Proceedings of the 5th International Conference on Literature, Language, and Culture Development
DOI: 10.54254/2753-7064/2026.33821

Articulation and Perception in Sound Change: A Critical Comparison of Explanatory Approaches

Chengze Xie

College of Arts and Law Edgbaston, University of Birmingham, Birmingham, UK
melanocyte @163.com

Abstract. This essay critically compares articulation-based and perception-based approaches to sound change. Articulation-based accounts explain sound change primarily through production pressures such as coarticulation, reduction, assimilation, lenition, and gestural timing, offering strong explanations for gradual and phonetically natural changes. In contrast, perception-based accounts emphasise the role of listener interpretation, normalisation, cue weighting, and reanalysis, making them especially useful for explaining changes such as dissimilation and perceptual misattribution. This essay evaluates the strengths and limitations of both perspectives in terms of explanatory scope, evidence, directionality, actuation, and diffusion. It argues that neither approach alone can fully explain the emergence, selection, and stabilisation of sound change. While production pressures often provide the phonetic source and direction of many gradual changes, perception-based mechanisms are essential for understanding listener reanalysis and certain non-reductive developments. The essay, therefore, supports an interactionist position: sound change is best explained through a type- and stage-sensitive model that integrates articulation, perception, lexical diffusion, and wider social and systemic factors.

Keywords: Sound change, Articulation-based approach, Perception-based approach, lenition, assimilation, dissimilation, coarticulation

1. Introduction

Sound change is commonly explained through two perspectives: articulation-based accounts, which prioritise biases arising in speech production, and perception-based accounts, which prioritise how listeners interpret and categorise the speech signal. Articulation-based explanations emphasise coarticulation, reduction, assimilation, and gestural timing, arguing that recurrent phonetic variation may become conventionalised as a community change. Drawing on Bybee's usage-based model [1], Ohala's listener-based theory, and related discussions by Beddor, Recasens, McMahon, and Millar and Trask [2-6]. Perception-based explanations instead highlight listener normalisation and reanalysis: change may be initiated when contextual effects are misattributed to a segment or when corrective strategies are misapplied. Although the approaches appear to focus on speaking versus hearing, they are linked by the speech chain, which connects articulation to acoustic transmission and listener perception. Production therefore supplies the variable input that perception must parse, and perceptual interpretations can shape what is learned and later reproduced.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7064/2026.33821

This essay compares the two approaches in terms of (i) the change types they explain best, (ii) the evidence they appeal to, and (iii) their explanatory limits. Following Bybee's usage-based synthesis [1] and McMahon's methodological cautions [5], this essay adopts a qualified interactionist stance: production pressures often best capture the phonetic source and direction of many gradual changes, but listener-based mechanisms are indispensable for some change types (notably certain dissimilations and reanalyses), and neither approach alone fully explains diffusion and stabilisation. The most defensible position is therefore articulation-leading but not articulation-only.

2. Explanation and criticism of the articulation-based approach

Articulation-based explanations locate the origins of sound change in production pressures. Speech is not produced as a sequence of invariant targets: speakers overlap gestures, undershoot articulatory extremes under time pressure, and reduce effort in frequent and predictable contexts. Such processes yield structured phonetic variation that may become entrenched if a community repeatedly encounters the same variants. Classic changes such as assimilation, epenthesis, and lenition are therefore often treated as speaker-side outcomes that reduce articulatory cost or improve coordination [6]. A key advantage of this approach is that it often motivates directionality: when gestures overlap, features may spread; when targets are undershot, constrictions may weaken; and when sequences are articulatorily complex, speakers may simplify them.

Lenition illustrates how articulatory explanations can motivate directional change with a clear phonetic basis. Lenition (weakening) refers to consonant changes that move sounds along strength-to-weakness scales. Millar and Trask present scales such as stop > fricative > approximant and voiceless > voiced, relating them to obstruction and tension: "stronger" consonants typically involve greater constriction and/or tension, while "weaker" consonants involve less constriction and thus less articulatory effort [6]. This yields a concrete account of directionality: complete closure may be replaced by narrow constriction and then by wider constriction.

Millar and Trask also note that lenition is especially common intervocalically, where weakening reduces articulator movement and produces more vowel-like consonants [6]; in this sense, intervocalic lenition can be analysed as assimilation towards surrounding vowels. More broadly, the articulatory approach can motivate other recurrent trajectories, such as palatalisation of velars before front vowels: the tongue-body gesture required by a following front vowel overlaps with the consonant gesture, shifting the place of articulation and reducing extreme transitions [6]. Taken together, these patterns support the view that production biases can generate repeatable variants rather than idiosyncratic errors.

Bybee modernises the articulatory approach by embedding it in a usage-based model with explicit diagnostics [1]. She argues that many changes are gradual both phonetically and lexically: words with identical phonemic strings may display different degrees of variation during a change in progress. This matters because it shifts attention from idealised segment inventories to the lexical distribution of variants, which is often where change first becomes visible.

Bybee further proposes that numerous reductive changes disproportionately affect high-frequency items, especially when frequency is defined as frequency of occurrence in the conditioning environment [1]. This yields an empirically tractable prediction: diffusion from high-frequency to low-frequency items is consistent with production-side automatization, because repetition strengthens motor routines and increases coarticulation and reduction. Her exemplar model provides a representational mechanism: repeated reduced tokens accumulate in a word's exemplar "cloud," reshaping word-specific routines and influencing more general articulatory routines, thereby providing a plausible route for diffusion beyond the original lexical items. Bybee also treats extreme

DOI: 10.54254/2753-7064/2026.33821

reductions in very frequent phrases as intensified outcomes of the same production pressures, supporting continuity between everyday coarticulation and more categorical-seeming outcomes.

Despite these strengths, articulation-based explanations face persistent criticisms. First, they can be overgeneralised if articulatory convenience is treated as a universal cause. McMahon notes that similar phonetic environments do not always produce the same historical outcomes, so "ease of articulation" cannot function as a strict, law-like explanation [5]; this connects directly to actuation, since production pressures do not explain why a change begins in one community at one time rather than another. In practice, articulatory pressures are best treated as probabilistic biases: they explain why certain outcomes are common and directionally stable, but they do not guarantee occurrence.

Second, some change types do not fit the typical "automation" profile. Dissimilation is a standard example: it is often sporadic and concentrated in particular segment classes (especially liquids), and some attested outcomes (e.g., Latin *arbor* > Spanish *árbol*) are not transparently motivated by effort reduction [6]. Bybee therefore proposes diagnostics: if a change is not phonetically gradual, not cross-linguistically common, not tied to coarticulation, or not led by high-frequency items, a production-automation source is less likely and perceptual sources become plausible [1]. Third, even where articulatory pressures plausibly generate variants, they do not by themselves explain why a particular variant is selected and stabilised rather than remaining an optional variation. This selection problem points to the need for listener interpretation, learning, and social transmission mechanisms, anticipating the perception-based accounts discussed next.

If articulation-based accounts are strongest at explaining the emergence and direction of many gradual variants, perception-based accounts aim to explain how such variants are interpreted, reanalysed, and in some cases transformed into stable categories.

3. Explanation and criticism of the perception-based approach

Perception-based explanations treat sound change as emerging from how listeners interpret and normalise a variable signal. Ohala frames the listener as a potential source of innovation at initiation, explicitly separating initiation from the later question of diffusion through the lexicon and community [2]. According to this view, listeners routinely correct for contextual distortion (coarticulation) in a normalisation process, but change can arise when correction is incomplete or misapplied. A listener may misconstrue what is contextual versus inherent, and a subset of such "mini-sound changes" may then enter circulation [2]. This framing is attractive because it offers a principled route from perception to phonological change: the listener's perceptual inference shifts, and later productions reflect that inference.

Ohala's hypercorrection model is particularly relevant for dissimilation. If listeners treat similarity as contextual distortion and attempt to factor it out, they may introduce greater differentiation [2], a pathway that complements the descriptive prominence of liquid dissimilation in historical surveys [6]. Importantly, this is not presented as a purely random error; it arises from systematic strategies of compensation in perception. Solé and Recasens also refine listener-based explanations beyond a single "mishearing" event [7]. Moreover, Beddor proposes perception grammars: listeners differ systematically in how they weight multiple coarticulatory cues, and innovative cue weightings can seed change if they become publicly manifested [3]. Perceptual variation can thus be structured and socially available, making it a plausible source of population-level change. Recasens similarly argues that perceptual outcomes depend on the relative prominence of cues in context and that a single change (e.g., developments affecting dark /l/ in Romance) may follow different paths depending on cue structure [4]. Together, these approaches strengthen perception-based explanations by tying them to concrete properties of the acoustic–auditory signal

DOI: 10.54254/2753-7064/2026.33821

and to systematic differences in listener inference, rather than to a generic notion of "misperception."

Perception-based accounts, however, face three recurring criticisms that limit their scope as standalone explanations. First, they often under-specify diffusion: even Ohala brackets how innovations spread, and initiation does not entail community stabilisation [2]. This limitation matters because sound change is, by definition, a community phenomenon. Second, the mapping from perception to production cannot be assumed. Beddor cautions against treating perception–production as isomorphic and stresses that perceptual grammars only matter for sound change if they are realised in production and alter the signal [3]. Grosvald and Corina provide an additional empirical caution: in long-distance coarticulation, they find no significant correlation between individuals' production and perception, although they note that change could still be initiated if minorities of highly coarticulatory speakers and highly sensitive listeners interact [8]. Third, causal ordering is difficult to establish. Bybee argues that perceptual similarity is often paralleled by articulatory similarity, so "perception-first" claims require independent diagnostics; she also notes that a pure misperception approach does not explain word-to-word diffusion [1]. Consistent with McMahon's broader caution, perception-based accounts are therefore strongest as partial, type-sensitive explanations rather than universal mechanisms [5].

These analyses suggest that the two approaches offer different explanatory leverage—one foregrounding production biases and diffusion signatures, the other foregrounding listener inference and cue structure—making a direct comparison necessary.

4. Comparison of articulation-based and perception-based approaches

The clearest contrast between the approaches is their explanatory focus within the speech chain. Articulation-based accounts prioritise production biases, and they are often most persuasive for gradual, phonetically natural changes such as lenition and many assimilations. Millar and Trask's lenition scales motivate directionality in terms of reduced obstruction and effort [6], while Bybee adds usage-based predictions: high-frequency items are more exposed to production pressures, and diffusion patterns can therefore support a production-automation source [1]. Perception-based accounts, by contrast, prioritise listener parsing and compensation strategies, and they are frequently invoked for changes that are harder to derive from effort reduction alone, particularly dissimilation [2, 5]. Crucially, the approaches also differ in the kinds of evidence they make central: articulation-based accounts often rely on phonetic naturalness and usage patterns, whereas perception-based accounts gain traction through models of cue attribution and, in some cases, experimentally motivated claims about cue weighting [3, 8].

A useful way to compare the approaches is by their predictions and diagnostics. Bybee's account predicts that many reductive changes will show diffusion signatures consistent with practice and routinisation (often high-frequency leading) [1]. Ohala's listener-based approach predicts characteristic outcomes for dissimilation under hypercorrection [2], while Beddor's perception-grammar approach predicts that systematic cue-weighting differences across listeners can, under some conditions, seed changes that later become socially available [3]. Where historical data allow, these predictions are best assessed against the distribution of variants during a change in progress rather than inferred only from endpoints. In practice, the most informative explanations triangulate across evidence types—phonetic conditioning, perceptual cue structure, and diffusion patterns—because any single strand can remain compatible with more than one causal story.

A second axis of comparison is explanatory scope. The two approaches can explain how structured variation might arise, but neither alone accounts for both actuation and diffusion.

DOI: 10.54254/2753-7064/2026.33821

McMahon cautions that "ease of articulation" cannot be law-like, since similar contexts do not guarantee identical outcomes, and argues for partial, probabilistic explanations; symmetrically, perception-based accounts struggle to demonstrate a stable pathway from perceptual reweighting to production change and to specify mechanisms of lexical spread [5]. Beyond initiation and diffusion, sound change can involve system-level restructuring (e.g., chain shifts), where segment-level production or perception mechanisms are unlikely to be sufficient. Millar and Trask emphasise that sound change can reorganise phonological systems [6], while Smith highlights, in the case of English, that explaining why changes happen when and where they do requires attention to historical and social embedding alongside phonetic mechanisms [9]. The comparison therefore supports a stage-sensitive model in which production biases, listener reanalysis, and diffusion processes interact rather than compete.

5. Conclusion

Articulation-based and perception-based explanations address different links in the speech chain rather than offering competing either/or theories. Production-side pressures often provide the primary impetus for many common, gradual sound changes, and Bybee's usage-based account strengthens this view by linking articulatory mechanisms to frequency effects and diffusion patterns [1]. Nonetheless, a production-only view is insufficient: listener normalisation, cue weighting, and reanalysis remain essential for some change types, particularly dissimilation [2, 3]. Moreover, as McMahon emphasises, explaining how variation arises is not the same as explaining actuation or community-level stabilisation [5]. The most convincing position is therefore articulation-leading but not articulation-only: a type- and stage-sensitive account must integrate production biases with perception, diffusion, and wider systemic and social factors.

References

[1] Bybee, J. (2012). Patterns of lexical diffusion and articulatory motivation for sound change. In M.-J. Solé & D. Recasens (Eds.), Initiation of sound change: Perception, production, and social factors (pp. 211–234). John Benjamins Publishing Company.
[2] Ohala, J. J. (2012). The listener as a source of sound change: An update. In M.-J. Solé & D. Recasens (Eds.), Initiation of sound change: Perception, production, and social factors (pp. 21–35). John Benjamins Publishing Company.
[3] Beddor, P. S. (2012). Perception grammars and sound change. In M.-J. Solé & D. Recasens (Eds.), Initiation of sound change: Perception, production, and social factors (pp. 37–55). John Benjamins Publishing Company.
[4] Recasens, D. (2012). A phonetic interpretation of the sound changes affecting dark /l/ in Romance. In M.-J. Solé & D. Recasens (Eds.), Initiation of sound change: Perception, production, and social factors (pp. 57–75). John Benjamins Publishing Company.
[5] McMahon, A. M. S. (1994). Understanding language change. Cambridge University Press.
[6] Millar, R. M., & Trask, R. L. (2015). Trask's historical linguistics. Taylor & Francis Group.
[7] Solé, M.-J., & Recasens, D. (Eds.). (2012). Initiation of sound change: Perception, production, and social factors. John Benjamins Publishing Company.
[8] Grosvald, M., & Corina, D. (2012). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In M.-J. Solé & D. Recasens (Eds.), Initiation of sound change: Perception, production, and social factors (pp. 77–100). John Benjamins Publishing Company.
[9] Smith, J. (2007). Sound change and the history of English. Oxford University Press. https: //doi-org.bham-ezproxy.idm.oclc.org/10.1093/acprof: oso/9780199291953.001.0001 (Accessed 16 Feb. 2026)