Journal of Phonetics 114 (2026) 101470

Contents lists available at ScienceDirect

Journal of Phonetics

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

Linguistic and cognitive functions of fine phonetic detail underlying sound systems and sound change

Taehong Cho [a,b,*], Sahyang Kim [c], Holger Mitterer [a,d]

[a] Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS), Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, the Republic of Korea
[b] Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, the Republic of Korea
[c] Department of English Education, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul 04066, the Republic of Korea
[d] Department of Cognitive Science, University of Malta, Msida MSD 2080, Malta

ARTICLE INFO

Article history:
Received 4 November 2025
Received in revised form 28 November 2025
Accepted 2 December 2025
Available online xxxx

Keywords:
Fine phonetic detail
Phonetic grammar
Speaker control
Prosody
Sound change

ABSTRACT

This special issue examines how fine phonetic detail participates in the shaping of sound systems. Across fourteen studies, the central theme is that subtle temporal, spectral, and articulatory patterns are not incidental by-products of articulation, but are systematically regulated aspects of speakers' phonetic knowledge. They provide the means through which phonological contrasts and prosodic structure are realized, maintained, and sometimes reorganized. The contributions show how languages allocate continuous phonetic parameters—such as timing, coordination, voice quality, and nasality—within prosodic domains (e.g., phrases, words, and syllables) and under general biomechanical and communicative pressures. Studies of Irish, Hawaiian, Japanese, and Mandarin illustrate how prosodic structure guides segmental and suprasegmental realization. Work on English, German, Danish, and Cantonese demonstrates how fine phonetic detail underlies patterns of variation and creates potential pathways for change. Production connects naturally to perception and learning: findings from English accent adaptation and Samoan iterated learning reveal how listeners stabilize or reinterpret detail, linking individual processing to community-level patterning. A set of studies on Italian, Korean, English, and L2 German show how prominence reorganizes cues across articulation, interaction, and acquisition, shaping how speakers signal and listeners recover linguistic structure. These studies converge on a view in which fine phonetic detail arises from a central phonetic component (or the phonetic grammar) of linguistic structure—controlled by speakers, shaped by universal motor and perceptual constraints, and continually adjusted through perception and learning. In this perspective, sound systems emerge from the interplay of these regulated patterns, which sustain contrasts, support communication, and open principled routes for change.

1. Introduction

One enduring question in linguistics concerns how abstract categories are realized in actual speech. Mid-20th-century approaches—epitomized by *The Sound Pattern of English* (Chomsky & Halle, 1968)—largely treated phonetics as a downstream epiphenomenon of the articulatory apparatus. A different picture began to take shape with the emergence of Laboratory Phonology, particularly in the first edited volume, *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Beckman & Kingston, 1990), which marked a shift toward treating phonetic detail as structured evidence for phonological organization and toward bridging grammatical theory with experimental phonetics. This shift was further articulated by Keating's (1985, 1990) proposal of a phonetic grammar: a component of linguistic knowledge that systematically shapes phonological representations into community-specific pronunciations by regulating fine-grained detail. This reframing positioned phonetics within the core architecture of grammar, as is also implied by the term *Phonetic Knowledge* (Kingston & Diehl, 1994). Subsequent work by Cho and Ladefoged (1999) emphasized *phonetic arbitrariness*, namely that languages select modal realizations for the "same" feature in ways not fully determined by universals—for

* Corresponding author at: Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS) and Department of English Language and Literature, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, the Republic of Korea.
E-mail address: tcho@hanyang.ac.kr (T. Cho).

https://doi.org/10.1016/j.wocn.2025.101470
0095-4470/

instance, adopting longer or shorter modal VOT for aspirated stops. Such patterns reinforce the idea that phonetic implementation is learned, conventionalized, and language-specific (Cho, Whalen, & Docherty, 2019).

Building on this reconceptualization, recent decades have seen phonetics and phonology converge toward a unified view of sound systems in which phonetic implementation is an integral, regulated component of linguistic structure—governing the deployment of fine detail—rather than arising from low-level automatic process. Multiple strands of research indeed demonstrate that subtle temporal, spectral, and phonatory adjustments are not random noise but the very mechanisms through which phonological contrast and prosodic organization are realized, stabilized, and, crucially, transformed over time. These include Articulatory Phonology and speech dynamics (e.g., Iskarous & Pouplier, 2022, and references therein), usage-based or exemplar models (e.g., Goldrick & Cole, 2023, and references therein), and phonetically driven phonological models (e.g., Hayes, Kirchner, & Steriade, 2004; see also Pater, 2009, and Flemming, 2021, on weighted-constraint Harmonic Grammar and its probabilistic extension, the Maximum Entropy/MaxEnt model). Seemingly non-contrastive variation—coarticulation, gestural overlap, micro-timing, energy distribution, pitch, and voice-quality nuance—systematically aligns with phonological, prosodic and other higher-order structures in learnable, language-specific ways and is actively deployed for communicative ends.

The challenge, then, is not to divide the categorical from the continuous but to explain how abstract structure and communicative stability *emerge from* and *evolve through* continuous control of phonetic realization operating across multiple timescales—from millisecond-level articulatory coupling to generational sound change. Within this perspective, the system-internal grammar governing phonetic implementation—the *phonetic grammar* (Keating, 1984, 1985; Cho & Ladefoged, 1999; see Cho, 2025 for *the extended model of phonetic grammar*)—functions as a dynamic control architecture that allocates and coordinates articulatory and acoustic resources in lawful, learned, and language-specific ways.

This reconceptualization motivates the theme of this special issue, *"Linguistic and cognitive functions of fine phonetic detail underlying sound systems and sound change."* The papers assembled here approach this theme from complementary perspectives, showing that fine phonetic detail functions as one of its organizing principles—linking phonological contrast, prosodic structure, communicative function, and cognitive/motor control. Collectively, they show that the phonetic grammar—understood here broadly, not as a specific model but as a family of mechanisms that govern fine phonetic detail in concert with prosodic, linguistic, and communicative factors—modulates phonetic implementation and serves as the interface through which sound systems maintain stability while remaining adaptable to change. The following section, *the role of fine phonetic detail in the architecture of structure and change*, develops this theme further through the fourteen studies presented in this special issue, which span *prosodic scaffolding*, articulatory regulation, synchronic variation, perceptual learning, second-language acquisition, and the realization of prominence, with several also addressing potential pathways of diachronic reorganization. Note that the term ***prosodic scaffolding*** is conceptualized as *a persistent structural framework within which a dynamical control architecture operates*, supporting the stabilization of phonological contrasts while remaining adaptable to phonetic variation. Thus, while *prosodic structure* refers to the hierarchical organization itself, *prosodic scaffolding* emphasizes the active, supportive role that this structure plays in constraining and guiding the dynamic allocation of phonetic resources. Taken together, these contributions demonstrate that fine phonetic detail provides the medium through which linguistic structure is constructed, transmitted, and continually reshaped across both individual and generational timescales.

2. Role of fine phonetic detail in the architecture of structure and change

Building on the premise that fine phonetic detail is integral to linguistic organization, what follows specifies how the control architecture operates: which prosodic domains license particular phonetic cues, how timing and laryngeal settings interact with positional structure, how listeners and learners recast these patterns, and how prominence reorganizes them in L1 use and L2 acquisition.

The subsections proceed in turn. *Section 2.1 (Prosodic scaffolding and phonetic regulation)* shows that prosodic organization at the syllable, word, and phrase levels orchestrates fine phonetic detail across Irish, Hawaiian, Japanese, and Mandarin—allocating articulatory and acoustic resources to privileged positions, matching enhancement to mechanical limits, and scaling cues with information load, so that prosodic structure functions as an active scaffold for phonetic implementation within general constraints of effort and efficiency. *Section 2.2 (From control to change)*, drawing on English, German, Danish, and Cantonese, demonstrates that the same control architecture regulating timing, coordination, and laryngeal settings also delimits the range of variation and potential reorganization: local mechanisms that sustain contrast can, over time, bias pathways of systemic and diachronic change. *Section 2.3 (Perceptual adaptation and phonetic biases in the evolution of sound systems)* turns to perception and learning, showing—via adaptation experiments in English and probabilistic modeling in Samoan—how perceptual plasticity and bias-driven learning can stabilize, reinterpret, or amplify fine phonetic variation, linking individual processing to speech-community patterning and historical reanalysis. *Section 2.4 (Implementing prominence through phonetic detail)* closes the sequence by examining prominence as a multidimensional interface between prosody and phonetic realization: evidence from Italian, Korean, English, and L2 German shows that prominence retimes, aligns, and redistributes cues across production, interaction, and acquisition, integrating articulatory, perceptual, and learning systems within a unified adaptive framework.

2.1. Prosodic scaffolding and phonetic regulation: toward a cross-linguistic economy of fine phonetic detail

This section opens with a shared question: How does prosodic structure across levels—syllable, word, and phrase—

shape the distribution of fine phonetic detail, and what linguistic or cognitive work do those details perform?

Before proceeding, however, it is useful to clarify *what "prosody" denotes* in this context. Within the present framework, prosody functions as a structured, hierarchical frame that organizes the deployment of fine phonetic detail (e.g., Beckman, 1996; Shattuck-Hufnagel & Turk, 1996). It provides *the articulatory and temporal scaffold* for speech, defining the positions and domains in which phonetic cues are licensed, enhanced, or suppressed (e.g., Keating & Shattuck-Hufnagel, 2002; Keating, 2006; Cho, 2016). In this role, prosodic structure feeds the phonetic grammar with an information-rich specification that encodes not only suprasegmental properties—intonation, rhythm, and prominence—but also positional conditions on segmental realization. The phonetics–prosody interface captures this process: abstract phonological representations are systematically shaped as a function of a language's prosodic architecture, with prosodic domains fine-tuning the timing, magnitude, and coordination of articulatory gestures. This "phonetic encoding of prosodic structure" (e.g., Keating, 2006; Cho, 2022) yields lawful, language-specific patterns of segmental and suprasegmental variation, reflecting both higher-order linguistic organization and universal utterance-level pressures such as respiratory and articulatory resets (e.g., Cho, 2025). In this sense, prosodic structure is not a passive hierarchy but a structured input to the phonetic grammar—an input through which the grammar implements these specifications to determine the final, fine-grained phonetic realization of the utterance.

The four contributions presented here, across Irish, Hawaiian, Japanese, and Mandarin, converge on a common view: prosodic organization provides the scaffold within which the phonetic grammar—here, the system-internal production control—dynamically regulates articulation. These studies invite us to see prosodic structure not as a static hierarchy but as an active scaffold for phonetic regulation, yielding balanced, efficiency-seeking allocations of fine phonetic detail under shared constraints. In our synthesis, three recurrent tendencies emerge: an allocation tendency, whereby prosodically prominent positions concentrate articulatory and acoustic resources where they most aid phonological and lexical processing; a constraint-matching tendency, whereby phonetic enhancement is limited by the biomechanical and aerodynamic constraints of the articulatory system; and an information-sensitivity tendency, whereby cue magnitude scales with the expected communicative load of a position.

Given this overall picture, each contribution highlights a different facet of how prosodic structure orchestrates fine phonetic detail: Bennett, Padgett, Ní Chiosáin, McGuire, and Bellik (2024) probe prosodic-structurally driven spatial constraints on secondary dorsals in Irish; Davidson and Parker Jones (2024) track position-dependent word-internal cue enhancement in Hawaiian stops; Sano and Guillemot (2025) attempt to understand information-sensitive scaling of Japanese vowel length across phrase-level prosodic contexts; and Huang (2025) teases apart tone–phrasing interactions in Mandarin voice quality. Together, these studies show that prosodic structure provides the scaffold and language-specific phonetic control regulates implementation within it; fine detail is the medium by which prosodic structure anchors contrast, signals structure, and manages clarity–effort trade-offs in an ecologically adequate way.

As a starting case, Irish shows how syllable position and place of articulation shape the separability of palatalization and velarization. Bennett et al. (2024) ask why these secondary dorsal contrasts—palatalization ([C^j], tongue body fronted) and velarization ([C^ɣ], tongue body retracted)—are typologically rare in coda position and on labials. Despite being phonemic, for example, these contrasts frequently disappear in codas (Bulgarian, Nenets) and on bilabials (Czech). The authors test whether these distributional asymmetries follow from articulatory constraints that limit how far secondary gestures can be kept apart in certain contexts.

The study hypothesizes that some positional (coda) and segmental (labial) environments compress the articulatory space for secondary gestures. As for the coda context, consonantal gestures overlap more with the preceding vowel, accommodating vowel-to-consonant coarticulation and narrowing the tongue-body range: After a front vowel like /iː/ a velarized coda [C^ɣ] is drawn forward toward [C^j], reducing distinctness. As for the labial context, compared to lingual consonant contexts, the tongue body is only weakly coupled to the lip-based primary constriction, resulting in greater variability and less spatial precision—conditions that make front–back contrasts harder to sustain stably. These articulatory factors, the authors argue, underlie typological gaps and perceptual fragility observed in secondary dorsal contrasts.

Using ultrasound tongue imaging, the authors analyze midsagittal tongue shapes in Irish for matched pairs of palatalized and velarized stops across place (labial, coronal, dorsal), position (onset vs. coda), and vowel context (/iː, uː, ɔː/). The results reveal a consistent pattern. First, secondary dorsal contrasts are markedly weaker in codas than onsets, with the greatest reduction for dorsals and labials; coronals remain relatively stable. Second, labials show the largest articulatory variability, reflecting weaker lingual anchor. Third, vowel-to-consonant coarticulation is greater in codas: after /iː/, coda [C^ɣ] fronts sharply, compressing the [C^ɣ]–[C^j] distance. In short, in precisely those contexts where secondary dorsal gestures show reduced spatial separation and increased susceptibility to contextual influence—namely codas and labials—the corresponding tongue-body targets for the secondary contrast are the hardest to keep apart across contexts.

These findings ground typological asymmetries in articulation: when the palatalization–velarization space is mechanically compressed, neutralization becomes likely. The data support a perception–articulation feedback account in which prosodic position and segmental makeup jointly govern the realizability of secondary gestures. The authors suggest that apparent coda/labial "weakness" reflects dynamic differences in gestural stiffness and timing, not categorical status: coda gestures are less stiff (Byrd, 1994), align with vowel–consonant transitions, and invite vowel-driven coarticulation; labials, lacking a lingual constriction, show parallel susceptibility. As such, the authors suggest that these patterns fit Articulatory Phonology and task-dynamic models (Browman & Goldstein, 1992; Goldstein, Byrd, & Saltzman, 2006; Iskarous & Pouplier, 2022) and position-specific coupling relations (Krakow, 1989, 1999). A corollary of Bennett et al.'s study is that motor-timing constraints make prosodic structure (onset vs. coda) a control regime for gestural resources, predicting prosodic sites of enhancement, heightened neutralization risk, and likely diachronic trajectories—thus explaining typological gaps.

Whereas the Bennett et al. study locates the limits of contrast realization in the articulatory domain, Davidson and Parker Jones (2024) extend the inquiry to how prosodic structure redistributes phonetic detail within those articulatory possibilities in Hawaiian, asking how word-level prosodic structure shapes the fine phonetics of Hawaiian oral stops and what functional work those details might do. Analyzing spontaneous speech from the Ka Leo Hawai'i radio archive (1970–80s), they measured VOT, closure duration, and lenition for /p, k/, annotating each token for prosodic-word position, lexical-word position, and stress. Prosodic words were defined following Parker Jones's stress/metrical analysis in such a way that a lexical word may contain multiple prosodic words.

Two main results stand out. First, Hawaiian /p, k/ are consistently unaspirated, falling within cross-linguistic unaspirated VOT ranges. Second, domain-initial strengthening (Fougeron & Keating, 1997; Cho & Keating, 2001, 2009) is quite robust at the word level. That is, prosodic-word-initial stops show longer VOT and closure and reduced lenition relative to prosodic-word-medial stops; lexical-word-initial position also lengthens VOT relative to medial position, but only when the syllable lacks primary stress, showing a primary-stress "ceiling" on VOT. Within (lexical) word-medial position, primary and secondary stress lengthen VOT relative to unstressed syllables. These patterns show that, in Hawaiian, even unaspirated stops are systematically modulated by word-internal prosody and prominence—unlike in languages where such stops are largely prominence-insensitive (e.g., Spanish, Blackfoot, Plains Cree, Sierra Norte de Puebla Nahuatl). Cross-linguistically, however, prominence does not uniformly lengthen VOT (e.g., Dutch can shorten it; Cho & McQueen, 2005), indicating language-specific implementations of strengthening that likely reflect differences in respiratory drive and laryngeal/articulatory force at domain edges.

Functionally, these fine-grained, subphonemic patterns appear to hyperarticulate word boundaries where other cues are sparse (cf. Fougeron & Keating, 1997). The authors suggest that prosodic-word-initial strengthening and stress-conditioned VOT may aid lexical segmentation in a system with a small consonant inventory and limited phonotactic boundary cues. On this view, prosodic organization redistributes subphonemic detail to support contrast as well as lexical parsing at privileged positions (e.g., word-initial). As such, beyond documenting rare facts about Hawaiian stop phonetics, Davidson and Parker Jones' study advances a principled link between word-level prosody and cue deployment, although perceptual consequences for lexical processing remain to be tested (see Cho, McQueen, and Cox (2007) for English; Warner (2023) for related discussion). A further question also remains about how these word-internal effects interact with higher-level influences (e.g., phrase-initial strengthening, sentence-level prominence) in continuous speech, since the present data could not fully isolate strictly word-level materials owing to both theoretical and practical constraints.

The two preceding studies thus show that prosodic structure both constrains and redistributes segmental phonetic resources at the word level. Drawing on the Corpus of Spontaneous Japanese (CSJ)-Relational Database (2012), Sano and Guillemot (2025) extend this account to the phrase level, testing whether the same redistributive control also governs categorical timing contrasts. Crucially, treating position and domain size as information-bearing variables, they test whether prosodic organization scales the short–long vowel contrast where recognition benefits most and tempers it where it does not.

Two robust patterns emerge. First, the short–long contrast in Japanese is enhanced word-initially and in shorter (vs. longer) words, but reduced word-finally. In other words, the duration gap between short and long vowels is largest where segments are likely to bear the greatest informational load for lexical access (initial position) and where each segment contributes a larger share of a word's information (short words). Second, at the IP level, both short and long vowels show phrase-final lengthening. However, long vowels lengthen by a smaller amount than short vowels, helping keep the short–long contrast stable at the boundary (i.e., over-exaggerating the length contrast would compromise its stability). As the authors explain, because phrase-final lengthening can introduce timing variability, limiting the expansion of already long vowels is likely a speaker-controlled way to keep durations within language-specific timing constraints. In addition, the fact that vowels are also longer in shorter IPs than in longer ones is attributed to respiratory capacity: shorter phrases allow more temporal expansion per segment, consistent with the Respiratory Code (e.g., Gussenhoven, 2016).

An important theoretical implication follows: prosodic structure shapes temporal realization in an information-sensitive way. Japanese speakers regulate duration by balancing communicative efficiency against phonetic cost—that is, making contrasts clear where recognition benefits most while avoiding unnecessary expansion where it buys little. Although the effects are modest, they are systematic and theoretically interpretable. The authors call for further experimental and cross-linguistic work to test listener sensitivity to these cues and to determine whether similar information-driven timing patterns emerge in other quantity systems.

Across the three studies discussed thus far, prosodic structure emerges as a regulator of fine phonetic detail—from articulatory spacing in Irish and cue enhancement in Hawaiian to information-sensitive timing in Japanese. Building on these patterns, the next contribution by Huang (2025) examines whether prosodic organization likewise governs voice quality. Focusing on Mandarin, Huang asks how tone and utterance-level phrasing jointly shape the distribution of two distinct creaky subtypes—period doubling (alternating glottal cycles, not inherently low-F0) and vocal fry (low, quasi-regular F0 with strong glottal constriction). Using a scripted read-speech corpus with simultaneous audio and EGG from 20 speakers, the study identifies each subtype by its acoustic/EGG signature and codes tokens for lexical tone (T1–T4) and phrasal position (utterance-initial, utterance-medial pre-/post-focal, utterance-final).

Two prominent results can be reiterated. First, the subtypes of voice quality differ sharply in tonal distribution. Period doubling is far more common overall than vocal fry and occurs with all tones. By contrast, vocal fry is rare in a high-tone context (i.e., Tone 1) and concentrates in low/compressed-F0 contexts (especially Tones 2–3), showing its stronger dependence on increased glottal constriction and low F0. Second, the subtypes display distinct prosodic distributions. Period doubling is frequent at both utterance edges and increases toward the utterance-final position, reflecting edge-related phrase-final phonatory instability that nonetheless provides a boundary cue. Vocal fry is predominantly at the right edge (utterance-final; seldom utterance-initial) and, in *prosodically weak* utterance-medial position, clusters in pre-focal and post-focal positions—i.e., outside the focal domain.

Crucially, these distributions reveal that phrasal structure can override lexical tone when the two exert competing pressures on phonation. In this corpus, utterance-final tokens are often Tone 1 (a high tone that would otherwise disfavor creak), whereas utterance-initial tokens are frequently Tone 3 (a low tone that would otherwise favor creak). Yet both subtypes occur more often utterance-finally than utterance-initially regardless of the lexical tones, indicating that utterance-level organization—and the degree of finality at boundaries—can be the dominant determinant of creaky realization. Crucially, within that overarching prosodic scaffold, the two subtypes appear to serve different subphonemic roles: period doubling as a boundary-marking cue with weaker tonal dependence; vocal fry as a low-pitch phonetic enhancer aligned with low/compressed-F0 tones and pre-/post-focal material.

The results thus motivate a fine-grained typology of creaky phonation in Mandarin in which tone and phrasing jointly regulate the occurrence of period doubling and vocal fry. More broadly aligning with the preceding contributions in this section, voice quality is placed within the same prosodic economy that governs other subphonemic cues, where boundary signaling, focus structure, and tonal realization interact to deploy phonation types where they best support interpretation. Thus, Huang's study shows that fine-grained analysis of voice quality reveals creaky subtypes—often treated as mere variants—perform distinct linguistic roles that systematically contribute to the grammar.

2.1.1. Summary and discussion

Viewed together, these patterns suggest a "prosodic economy": languages allocate articulatory and acoustic resources where communicative return is highest and retract them where mechanics or payoff are unfavorable. The cross-linguistic evidence lines up accordingly—Irish (Bennett et al., 2024) shows spatial separability of $[C^j]$–$[C^\gamma]$ tracking positional constraints; Hawaiian (Davidson & Parker Jones, 2024) channels VOT/closure/lenition to word-initial positions to support lexical segmentation; Japanese (Sano & Guillemot, 2025) scales duration with informational load while tempering phrase-final expansion to preserve the short–long (vowel quantity) contrast; and Mandarin (Huang, 2025) differentiates creaky subtypes, so that boundary signaling and low-F0 enhancement do not conflict. This economy—the efficient deployment of phonetic detail across prosodic positions—is consistent with positional faithfulness and licensing by cue (Beckman, 1997, 1998; Steriade, 1999) without presupposing them: contrasts are maintained and enriched where cues and mechanics jointly support them in a given prosodic configuration, and down-weighted where they do not.

On this view, apparent "idiosyncrasies" are language-specific implementations of a common allocation problem under mechanical and informational constraints. As noted in the introduction, this accords with "variation within universals": while the constraints and organizing principles governing phonetic realization are shared across languages, the solutions they yield are language-specific, implemented through the system-internal grammar of phonetic implementation—*the phonetic grammar* (Keating, 1984, 1985; Cho & Ladefoged, 1999), which operates in continuous reference to prosodic structure (the *extended model of phonetic grammar,* Cho, 2025). Prosodic domains provide the scaffold through which that grammar distributes articulatory and acoustic effort, determining which positions are prioritized, how contrasts are maintained, and where reduction or redistribution occurs (see Cho, 2025, for the extended model). The four studies in this section therefore support a view in which cross-linguistic variability arises from systematic adaptations within a universal prosodic–phonetic framework—a grammar that negotiates mechanical limits, perceptual goals, and communicative efficiency within the organizing geometry of prosody.

2.2. From control to change: how fine phonetic detail shapes sound systems

Building on Section 2.1—where prosodic structure provided the scaffold and the phonetic grammar regulated fine detail across languages—this section turns from control to its consequences for the sound system. If prosodic organization allocates and shapes phonetic resources in real time, what follows for within-language variation and for sound change? The four contributions in this section converge on a common claim: micro-scale timing, coordination, and laryngeal/supralaryngeal settings are system-internal targets that speakers control in production and that listeners monitor in real time—patterns that, when accumulated across speakers, can also shape community-level restructuring.

At the same time, they are complementary in scope. At the representational level, Strycharczuk, Kirkham, Gorman, and Nagamine (2024) show that Northern Anglo-English long vowels use a two-target architecture: nucleus–offglide spacing regulates the monophthong–diphthong continuum, preserving categories while enabling shift. Pouplier et al. (2024) find that anticipatory nasality is language- and position-specific, shaped by boundaries and speaker variation—evidence for system-internal settings rather than mechanical lag. Puggaard-Rode (2024) links Jutland Danish cue orientation (voicing vs. aspiration in strong onsets) to the direction of weak-position gradation, showing how community cue weights bias change. Yu, Li, and Mok (2025) track a suspended Cantonese tonal contrast across generations, where subtle F0 asymmetries persist, erode, or reorganize under contextual and paradigmatic pressures—mapping fine bias to merger paths. Together, micro-adjustments in spacing and timing constitute the means by which the phonetic grammar sustains contrasts, governs variability, and biases change—an economy of regulation (Section 2.1) calibrated by position and practice and propagated over time.

Specially, as the first contribution in this section, Strycharczuk et al. (2024) ask whether Northern Anglo-English long vowels are best modeled as discrete monophthongs and diphthongs, or as a graded two-target system in which small dynamical parameter shifts yield surface diphthongization. Although the paper focuses on the synchronic control of long vowels, the architecture it tests has broader implications—also noted by the authors—for diachronic pathways. A unified two-target system offers a mechanism through which small adjustments in control parameters can gradually accumulate, yielding the kinds of monophthongization and diphthongization shifts historically attested in English. They combine articulatory and acoustic data from six Northern Anglo-English speakers to examine whether long vowels cluster as monophthongs, diphthongs, or intermediate types. Using unsupervised clustering to assess emergent groupings and a modified task-dynamic model (after Sørensen & Gafos, 2016) to test whether a unified two-target architecture can generate the observed patterns, they evaluate how small changes in control parameters produce the continuum of surface realizations.

Two empirical points underpin the account. First, canonical diphthongs (e.g., *buy, boy*) traverse clearly separated targets and show larger displacements than canonical monophthongs (e.g., *bar, burr* in this non-rhotic variety), which cluster around a single articulatory locus. Second, the distinction is not dichotomous: across speakers and measures, intermediate vowels (*bee, boo, beer, beau*) show partial target separation and mixed velocity profiles, indicating the long vowels of English do not fall into two clearly distinct categories (monophthongs vs. diphthongs), but rather exist on a continuum. Clustering typically yields two or three groups—monophthongs, diphthongs, and an intermediate cluster with limited, speaker-dependent diphthongization.

The Task Dynamic (TD) Model simulations are used to test the predictions of the Articulatory Phonology and TD model, a formal mechanism of the grammar (e.g., Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Goldstein et al., 2006; Iskarous & Pouplier, 2022). Modeling generalizes from these data: a uniform two-target control architecture—both diphthongs and long monophthongs are modeled with two vowel gestures coupled anti-phase; the *difference* between them comes from the spatial specifications of the two targets (identical articulatory targets for long monophthongs vs. different targets for diphthongs), not from a contrast between anti-phase and in-phase coupling or small adjustments in gestural timing. The authors explicitly note that treating diphthongs as in-phase (e.g., Marin, 2007) would make them phonologically light, which is incorrect for English off-gliding diphthongs, and therefore adopt anti-phase coordination for both classes in their preferred account.

The theoretical upshot is a compositional representation in which all long vowels share two potential targets; whether they surface as monophthongs or diphthongs depends on *fine control settings*, not on different underlying types. On this view, “monophthong” vs. “diphthong” are gradient phonetic descriptors, while phonological weight is unified (long vowels pattern as heavy, paralleling coda consonants in Articulatory Phonology). The same architecture accounts for diachronic shifts between monophthongization and diphthongization. The authors argue that if monophthongs and diphthongs had different numbers of gestures, change would require inserting or deleting a target—a major restructuring—yet in English such shifts are common and often leave phonological patterning intact. This motivates a single architecture in which minor parameter tweaks (e.g., control settings in target location/degree) yield swings between more monophthongal and more diphthongal realizations without reconfiguring the plan. Overall, small, learnable adjustments to shared gestural targets regulate synchronic variation within language-specific bounds—exactly the kind of fine phonetic detail this section treats as governed by the phonetic grammar or controlled—and provide a ready pathway for diachronic change within a unified two-target, anti-phase AP/TD framework, moving the system along the monophthong–diphthong continuum while preserving category weight.

Building on a unified control architecture in the speech production system, the next study shifts from representational settings to coarticulatory regimes. Pouplier et al. (2024) test whether the temporal reach and magnitude of anticipatory nasal coarticulation are language-specific (English, French, German) and position-dependent, and how these patterns interact with individual variability. Thirty speakers each of American English and German, and twenty-seven of French, produced minimal pairs differing by a nasal vs. oral consonant in initial (e.g., *mat–pat*) or non-initial positions (e.g., *rhymer–riper; ran–rat*). A nasometer tracked oral–nasal acoustic energy, and the onset of anticipation was defined as the time point where nasal and oral intensity curves diverged statistically.

Two positional profiles emerge. In word-initial position (with a preceding prosodic boundary), the temporal extent of anticipation converges across languages, suggesting that boundaries attenuate language-specific spreading. In non-initial position, robust differences surface: English shows earlier, wider anticipation than French and German, which pattern together. Strikingly, across all three languages the onset of nasality often precedes the pre-nasal vowel, indicating coarticulation beyond the nearest vowel nucleus and operating on a separate articulatory tier, with velum lowering mediated via the nasal channel. The available anticipatory window predicts earlier divergence cross-linguistically (a longer window invites earlier nasalization), contradicting purely mechanical accounts that would otherwise predict time-locked, fixed-lag velum opening.

English also shows larger nasal-intensity differences, consistent with greater velic opening. Variability patterns also track grammatical specification: German exhibits higher within- and between-speaker dispersion than English and French, aligning with the absence of phonological nasality specifications that would otherwise channel production. Theoretically, these results challenge a simple contrast-suppression view. On one hand, despite a phonological nasal–oral vowel contrast, French still displays substantial anticipatory nasalization (though less than English; but see Cohn, 1990, 1993, for evidence of more stringent anticipatory patterns in French). On the other hand, although neither German nor English contrasts vowel nasality, they differ markedly in domain and degree; English, as the authors suggest, is compatible with phonologized contextual nasalization, where the pre-nasal context serves as a source rather than a passive recipient of coarticulation. The widespread early onset of nasality also challenges models that preclude long-distance coarticulation, pointing instead to broader planning domains—plausibly supported by the relative independence of velic gestures from lingual gestures.

Pouplier et al. situate these results in the view that languages can differ in the numeric phonetic range to which a given phonological category is mapped, with that range regulated by individual speaker grammars (Beddor, 2009; Beddor, Coetzee, Styler, McGowan, & Boland, 2018; Zellou, 2017). In other words, although the study examines synchronic coarticulatory regimes, its findings have implications for diachronic pathways: language-specific differences in the timing and magnitude of anticipatory nasalization constitute precisely the kinds of structured phonetic biases that can accumulate across speakers and generations, shaping well-attested trajectories of nasal vowel development and contextual nasalization as they differ across languages. This perspective accords with the phonetic grammar framework, and supports a system-specific view: anticipatory coarticulation is neither universal nor purely mechanical, but reflects how each grammar recruits nasality as a cue—shaping trajectories of sound change and bounding within-community variability, and hence cross-linguistic variation. At the same time, as discussed above, these language-specific implementations rest on shared human constraints—velopharyngeal biomechanics, aerodynamics, and perceptual demands—which constitute phonetic universals and provide the substrate on which grammars calibrate vowel nasalization.

While Pouplier et al. (2024) examine controlled cross-linguistic variation within phonetic universals, Puggaard-Rode (2024) turns to within-language variation, testing whether cue orientation across Jutland Danish—aspiration vs. voicing in strong onsets—predicts weak-position realizations. The study asks whether fine laryngeal implementation in strong positions (onsets before full vowels) versus weak positions (onsets before neutral vowels and all codas) covaries with regional outcomes of stop gradation, thus linking micro-phonetic orientation (aspiration vs. closure voicing) to macro-level change. Jutland Danish shows classic gradation: in strong onsets, /p t k/ are aspirated [pʰ tʰ kʰ] and /b d g/ surface as unaspirated [p t k]; in weak onsets and codas, fortis loses aspiration and lenis lenites (glides/approximants or Ø). Using a large legacy corpus of rural interviews (1971–76), the analysis models fortis VOT and the probability of continuous closure voicing in lenis stops with spatial GAMMs.

The spatial models reveal gradual north–south gradients rather than sharp isoglosses: fortis VOT (aspiration) shortens toward the north while lenis closure voicing increases, with a non-linear transition between them. Consequently, the laryngeal contrast is aspiration-oriented in the south and voicing-oriented in the north—what the author terms contrast orientation—and these orientations co-vary with weak-position gradation, yielding more sonorous outcomes where voicing is weighted more and less sonorous outcomes where aspiration dominates. This pattern indicates sensitivity to fine laryngeal settings beyond a simple voicing–aspiration dichotomy: phonetic implementation in prosodically strong onsets biases how the same stops are realized in weak positions. Communities that weight voicing more tend to produce more sonorous weak-position realizations, whereas aspiration-leaning communities produce less sonorous ones—evidence of system-wide cue balancing that maintains regional coherence within variation.

In broader perspective, these gradient cue orientations provide a plausible bridge from synchronic variation to diachronic divergence: when communities sustain stable preferences for distributing cues across strong versus weak positions, small, consistent biases can accumulate and crystallize into distinct phonetic–phonological systems. Thus, regional phonetic tendencies—rooted in everyday articulatory implementation—offer a concrete mechanism through which gradual, locally conditioned variation propagates into sound change via dialect contact, a process reflected in individual speaker grammars that, over time, cohere into a community's shared phonetic grammar presumably through dialect leveling or accommodation (see Kendall, Pharao, Stuart-Smith, & Vaughn, 2023, for a general review).

Extending the move from regional bias to category stability, Yu et al. (2025) examine a tonal system in flux to ask how subtle, subphonemic distinctions persist or erode across generations. Focusing on *pinjam* (a process where a syllable is realized with a tone other than its lexical (canonical) tone) in Hong Kong Cantonese, they analyze *suspended contrast*—cases where change brings two phonemes close phonetically without full merger, so speakers perceive little or no semantic difference. The persistence of such near-mergers challenges interface accounts assuming categorical lexical representations and functional views predicting rapid loss of non-communicative differences. It also allows the authors to test how contextual and paradigmatic factors can stabilize—or hasten the erosion of—fine phonetic detail as broader sound change progresses within the Hong Kong Cantonese speech community.

The suspended contrast (*pinjam*) under investigation involves the lexical high-rising tone "LR" (Tone 2/25) versus a morphologically derived high-rising tone "DR" (Tone 2/25 from underlying T33/T23/T22/T21 in contexts signaling familiarity, perfective, or note-worthiness). Using an apparent-time design, the study compares LR and DR in near-minimal pairs embedded in carrier phrases, analyzing F0 contours (via GAMMs) and rime duration (via linear mixed-effects) for two cohorts of Hong Kong Cantonese speakers: 20 older (64–85) and 20 younger (21–27).

Overall, DR had a consistently higher F0 than LR across most of the rime, with no durational difference between them (apart from age-related slowing). The LR-DR distinction was most evident among older participants, particularly older men. Among younger participants, the distinction appeared to have largely merged, especially among female speakers. These findings suggest that the contrast is disappearing, with women potentially leading the change, as observed in other sound changes (e.g., Gordon & Heath, 1998; see Kendall et al., 2023, for review). In addition, the contextual effects from the preceding tone were observed among speakers who maintained the distinction. Specifically, high or rising preceding tones (T55, T25) facilitated the preservation of DR being higher than LR, while lower-pitched preceding tones tended to obscure the distinction.

Yu et al.'s apparent-time study provides strong evidence that Cantonese's long-standing suspended contrast has persisted, directly supporting Labov (1987) hypothesis that subphonemic differences can endure for a "considerable period of time" without disrupting word classes or phonemic systems. At the same time, the contrast is trending toward merger—particularly among younger speakers and women—offering a concrete case of evolving phonetic detail. Crucially, the maintenance or loss of subtle F0 distinctions is not random but systematically conditioned by linguistic factors (e.g., preceding tonal contexts and morphologically related neighboring tones), illustrating structured sound change across generations.

2.2.1. Summary and discussion

The four studies in this section show how fine phonetic detail mediates between real-time control and longer-term restructuring. In Northern Anglo-English, Strycharczuk et al. (2024) model long vowels with a shared two-target architecture whose subtle control adjustments yield gradient dispersion that supports both stability and shift; in a cross-linguistic articulatory dataset, Pouplier et al. (2024) demonstrate that anticipatory nasality is language- and position-specific, quantitatively regulated within shared physiological limits; in Jutland Danish, Puggaard-Rode (2024) links cue orientation in strong onsets to weak-position outcomes, tying micro-laryngeal settings to diachronic divergence; and in Hong Kong Cantonese, Yu et al. (2025) trace how a suspended tonal contrast can persist, reorganize, or collapse across generations in systematic ways.

Viewed together, these findings suggest that the phonetic grammar encodes gradient, relational distributions among articulatory targets, coarticulatory timing domains, and cue weights—calibrated by biomechanical constraints and community practice yet transmissible across generations. On this view, sound change does not reflect any loss or weakening of control; rather, it emerges through the continued operation of language-specific phonetic biases embedded in the phonetic grammar itself. The ecological regulation characterized in Section 2.1—balancing effort, clarity, and positional privilege—thus extends naturally into an ecology of change, where small, law-governed adjustments in fine detail cumulatively reshape the system.

2.3. Perceptual adaptation and phonetic biases in the evolution of sound systems

This subsection turns to the perceptual and learning mechanisms through which fine phonetic detail is interpreted, generalized, and transmitted, clarifying how immediate perceptual processing can shape longer-term outcomes. If articulatory micro-variation provides the raw material, then perceptual and grammatical learning determine whether that variation is stabilized or reanalyzed. The two studies approach this interface from complementary perspectives. Melguy and Johnson (2025) test real-time adaptation to accented realizations, with implications for how dialect contact may shift segmental category boundaries and, cumulatively, reshape phonological systems. Kuo (2024) then models how learners internalize perceptually grounded constraints via iterated learning, providing a mechanism for historical reanalysis in Samoan morphophonology. Together they outline the cognitive counterpart to production-side control: listeners and learners structure the phonetic landscape they encounter in ways that plausibly condition how sound systems change under sustained exposure.

Melguy and Johnson (2025) specifically ask how listeners adapt to unfamiliar (accented) pronunciations and whether brief accent exposure reshapes phonetic categories. Targeting English /θ/, they induce an artificial accent in which /θ/ is realized ambiguously between [θ] and [s] (=[θ/s]), mimicking a dialect unfamiliar to the listener. The core question is whether short-term learning through exposure to the accented form simply retunes the perceptual boundary for /θ/ (a narrow, speaker-specific adjustment tied to the exposure) or broadens the category so that untrained variants are also accepted as /θ/—a flexibility that, under dialect contact, could cumulatively condition sound change.

Two cross-modal priming experiments tested this. In both, American English listeners first completed an identical exposure phase under one of three conditions: accented ([θ/s]), canonical ([θ]), or no /θ/ (control). In Experiment 1, a lexical-decision task assessed whether spoken primes containing the learned ambiguous [θ/s] or a more deviated [s] (replacing /θ/) facilitated recognition of /θ/-words. In Experiment 2, primes used a new ambiguous [θ/ʃ] (not heard during exposure) and a related [ʃ] (replacing /θ/). If listeners simply retune to the exposed form, hearing [θ/s] in *hypo*[θ/s]*etical* should speed recognition of *hypothetical* relative to an unrelated prime; if adaptation broadens the category, even the untrained [θ/ʃ] should also yield priming.

The results showed that exposure indeed broadened what listeners accepted as /θ/. Crucially, the listeners trained on either accented or canonical tokens ([θ/s] and [θ], respectively) not only recognized the untrained accented forms ([θ/ʃ]) faster but also generalized to a related [ʃ]. This indicates category expansion, not mere perceptual boundary retuning. Even listeners trained only on canonical [θ] during exposure quickly adjusted once [ʃ]-like realizations appeared in the test phase, suggesting that adaptation can occur on the fly through contextual exposure rather than prolonged training.

These findings thus demonstrate that even short-term experience relaxes the perceptual criteria for /θ/, widening the "acceptance window" for its pronunciation. The extent of this generalization appears to depend in part both on acoustic similarity and on lexical statistics. The authors suggest that given the phonetic similarities of [θ/s] to [θ/ʃ] and [ʃ] in perceptual space, participants when trained with [θ/s] can accommodate [θ/ʃ] and [ʃ]. Moreover, when many /s/-words are present during exposure, listeners resist treating [s] tokens as /θ/ (there's no priming from the related [s]-like form). By contrast, when /ʃ/-words are scarce in exposure, listeners more readily accept /ʃ/-like realizations of /θ/, showing priming effects. Fine phonetic detail thus guides perceptual flexibility, while lexical distributions shape how far it extends.

The theoretical implications are twofold. First, listeners' phonetic categories are not fixed templates but dynamically reweighted distributions—consistent with exemplar models (though not the authors' focus) in which categories consist of stored exemplars whose activation weights shift with exposure, recency, and context, thereby expanding acceptance regions (see Goldrick & Cole, 2023, for review). Second, micro-level perceptual accommodation—learning to interpret an accented variant in the short term—may scale up to macro-level change. It is then plausible that when listeners repeatedly generalize across variants, their broadened perceptual space can influence production through imitation or convergence, gradually shifting community norms. In this way, Melguy and Johnson link rapid perceptual learning with potential longer-term restructuring, suggesting how fine phonetic sensitivity can seed system change.

Kuo (2024) crystallizes this premise by shifting to a slower timescale and modeling how fine phonetic biases become grammatical constraints that propagate across generations. Using an iterated Maximum Entropy grammar for Samoan morphophonology, the study shows how phonetic naturalness—encoded in gradient similarity among sounds—can be internalized and propagated, producing systematic reanalysis over time. The key question is how the reorganization of stem–ergative alternations (e.g., *fulu* ∼ *fulu-a*, *pulu* ∼ *pulu-tia*, *laka* ∼ *laka-sia*) reflects not just lexical frequency but phonotactic pressures grounded in perceptual similarity, particularly the Obligatory Contour Principle for Place of Articulation (OCP-place), which disfavors nearby homorganic consonants because their acoustic similarity reduces perceptual distinctiveness. On this view, graded similarity drives markedness, and learning amplifies these biases into diachronic change.

The model implements a probabilistic learning grammar that assigns weights to constraints rather than categorical rules. Through iterated learning, small phonotactic preferences accumulate across "generations," simulating how minute biases become entrenched. The model is trained on 593 Proto-Oceanic–to-modern Samoan stem–ergative pairs and compares three phonotactic designs: OCP-place (avoid homorganic C–C pairs), BIGRAM (frequency-based pairwise constraints for every C1–C2 sequence), and ACTIVE-CLASS (constraints over induced natural classes such as *CORONAL…CORONAL-sonorant*). Importantly, phonetic similarity provides the model's grounding. Using Mel-Frequency Cepstral Coefficients (MFCCs), Kuo quantifies spectral distance between consonant pairs, correlating acoustic similarity with constraint weighting. The results confirm that the more acoustically similar two consonants are, the more strongly they are disfavored. Among coronals, for example, [*n…l*] (two sonorants) is more marked than [*n…s*] (a sonorant and a fricative), reflecting a phonetically grounded perceptual hierarchy within the grammar.

Three main findings emerge. First, frequency alone cannot explain the modern pattern; models incorporating phonotactic constraints perform better, with OCP-place consistently yielding the best fit. Second, the acoustic data validate the naturalness hypothesis: greater spectral similarity predicts stronger avoidance of the two consonants across syllables. Third, in the iterated-learning simulations, these biases amplify across generations—forms violating OCP-place gradually lose probability, while perceptually distinct forms stabilize—reproducing observed modern outcomes without stipulating exceptions. Kuo thus demonstrates that fine-grained phonetic (dis)similarity feeds directly into grammatical learning: learners internalize perceptually motivated constraints and use them in paradigm restructuring, so reanalysis proceeds not merely by frequency but via naturalness-sensitive regularization. This account reconciles seemingly unpredictable alternations (e.g., the $\varnothing \sim$ C alternations in Samoan stem–ergative paradigms) with systematic, phonetically grounded pressures, indicating that diachronic restructuring follows gradient perceptual logic.

More broadly, by integrating probabilistic grammar, iterated learning, and acoustic metrics, this phonology-oriented study delineates a mechanistic pathway by which subtle phonetic biases accumulate into morphophonological change, providing a concrete bridge from synchronic variation to diachronic evolution. It is conducted within a theoretical framework consistent with phonetically driven phonology (e.g., Hayes, Kirchner, & Steriade, 2004) and weighted-constraint approaches (Harmonic Grammar and its probabilistic MaxEnt extension; see Pater, 2009; Flemming, 2021), and serves as a paradigmatic example of how phonetics and phonology can be integrated into a unified, phonetically grounded model of speech production.

2.3.1. Summary and discussion

Across these two studies, fine phonetic detail emerges as the perceptual–cognitive counterpart to production-side control. Melguy and Johnson (2025) show that brief exposure to accented variants flexibly expands /θ/ categories—perceptual learning is gradient, context-sensitive, and shaped by lexical distributions—and, if repeated in interaction and dialect contact, can feed into community-level change. Extending this logic to transmission, Kuo (2024) uses iterated learning to show how learners internalize phonetically grounded constraints (e.g., OCP-place) and how these biases, amplified across generations, yield systematic morphophonological restructuring.

Together, these studies indicate that perception and learning are active components of the phonetic grammar, linking individual experience to population-level outcomes by channeling fine detail into community-shared patterns. On this view, the phonetic grammar is bi-directional: production generates structured variation, while perception and learning shape how that variation is interpreted, stabilized, or reanalyzed. The ecological regulation traced in Section 2.1 and the articulatory-to-diachronic pathways in Section 2.2 meet here, showing that sound change may emerge from the continuous circulation of fine phonetic detail across articulatory control, perceptual adaptation, and grammatical learning (see Beddor, 2023, for related discussion on impacts of a perception-production loop on sound change).

2.4. Implementing prominence through phonetic detail: from articulation to interaction and learning

Having examined how phonetic control operates across prosodic scaffolds (Section 2.1), drives synchronic variation and sound change (Section 2.2), and interfaces with perception and learning (Section 2.3), we now turn to prominence, which is itself a core component of prosodic structure. Rather than marking a contrast with the preceding sections, this section extends the same prosodic-scaffolding perspective into interactional and learning domains. Prominence—typically realized through a language's intonational grammar and prosodic structuring—provides a primary context in which fine phonetic detail is recruited, redistributed, and learned. The four studies collected here trace this continuum from production to acquisition. Shao, Hermes, Buech, and Giavazzi (2025) model stress as an articulatory control gesture that retimes and retargets overlapping movements in Italian; Hatcher, Joo, Kim, and Cho (2024) extend this to phrasal focus in Korean, showing that prominence can be realized without prosodic restructuring; Nielsen and Scarborough (2024) examine how interlocutors recalibrate prominence through phonetic convergence; and Sbranna, Albert, and Grice (2025) explore how second-language learners internalize the phonetic correlates of prominence. Taken together, the studies portray prominence as a kind of prosodic regulator realized through fine phonetic detail: it modulates articulatory timing, shapes acoustic cues, guides listener interpretation, adapts dynamically in interaction, and is internalized in second-language acquisition.

We begin with Shao et al. (2025) who investigate a word-level effect, asking how lexical-stress prominence shapes coarticulation—specifically, whether a stressed vowel shifts the timing and target of a following velar stop that would otherwise tend to palatalize before /i/ in Italian. Using articulatory data from 15 speakers of Northern Italy, the authors compared two stress-defined contexts: in CLOSE, the velar occurred immediately after a stressed vowel; in FAR, it occurred farther from stress. Tongue-dorsum movement across the vowel–consonant sequence was modeled to assess when the velar reached its constriction target and how fronted or retracted that target was.

Two robust effects emerged. The first is a temporal effect: in the final syllable, the velar reached its constriction target later in CLOSE than in FAR, indicating that post-tonic stops immediately after accentuation show longer closures. The delay is, as the authors suggest, plausibly due to the stressed vowel's larger, longer movement, which expands the vocalic gesture and pushes back the onset of the consonant's gesture. The second is a spatial effect: in CLOSE, the velar's place of articulation shifted posteriorly (i.e., it was less palatalized) than in FAR. Thus, stress on the preceding vowel not only stretches time, it also retunes where the following velar is aimed in the vocal tract. These findings show that stress-related strengthening "spills over" beyond the stressed syllable, reshaping the subsequent consonant in both when (later target achievement) and where (more posterior target). This effect is reminiscent of Turk and colleagues' (e.g., Turk & Sawusch, 1997; Dimitrova & Turk, 2012) studies which define the possible scope of accentual lengthening when a lexical stress syllable receives a phrase-level pitch accent. Functionally, this acts as a resistance to coarticulatory palatalization in post-tonic position—providing an articulatory basis for the well-known blocking of phonological palatalization after stress in Italian.

The authors interpret their results within the μ-gesture (modulation-gesture) framework (e.g., Saltzman, Nam, Krivokapić, & Goldstein, 2008; Katsika, Krivokapić, Mooshammer, Tiede, & Goldstein, 2014; Katsika, 2016): a prosodic control gesture that does not shape the vocal tract directly but modulates the magnitude, timing, and stiffness of overlapping segmental gestures to realize prominence. On this view, lexical stress enlarges the vowel gesture, and—because gestures overlap—this spills over to the following velar, delaying target attainment and retracting place. Stress thus behaves as a temporally extended event that dynamically couples with neighboring gestures. The outcomes support a view in which prominence shapes articulation through gradient biomechanical coordination across gestures—not a categorical, syllable-bounded label—yielding fine-grained acoustic–phonetic consequences. They indicate that dynamic control, consistent with the phonetic grammar, interacts with the language's higher-order prominence system.

Moving from word-level to phrase-level prominence, Hatcher et al. (2024) turn to phrasal focus in Korean. They reexamine the long-standing claim that in Seoul Korean—an edge-prominence system where phrasing cues prominence—narrow/contrastive focus must trigger prosodic restructuring, typically by launching a new Accentual Phrase (AP) or a higher unit (intermediate phrase, ip; Intonational Phrase, IP) (Jun, 2005, 2007, 2011). They test this assumption using short utterances to determine whether focus necessarily induces new prosodic junctures at the phrase level or can instead be expressed through fine phonetic adjustments within an existing prosodic frame. Fourteen native speakers produced short sentences with monosyllabic targets (/pam/, /pap/) in IP-initial, medial, and final positions. Using generalized additive mixed models, the authors modeled continuous F0 trajectories to examine how the alignment and scaling of Accentual Phrase (AP) tones (L and H) vary with position and focus. In Seoul Korean, the left edge of an AP is canonically marked by an L tone followed by an H tone (LH), providing a key tonal cue to phrasing.

Phrase-initially, focus elicited the expected AP-level LH rise, with pitch-range expansion driven by a higher H peak while L remained stable. Phrase-medially, however, there was no pitch-range expansion and no clear tonal evidence of a new phrasal boundary, indicating that focus can be realized within an existing phrase via local tonal adjustment rather than by initiating a new phrase. Phrase-finally, focus produced a continuous F0 fall compressed by the boundary tone (L%), with a brief F0 sustain that preserved only a phonetic trace of the focal LH pattern—again without evidence of rephrasing. Segmental measures reinforced the tonal results: focused targets were longer, initial lenis stops showed increased VOT, and lenition between sonorants was blocked—enhancements typical of Korean phrase edges, here realized as localized hyperarticulation, paralleling effects reported for head-prominence languages like English (de Jong, 1995). Crucially, these edge-like strengthening effects occurred without the creation of a new phrase, showing that focus heightens articulatory and perceptual prominence within the existing prosodic constituent, independent of phrasing.

These results challenge the view that focus in a so-called edge-prominence languages such as Korean necessarily triggers prosodic restructuring. This extends cross-linguistic perspectives on prominence beyond stress-based systems: even in an edge-prominence language, prominence is realized through gradient, fine-grained phonetic control embedded in the intonational grammar. In dynamical terms, Hatcher et al.'s focus effects can be read—though not framed that way in the paper—as a higher-tier counterpart to Shao et al. In Italian, lexical stress is instantiated by the μ-gesture that modulates spatial and temporal realizations of overlapping segmental gestures; in Korean, focus can be modeled as a μ-gesture that modulates segmental hyperarticulation and likely LH scaling/alignment within an existing phrase, yielding expansion in both segmental and tonal realization. Both are non-tract-variable control gestures; their distinct phonetic signatures (reduced post-tonic palatalization vs. LH alignment and H-peak scaling without rephrasing) instantiate the same mechanism: prominence arising from graded coupling among overlapping gestures.

Taken together, Shao et al. (2025) and Hatcher et al. (2024) outline a continuum of prominence control—from local gesturing timing shifts caused by stress to focus realized through tonal alignment and segmental hyperarticulation within an existing phrase. Nielsen and Scarborough (2024) then ask whether prominence itself is a target of phonetic convergence—whether interlocutors recalibrate these controls in interaction, shifting toward one another's patterns. From a dynamical perspective, though the authors do not use this vocabulary, they test whether imitation is driven by low-level acoustics or by abstract, relational representations of prominence that govern internal control gestures (or, descriptively, the phonetic grammar), thereby linking micro-mechanics to the system's adaptive plasticity. Specifically, Nielsen and Scarborough examine how talkers adapt their productions to recently heard patterns and whether such adaptation reflects sensitivity to linguistic structure rather than mere acoustic mimicry.

Across two implicit-imitation experiments in North American English, participants shadowed and retold sentences produced by a model talker with low mean F0 and a fast rate, while the critical manipulation contrasted a salient contrastive pitch accent (L + H*) with a less informative default accent (H*). The question was whether convergence tracks absolute acoustics—drifting toward the model's low baseline and rapid tempo—or relational/phonological structure—enhancing cues to prominence even when that entails diverging from the model's absolute values. Accordingly, the authors analyzed imitation at two levels: a phonological/categorical level (accent type, the relative L + H* peak-to-baseline rise, and the proportion of the accented region within its carrier phrase) and an acoustic–phonetic level (absolute carrier-phrase F0 and duration).

When participants heard an exaggerated contrastive accent (L + H*) from a low-pitched model talker, they enhanced the contrast by lowering the carrier-phrase baseline rather than raising the peak—thus increasing the relative rise without changing the peak's absolute height. This relational adjustment persisted beyond shadowing, consistent with consolidating a normalized linguistic target rather than copying the model's values. Timing showed a parallel pattern: speech accelerated overall, but the carrier phrase shortened more than the accented region, increasing the target's relative durational prominence. By contrast, with a default H* (no exaggerated prosody associated with L + H*), participants converged on tempo but did not uniformly lower baseline F0. This confirms that baseline f0 adjustment was prominence-driven in relative terms, not mere imitation of a low voice, so that the target was linguistically interpretable with respect to tonal realization that captures its phonological status in the intonational grammar.

Nielsen and Scarborough (2024) thus show that convergence is multi-tiered. Speakers accommodate some low-level acoustic–phonetic properties—most reliably by accelerating the carrier phrase—but the more diagnostic adjustments target higher-level, linguistically normalized relations: accent category, relative peak rise, and relative target-to-carrier duration. Low-level changes to surrounding material (e.g., baseline lowering, durational redistribution) are recruited to realize these higher-level goals. This, as the authors imply, helps reconcile mixed findings in the convergence literature: imitation is guided primarily by abstract, relational representations while still permitting limited accommodation to raw acoustics. In relation to Shao et al. (2025) and Hatcher et al. (2024), Nielsen and Scarborough's (2024) study—if cast in the same dynamical terms—extends prosodic control from individual articulation to interactive adaptation, where internal control gestures are retuned through social and perceptual coupling. More broadly, these studies together suggest that prominence functions not only as a production-internal control signal but also as a shared, adaptive parameter in communication.

Imitation has often been regarded as an important mechanism in language acquisition (cf. Meltzoff & Moore, 1977; Kuhl & Meltzoff, 1996). The next question, then, is how learners internalize the notion of prominence itself. Sbranna et al. (2025) tackle this by asking how L2 speakers map fine-grained phonetic detail onto information-structural prominence—another manifestation of the prosodic scaffold discussed earlier—expressed in German via pitch accents, and how those mappings diverge from both native German and the learners' L1. They examine Italian learners of German producing two-word noun phrases that vary in information status (Given–New vs. New–Given), such as *braune WELLE* vs. *BRAUNE welle* (lit. "brown WAVE" vs. "BROWN wave"). The languages encode this contrast differently: German typically deaccents given material, whereas Italian disfavors deaccentuation within noun phrases and often signals information status by word order, frequently accenting the final word regardless of status. Prior work, as the authors note, points to an asymmetry in learnability: German learners can adopt the Italian pattern, whereas Italian learners struggle to produce German-style deaccentuation.

Departing from such categorical division, Sbranna et al. probe the phonetic detail of interlanguage prosody using the ProPer workflow (Albert, Cangemi, Ellison, & Grice, 2020), which quantifies *Mass* (prosodic weight, computed as the integral over duration and power of the periodic-energy curve), *Synchrony* (within-syllable F0 slope), and ΔF0 (F0 change across syllables). These measures capture continuous modulation of pitch and prominence across two-word noun phrases produced by forty Italian learners of German and eighteen native German speakers in a semi-spontaneous board-game task. The results show that Italian learners distinguished Given–New and New–Given sequences by F0 timing—placing the fall earlier or later in the phrase—mirroring L1 Italian (but within a reduced pitch range, often found in L2 speech). They did not, however, redistribute prosodic weight: the second word was consistently weakened in both contexts, unlike native German, which strengthens it in Given–New and weakens it in New–Given. This pattern—stable across proficiency levels—suggests that learners acquire tonal timing relatively early, whereas redistributing prosodic weight is a later and more difficult attainment. The authors attribute this asymmetry in part to markedness: Mass-related weighting is the more marked option and thus slower to emerge. A complementary account, in our synthesis, is that temporal control is broadly shared across languages: in a given speech system, speakers continually coordinate segmental and suprasegmental events while adapting to prosodic and rate demands; consequently, L2 learners internalize timing more readily than spectral properties or amplitude/energy distributions (see Choi, Kim, & Cho, 2016, and references therein).

Overall, Sbrann et al's study demonstrate that continuous phonetic analysis of prosodic patterns can uncover subtle aspects of L2 prominence marking that categorical approaches tend to obscure. The findings reveal that prosodic acquisition cannot be reduced to a binary of "transfer" versus "no transfer." Instead, they show how learners' cognitive systems actively attempt to map new phonetic forms onto familiar linguistic functions—even when those mappings are imperfect. (See below for further discussion.).

2.4.1. Summary and discussion

Across the four studies, prominence emerges as a multidimensional interface coordinating phonetic realization across articulatory, prosodic, interactive, and acquisitional domains. Shao et al. (2025) show that Italian lexical stress acts as an articulatory control gesture that retimes and retargets overlapping movements, modulating coarticulation. Hatcher et al. (2024) extend this to Korean phrasal focus, demonstrating realization via continuous tonal and segmental adjustments within an existing phrase—independent of phrasing. Nielsen and Scarborough (2024) find that in interaction, speakers recalibrate prominence relations through phonetic convergence guided by abstract relational targets rather than raw acoustics. Sbranna et al. (2025) complete the arc to L2 acquisition, showing that learners internalize prominence correlates on distinct timelines: tonal timing is acquired relatively early, whereas redistribution of prosodic weight proves more resistant.

Viewed in the broader trajectory of this section, Sbranna et al. (2025) complement Shao et al. (2025) (local articulatory control via stress) and Hatcher et al. (2024) (phrasal prominence without obligatory rephrasing) by showing that prosodic control regimes are only partially internalized in L2 and unfold along separable phonetic dimensions—plausibly depending on whether a given dimension is more universally available or less marked. Together with Nielsen and Scarborough, who demonstrate interactive retuning of prominence, these results support a unified view of prominence as a learnable, adaptive control system operating across multiple phonetic levels. Its components (e.g., timing vs. weighting) are retimed, rescaled, and recalibrated through social and linguistic experience, but they adapt at different rates. A promising direction for future work is to compare phonetic convergence with L2 acquisition to determine which components of prominence are more readily accommodated and which are more resistant to change. Such a control system may be modeled in a dynamical system in which a separate modulation gesture may operate in controlling the prominence-induced phonetic variation in both segmental and suprasmegntal dimensions, and even in the L2 system.

3. Conclusion

Across fourteen contributions, a shared insight comes into view: fine phonetic detail is not an epiphenomenon of linguistic structure but the medium through which structure itself is created, maintained, and reshaped. It functions as the control layer of speech—the continuous system by which languages realize contrasts, keep them intelligible in communication, and steer them toward change. Two themes stand out.

The first is that speaker control lies at the heart of these processes, not mechanical necessity alone. Across languages and methodologies, the contributions converge on dynamical accounts of production: the fine-tuning of overlapping gestures, the coordination of vowel targets, and the precise timing of articulatory and acoustic cues. In studies of prominence—whether through segmental retiming or phrase-level scaling—these patterns reveal speech as an ongoing negotiation of timing and effort, reasonably well captured by dynamical models of continuous coordination such as Articulatory Phonology and Task Dynamics.

The second theme is that control is built into the linguistic system itself. As several papers imply, the notion of *the phonetic grammar* (Keating, 1984, 1985; Cho & Ladefoged, 1999; Cho, 2025; cf. Kingston & Diehl, 1994) provides a framework for understanding how fine detail is language-specifically fine-tuned, and interacts with prosodic, linguistic, and communicative structure. This grammar modulates phonetic implementation so that sound systems remain stable yet adaptable within the limits imposed by human biomechanics and language-specific organization. Complementary studies on perception and learning show how listeners continually adjust cue weights in real time, and how these local adaptations, accumulating across speakers and generations, shape the long-term evolution of sound systems.

This perspective supports what might be called a *variation-within-universals* view of sound systems. Across the typological range represented in this issue, cross-linguistic diversity arises not from arbitrary difference but from systematic adaptations to shared biomechanical, perceptual, and informational constraints. The phonetic grammar encodes these adjustments as continuous relations among targets, timing, and cues rather than as categorical rules. Micro-adjustments in gestures, nasality, or voicing gradually crystallize into typological patterns and sound change, without the need for wholesale restructuring.

Taken together, these studies identify the prosody–phonetics interface as a locus of control: continuous parameters are actively managed to maintain contrast and patterning, and perception and learning recalibrate those relations across developmental, social, and historical scales—linking immediate adaptation to population outcomes. As the studies here make clear, prosody provides the primary scaffold for fine phonetic detail—that is, *prosodic scaffolding* understood as a persistent organizational framework that constrains and guides the dynamic allocation of phonetic resources, stabilizing contrasts while remaining adaptable to contextual and developmental variation. Segmental realization occurs within a prosodic frame from early acquisition onward, and prosodic organization mediates interaction with higher-order structure (morphology, syntax, information structure). At the same time, as the extended model suggests, phonetic grammar can also interact directly with these domains. The primacy of prosody in speech production is evident in both developmental and structural domains. From the earliest stages of life, infants are immersed in speech sounds intertwined with prosodic patterns—even in utero, where they become sensitive to the rhythmic and intonational contours of the ambient language—well before abstract phonological or syntactic categories emerge. As phonological categories gradually crystallize during acquisition, children fine-tune their speech production to align with them, just as other members of their speech community do. Conceptually, the contributions in this issue accord with this view: the phonetic grammar functions as a system-internal control architecture linking phonetics with phonology and prosody, with prosody anchoring fine-grained phonetic regulation. Through prosodic mediation, and sometimes more directly, this control extends to higher-order structures, yielding a language's distinctive phonetic hallmarks and shaping patterns of synchronic variation and diachronic change.

As such, beyond phonetics, the findings also motivate phonological models that integrate continuous control into the grammar rather than treating it as post-hoc implementation. They also identify prominence and convergence as tractable domains for examining control and alignment in interaction and second-language acquisition. Methodologically, a cross-linguistic program integrating production, perception, and modeling provides a concrete test bed for explaining how abstract structure emerges from, and is sustained by, continuous phonetic control. All in all, the contributions of this special issue present sound systems as stable yet adaptable outcomes of coordinated control among prosody, articulation, perception, and learning, with fine phonetic detail at the center—realizing structure, sustaining communication, and channeling principled change, as emergent properties of coordinated control.

CRediT authorship contribution statement

Taehong Cho: Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization. **Sahyang Kim:** Writing – review & editing, Writing – original draft, Validation, Conceptualization. **Holger Mitterer:** Writing – review & editing, Writing – original draft, Validation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

We extend our sincere thanks to all the contributors whose insightful work and dedication have shaped and enriched this special issue. We are also grateful to Marianne Pouplier and an anonymous reviewer for their thoughtful and constructive comments, which significantly improved the final product. We also thank Jiyoung Jang at HIPCS for her helpful comments on an earlier version of this study. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884).

References

Albert, A., Cangemi, F., Ellison, T. M., & Grice, M. (2020). 'ProPer: PROsodic analysis with PERiodic energy' [Computer Software]. *OSF*. https://doi.org/10.17605/OSF.IO/28EA5.

Beckman, J. N. (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology, 14*(1), 1–46. https://doi.org/10.1017/S0952675797003308.

Beckman, J. N. (1998). *Positional Faithfulness* (Doctoral dissertation, University of Massachusetts Amherst). Rutgers Optimality Archive (ROA–234). http://roa.rutgers.edu.

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., & Kingston, J. (Eds.). (1990). *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*(4), 785–821. https://doi.org/10.1353/lan.0.0165.

Beddor, P. S. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change. *Journal of Phonetics, 97*, 101228. https://doi.org/10.1016/j.wocn.2023.101228.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. G., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language, 94*(4), 931–968. https://doi.org/10.1353/lan.2018.0051.

Bennett, R., Padgett, J., Ní Chiosáin, M., McGuire, G., & Bellik, J. (2024). Effects of syllable position and place of articulation on secondary dorsal contrasts: An ultrasound study of Irish. *Journal of Phonetics, 107*, 101368. https://doi.org/10.1016/j.wocn.2024.101368.

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

Byrd, D. M. (1994). *Articulatory timing in English consonant sequences* (Doctoral dissertation). Los Angeles: University of California.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass, 10*(3), 120–141.

Cho, T. (2022). The phonetics-prosody interface and prosodic strengthening in Korean. In S. Cho & J. Whitman (Eds.), *The Cambridge Handbook of Korean Linguistics* (pp. 248–293). Cambridge University Press.

Cho, T. (2025). Advancements of phonetics in the 21st century: Phonetic universals, language variation, and phonetic grammar. *Journal of Phonetics, 111*, 101426. https://doi.org/10.1016/j.wocn.2025.101426.

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., & 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., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094.

Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent, and lexical stress. *Journal of Phonetics, 33*(2), 121–157. https://doi.org/10.1016/j.wocn.2005.01.001.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain–initial strengthening in English. *Journal of Phonetics, 35*(2), 210–243. https://doi.org/10.1016/j.wocn.2006.03.003.

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.

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.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Cohn, A. C. (1990). Phonetic and phonological rules of nasalization. *UCLA Working Papers in Phonetics, 76*.

Cohn, A. C. (1993). Nasalisation in English: Phonology or phonetics? *Phonology, 10*(1), 43–81.

Davidson, L., & Parker Jones, O. (2024). Effects of word–level structure on oral stop realization in Hawaiian. *Journal of Phonetics, 107*, 101367. https://doi.org/10.1016/j.wocn.2024.101367.

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. https://doi.org/10.1121/1.412275.

Dimitrova, S., & Turk, A. (2012). Patterns of accentual lengthening in English four–syllable words. *Journal of Phonetics, 40*, 403–418.

Flemming, E. (2021). Comparing MaxEnt and noisy harmonic grammar. *Glossa: A Journal of General Linguistics, 6*(1), 28.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*(6), 3728–3740. https://doi.org/10.1121/1.418332.

Goldrick, M., & Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics, 99*, 101254. https://doi.org/10.1016/j.wocn.2023.101254.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge University Press. https://doi.org/10.1017/CBO9780511541599.008.

Gordon, M., & Heath, J. (1998). Sex, sound symbolism, and sociolinguistics. *Current Anthropology, 39*(4), 421–449.

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science, 8*, 425–434.

Hatcher, R., Joo, H., Kim, S., & Cho, T. (2024). Focus–induced tonal distribution in Seoul Korean as an edge–prominence language. *Journal of Phonetics, 107*, 101353. https://doi.org/10.1016/j.wocn.2024.101353.

Hayes, B., Kirchner, R., & Steriade, D. (Eds.). (2004). *Phonetically based phonology*. https://doi.org/10.1017/CBO9780511486401.

Huang, Y. (2025). The role of tone and phrasing in the occurrence of period doubling and vocal fry in Mandarin. *Journal of Phonetics, 111*, 101416. https://doi.org/10.1016/j.wocn.2025.101416.

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. https://doi.org/10.1016/j.wocn.2022.101195.

Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press.

Jun, S.-A. (2007). The intermediate phrase in Korean: Evidence from sentence processing. In T. Riad & C. Gussenhoven (Eds.). *Tones and tunes: Experimental studies in word and sentence prosody* (Vol. 2). De Gruyter Mouton.

Jun, S.-A. (2011). Prosodic markings of complex NP focus, syntax, and the pre–/post–focus string. In *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* (pp. 214–230).

Katsika, A. (2016). The role of prominence in determining the scope of boundary–related lengthening in Greek. *Journal of Phonetics, 55*, 149–181. https://doi.org/10.1016/j.wocn.2015.12.003.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and their interaction with prominence. *Journal of Phonetics, 44*, 62–82. https://doi.org/10.1016/j.wocn.2014.03.003.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*(2), 286–319. https://doi.org/10.2307/413642.

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

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics, 18*(3), 321–334. https://doi.org/10.1016/S0095-4470(19)30377-8.

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

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

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*(3), 419–454. https://doi.org/10.1353/lan.1994.0023.

Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures* (Doctoral dissertation). Yale University.

Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics, 27*(1), 23–54.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*, 2425–2438.

Kuo, J. (2024). Phonetic naturalness in the reanalysis of Samoan thematic consonant alternations. *Journal of Phonetics, 107*, 101355. https://doi.org/10.1016/j.wocn.2024.101355.

Labov, W. (1987). The overestimation of functionalism. In R. Dirven & V. Fried (Eds.), *Functionalism in linguistics* (pp. 311–332). John Benjamins.

Marin, S. N. (2007). *Vowel–to–vowel coordination, diphthongs and Articulatory Phonology* (Doctoral dissertation). Yale University.

Melguy, Y. V., & Johnson, K. (2025). What are you sinking about? Experience with unfamiliar accent produces both inhibition and facilitation during lexical processing. *Journal of Phonetics, 109*, 101401. https://doi.org/10.1016/j.wocn.2025.101401.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 75–78.

National Institute for Japanese Language and Linguistics (2012). *CSJ–RDB* [Database].

Nielsen, K., & Scarborough, R. (2024). On the target of phonetic convergence: Acoustic and linguistic aspects of pitch accent imitation. *Journal of Phonetics, 107*, 101372. https://doi.org/10.1016/j.wocn.2024.101372.

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science, 33*(6), 999–1035. https://doi.org/10.1111/j.1551-6709.2009.01047.x.

Pouplier, M., Rodriquez, F., Lo, J. J. H., Alderton, R., Evans, B. G., Reinisch, E., & Carignan, C. (2024). Language–specific and individual variation in anticipatory nasal coarticulation: A comparative study of American English, French, and German. *Journal of Phonetics, 107*, 101365. https://doi.org/10.1016/j.wocn.2024.101365.

Puggaard-Rode, R. (2024). Regional covariation of stop gradation and laryngeal contrasts in Jutland Danish. *Journal of Phonetics, 106*, 101354. https://doi.org/10.1016/j.wocn.2024.101354.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333–382.

Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task–dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody. Campinas, Brazil* (pp. 175–184).

Sano, S., & Guillemot, C. (2025). Contrast enhancement and the distribution of vowel duration in Japanese. *Journal of Phonetics, 108*, 101386. https://doi.org/10.1016/j.wocn.2024.101386.

Sbranna, S., Albert, A., & Grice, M. (2025). Investigating interlanguages beyond categorical analyses: Prosodic marking of information status in Italian learners of German. *Journal of Phonetics, 108*, 101377. https://doi.org/10.1016/j.wocn.2024.101377.

Shao, B., Hermes, A., Buech, P., & Giavazzi, M. (2025). Articulatory consequences of lexical stress on post–tonic velar plosives in Italian. *Journal of Phonetics, 108*, 101371. https://doi.org/10.1016/j.wocn.2024.101371.

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

Sørensen, T., & Gafos, A. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology, 28*, 188–215. https://doi.org/10.1080/10407413.2016.1230368.

Steriade, D. (1999). Phonetics in phonology: The case of laryngeal neutralization. In M. Gordon (Ed.), *Papers in phonology 3 (UCLA working papers in linguistics No. 2)* (pp. 25–146). UCLA Department of Linguistics.

Strycharczuk, P., Kirkham, S., Gorman, E., & Nagamine, T. (2024). Towards a dynamical model of English vowels: Evidence from diphthongisation. *Journal of Phonetics, 107*, 101349. https://doi.org/10.1016/j.wocn.2024.101349.

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics, 25*, 25–41.

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. https://doi.org/10.1016/j.wocn.2023.101275.

Yu, A. C. L., Li, V. G., & Mok, P. P. K. (2025). Contextual and paradigmatic effects on suspended contrast across generations: The case of Cantonese pinjam revisited. *Journal of Phonetics, 110*, 101412. https://doi.org/10.1016/j.wocn.2025.101412.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics, 61*, 13–29.