magic moment - Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics, 44*, 1–7.
Myers, S. (2007). The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language & Linguistic Theory, 25*, 157–193.
Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics, 37*(1), 29–45.
Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2253–2256).
Napoleão de Souza, R. (2023). Segmental cues to IP-initial boundaries: Acoustic data from English, Spanish, and Portuguese. In F. Schubö, S. Hanne, I. Wartenburger, & S. Zerbian (Eds.), *Prosodic Boundary Phenomena (Studies in Laboratory Phonology Series)* (pp. 35–86). Berlin: Science Language Press.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*(1), 1–19.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological cybernetics, 46*, 135–147.

Nespor, M., & Vogel, I. (1986). Prosodic Phonology. *Foris*. Publications.

Nooteboom, S. (1997). The prosody of speech: Melody and rhythm. *The Handbook of Phonetic Sciences, 5*, 640–673.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association, 31*(1), 139–152.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics, 39*, 59–67.

Ohala, J. J. (1978). The production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). Academic Press.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). Springer-Verlag.

Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication, 13*(1–2), 155–161.

Ohala, J. J. (1997). Aerodynamics of phonology. In *Proceedings of the Seoul International Conference on Linguistics* (Vol. 92, p. 97). Seoul: Linguistic Society of Korea.

Ohala, J. J., & Eukel, B. W. (1976). Explaining the intrinsic pitch of vowels. *The Journal of the Acoustical Society of America, 60*, S44.

Ohala, J. J., & Riordan, C. J. (1979). Passive vocal tract enlargement during voiced stops. *The Journal of the Acoustical Society of America, 65*(S1), S23–S.

Öhman, S. E. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America, 41*(2), 310–320.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Volume 1. Production* (pp. 93–112). Academic Press.

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Psychology Press.

Onaka, A. (2003). Domain-initial strengthening in Japanese: An acoustic and articulatory study. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2091–2094). Barcelona.

Onaka, A., Watson, C., Palethorpe, S., & Harrington, J. (2003). An acoustic analysis of domain-initial strengthening effect in Japanese. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 201–206). Sydney.

Pan, H. H. (2009). Initial strengthening of lexical tones in Taiwanese Min. *Experimental Studies in Word and Sentence Prosody, 12*(2), 271–292.

Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics, 94*, 101179.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693–703.

Peterson, G. E., & Lehiste, I. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*(4), 419–425. https://doi.org/10.1121/1.1908681.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 90–117). Cambridge University Press.

Piroth, H. G., & Janker, P. M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics, 32*, 81–109.

Pompino-Marschall, B., & Żygis, M. (2010). Glottal Marking of Vowel-Initial Words in German. *ZAS Papers in Linguistics, 52*, 1–17.

Pycha, A., & Dahan, D. (2016). Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /aɪ/. *Journal of Phonetics, 56*, 15–37.

Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni & R. Remez (Eds.), *Handbook of Speech Perception* (pp. 182–206). Blackwell.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics, 29*(4), 407–429. https://doi.org/10.1006/jpho.2001.0143.

Riordan, C. J. (1980). Larynx height during English stop consonants. *Journal of Phonetics, 8*(3), 353–360.

Rochet-Capellan, A., & Fuchs, S. (2013). Changes in breathing while listening to read speech: The effect of reader and speech mode. *Frontiers in Psychology, 4*, 906. https://doi.org/10.3389/fpsyg.2013.00906.

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics, 43*, 11–25.

Sanker, C. (2019). Influence of coda stop features on perceived vowel duration. *Journal of Phonetics, 75*, 43–56.

Scholz, F., & Chen, Y. (2014). Sentence planning and f0 scaling in Wenzhou Chinese. *Journal of Phonetics, 47*, 81–91. https://doi.org/10.1016/j.wocn.2014.09.004.

Seifart, F., Paschen, L., & Stave, M. (2022). *Language Documentation Reference Corpus (DoReCo) 1.0*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2). https://doi.org/10.34847/nkl.7cbfq779.

Selkirk, E. O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. MIT Press.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3*, 371–405.

Seo, J., Kim, S., Kubozono, H., & Cho, T. (2019). Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter? *Journal of the Acoustical Society of America, 146*(3), 1817–1823. https://doi.org/10.1121/1.5122191.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.

Shepherd, M. A. (2008). The scope and effects of preboundary prosodic lengthening in Japanese. *USC Working Papers. Linguistics, 4*, 1–14.

Shin, J. E. J., Kim, S., & Cho, T. (2023). Alignment of prosodic and syntactic junctures and vowel-initial glottalization in syntactic disambiguation in English. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1504–1508).

Silpachai, A. (2024). The boundary-induced modulation of obstruents and tones in Thai. *Journal of Phonetics, 102*, 101291.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology, 23*(2), 287–308. https://doi.org/10.1017/S0952675706000911.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36*(1), 91–113.

Solé, M.-J. (2007). Controlled and mechanical properties in speech: A review of the literature. In M.-.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology* (pp. 302–321). Oxford University Press.

Solé, M. J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. *Papers in laboratory phonology, 10*, 607–655.

Son, M., Kim, S., & Cho, T. (2011). Supralaryngeal Articulatory Characteristics of Coronal Consonants /n, t, th, t*/ in Korean. *Phonetics and Speech Sciences, 3*(4), 33–43.

Sonesson, B. (1982). Vocal fold kinesiology. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control* (pp. 113–117). Pergamon Press.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology: Volume 1. Production (pp. 73–92). Academic Press.

Stark, R. E. (1981). Infant vocalization: A comprehensive view. *Infant Mental Health Journal, 2*(2), 118–128.

Stathopoulos, E. T., & Weismer, G. (1983). Closure duration of stop consonants. *Journal of Phonetics, 11*(4), 395–400.

Steffman, J. (2023). Vowel-internal cues to vowel quality and prominence in speech perception. *Phonetica, 80*(5), 329–356.

Steffman, J., Kim, S., Cho, T., & Jun, S. A. (2022). Prosodic phrasing mediates listeners' perception of temporal cues: Evidence from the Korean Accentual Phrase. *Journal of Phonetics, 94*, 101156.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*(1–2), 3–45.

Stevens, K. N. (1997). Articulatory-acoustic-auditory relationships. *The Handbook of Phonetic Sciences*, 462–506.

Storme, B. (2019). Contrast enhancement as motivation for closed syllable laxing and open syllable tensing. *Phonology, 36*(2), 303–340. https://doi.org/10.1017/S0952675719000149.

Straka, G. (1963). *La divison de sons du language en voyelles et consonnes peut-elle être justifiée? Travaux de Ling. et de Littérature, 1*, 17–99.

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America, 113*, 2834–2849.

Tabain, M., & Perrier, P. (2005). Articulation and acoustics of/i/in preboundary position in French. *Journal of Phonetics, 33*(1), 77–100.

Tabain, M., & Perrier, P. (2007). An articulatory and acoustic study of /u/ in preboundary position in French: The interaction of compensatory articulation, neutralization avoidance and featural enhancement. *Journal of Phonetics, 35*(2), 135–161. https://doi.org/10.1016/j.wocn.2006.05.001.

'tHart, J. T., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge University Press.

Tiede, M., Espy-Wilson, C., Goldenberg, D., Mitra, V., Nam, H., & Sivaraman, G. (2017). Quantifying kinematic aspects of reduction in a contrasting rate production task. *Journal of the Acoustical Society of America, 141*(5), 3580. https://doi.org/10.1121/1.4987629.

Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A. (2015). Vowel height and duration. *Proceedings of the 32nd west coast conference on formal linguistics* (Vol. 32, pp. 64–71). Somerville, MA: Cascadilla Proceedings Project.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*, 445–472.

Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics, 39*(2), 168–177.

Vatikiotis-Bateson, E., & Fowler, C. (1988). Kinematic analysis of articulatory declination. *Journal of the Acoustical Society of America, 84*. S128(A).

Vayra, M., & Fowler, C. (1992). Declination of supralaryngeal gestures in spoken Italian. *Phonetica, 49*, 48–60.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61*(2), 397–445.

Wang, C., Xu, Y., & Zhang, J. (2019). Mandarin and English use different temporal means to mark major prosodic boundaries. *Corpus, 2*, 1.

Wang, Y.-T., Green, J. R., Nip, I. S. B., Kent, R. D., & Kent, J. F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. *Folia Phoniatrica et Logopaedica, 62*(6), 297–302. https://doi.org/10.1159/000316976.

Warner, N. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues of spoken word recognition in phonetic research. *Journal of Phonetics, 101*, 101275.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*, 251–276.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics, 34*(2), 285–293. https://doi.org/10.1016/j.wocn.2004.11.003.

Watson, P. J., Ciccia, A. H., & Weismer, G. (2003). The relation of lung volume initiation to selected acoustic properties of speech. *The Journal of the Acoustical Society of America, 113*(5), 2812–2819.

Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. *Journal of Phonetics, 8*(4), 427–438.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.

Westbury, J. R. (1994). On coordinate systems and the representation of articulatory movements. *The Journal of the Acoustical Society of America, 95*(4), 2271–2273.

Westbury, J. R., & Keating, P. A. (1980). Central representation of vowel duration. *The Journal of the Acoustical Society of America, 67*(S1), S37–S.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics, 23*(3), 349–366.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1998). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics, 27*(2), 125–142.

Whalen, D. H., Giulivi, S., Goldstein, L. M., Nam, H., & Levitt, A. G. (2011). Response to MacNeilage and Davis and to Oller. *Language Learning and Development, 7*(3), 243–249. https://doi.org/10.1080/15475441.2011.578547.

White, L., Benavides-Varela, S., & Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics, 81*, 100982. https://doi.org/10.1016/j.wocn.2020.100982.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91*, 1707–1717.

Wilson, C., & Chodroff, E. (2017). Uniformity of inherent vowel duration across speakers of American English. *The Journal of the Acoustical Society of America, 142*(4_Supplement), 2580. https://doi.org/10.1121/1.5014434.

Xu, Y., & Xu, A. (2021). Consonantal F0 perturbation in American English involves multiple mechanisms. *The Journal of the Acoustical Society of America, 149*(4), 2877. https://doi.org/10.1121/10.0004239.

Yoo, K., & Nolan, F. (2020). Sampling the progression of domain-initial denasalization in Seoul Korean. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), 22. https://doi.org/10.5334/labphon.203.

Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics, 5*(1), 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815.

Zellou, G. (2022). *Coarticulation in Phonology*. Cambridge University Press. https://doi.org/10.1017/9781009082488.

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 157–190). Cambridge University Press.