ANGLICA

An International Journal of English Studies
35/2 ▪ 2026

From Generative to Laboratory Phonology: The Phonetics-Phonology Overlap

Guest Editor
Artur Kijak

UNIVERSITY OF WARSAW

ANGLICA

An International Journal of English Studies

From Generative to Laboratory Phonology:
The Phonetics–Phonology Overlap

35/2 2026

GUEST EDITOR
Artur Kijak [artur.kijak@us.edu.pl]

EDITORS-IN-CHIEF
Anna Rędzioch-Korkuz [annaredzioch@uw.edu.pl]
Przemysław Uściński [p.uscinski@uw.edu.pl]

ASSOCIATE EDITORS
Silvia Bruti [silvia.bruti@unipi.it]
Lourdes López Ropero [lourdes.lopez@ua.es]
Martin Löschnigg [martin.loeschnigg@uni-graz.at]
Jerzy Nykiel [jerzy.nykiel@uib.no]

ASSISTANT EDITORS
Dominika Lewandowska-Rodak [d.lewandowska10@uw.edu.pl]
Bartosz Lutostański [b.lutostanski@uw.edu.pl]
Anna Łukaszewicz [a.b.lukaszewicz@uw.edu.pl]
Ewa Wałaszewska [e.walaszewska@uw.edu.pl]

ENGLISH LANGUAGE EDITOR
Mary Newbould [marynewbould23@gmail.com]

ADVISORY BOARD
Michael Bilynsky, University of Lviv
Andrzej Bogusławski, University of Warsaw
Mirosława Buchholtz, Nicolaus Copernicus University, Toruń
Jan Čermák, Charles University, Prague
Edwin Duncan, Towson University
Jacek Fabiszak, Adam Mickiewicz University, Poznań
Elżbieta Foeller-Pituch, Northwestern University, Evanston-Chicago
Piotr Gąsiorowski, Adam Mickiewicz University, Poznań
Keith Hanley, Lancaster University
Andrea Herrera, University of Colorado, Colorado Springs
Christopher Knight, University of Montana,
Marcin Krygier, Adam Mickiewicz University, Poznań
Krystyna Kujawińska-Courtney, University of Łódź
Brian Lowrey, Université de Picardie Jules Verne, Amiens
Zbigniew Mazur, Maria Curie-Skłodowska University, Lublin
Rafał Molencki, University of Silesia, Sosnowiec
John G. Newman, University of Texas Rio Grande Valley
Jerzy Rubach, University of Iowa/University of Warsaw
Piotr Ruszkiewicz, Pedagogical University, Cracow
Krystyna Stamirowska, Jagiellonian University, Cracow
Merja Stenroos, University of Stavanger
Jeremy Tambling, SWPS University, Warsaw
Peter de Voogd, University of Utrecht
Anna Walczuk, Jagiellonian University, Cracow
Jean Ward, University of Gdańsk
Jerzy Wełna, University of Warsaw
Florian Zappe, independent scholar

GUEST REVIEWERS
Csaba Csides, Károli Gáspár University
Bartłomiej Czaplicki, University of Warsaw
Stephen Nichols, University of Edinburgh
Marc van Oostendorp, Radboud Universiteit
Krisztina Polgárdi, Hungarian Academy of Sciences
Geoffrey Schwartz, Adam Mickiewicz University
Patrycja Strycharczuk, University of Manchester
Péter Szigetvári, Eötvös Loránd University
Mateusz Urban, Jagiellonian University
Islam Youssef, University of South-Eastern Norway
Markéta Ziková, Masaryk University

UNIVERSITY OF WARSAW | INSTITUTE OF ENGLISH STUDIES UNIVERSITY OF WARSAW

ANGLICA
An International Journal of English Studies

COVER DESIGN
Anna Gogolewska

COMMISSIONING EDITOR
Anna Kołtunowska

PROOFREADING
Monika Mazowiecka

ISSN 0860-5734 (Print)
ISSN 2957-0905 (Online)
DOI: 10.7311/Anglica/35.2

www.anglica-journal.com

License:
Copyright © by Authors, 2026
Publication under the Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0) license
(full text available at: https://creativecommons.org/licenses/by-sa/3.0/)

Publisher:

University of Warsaw Press
02-678 Warsaw, 5/7 Smyczkowa Str.
wuw@uw.edu.pl
Sales Department: tel. (48 22) 55-31-333
e-mail: dz.handlowy@uw.edu.pl
Online bookstore: www.wuw.pl

Typesetting:
Beata Stelęgowska

Print: POZKAL

TABLE OF CONTENTS

Artur Kijak
From Generative to Laboratory Phonology: The Phonetics–Phonology Overlap 5

Markus A. Pöchtrager
Tales of the Unexpected (Phonotactic Dependencies) 15

Kuniya Nasukawa
The Representation of *s*C Clusters in English: A PfP Perspective 37

Sławomir Zdziebko
A Substance-free Autosegmental Account of the Coronal Mutation in English 59

Bert Botma, Hamza Khwaja and Joseph Smith
A Stratal Approach to English Affricates 87

Silke Hamann and Radosław Święciński
How Language-specific Perceptual Cues Can Explain Differences in Onset Clusters: The Case of /tr/ in Polish and English 105

Karolina Broś
Lip Tracking and Acoustics at the Phonetics–Phonology Interface: Coronal Stop Deletion in American English 129

Katalin Balogné Bérces and Erika Sajtós
Correlations and Functional Explanations in Laryngeal Phonetics and Phonology: Evidence from New Zealand English 167

Vencel Wutka
Rhapsodic Revisions: Voiceless Obstruent Clusters in English 193

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Artur Kijak, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.01

Artur Kijak
https://orcid.org/0000-0002-5048-9022
University of Silesia in Katowice

From Generative to Laboratory Phonology: The Phonetics–Phonology Overlap

Abstract: This introductory chapter situates the present volume within ongoing debates on the relationship between phonetics and phonology. It outlines the historical division between the two domains and highlights contemporary disagreements concerning their respective explanatory roles, ranging from strongly modular approaches to more integrative frameworks such as Laboratory Phonology. Drawing on recent scholarship, the chapter emphasizes both the persistence of theoretical tension and the growing interest in bridging the phonetics–phonology divide. Against this background, the volume brings together contributions that explore sound structure from both theoretical and empirical perspectives, primarily with reference to English. Taken together, the studies reflect the plurality of current approaches and confirm that the phonetics–phonology interface remains a central area of inquiry in contemporary linguistics.

Keywords: phonetics–phonology interface, Laboratory Phonology, generative phonology, English phonology

The study of speech sounds has long occupied a central position in linguistic inquiry, yet it has rarely constituted a unified domain. For much of the twentieth century, research into sound structure was divided into two complementary but conceptually distinct areas: phonetics and phonology. While both fields aim to explain how languages encode and organize sound, they have traditionally differed in their assumptions and analytical focus. Phonology has been associated with abstract representation and categorical patterning, whereas phonetics has concentrated on the

physical realization of speech across articulatory, acoustic and perceptual dimensions. This division, rooted in structuralist thought, has proven both productive and problematic, generating theoretical advances while raising persistent questions about the boundary between the two domains.

In recent decades, this boundary has come under renewed scrutiny. Advances in theory and experimental methodology have not produced a unified view of the phonetics–phonology relationship but have instead highlighted the complexity of their interaction. Sound patterns are commonly understood as involving both abstract structure and gradient phonetic phenomena. Although the two domains rely on different tools and assumptions, they often address similar empirical issues, giving rise to ongoing disagreements about their respective explanatory roles. The perspective adopted here is informed by recent discussions of this evolving relationship, particularly Cyran (2025) and Żygis (2025), which trace developments from generative approaches to Laboratory Phonology.

These disagreements are reflected in competing approaches to sound structure. Some frameworks prioritize phonological computation within a formal system, limiting the role of phonetics to a narrow range of effects. Others emphasize the explanatory potential of phonetic factors, arguing that patterns traditionally treated as phonological may instead follow from physical and perceptual constraints. Rather than converging on a single perspective, the two domains often advance partially competing accounts of the same phenomena.

At the same time, efforts to bridge this divide have become increasingly prominent. In response to the limitations of strictly separational or reductionist views, a number of approaches seek either to integrate the two domains or to define their interaction more precisely. Laboratory Phonology has been particularly influential in this respect, combining formal analysis with experimental methods to test phonological claims. While such approaches do not eliminate theoretical tensions, they shift attention toward the interface between abstract structure and physical realization.

The growing visibility of these issues is reflected not only in theoretical debate but also in the organization of scholarly exchange. The prominence of integrative perspectives can be observed in conferences, thematic panels and edited volumes (e.g. Cyran and Szpyra-Kozłowska 2014) that explicitly foreground the interaction between phonetics and phonology. Such developments underscore a broader tendency to treat the interface as a central object of investigation.

Despite these developments, the precise nature of the phonetics–phonology relationship remains unresolved. Some approaches maintain a clear division between levels of analysis, treating phonology as largely autonomous, while others argue for a more direct role of phonetic substance in shaping phonological structure. Between these positions lies a range of intermediate proposals that acknowledge both abstraction and phonetic detail (e.g. Laboratory Phonology).

The present volume is situated within this landscape. Its primary aim is to explore the phonetics–phonology interface through contributions reflecting a wide spectrum of current research. Although English constitutes the primary empirical domain, consistent with the profile of the journal, the volume incorporates a range of varieties and, in selected cases, cross-linguistic comparison, most notably with Polish. This combination allows the volume to remain within its empirical scope while contributing to broader theoretical discussions.

The contributions collected in this volume illustrate the multiplicity of approaches that characterize contemporary research on speech sounds. Some chapters are primarily theoretical, advancing or refining models of phonological representation, while others are grounded in empirical investigation, employing experimental methods to examine specific phenomena. In many cases, however, this distinction is not clear-cut, as formal analysis and experimental evidence are combined within a single study. Across these contributions, several recurring themes emerge, notably the representation of segments or the explanation of some exceptional sound patterns and the role of phonetic detail in shaping phonological structure. These issues are central to current debates on the status of phonological categories and the mechanisms underlying their distribution and realization.

More broadly, the volume engages with the question of how linguistic theory should balance empirical coverage and explanatory power. While models must account for the variability of observed data, they must also provide principled explanations. This tension is particularly evident at the phonetics–phonology interface, where approaches differ in the extent to which they privilege abstraction or phonetic grounding.

The contributions also reflect a commitment to methodological diversity. The study of sound structure draws on multiple types of evidence, including introspective judgments, corpus data and experimental results. Their combination allows for more robust and nuanced analyses. Against this background, the volume brings together a set of studies that differ in their theoretical assumptions, methodological orientations and empirical focus, yet share a concern with the organization of speech sounds. This diversity reflects the current state of the field, in which competing models coexist alongside attempts at integration.

A useful way to approach the volume is to distinguish, albeit loosely, between contributions that are primarily theoretical and those that incorporate a stronger empirical component, while recognizing that many papers occupy an intermediate position. The former group includes studies that refine phonological frameworks, such as those by Nasukawa, Pöchtrager, Zdziebko and Botma, Khwaja and Smith, which refine phonological frameworks and address issues of representation and formal structure within different theoretical traditions. The latter group places greater emphasis on experimental evidence, as in the work by Hamann and Święciński, Broś, Balogné Bérces and Sajtós and Wutka, where phonetic data play a central role in shaping phonological interpretation.

A number of papers also address issues of segmental complexity, revisiting the status of structures such as affricates (Botma et al.), *s*C clusters (Nasukawa) and super-heavy rhymes (Pöchtrager). These analyses propose alternative representations that challenge traditional distinctions between segments and clusters, exemplifying a broader move toward more abstract accounts of phonological structure.

Within this broader landscape, individual contributions address specific empirical and theoretical problems. **Markus A. Pöchtrager** investigates a previously underexplored phonotactic dependency in Southern British English, focusing on superheavy rhymes composed of a branching nucleus followed by a nasal–stop cluster. Drawing on a restricted but revealing dataset, the author identifies a systematic correlation between vowel quality and the fortis/lenis nature of the final stop, yielding three generalizations: vowels (diphthongs and long monophthongs) lacking the element |A| in the head position co-occur only with lenis stops, vowels combining |A| with other elements select fortis stops and vowels containing only |A| permit both. This pattern, largely unnoticed in earlier literature, presents a challenge to standard assumptions in Government Phonology (GP), as it satisfies the Minimality Hypothesis (Kaye 1995) but appears opaque under the Non-Arbitrariness Principle (Kaye et al. 1990).

To resolve this tension, the author adopts the framework of GP 2.0 (Pöchtrager 2006; Kaye and Pöchtrager 2013), which reconceptualizes traditionally melodic elements (notably |ʔ|, |A| and |H|) as structural configurations. Within this approach, phonotactic dependencies emerge from interactions between skeletal positions rather than from melodic properties (elements). The analysis proposes that coronals and non-high vowels, both structurally "larger" (as containing more empty positions), facilitate superheavy configurations by providing additional positions that can be exploited by adjacent segments. This structural reinterpretation accounts for the observed distributional patterns and offers a principled explanation for the interaction between vowel quality and consonantal phonation in superheavy rhyme structures, advancing a more constrained and explanatory model of English phonotactics.

Kuniya Nasukawa offers a theoretically informed reanalysis of English word-initial *s*C clusters within the model of Precedence-free Phonology (PfP) (Nasukawa 2017; Backley and Nasukawa 2020). The study situates itself against the backdrop of Government Phonology (GP) (Kaye et al. 1985; 1990), where *s*C sequences are traditionally treated as heterosyllabic structures involving empty nuclei and syllabic constituency (Kaye 1992). While acknowledging the descriptive success of GP, the author critically evaluates its reliance on abstract licensing mechanisms and the lack of direct empirical support for word-initial empty nuclei in English.

In contrast, the paper advances the proposal that *s*C clusters are best analyzed as complex singleton segments formed through hierarchical dependency relations

among phonological elements. By eliminating both prosodic categories and linear precedence, PfP derives surface ordering effects from structural asymmetries alone. This approach accounts for the exceptional phonotactic behavior of *s*C sequences, including their resistance to sonority constraints and their distinct distributional patterns relative to both branching onsets and word-final clusters.

More broadly, the article argues for the explanatory and theoretical economy of a precedence-free model, proposing PfP as a viable alternative framework capable of capturing core generalizations in English phonology while aligning more closely with general principles of grammatical (syntactic) computation.

Sławomir Zdziebko, in turn, develops a substance-free autosegmental analysis of coronal mutation in English, focusing on alternations affecting coronal obstruents before the suffixes *-ion*, *-ian* and *-ious*. The study is embedded in a restrictive model of subsegmental structure, in which phonological representations consist of strictly binary-branching nodes (Pöchtrager 2015; Schwartz 2010; Nasukawa 2017; 2018; van der Hulst 2020) and privative, emergent features, interpreted at the phonology–phonetics interface via a bottom-up procedure (Hale and Reiss 2008; Chabot 2022; Scheer 2022).

Empirically, the study surveys a range of morphophonological alternations (e.g. /t/ > /ʃ/, /s/ > /ʃ/, /d/ > /ʒ/, as well as conditioned variants, e.g. /t/ > /ʧ/ after /s/) and identifies regularities in the behavior of coronal obstruents, alongside systematic cases of resistance. The central proposal is that mutation is triggered by a floating structural node associated with the relevant affixes, which docks onto the stem-final consonant and induces structural reconfiguration.

The outcomes of mutation are determined by the interaction of this floating material with the internal structure of segments and by the Featural Complexity Parameter, which limits English consonants to maximally three features. Cases of resistance and apparent irregularity are accounted for through structural incompatibility and the last-resort operation of node epenthesis, whereby the floating node is realized on empty positions. Overall, the paper argues that coronal mutation follows from independently motivated architectural constraints on phonological representation.

Issues of segmental complexity are further explored by **Bert Botma**, **Hamza Khwaja** and **Joseph Smith**, who propose a stratal account of English affricates within Stratal Optimality Theory (Bermúdez-Otero 2012; 2018). Their paper addresses a long-standing inconsistency in the treatment of /ʧ/, which is typically analyzed as a unitary segment but exhibits cluster-like behavior in certain phonological processes, notably glottalization in British English (Cruttenden 2001).

The authors demonstrate that while /ʧ/ patterns as a single segment in linguistic performance and native speaker intuitions, its phonological behavior – particularly its susceptibility to glottal reinforcement in intervocalic contexts – aligns with that of consonant clusters. To reconcile this discrepancy, they propose that /ʧ/ has a dual status across strata: it is represented as a cluster at the stem level but undergoes coalescence to a monosegmental affricate at the word level.

This account is further supported by an analysis of gemination patterns, whereby heterosyllabic cluster structures yield geminate affricates under specific prosodic conditions. In contrast, the lenis counterpart /dʒ/ consistently behaves as a single segment across all levels of representation. The study thus argues for a stratified model of phonological representation in which apparent contradictions in segmental status follow from level-specific structural configurations.

A shift toward empirically grounded approaches is evident in the contributions that foreground phonetic evidence. **Silke Hamann** and **Radosław Święciński** examine cross-linguistic differences in the realization of /tr/ onset clusters in English and Polish, arguing that these differences are phonetic rather than phonological in nature. Rejecting accounts that attribute the divergence to distinct gestural timing or underlying representations, the authors propose that the two languages rely on different language-specific perceptual cues in cluster interpretation.

Within the framework of Bidirectional Phonetics and Phonology (Boersma 2007; 2011), the study models how auditory input is mapped onto phonological structure via cue constraints that are part of the speaker's grammatical knowledge. The key proposal is that English listeners identify plosives in clusters primarily through formant transitions in following sonorants, allowing greater gestural overlap, whereas Polish requires an audible release burst to recover place of articulation, reflecting its richer cluster inventory.

The authors formalize these differences and test their predictions against existing perceptual data, demonstrating that cross-linguistic variation in cluster realization follows from differences in perceptual grammars rather than phonological structure. The analysis thus supports a strict separation between phonology and phonetics while integrating perception into the grammatical system.

Karolina Broś investigates the status of coronal stop deletion in American English, focusing on whether the apparent absence of /t/ and /d/ in consonant clusters constitutes categorical deletion or gradient phonetic weakening. Adopting the framework of Articulatory Phonology (Browman and Goldstein 1990), the study assumes that deletion reflects reduction in gestural magnitude and increased overlap rather than complete removal of articulatory gestures.

Methodologically, the paper introduces a novel application of motion capture techniques to track lip movements as an indirect measure of lingual activity, combined with acoustic analysis of duration, intensity and harmonics-to-noise ratio. By examining sequences in which coronal stops occur between labials, the study tests whether differences in articulatory coordination and acoustic cues can distinguish between retained, weakened and apparently deleted consonants.

The results reveal systematic differences between these conditions, including gradient variation in both articulatory and acoustic parameters, even where no audible trace of /t/ or /d/ is present. These findings support the view that so-called deletion is phonetically incomplete and emerges from gestural overlap and masking effects.

The paper argues that coronal stop deletion should be reinterpreted as a gradient phenomenon at the phonetics–phonology interface, and proposes that integrated articulatory–acoustic methodologies are essential for capturing its true nature.

Two further contributions engage directly with laryngeal phenomena. **Katalin Balogné Bérces** and **Erika Sajtós** investigate the relationship between phonetic detail and phonological patterning in the domain of laryngeal contrasts in English, with particular reference to New Zealand English. Focusing on two well-known correlates of laryngeal specification – Voice Onset Time (VOT) in plosives and pre-fortis vowel shortening (the Voicing Effect) – the authors examine whether these phenomena are systematically related and how such correlations can be functionally motivated.

Empirically, the study analyzes data from Pākehā and Māori New Zealand English, two varieties that share the same vowel system, thereby controlling for phonological differences in vowel length. Acoustic measurements and statistical modeling demonstrate that correlations between VOT and vowel duration persist independently of phonological vowel quantity, suggesting a robust interaction between these phonetic parameters.

Theoretically, the article situates these findings within laryngeal realism (Honeybone 2005; Iverson and Salmons 2008; Cyran 2014) and broader discussions of phonological typology, arguing that phonetic detail reflects functional pressures shaping sound systems. The authors propose that the observed patterns are best understood as emerging from the interaction between systemic phonological structure and phonetically grounded tendencies, thereby contributing to ongoing debates on the role of phonetic substance in phonological representation.

Finally, **Vencel Wutka** presents an empirical investigation of voiceless obstruent clusters (VOCs) in English, testing the validity of Szigetvári's (2020) *FortisFortis (*FF) phonotactic constraint, which prohibits adjacent fortis obstruents within words. The study challenges the traditional four-way classification of obstruent clusters (LL, LF, FL, FF), arguing that the apparent existence of FF clusters may reflect transcriptional convention rather than phonological reality.

The analysis is based on acoustic data from selected lexical items (*absolute, rhapsody* and *knapsack*), sourced from YouGlish and examined in Praat. Multiple phonetic correlates are measured, including aspiration, segment duration, vowel ratios and voicing, in order to determine the laryngeal identity of the relevant clusters.

The results show that the cluster in *rhapsody* patterns with that in *absolute*, supporting a reinterpretation as a lenis–fortis sequence /bs/ rather than a fortis–fortis cluster /ps/. In contrast, *knapsack* behaves differently, consistent with its morphological structure. These findings provide empirical support for the *FF constraint and suggest that English phonotactics systematically avoid adjacent fortis obstruents.

The analysis adopts an empirically grounded perspective in which detailed phonetic evidence is used to inform phonological interpretation. Rather than

treating phonetic detail as incidental, the study systematically examines acoustic correlates of laryngeal contrast in obstruent clusters and demonstrates their relevance for determining underlying phonological structure.

On this basis, Vencel Wutka argues that careful phonetic analysis can reveal systematic regularities that are not captured by standard transcription practices. The article thus advances a data-driven approach in which fine-grained phonetic evidence plays a crucial role in evaluating phonotactic constraints and motivates a more restrictive account of obstruent clustering in English.

Taken together, the contributions do not converge on a single view of the phonetics–phonology relationship. Rather, they offer a spectrum of perspectives, ranging from strongly modular accounts to approaches that seek closer integration of phonetic and phonological knowledge. This plurality of viewpoints, combined with a shared concern for both empirical adequacy and theoretical explanation, defines the present collection and situates it within ongoing debates in contemporary phonology.

References

Backley, Philip, and Kuniya Nasukawa. 2020. "Recursion in Melodic-prosodic Structure." *Morpheme-internal Recursion in Phonology*. Ed. Kuniya Nasukawa. Boston & Berlin: Mouton de Gruyter. 11–35.

Bermúdez-Otero, Ricardo. 2012. "The Architecture of Grammar and the Division of Labor in Exponence." *The Morphology and Phonology of Exponence*. Ed. Jochen Trommer. Oxford: Oxford University Press. 8–83.

Bermúdez-Otero, Ricardo. 2018. "Stratal Phonology." *The Routledge Handbook of Phonological Theory*. Ed. S. J. Hannahs, and Anna R. K. Bosch. Abingdon: Routledge. 100–134.

Boersma, Paul. 2007. "Some Listener-oriented Accounts of h-aspiré in French." *Lingua* 117.12: 1989–2054. DOI: 10.1016/j.lingua.2006.11.004

Boersma, Paul. 2011. "A Programme for Bidirectional Phonology and Phonetics and Their Acquisition and Evolution." *Bidirectional Optimality Theory*. Ed. Anton Benz, and Jason Mattausch. Amsterdam & Philadelphia: John Benjamins. 33–72.

Browman, Catherine, P., and Louis Goldstein. 1990. "Tiers in Articulatory Phonology, with Some Implications for Casual Speech." *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Ed. John Kingston, and Mary E. Beckman. Cambridge: Cambridge University Press. 341–397.

Chabot, Alex. 2022. "On Substance and Substance-free Phonology: Where We Are at and Where We Are Going." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4: 429–443.

Cruttenden, Alan. 2001. *Gimson's Pronunciation of English*. London: Edward Arnold.

Cyran, Eugeniusz. 2014. *Between Phonology and Phonetics – Polish Voicing*. Boston & Berlin: De Gruyter Mouton.

Cyran, Eugeniusz. 2025. "Empirical Coverage vs. Theoretical Implications. In Search of Phonology in Speech Sound Systems." *Contemporary Linguistics: Integrating Languages, Communities, and Technologies*. Ed. David Bradley, Katarzyna Dziubalska-Kołaczyk, Camiel Hamans, Ik-Hwan Lee, and Frieda Steurs. Leiden: Brill. 81–91.

Cyran, Eugeniusz, and Jolanta Szpyra-Kozłowska, eds. 2014. *Crossing Phonetics–Phonology Lines*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise.* Oxford: Oxford University Press.

Honeybone, Patrick. 2005. "Diachronic Evidence in Segmental Phonology: The Case of Obstruent Laryngeal Specifications." *The Internal Organization of Phonological Segments*. Ed. Marc van Oostendorp, and Jeroen van de Weijer. Berlin & Boston: Mouton de Gruyter. 319–354.

Hulst, Harry van der. 2020. *Principles of Radical CV Phonology: A Theory of Segmental and Syllabic Structure*. Edinburgh: Edinburgh University Press.

Iverson, Gregory K., and Joseph C. Salmons. 2008. "Germanic Aspiration: Phonetic Enhancement and Language Contact." *Sprachwissenschaft* 33: 257–278.

Kaye, Jonathan. 1992. "Do You Believe in Magic? The Story of s+C Sequences." *SOAS Working Papers in Linguistics and Phonetics* 2: 293–313.

Kaye, Jonathan. 1995. "Derivations and Interfaces." *Frontiers of Phonology: Atoms, Structures, Derivations*. Ed. Jacques Durand, and Francis Katamba. London & New York: Longman. 289–332.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1985. "The Internal Structure of Phonological Elements: A Theory of Charm and Government." *Phonology Yearbook* 2: 305–328.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1990. "Constituent Structure and Government in Phonology." *Phonology* 7: 193–231.

Kaye, Jonathan, and Markus A. Pöchtrager. 2013. "GP 2.0." *SOAS Working Papers in Linguistics and Phonetics* 16: 51–64.

Nasukawa, Kuniya. 2017. "Extending the Application of Merge to Elements in Phonological Representations." *Journal of the Phonetic Society of Japan* 21.1: 59–70.

Nasukawa, Kuniya. 2018. "The Phonological Shape of Morphemes Constructed by Merge." *Phonological Externalization*. Ed. Hisao Tokizaki. Volume 3. Sapporo: Sapporo University. 1–8.

Pöchtrager, Markus A. 2006. "The Structure of Length." PhD diss, University of Vienna.

Pöchtrager, Markus A. 2015. “Binding in Phonology.” *Representing Structure in Phonology and Syntax*. Ed. Marc van Oostendorp, and Henk van Riemsdijk. Berlin: Mouton de Gruyter. 255–275.

Scheer, Tobias. 2022. “3 x Phonology.” *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4: 444–99. DOI:10.1017/cnj.2022.22

Schwartz, Geoffrey. 2010. “Phonology in the Speech Signal – Unifying Cue and Prosodic Licensing.” *Poznań Studies in Contemporary Linguistics* 46.4: 499−518.

Szigetvári, Péter. 2020. “Emancipating Lenes: A Reanalysis of English Obstruent Clusters.” *Acta Linguistica Academica* 67.1: 39–52. DOI:10.1556/2062.2020.00004

Żygis, Marzena. 2025. “Phonetics and Phonology Closer Than Ever Before? From Generative to Laboratory Phonology.” *Contemporary Linguistics: Integrating Languages, Communities, and Technologies*. Ed. David Bradley, Katarzyna Dziubalska-Kołaczyk, Camiel Hamans, Ik-Hwan Lee, and Frieda Steurs. Leiden: Brill. 361–375.

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Markus A. Pöchtrager, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.02

Markus A. Pöchtrager
https://orcid.org/0000-0002-8743-8165
University of Vienna

Tales of the Unexpected (Phonotactic Dependencies)

Abstract: This contribution looks at a particular type of superheavy rhyme in English: a long vowel/diphthong followed by a nasal-stop cluster. Such sequences show a complex interaction of vowel quality and phonation of the final stop; an interaction that remains obscure under current assumptions about the internal structure of vowels and the representation of phonation. I argue that Government Phonology 2.0, a model where many melodic properties are replaced by structural relations, provides the tools to understand (at least part of) the pattern.

Keywords: superheavy rhymes, coronals, Government Phonology, aperture, elements

1. Introduction

This article deals with an unexpected phonotactic dependency in Southern British English within the framework of Government Phonology (GP; Kaye et al. 1985; 1990; Kaye 1990), more precisely the flavour referred to as GP 2.0 (references throughout the text). The relevant data are given in (1), based on Pöchtrager (2006, 60). All words involve superheavy rhymes, i.e. a branching nucleus followed by a coda. (1) is limited to one subtype of superheavy rhymes, viz. those where the coda is nasal and the following (word-final) consonant is a stop. This extreme limitation has its drawbacks, as some cells are filled with only one example, but allows us to see a surprising pattern with some clarity. That stop (treated as an onset, as are all final consonants in Standard GP, see below) is either lenis [d] or fortis [t] (terms more accurate for English than "voiced/voiceless"), i.e. we find words like

fiend [fiːnd] or *paint* [peɪnt].[1] All forms are morphologically simplex, with the exception of *don't* and *won't*, which I take to involve non-analytic morphology (Kaye 1995), i.e. morphology that is not visible to phonology. As a result, all forms can be treated as single phonological domains. (Removal of *don't* and *won't* from the chart reduces the number of examples, but leaves the generalisation made below unaffected.)

(1)

	lenis	fortis (|H|)		branching nucleus	
				head	complement
a.	*fiend*	*	[iː]	|I|	|I|
b.	*wound*	*	[uː]	|U|	|U|
c.	*	*taint*, *paint*, *saint*, etc.	[eɪ]	|A|, |I|	|I|
d.	*	*don't*, *won't*, *wont*	[oʊ]	|A|, |U|	|U|
e.	*	*taunt*, *haunt*, *daunt*, etc.	[ɔː]	|A|, |U|	|A|, |U|
f.	*command*, *demand*, *remand* etc.	*aunt*, *chant*, *grant* etc.	[ɑː]	|A|	|A|
g.	*round*, *bound*, *sound* etc.	*count*, *mount*, *amount* etc.	[aʊ]	|A|	|U|
h.	*kind*, *find*, *mind* etc.	*pint*	[aɪ]	|A|	|I|
i.	*	*joint*, *point*, *disappoint* etc.	[oɪ]	|A|, |U|	|I|

The chart is organised by vowel quality and phonation of the final stop. (1g–i) contain diphthongs whose two members share no melody, like [aʊ], while (1a–f) give words with monophthongs, or diphthongs where some melody is shared between the two members, like [eɪ]. The chart lists the elements that many practitioners of GP (e.g. Backley 2011) would employ to characterise (i) the fortis/lenis contrast (presence/absence of |H|) and (ii) the internal composition of vowels in

terms of elements. Here I abstract away from further details about that internal composition, in particular the question which element is head, as those issues will not be relevant. What is important is what kind of melody is attached to the head of the branching nucleus, which is why the two halves of the branching nucleus (head and complement) are listed separately. (Listing the same melody for both halves in the case of monophthongs, e.g. |I| for the head and another |I| for the complement, does not imply that there are two independent instances of that element; |I| might well be shared between the two positions. Since that question is again irrelevant, I choose to be cavalier about it.)

The chart in (1) reveals a strikingly regular interdependency between the quality of the vowel and whether the following nasal-stop cluster ends in lenis [d] (no |H|) or fortis [t] (containing |H|). That is, while there are morphologically simplex words like *taint* [teɪnt] or *paint* [peɪnt], there are none like **taind* [teɪnd] or **paind* [peɪnd]. (That sequence does arise in morphologically complex forms of course, *pain-ed*, etc.) Likewise, we find *fiend* [fiːnd] or *wound* [wuːnd], but no **fient* [fiːnt] or **wount* [wuːnt]. By focusing on the head of the branching nucleus, the generalisations in (2) can be extracted.

(2)
a. Head contains no |A| → only lenis [d] (no |H|), see (1a–b).
b. Head with |A| plus another element → only fortis [t] (containing |H|), see (1c–e, i).
c. Head with |A| by itself → both lenis [d] or fortis [t] (with or without |H|), see (1f–h).

Words with superheavy rhymes have been discussed extensively in the literature (see Section 4), but to the best of my knowledge, the pattern in (2) has mostly gone unnoticed so far. Not even Fudge (1969), in his rather extensive list of all kinds of interdependencies between various positions in the English syllable, lists them.[2] The pattern is mentioned in Pöchtrager (2006, 60), where it is presented as an indication that the element |A|, playing a crucial role in (2), should be understood as structural, but the pattern itself is never analysed in any detail. In Pöchtrager (2010; 2021b), this idea of |A| being structural, in a sense to be made precise in Section 4, is fleshed out some more, but again without addressing the exact set of generalisations in (2). The present contribution is meant to fill that gap. In order to do that, we first need to look at some meta-theoretical assumptions within GP in Section 2, and why the data in (1) are unexpected. Sections 3 and 4 present some reasons why GP 2.0 has reinterpreted various elements as structural configurations, and how this can help us understand superheavy structures in general, which are the focus of Section 4. In Section 5, I will look at some simple phonological structures in English from the perspective of GP 2.0, and from there move into the analysis of superheavy rhymes. Section 6 concludes.

2. Meta-theoretical assumptions

In GP, truly phonological phenomena (processes and distributional restrictions) have to conform to both the Non-Arbitrariness Principle (NAP, Kaye et al. 1990, 194f) and the Minimality Hypothesis (Kaye 1995; Pöchtrager 2014). The NAP demands that there be a connection between the phenomenon in question and its environment: Why do things happen in exactly the place where they happen? Palatalisation of a consonant preceding a front vowel is an illustration in point, since both frontness of a vowel and palatality of a consonant are expressed by the same object (the element |I|), which in this case is simply shared between vowel and consonant. Similarly, processes of final devoicing are typically linked to the final empty nucleus that follows (and therefore licenses) the alleged final consonant. (Audibly final, but not truly so with respect to the phonological representation.) Being empty and unrealised, that final empty nucleus is assumed to possess weaker licensing powers for its onset, which leads to the loss of elements responsible for voicing from that onset. Obviously, numerous questions of detail remain in order to fully satisfy the NAP in both examples: What is the relationship holding between the nucleus and onset and why does it lead to the sharing of an element, and of that particular element? What is it about phonation that makes it harder for an empty nucleus to license it in an onset, as opposed to, say, place? etc. Those are important issues that deserve to be treated with care.

Furthermore, not just any connection one can conceivably construe will do. Ideally, the link should be independently testable and allow for the explanation of unrelated phenomena. For example, taking final consonants as onsets (Kaye 1990), which are in turn followed by empty nuclei, also explains why the set of final clusters in English matches that of intervocalic coda-onset clusters nearly perfectly (cf. *winter* and *hint*): Those final clusters are coda-onset clusters, only that the nucleus following the onset is not realised (*hint*Ø). It can be made visible, though, thereby explaining yet another phenomenon, viz. epenthesis in *hint-ed*. By making a theoretical construct responsible for the explanation of several phenomena simultaneously ("triangulation"), circularity can be avoided.

Conformity to the NAP is a necessary, but not a sufficient condition for a phenomenon to count as truly phonological. GP also requires that phonological events be exceptionless, as stated in the so-called Minimality Hypothesis (MH, Kaye 1995). The two principles are assumed to operate in tandem: Truly phonological phenomena will conform to both of them, non-phonological ones to neither. Together, the NAP and the MH put considerable restrictions on what is to be counted as phonological.

The importance of those criteria for the empirical content of the theory cannot be overstated: They allow the theory to move away from simple rewrite rules with no connection between target and trigger (the variables in $A \rightarrow B \,/\, C __ D$ can stand in for anything and there is no necessary connection between any of them),

to a much more constrained set of possible phonological events. Thus, it is of crucial importance to consider with care those cases that seem to run counter to the NAP or the MH, as they can be indicative of where the theory is in need of improvement. For example, if there is a phonological pattern that is exceptionless (i.e. meets the MH) but not amenable to the NAP (i.e. where the connection between phenomenon and context remains obscure), then either there is a problem with the phenomenon itself (not all data are correctly understood) or with the theory (since it cannot express in a non-arbitrary fashion what seems to be a bona fide phonological pattern).

I take the data in (1), along with the generalisations in (2), to provide a case in point, since the interdependency between vowel quality and phonation of the final consonant is completely opaque from the point of view of the NAP: Why should the melodic composition of vowels matter for whether the final consonant is fortis or lenis? Yet at the same time, the pattern is exceptionless (meets the MH), i.e. the NAP and the MH are in conflict. I hasten to add that this conclusion might be rash, since the data set is seriously limited. If this were all there was, I might indeed be chasing ghosts.

However, there are other indications that we need to rethink how the relevant ingredients (|A|, the fortis/lenis contrast) are commonly conceptualised. This is where GP 2.0 (Pöchtrager 2006; 2016; 2018; 2020; 2021a; 2021b; 2023; 2024; Kaye and Pöchtrager 2013; Živanović and Pöchtrager 2010) comes in. GP 2.0 holds that several properties commonly understood as melodic (expressed by elements) should be reinterpreted in structural terms (by the number of skeletal positions and the relations they contract). This is comparable to the (by now long established) replacement of the paradigmatic feature [±long] by a syntagmatic representation of length in terms of positions (on which see Fox 2000, ch. 2). The elements that have eventually been replaced by structure in GP 2.0 are |ʔ| (stop), |H| (voicelessness/high tone), and |A| (openness/coronality).[3] This will be the subject matter of the next two sections.

3. From melody to structure

The individual motivations for replacing |ʔ|, |H|, and |A| were somewhat different, but all shared the observation that those (former) elements interacted in unexpected ways with structure. For |ʔ| characterising stops, Jensen (1994) had argued that occlusion was an automatic side effect of a governing relationship contracted between an onset and a preceding coda. This was motivated by languages like Pulaar, where continuants automatically become non-continuants when forced to govern another position (e.g. in a geminate). Instead of saying that governing onsets simply "get assigned" |ʔ| (out of nowhere), Jensen proposed that occlusion *is* that governing relationship. That particular reinterpretation of |ʔ| as a relation between

consonantal positions also explains why that element seems to be barred from vowels. However, since not every stop is an onset and not every (governing) onset is a stop, Jensen's proposal was at odds with the then current theory of constituent structure. GP 2.0 integrated his idea that occlusion is mediated by a relationship between positions rather than by an element, but decoupled it from codas and onsets. An example of this will be given in Section 5.1.

Similarly, the former element |H| of GP, which expressed (amongst other things) voicelessness, has been given a structural reinterpretation in GP 2.0, starting with Pöchtrager (2006), which presented a detailed comparison of length in English and Estonian – two systems that have more in common than first impression might suggest. Part of the reasoning for the replacement of |H| can be illustrated with English pairs like *bid*/*bit* or *bead*/*beat*, which show a difference in the final consonant (lenis/fortis) that interacts with the length of the preceding vowel. (This phenomenon is known as pre-fortis clipping in the literature, and not all of its details are relevant here; the reader is referred to Pöchtrager (2006).) A more accurate transcription of those words would thus be [bɪːd]/[bɪt] and [biːːd]/[biːt] (still not ideal, as shown in a moment). That a longer vowel should precede a lenis consonant and a shorter vowel a fortis consonant remains mysterious as long as lenis/fortis is interpreted as the absence/presence of |H|: Why should vowel length (expressed structurally, by reference to the number of skeletal positions) be sensitive to a melodic property (|H|), and why to that property in particular? GP 2.0 has therefore chosen an alternative approach: It has abandoned |H| and replaced it by a structural configuration, where lenis consonants (like [d]) differ from fortis ones (like [t]) in that they contain an unused skeletal position, and this unused position can be claimed by a preceding vowel. The vowel "borrows" length from the following consonant, so to speak. Again, examples of this will be found in Section 5.1. In contrast, fortis consonants do not have such an unused position, so there is nothing for the preceding vowel to claim from the consonant. That is, *bid*/*bit* and *bead*/*beat* are reinterpreted as a trade-off in length, where in each pair it is either the vowel that is longer (extends over more positions) or the consonant. A transcription accurately reflecting this underlying phonological difference would thus be [bɪːd]/[bɪdː] for *bid*/*bit*, and [biːːd]/[biːdː] for *bead*/*beat*, deviating even more from what is considered orthodox than the transcription in the first half of this paragraph. The two properties, vowel length and the fortis/lenis contrast, can interact because both are expressed in the same "currency," viz. skeletal positions. For the sake of readability, I will keep using the terms fortis/lenis and transcriptions like [t]/[d] on the understanding that this is really a difference in the amount of skeletal positions involved. Obviously, the implementation of that idea will necessitate more complex phonological structures than assumed in other models. If lenis [d] in *bid* contains an unused position, then it will have to have more than one. (As we shall see, it has several.) The assumption that singleton consonants are linked

to one skeletal position (in my view, a leftover from earlier phonemic thinking) will have to be discarded.

This leaves us with |A|. In order to understand its quirky behaviour and why it, too, deserves to be reinterpreted structurally, we will now return to superheavy rhymes.

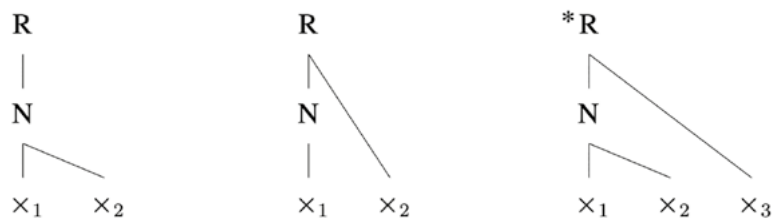
4. Superheavy rhymes

As stated before, by superheavy rhymes I am referring to a branching nucleus followed by a coda (in standard GP parlance). That coda will always be followed by an onset, as required by the principle of coda licensing (Kaye 1990). English is, in fact, quite rich in such structures, much more than, say, German (Pöchtrager 2010; 2012). In addition to the cases in (1), we also find the ones in (3); again, there is no morphological boundary contained within the string in question.[4]

(3)
a. final position: *mount*, *paste*, *east*, *field*, *bold*, *beard*, etc.
b. non-final position: *shoulder*, *mountain*, *cauldron*, *boulder*, *poultry*, etc.

As mentioned before, GP would treat *nt* as coda-onset both finally (*count*) and non-finally (*mountain*), and the preceding diphthong would require a branching nucleus. However, the Binarity Theorem of GP excludes exactly such a sequence of branching nucleus and coda-onset cluster (4c) as universally ungrammatical. There can either be a branching nucleus as in (4a) or a branching rhyme (housing the coda) as in (4b), but not both (4c).[5] Accordingly, the words in (1) and (3) should not exist.

(4)
a. branching nucleus b. branching rhyme c. branching nucleus & rhyme



For the most part, English conforms to the Binarity Theorem: The words in (5a–b) all have a branching nucleus followed by an onset, which in turn is followed by an empty nucleus (5a) or a realised one (5b), while those in (5c–d) have a branching rhyme plus onset, again followed by an empty or realised nucleus, respectively.

(5)
a. Branching nucleus, no coda (4a): *beat*, *boat*, *boot*, *bait*, *keen*, *soon*, etc.
b. Branching nucleus, no coda (4a): *meter*, *motor*, *cater*, *tuna*, *sonar*, etc.
c. Branching rhyme, coda (4b): *wimp*, *link*, *tent*, *tend*, *belt*, etc.
d. Branching rhyme, coda (4b): *tamper*, *tinker*, *winter*, *render*, *shelter*, etc.

The offending sequences in (1) and (3) all involve clusters where both members are coronal, with the coda a sonorant and the onset an obstruent.[6] The important role played by coronals in such structures is well-known in the literature, going back at least to Fudge (1969). He only considers monosyllabic words, and both consonants of the cluster belong to the same syllable for him. According to Fudge, the second consonant being coronal is the crucial factor, and that consonant is assigned to a special position, the termination. A similar strategy is followed in Selkirk (1982), where the second of two coda positions is reserved for coronals. (Again, all the words in (3a) are treated as monosyllabic.) Similar remarks can be made about the appendix (Vaux and Wolfe 2009). And also within GP, where there is no constituent 'syllable' and where constituent structure therefore looks somewhat different from other models anyway, a similar strategy has been pursued. For example, Harris (1994, 77) assumes that the structure in (4c) is licit after all, but typically only if both members of the cluster are coronal.[7] That is, coronals somehow license what is otherwise excluded.

I agree with Harris that coronals license superheavy rhymes, but that statement alone is not enough to explain why they do. In fact, all accounts mentioned above, including that by Harris, suffer from (a variant of) the same problem: Relaxing a model of constituent structure just for coronals, or creating a special position for coronals. Neither one is an explanation for why coronals have that special property in the first place. It is just another weird property on a long list of idiosyncratic quirks. Why should coronals allow for bigger structures than otherwise possible?

This is of course the aspect that GP 2.0 capitalises on. A property's consistent interaction with structure is a strong indication that that property is itself structural. As mentioned above, GP 2.0 follows other scholars in assuming that openness in vowels and coronality in consonants is characterised by the same phonological entity, only that that entity is not melodic (the element |A|), but rather structural. Its presence both (i) in vowels and (ii) in consonants leads to unexpected interactions with structure, which is atypical for a melodic prime. The words in (1) and (3) are an illustration of (ii): Superheavy structures are made possible by coronal consonants.[8] Southern British English provides further clues that openness in vowels is also conducive to super-heavy rhymes, thus illustrating (i): Words like *draft* [dɹɑːft], *task* [tɑːsk], or *clasp* [klɑːsp] violate the condition that both members of the cluster have to be coronal, because only one is. But notice how the vowel is [ɑː] in all these cases. In a theory with the old element |A|, this vowel would simply be characterised as |A| without any other elements. One can then go on to say that if only one

member of the cluster is coronal, a superheavy rhyme is still possible, provided the vowel itself contributes |A|.[9] Polgárdi (2003, 61) shows that in Hungarian, the only long vowels that are able to precede (any kind of) clusters are long *é*/*á*, both of which at least contain |A|. These are not all the Hungarian vowels containing |A|, and the distributional quirk is not limited to coda-onset clusters, but it is the case that whenever an exceptionally long structure occurs, the vowel contains |A|, which seems like an odd coincidence. But notice that in all those cases, reference to |A| leaves unclear why it should be that element in particular that makes superheavy (or other large) structures possible. Once |A| as an element is replaced by structure, this relation becomes less of a mystery.

|A| is replaced by structure as follows in GP 2.0: Expressions which (in the old theory) were thought to contain |A| are taken to contain more empty positions than those without. As a result, those structures will also be relatively big, but it is the number of empty positions that is crucial, cf. in particular Pöchtrager (2021a; 2023). That is, [ɑ] will contain more empty positions than [u], and [t] more than [p]. Since |A| is replaced by structure, as are many other elements, there will be several types of sound that contain no elements whatsoever ([ɑ], [t], [s], etc.) For superheavy structures, those empty positions are relevant in two ways: They can be claimed by adjacent segments under conditions to be specified further, and also, in those bigger structures (of a [t], say) there will be room to embed other consonants (nasals). This is how superheavy structures become possible: It is not the case that there is a privileged class of segments (like coronals) that are assigned to designated syllabic positions. Rather, those segments themselves are such that they contain within themselves, as part of their phonological identity, empty positions that are available for other segments to borrow, or positions that can provide a home for other segments.[10]

The idea that (more) open vowels and coronals are structurally bigger also makes interesting (and correct) predictions about lenition and vowel reduction, see Pöchtrager (2016; 2018; 2024). But for the purposes of this paper, we now need to see how all of this can be applied to the problem we started out with, i.e. the data set in (1).

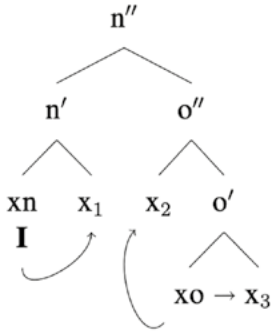
5. GP 2.0 and nasal-stop clusters

5.1. Some theoretical background

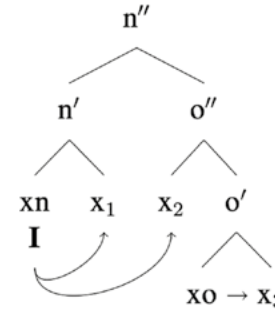
This paper cannot provide a detailed introduction to GP 2.0, the reader is referred to the literature given above. I will just look at some example representations to address the main points that are relevant for the present topic. (6) gives two simple representations, those of the relevant portions of *leak* and *league*. They focus on the replacement of |H| by a structural relationship between skeletal positions.

(6)

a. [iːk] (relevant part of *leak*)



b. [iːːg] (relevant part of *league*)



(6) employs three types of skeletal slots: Onset heads (xo), nuclear heads (xn), and non-head positions (x, numbered for ease of reference).[11] Heads may project up to twice in accordance with X-bar theory. Both words contain a long vowel: The nuclear head xn (annotation with |I| expresses frontness) has a non-head complement x_1 which it melodically commands, or "m-commands" for short. (This is not to be confused with its syntactic namesake, where "m" stands for "maximal," i.e. m-command here is not bound by the maximal projection.) M-command is indicated by curved arrows throughout; it passes on the interpretation of melodic properties (here: frontness) from m-commander to m-commandee; xn and x_1 together make up the long vowel, fully contained within n′. This n′ combines with the final consonant o″. Its head xo bears no annotation (i.e. there are no elements), since in both cases the final consonant is velar, cf. also Pöchtrager (2006, 83f) for discussion of velars. The (straight) arrow between xO and x_3 represents control and is the formal expression of stopness, i.e. the successor of the old |ʔ| element. The specifier position x_2 plays a crucial role in expressing the difference between fortis [k] and lenis [g]. In the former case, xo m-commands x_2, unlike in the latter. Since x_2 in lenis [g] is unclaimed by xo, the nuclear head xn claims (m-commands) it. This gives rise to a vowel that is even longer before a final lenis consonant.

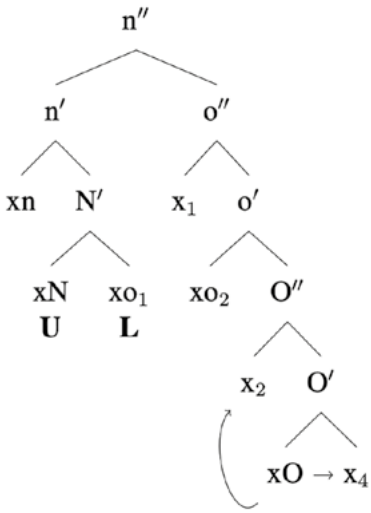
Since GP 2.0 has also replaced |A| by structure, representations where the old theory would have employed that element will look fairly complex. (7) gives the relevant parts of *font* ([ɒnt]) and *bond* ([ɒnːd]). Each word contains a non-high vowel (earlier expressed by |A|) and ends in a coronal cluster (coronality earlier expressed by |A|).

Here, both the vowel and the consonant are more complex than in the previous example. Most notably, we see a second set of heads, xN and xO. GP 2.0 assumes that a vowel will consist of maximally two nuclear heads, xn and xN, and a consonant of maximally two non-nuclear heads, xo and xO. The heads xn and xo can (but do not have to) select a projection of xN and xO, respectively, thus effectively

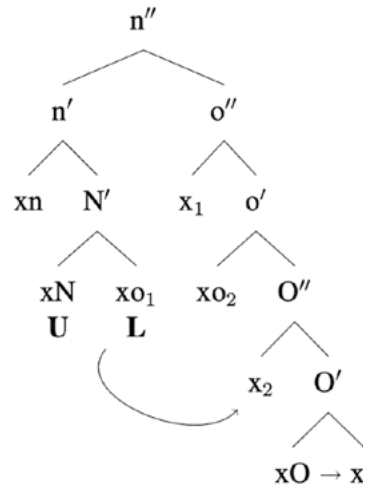
embedding the projection of xN and xO inside that of xn and xo. (7) illustrates that for both nuclear and non-nuclear heads.

(7)

a. [ɒnt]



b. [ɒnːd]



Now, why would we need such a second set of heads? If a head can project maximally twice, then more than one head will be needed to incorporate the idea that |A| is replaced by a more empty structure. Firstly, let us look at vowels and briefly go back to (6), where we are dealing with a high vowel. Its head xn projects to its maximum. Any vowel more open than a high vowel cannot therefore be expressed by a single head. Going back to (7a–b), each with the vowel [ɒ], makes clear that embedding (the projection of) another nuclear head is the only way to express differences in vowel aperture structurally, since xn has already reached its maximum in all those examples. Thus, xn has as its complement N′, a projection of xN. The sister of the lower nuclear head xN is a simple xo_1 which does not project itself, so it is both a head and a complete projection by itself. I assume xo_1 with its annotation (the nasal element |L|) to be the representation of the nasal [n].[12] Secondly, the internal structure of the final coronal stop is fairly complex, since we have two non-nuclear heads, xo_2 and xO, each with a complement and a specifier. Other, non-coronal stops would not have this additional head xO and its projection. I will assume that the specifier of the lowest onset head normally serves to express the fortis/lenis contrast, i.e. in this case x_2.[13] That is, if x_2 is m-commanded by xO, the final consonant will be fortis as in (7a). Otherwise, it will be lenis as in (7b). As before in (6), if x_2 is not m-commanded by its head xO, it can be m-commanded by someone else, in this case the nasal [n] represented by xo_1.

This provides most of the theoretical background that is needed in order to understand superheavy rhymes, to which I return now.

5.2. Confronting the mystery

In (8), the generalisations from (2) dealing with superheavy structures ending in nasal plus stop are repeated.

(8)
a. Head contains no |A| → only lenis [d] (no |H|).
b. Head with |A| plus another element → only fortis [t] (containing |H|).
c. Head with |A| by itself → both lenis [d] or fortis [t] (with or without |H|).

Both relevant properties (the presence and role of |A| and |H|, respectively) figuring in these generalisations have been replaced by structure in GP 2.0, i.e. by skeletal positions and the relationships they contract. It is a structure that will serve as the common currency needed in order to make inroads into understanding the pattern. (9) paraphrases the statements in (8) in a way that will allow us to start proposing actual representations. While (9a) can be accounted for rather easily, (9b) will turn out to be more thorny.

(9)
a. A final fortis consonant requires |A| in the nuclear head.
b. Melody in a nuclear head in addition to |A| makes a final lenis consonant impossible.

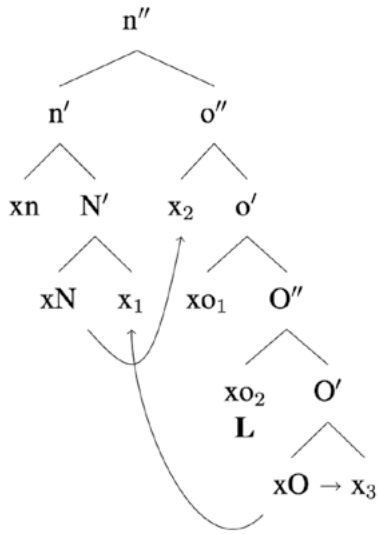
Consider (9a): Normally, a fortis stop differs from its lenis counterpart in whether (or not) the (lowest) onset head m-commands its specifier. But assume that this cannot happen in *aunt* or *paint*, or in fact any of the words in (1) ending in a fortis [t], because here the nasal sits in exactly that specifier position and thus blocks any m-command relationship. If that is correct, then how could there ever be words like *aunt* or *paint*, which clearly *do* have a final fortis stop? Which position does the onset head m-command in that case in order to guarantee fortis [t]? I submit that it m-commands one of the positions contained in the vowel, specifically a position contained in the projection of the lower head of the vowel. This is why there is an interaction between the final stop (whether it be fortis or lenis) and vowel quality, a relationship that is otherwise rather unexpected. (10) gives the structures I propose for [ɑːnt] (as in *aunt*) and [ɑːːnd] (as in *demand*), respectively, to illustrate this. The difference in vowel length is of course due to whether the final consonant is fortis or lenis.

Let us first focus on parallels between (10a–b). Both structures are identical in terms of number, type, and arrangement of skeletal positions. The root of the tree is n″, a projection of the nuclear head xn. Within the projection of xn, two other

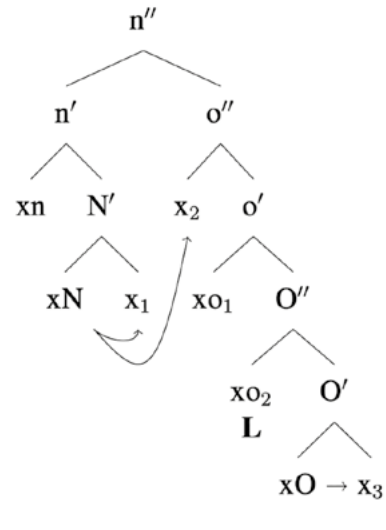
projections are embedded: N′ as well as o″. The former is very simple in terms of the internal structure, the latter is more complex. The node o″ is a projection of xo_1, and it contains yet another projection, that of xO, which in turn contains another head, xo_2, in the specifier position of xO. That head xo_2 with its annotation |L| represents the nasal, as before.

(10)

a. [ɑːnt]



b. [ɑːːnd]



The two structures differ in which heads claim which non-head positions. (10a) gives [ɑːnt]. As stated before, xO cannot m-command its specifier, as that would require an empty, unannoted non-head position, but instead we find the nasal, an annotated head. The only chance of representing a fortis consonant is then by making use of one of the positions within the vowel, viz. x_1, thus making the final fortis consonant contingent on vowel quality (openness, more specifically). Now, why would xO m-commanding x_1 be the only option? There is another potential candidate, viz. x_2, the specifier of xo_1. The preference of x_1 over x_2 cannot be an issue of the distance between the positions involved, since tracing the edges of the tree makes clear that x_1 is further removed from xO than x_2 is. Instead, what I will assume is that there is a difference between lower heads (xN, xO) and higher heads (xo, xn), cf. Pöchtrager (2025) for the same partition. Furthermore, I propose that a lower onset head can only borrow from the projection of another lower head, but not from a higher head. This bars an m-command relationship between xO and x2 (going from a lower head to the specifier of a higher head). I will call this restriction "lower-to-lower," which at this point is a stipulation, an attempt to limit the extension of m-command and therefore to restrict the power of the theory. However, in the discussion of (11) below, we shall see that it also helps us to understand another

part of the puzzle. As (10a) shows, lower-to-lower does not apply to nuclear heads, since xN (a lower head) is free to m-command x_2 (the specifier of a higher head, xo1). However, xN can only m-command x_2 but not x_1, since the latter position is needed for the final fortis consonant. In total, the vowel is long, not overlong.[14]

Matters are different in (10b), where the final consonant is lenis. Accordingly, xO does not have to "borrow" a position from the vowel and m-command x_1. The nuclear head xN can m-command both x_1 and x_2, giving rise to an overlong vowel.

The structures in (10) explain several things at the same time: The specifier of the lowest onset head normally serves to express the fortis/lenis contrast. Coronal stops, unlike other stops, possess a lower head xO in addition to the higher head xo. The specifier position of xO can express a fortis/lenis contrast, but can also be used to embed the nasal. Since non-coronals do not possess xO, they will not be able to embed another consonant inside them. Only coronals can, which is why they are repeatedly involved in superheavy structures; that way they free up positions within the projection of the nuclear heads, so to speak. But if the specifier of xO is occupied by a nasal, it cannot be used to express a fortis stop. That is, the location of [n] inside the final stop not only explains how superheavy structures can exist at all and how they depend on coronality, it also explains why the final stop needs to borrow from the nucleus in order to be fortis, and why a final fortis consonant in such structures depends on the openness of the vowel in particular.

It is worth stressing that this ability of a coronal consonant to embed another consonant does not imply the obligation to embed: The superheavy structures *aunt* and *demand* with their long vowels are well-formed, just like (non-superheavy) *rant* or *land* with short vowels are. In the former two words, the nasal will be embedded inside the final stop. In the latter two, it will not be embedded in the final stop, but rather as a sister to the lowest nuclear head, parallel to the representations of *font* and *bond* in (7). Furthermore, I do assume that there is some asymmetry here in the sense that the former option (embedding a consonant inside another consonant) is universally more marked than the latter. I will return to this issue in the conclusions.[15]

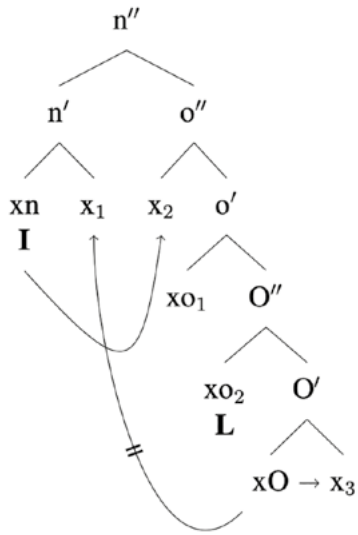
A final comment on the difference between *font*/*bond* in (7) and *aunt*/*demand* in (10): in the latter case, the vowel could borrow space from the consonant, unlike in the former. This must be linked to the position of the nasal, in the sense that lodging the nasal inside the final stop clears the path for the nuclear head to claim the higher specifier of the consonant. In (7), in contrast, the nasal as the sister of the nuclear head blocks such a relationship.

With this, we can move on to the next case. In the course of the discussion, I introduced the notion lower-to-lower, the idea that a lower onset head can only borrow from the projection of a lower head. This allowed us to make the final consonant in *aunt*/*demand* dependent on the vowel. The very same principle immediately explains why [fiːːnd], [wuːːnd], etc. exist, but not *[fiːnt] or *[wuːnt],

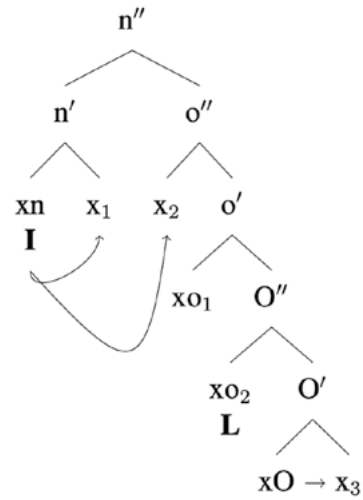
as stated in (8a). Even though there is only a single instance of each, we get the asymmetry in distribution for free. Consider (11).

(11)

a. *[iːnt]



b. [iːːnd]



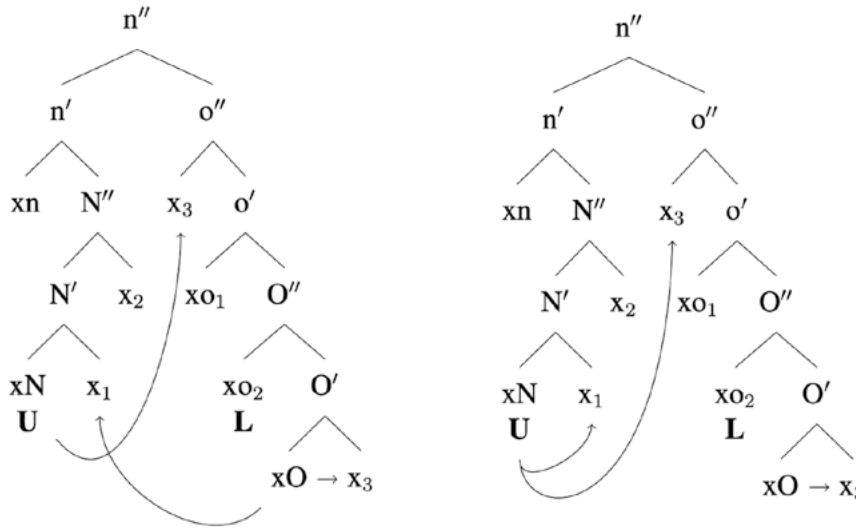
The high vowels [i] and [u] do not contain a projection of a lower head xN. A single unannoted skeletal position (x_1) serves as the complement to xn. Lower-to-lower immediately rules out the structure in (11a), since here xO (a lower head) tries to m-command x_1, which belongs to the projection of a higher head, xn. An m-command relationship from the xO to x_2 would be ruled out for the same reason. Accordingly, (11a) is not salvageable and out. (Putting the nasal xo_2 in the position where x_1 sits would save the structure, but then we are dealing with a short vowel.) In contrast, (11b) suffers from no such problem. Nuclear heads are not bound by lower-to-lower and xn can freely m-command x_1 and x_2. With the nasal lodged in the specifier of xO and the final consonant being lenis, this gives us [iːːnd].

Let us now come to the final case, characterised by (9b): words like *paint*, *haunt*, etc., all of which contain mid vowels and a final fortis consonant. (12a) gives the relevant part of *haunt*, (12b) the inexistent string *[ɒːːnd]. (Inexistent in final position monomorphemically.)

(12a) fits in perfectly with what has been said so far: The specifier of xO is taken up by the nasal and therefore cannot be m-commanded by xO. The only way for the final stop to be fortis is thus by m-commanding x_1, i.e. again by using up a position inside the vowel.[16] This is entirely parallel to *aunt* in (10a) and in accord with (9a). On the other hand, nothing in the setup of the theory predicts the ungrammaticality of (12b), i.e. the lack of words like **taund* *[tɒːːnd].

(12)

a. [ɒːnt] b. *[ɒːːnd] (unexpectedly ungrammatical)



The case of *demand* in (10b) makes clear that xN is free to m-command both its complement and the specifier of xo_1, i.e. exactly the same set of m-command relationships we see in (12b). The only difference is that the vowel [ɑːː] in *demand* does not contain any elements, while (12b) does: xN is annotated with |U|.

(13)

	lenis	fortis (|H|)		branching nucleus	
				head	complement
a.	*yield*, *field*, *shield*, etc.	*	[iː]	|I|	|I|
b.	*	*	[uː]	|U|	|U|
c.	*	*	[eɪ]	|A|, |I|	|I|
d.	*bold*, *hold*, etc.	*bolt*, *revolt*, etc.	[oʊ]	|A|, |U|	|U|
e.	*bald*, *scald*, etc.	*halt*, *fault*, etc.	[ɔː]	|A|, |U|	|A|, |U|
f.	*	*	[ɑː]	|A|	|A|
g.	*	*	[aʊ]	|A|	|U|
h.	*wild*, *mild*, etc.	* *	[aɪ]	|A|	|I|
i.	*	?*spoilt*, ?*boilt*	[ɔɪ]	|A|, |U|	|I|

How exactly melody interferes here must remain an open question for the time being; however, there are indications that the nasal plays a major role. One such indication is provided by superheavy rhymes with a liquid rather than a nasal. Thus, as a kind of promissory note, let me expand the domain of investigation a little bit and look at (13).

(13) applies the same grouping as (1): the difference is that here the final cluster involves a liquid rather than a nasal. At the same time, we find a different distribution of vowels. High vowels require |I| (and only final lenis stops are found), while the mid-vowels require |U| (and either stop finally). Low vowels, lacking any melody, do not exist here. As for the diphthongs, [aɪ] seems to pattern with |I| (due to its complement?) in going together with a final lenis stop, while oI shows the reverse and also raises the question whether *spoilt* and *boilt* as clearly morphologically complex forms should be included, though they could be if the suffix is treated as non-analytic, i.e. if the entire word forms a domain. In other words, there is a complex interaction between the liquid, the melody of the vowel, and potentially the final fortis/lenis contrast. That is, unexpected interactions with melody exist apart from the cases with nasals. Asymmetries between |I| and |U| have received some attention within GP 2.0 (Pöchtrager 2015; Živanović and Pöchtrager 2010), though none of the insights from those papers are immediately applicable to the case at hand.

6. Conclusion and open questions

One reviewer expressed unease that the proposal set out here (i) "normalises" superheavy rhymes, ridding them of the marked status they seem to merit, and (ii) that we should therefore expect to find them more often. I disagree with both worries. As for (ii): Words with superheavy rhymes have to have *some* representation; any theory deeming them ungrammatical is incomplete and at least partially wrong. But any attempt to remedy such a deficiency by providing a suitable representation will automatically mean that a language can avail itself of that particular representation. I see no way to predict that, say, the number of superheavy rhymes of the type discussed in this paper is higher in English than of that very same type in German (Pöchtrager 2012), neither under the proposal here nor under any other. The fact that closely related Danish does not seem to have such superheavy structures in morphologically simplex forms (Ruben Schachtenhaufen p.c.) leads us to (i), the fact that superheavy rhymes constitute the more marked structure that presupposes the existence of simpler ones. That is still expressed in the present proposal, since for one thing superheavy rhymes require a certain amount of empty structure that can only be provided by non-high vowels or coronals, and not just any other segment type. In addition, embedding a consonant inside another consonant and/or borrowing from certain positions is still considered a marked option and might be prohibited in a given language altogether.

The types of superheavy rhymes discussed here are but the proverbial tip of the iceberg, and one might well ask why such a narrow focus is adopted. Besides obvious reasons of space, there is justification coming from how the data break down. The various dependencies in superheavy rhymes are fairly complex and change considerably once one ingredient is exchanged for another, as comparison of nasal-initial and liquid-initial clusters at the end of the last section showed. Similarly, if instead of nasal-stop clusters one looks at combinations of nasal plus fricative or nasal plus affricate, matters change again: except for names like *Caunce*, the only long vowel we find before final [ns] is [ɑː], as in *dance*, *prance*, etc. Likewise, while there is no final monomorphemic *[eɪnd] (vs *paint*), there certainly is [eɪnd͡ʒ] (*range*, *change*). But since the final affricate contains the element |I|, the situation changes subtly.

In addition, many more subtypes exist, such as those where the first member is an obstruent, brought up in Section 4: *draft*, *task*, *clasp*, etc. Yet another group that has gone completely unmentioned so far is the one where, instead of a long vowel/diphthong and a coda-onset cluster, we find a cluster of three consonants, such as *next*, *against*, *prompt*, *distinct*, etc. Again, those can be broken down further, by whether they contain recalcitrant *s*C as a substring (the first two examples) or not (the last two). This type of breakdown is followed in Polgárdi (2022, 14), where the author treats final and word-internal cases as one and the same group. This is relevant insofar as the word-internal context allows other vowels preceding the cluster, as in *laundry* or *jaundice* with [ɒː] plus nasal plus lenis stop. I stay away from word-internal position for two reasons: Firstly, in the interest of eliminating another potentially confounding factor, since internal superheavy rhymes could always be explained away by claiming that they contain an empty nucleus: *launØdry*, *jaunØdice* (at the expense of losing any explanation for homorganicity). Though rare in English, such cases do exist (*bowdØlerise*). Secondly, GP 2.0 differs from other versions of GP in that word-internal and word-final coda-onset clusters are not treated exactly alike in terms of constituency (cf. note 2); in that sense, GP 2.0 occupies a position intermediate between classic GP and mainstream versions of the syllable (for details on the structures and the reasoning behind them, cf. Pöchtrager 2006, ch. 3).

This contribution has proceeded from the assumption that it makes sense to take a strategy of "divide and conquer" even further, given the complexities of each type: Even minimal differences in the cluster (nasal vs liquid, at the end of the last section) can have an effect on the preceding vowel. By keeping the data uniform, we stand a chance to extract generalisations and capture them through some formalism. Obviously, a complete theory of all problematic structures in all positions of a word might show that only a subset of the proposals made here are actually relevant, while others might simply be artefacts of the data being too narrowly delimitated. Such an assessment seems to me to require a more advanced theory of superheavy structures. I remain convinced, however, that even if some (or many)

generalisations fall by the wayside, the replacement of |A| by structure will remain central, since it allows us to link the contribution made by non-high vowels and coronal consonants to the existence of superheavy structures, even if there are cases that do not involve both of them (like *fiend* and also the aforementioned *chamber*, etc.). Understanding their constituent structure remains a challenge, but I remain hopeful that the present contribution is a step in the right direction.

Notes

1 The stylesheet for this article requires // for phonemic representations and [] for phonetic ones, which is of course problematic for a generative model like GP that explicitly rejects the phoneme and makes no such distinction. I will use [] throughout, since most transcriptions reflect realisation anyway. Similarly, elements are expected to be enclosed in ||, which is a practice taken over from Dependency Phonology without having been given a proper (re-)definition within GP. Naturally, I object.

2 The syllable has no formal status in most versions of GP, and the final clusters in (1) are split between coda and onset. In GP 2.0 the grouping deviates somewhat from other versions of GP, but again, there is no syllable in the traditional sense.

3 Harris and Lindsey (1995) as well as Backley (2011) provide valuable overviews of the inventory of elements and their interpretation. That |A| expresses both openness (in vowels) and coronality (in consonants) is not universally accepted, see again Backley (2011) for arguments that coronals contain |I|, possibly alongside |A|. (Even earlier versions of GP had a dedicated coronal element |R|, see Rubin (2001).) Since openness and coronality share similar distributional quirks, I side with Broadbent (1991), Cyran (1997), Goh (1997), Lee (1998), or Pöchtrager (2013) and assume that they are expressed by the same phonological entity, only that that entity is structure, not melody, see main text.

4 Whether *beard* contains a coda-onset cluster depends on the fate and interpretation of the rhotic.

5 Binarity could be circumvented by assuming that the long vowel/diphthong is really a sequence of two non-branching nuclei separated by an empty onset [aØʊ] and/or the cluster a sequence of two independent onsets separated by an empty nucleus [nØt]. In my view, that is no viable solution, since separation by an empty nucleus would leave unexplained why there are phonotactic restrictions within the long vowel/diphthong and/or the cluster. In addition, an empty nucleus in a cluster would cause problems in the final position [nØtØ] since the Empty Category Principle would require the realisation of the first one, unless special provisions are made. The latter problem does not arise in morphologically complex forms like *scream-ed* […iːmØdØ], where stem and suffix are in different domains and thus each can be followed by its own domain-final nucleus (Kaye 1995).

6 There are a few exceptions: *sample*, *example*, *chamber*, *cambric*, and names like *James*, see also note 9. It is of course quite conspicuous that they all involve *m* and a non-high vowel like [ɑː] or [eɪ].

7 "Typically" because Harris speaks of coronal being the "favoured place," i.e. the condition is not absolute, presumably to allow cases like *sample*, etc. mentioned in

note 6. Cases like *draft* [dɹɑːft] (discussed below in the main text) are also licit because he lists no additional conditions when the first consonant is an obstruent, which raises the question why *[dɹɑːfk], etc. should be out.

8 This is also true of German, Finnish, and Spanish (Pöchtrager 2012), but English has the most complex patterns.

9 I will come back to this restriction, that there cannot be (any) other elements, in Section 5.2. If a vowel is allowed to provide the relevant ingredients for a superheavy rhyme all by itself, then that might also explain words like *sample* (note 6), though here we also find [eɪ] in addition to [ɑː]. Lastly, note that in *draft*, *task*, and *clasp* the first consonant in the cluster is a fricative, the potential relevance of which is still unclear to me.

10 "Segment" is something of a misnomer in a theory where positions can be borrowed by neighbouring "segments" and phonological information is thus distributed over larger parts of a representation.

11 The term "onset" is almost tantamount to "consonant" in GP 2.0 and differs in that both from Standard GP (Kaye 1990; Harris and Gussmann 1998) and from mainstream theories of syllable structure. The traditional distinction between onset and coda is captured by position in a phonological representation.

12 In Pöchtrager (2006), it is assumed that nasals have a more complex internal structure, like any other stop. This is called into question in Pöchtrager (2010) for consonants forming (the non-head) part of a cluster, such as the nasal in this example. In order to keep the discussion simple, I follow the second line of reasoning and represent the nasal as a single position, though it might well turn out that the representation of the nasal in (7) is too simplistic.

13 This deviates from Pöchtrager (2021b), who assumes, though without argument, that it is x_1 in coronal stops which expresses the fortis/lenis contrast. Later on, I will suggest that x_1 is categorically inaccessible to xO.

14 For a similar cross-over of m-command relationships as in (10a) cf. Pöchtrager (2006, 190).

15 This ability of coronals to embed other consonants finds a parallel in non-high vowels, which is of course to be expected if both consist of more complex structures (the successor of old |A|): Diphthongs usually have a mid or open vowel as their most salient member, i.e. vowels with enough positions to allow the embedding of on- or offglides (basically, another high vowel). Note again that the ability to embed does not imply the obligation to embed: A non-high vowel does not *have* to embed another vowel (it does not have to form a diphthong), and neither does a coronal have to embed another consonant.

16 At this point, it is unclear to me whether x_2 would be accessible, or if not – why not.

References

Backley, Phillip. 2011. *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Broadbent, Judith. 1991. "Linking and Intrusive *r* in English." *UCL Working Papers in Linguistics* 3: 281–301.

Cyran, Eugeniusz. 1997. *Resonance Elements in Phonology. A Study in Munster Irish*. Lublin: Wydawnictwo Folium.

Fox, Anthony. 2000. *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford: Oxford University Press.

Fudge, Erik C. 1969. "Syllables." *Journal of Linguistics* 5: 253–286.

Goh, Yeng-Seng. 1997. *The Segmental Phonology of Beijing Mandarin*. Taipei: Crane.

Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.

Harris, John, and Edmund Gussmann. 1998. "Final Codas: Why the West Was Wrong." *Structure and Interpretation. Studies in Phonology*. Ed. Eugeniusz Cyran. Lublin: Wydawnictwo Folium. 139–162.

Harris, John, and Geoff Lindsey. 1995. "The Elements of Phonological Representation." *Frontiers of Phonology: Atoms, Structures, Derivations*. Ed. Jacques Durand, and Francis Katamba. London & New York: Longman. 34–79.

Jensen, Sean. 1994. "Is ʔ an Element? Towards a Non-segmental Phonology." *SOAS Working Papers in Linguistics & Phonetics* 4: 71–78.

Kaye, Jonathan. 1990. "'Coda' Licensing." *Phonology* 7.2: 301–330.

Kaye, Jonathan. 1995. "Derivations and Interfaces." *Frontiers of Phonology: Atoms, Structures, Derivations.* Ed. Jacques Durand, and Francis Katamba. London & New York: Longman. 289–332.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1985. "The Internal Structure of Phonological Elements: A Theory of Charm and Government." *Phonology Yearbook* 2: 303–328.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1990. "Constituent Structure and Government in Phonology." *Phonology* 7.2: 193–231.

Kaye, Jonathan, and Markus A. Pöchtrager. 2013. "GP 2.0." *SOAS Working Papers in Linguistics & Phonetics* 16: 51–64.

Lee, Duck-Young. 1998. *Korean Phonology. A Principle-based Approach*. München: Lincom.

Pöchtrager, Markus A. 2006. "The Structure of Length." PhD diss, University of Vienna.

Pöchtrager, Markus A. 2010. "The Structure of A." Paper presented at the 33rd GLOW Colloquium, Wrocław, Poland, April 13–16.

Pöchtrager, Markus A. 2012. "Deconstructing A." Paper presented at the MFM Fringe Meeting on Segmental Architecture, Manchester, UK, May 23.

Pöchtrager, Markus A. 2013. "On A." Paper presented at the Workshop on Melodic Representation, London, UK, March 12.

Pöchtrager, Markus A. 2014. "Alternations: The Vipers in Our Bosom." *Dilbilim Araştırmaları* Special Number/2: 147–164.

Pöchtrager, Markus A. 2015. "Binding in Phonology." *Representing Structure in Phonology and Syntax.* Ed. Marc van Oostendorp, and Henk van Riemsdijk. Berlin: Mouton de Gruyter. 255–275.

Pöchtrager, Markus A. 2016. “It’s All about Size.” *70 Snippets to Mark Ádám Nádasdy’s 70th Birthday.* Ed. Péter Szigetvári. http://seas3.elte.hu/nadasdy70/pochtrager.html.

Pöchtrager, Markus A. 2018. “Sawing Off the Branch You Are Sitting on.” *Acta Linguistica Academica* 65.1: 47–68.

Pöchtrager, Markus A. 2020. “Tense? (Re)Lax! A New Formalisation for a Controversial Contrast.” *Acta Linguistica Academica* 67.1: 53–71.

Pöchtrager, Markus A. 2021a. “English Vowel Structure and Stress in GP 2.0.” *Perspectives on Element Theory*. Ed. Sabrina Bendjaballah, Ali Tifrit, and Laurence Voeltzel. Berlin: Mouton de Gruyter. 157–183.

Pöchtrager, Markus A. 2021b. “Towards a Non-arbitrary Account of Affricates and Affrication.” *Glossa: A Journal of General Linguistics* 6.1: 1–31.

Pöchtrager, Markus A. 2023. “The Unbearable Lightness of Being High: Openness as Structure and the Consequences for Prosody.” *Representing Phonological Detail. Part I: Segmental Structure and Representations*. Ed. Jeroen van de Weijer. Berlin: De Gruyter. 71–89.

Pöchtrager, Markus A. 2024. “Vowel Reduction in European Portuguese and the Removal of Structure.” *Isogloss* 10.3.2: 1–22.

Pöchtrager, Markus A. 2025. “A Neo-Jakobsonian Merger of Aperture, [−ATR], Lowering, Emphaticness, and Retroflexion.” *Acta Linguistica Academica* 72.1: 20–45.

Polgárdi, Krisztina. 2003. “Hungarian as a Strict CV Language.” *The Phonological Spectrum. Volume II: Suprasegmental structure*. Ed. Jeroen van de Weijer, Vincent J. van Heuven, and Harry van der Hulst. Amsterdam & Philadelphia: John Benjamins. 59–79.

Polgárdi, Krisztina. 2022. “The Representation of *s*C-Clusters: A Strict CV Analysis.” *Glossa: A Journal of General Linguistics* 7.1: 1–29.

Rubin, Dominic. 2001. “The Licensing and Interpretation of Coronality.” PhD diss, University of London/SOAS.

Selkirk, Elisabeth O. 1982. “The Syllable.” *The Structure of Phonological Representations*. Ed. Harry van der Hulst, and Norval Smith. Volume II. Dordrecht/Cinnaminson: Foris. 337–383.

Vaux, Bert, and Andrew Wolfe. 2009. “The Appendix.” *Contemporary Views on Architecture and Representations in Phonology*. Ed. Eric Raimy, and Charles E. Cairns. Cambridge, MA: MIT Press. 101–143.

Živanović, Sašo, and Markus A. Pöchtrager. 2010. “GP 2.0 and Putonghua, too.” *Acta Linguistica Hungarica* 57.4: 357–380.

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Kuniya Nasukawa, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.03

Kuniya Nasukawa
https://orcid.org/0009-0003-0521-2731
Tohoku Gakuin University

The Representation of *s*C Sequences in English: A PfP Perspective

Abstract: This paper re-examines the phonological status of English /s/+stop sequences (/sp, st, sk/) within the framework of Precedence-free Phonology (PfP). While these sequences are traditionally analysed as consonant sequences, they exhibit phonotactic properties that distinguish them from other *s*C combinations, including their restricted distribution across word positions and their apparent violation of the Sonority Sequencing Principle. Previous analyses within Government Phonology and the Strict CV framework have offered important insights into these patterns, proposing accounts based on empty nuclei or melodic spreading. Building on these observations, the present study argues that /s/+stop sequences are best analysed as singleton segments with internally complex structure. Within PfP, phonological representations are defined in terms of hierarchical dependency relations among elements, without recourse to linear ordering. Under this view, /sp, st, sk/ are interpreted as single |ʔ|-headed domains containing multiple resonance elements, whose phonetic realisation involves the temporal staggering of acoustic cues. This analysis provides a unified account of their distributional behaviour and offers a way of relating these patterns to their phonetic properties. More broadly, the findings are consistent with a precedence-free approach to phonological representation and suggest a possible perspective on the structure of *s*C sequences in English.

Keywords: *s*C sequences, elements, asymmetric relations, English, Precedence-free Phonology

1. Introduction

This paper focuses on *s*C sequences, specifically /sp, st, sk/. The reason for focusing on these sequences is that they represent the clearest and most canonical cases of

the special phonotactic behaviour associated with *s*-initial sequences in English. As Polgárdi (2022) shows, word-initial *s* differs strikingly from other consonants in that it may combine with stops in forms such as *spoil* /spɔɪl/, *stout* /staʊt/ and *school* /skuːl/, whereas ordinary initial consonants are much more restricted in the combinations they permit. More generally, *s*-initial sequences pattern differently from canonical branching onsets (e.g. /bl, tr, kw/) and therefore require a separate structural treatment. For this reason, /sp, st, sk/ provide a particularly revealing domain for investigating the representation of *s*C sequences.

In pretonic word-initial positions, i.e. in prosodically strong environments, the laryngeal contrast among English obstruents is commonly analysed in terms of a fortis–lenis opposition. The former (fortis) is typically realised as aspirated voiceless stops, while the latter (lenis) surfaces as unaspirated voiceless stops. That is, the contrast is conventionally described as /pʰ, tʰ, kʰ/ versus /p, t, k/. Crucially, it is the latter, i.e. the unaspirated stops /p, t, k/, that appear in the consonantal position of so-called *s*C sequences.

(1) The laryngeal contrast among English stops
a. /pʰ, tʰ, kʰ/ (aspirated stops)
b. /p, t, k/ (unaspirated stops)
c. /sp, st, sk/ (traditionally analysed as sequences; e.g. *spoil* /spɔɪl/, *stout* /staʊt/, *school* /skuːl/)

While (1a) and (1b) are typically analysed as singleton segments, and (1c) as sequences of two segments, the present paper adopts a different view. Based on the phonotactic behaviour of /sp, st, sk/, it is argued that (1c) should also be analysed as singleton segments, on a par with (1a) and (1b). Under this approach, the difference between (1a) and (1c) lies not in their segmental status, but in their internal melodic complexity: (1a) contains a single place element, whereas (1c) contains multiple place elements within the framework of Precedence-free Phonology (PfP: Nasukawa 2020, among others).

The structure of this paper is as follows. Section 2 reviews the distributional behaviour of /s/+stop sequences in English and highlights the empirical patterns that motivate the present analysis. Section 3 examines the treatment of *s*C sequences in Government Phonology, focusing on the role of empty nuclei and licensing mechanisms. Section 4 then considers the Strict CV approach, particularly its account in terms of melodic spreading. Section 5 introduces the framework of Precedence-free Phonology (PfP) and develops an analysis of /s/+stop sequences as complex singleton segments. Finally, Section 6 concludes the paper by summarising the main findings and discussing their implications.

2. The distributional peculiarities of *s*+stop sequences

This section reviews the distributional behaviour of *s*+stop sequences (i.e. /sp, st, sk/) in English. These sequences exhibit a set of well-known phonotactic properties that distinguish them from other types of *s*C combinations. In particular, their distribution shows systematic asymmetries across word-initial, word-medial, and word-final positions, which are not observed with other *s*C sequences. The empirical motivation for the present proposal derives from these distributional peculiarities. The relevant patterns can be summarised as follows:

(2) Empirical patterns of /s/+stop sequences

a. Word-initial three-consonant sequences (#$C_1C_2C_3$-)
When three consonants occur word-initially, only /sp, st, sk/ may appear in the C_1C_2 position, as in *split* /splɪt/, *strike* /stɹaɪk/ and *squeeze* /skwiːz/. Other combinations are unattested.

b. Word-initial two-consonant sequences (#C_1C_2-)
Only when C_1 = /s/ do we observe violations of the Sonority Sequencing Principle – according to which, in syllable-initial position, sonority is expected to rise from less sonorous to more sonorous segments – as in /sp, st, sk/ (*spoil* /spɔɪl/, *stout* /staʊt/, *school* /skuːl/).

c. Word-final two-consonant sequences (-C_1C_2#)
In non-inflected, non-derived words, final consonant sequences are overwhelmingly restricted to /sp, st, sk/, as in *lisp* /lɪsp/, *mist* /mɪst/ and *ask* /ɑːsk/. By contrast, sequences such as /bl, tr, kw/, which are permitted word-initially (*blue* /bluː/, *tree* /triː/, *quick* /kwɪk/), do not occur word-finally.

d. Word-medial position (-C_1C_2- in non-inflected/non-derived words)
When attention is restricted to non-inflected, non-derived forms, medial *s*C sequences are also heavily skewed towards /sp, st, sk/. Although combinations such as *sm* or *sf* are attested (e.g. *Asmara* /æsˈmɑːrə/, *asphalt* /ˈæsfælt/), they are rare and lexically restricted. By contrast, /sp, st, sk/ are robustly attested (e.g. *whisper* /ˈwɪspə/, *aster* /ˈæstə/, *basket* /ˈbɑːskɪt/).

These asymmetries are difficult to capture if /sp, st, sk/ are treated as ordinary consonant sequences. In order to account for the distributional peculiarities of /s/+stop sequences, a number of phonological representations have been proposed in the literature. These may be broadly classified into five types:

(3) Five representations of /s/+stop sequences

a. a branching onset analysis, in which /s/ occupies the first onset position and the following stop the second (as assumed in most textbooks);

b. an appendix analysis, in which /s/ is adjoined at a level above the syllable, external to the onset, while the stop occupies the onset proper (Rialland 1994, among others);

c. a classical Government Phonology (GP) analysis, in which /s/ occupies the rhymal complement position following an empty nucleus, and the stop occupies the onset of the following unit (Kaye 1992, among others);
d. a Strict CV analysis, in which /s/ is associated simultaneously with both C and V positions, and the following stop is linked to the subsequent C position (Polgárdi 2022); and
e. an analysis in which /s/+stop does not constitute a sequence of two independent segments, but rather forms a singleton segment (van de Weijer 1996, among others).

Among these approaches, the present paper focuses on (3c) and (3d), presenting and critically evaluating their analyses. On the basis of this comparison, it argues in favour of the representation in (3e), namely that /s/+stop sequences are best analysed as singleton segments. Building on this conclusion, the paper then proposes a formal representation of these structures within the framework of Precedence-free Phonology (PfP).

3. The representation of *s*C sequences in Government Phonology

3.1. *s*C as a rhymal complement–onset sequence

Within GP, empty categories – and particularly, empty nuclei – have long been regarded as indispensable representational devices. They are used to capture generalisations about prosodic structure, the licensing of consonant sequences, and the alternation between the presence and absence of vowels. In GP, each segmental position in the phonological representation corresponds to a potential onset–nucleus pair. When a nucleus position lacks segmental material but is nevertheless structurally present, it is identified as an empty nucleus. Such nuclei can remain phonetically unrealised if certain licensing conditions are met, but otherwise they must surface as audible vowels.

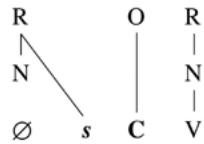
The explanatory power of this device is well illustrated by a range of alternations in English and other languages. For example, GP analyses the formation *kiss* + *-es* → *kiss*[ɪ]*s* (or *kiss*[ə]*s*) as involving the phonetic realisation of an empty nucleus in stem-final position.

The presence of this nucleus prevents the stem-final and suffix-initial consonants from surfacing as an otherwise illicit cluster. The existence of empty categories accounts for vowel-zero alternations and clarifies the conditions under which syllabic material surfaces, without invoking any arbitrary vowel insertion processes.

In GP, empty nuclei are central to explaining the distributional behaviour of *s*C sequences. In the word-initial context, GP assigns the *s* in sequences such as *spin*, *step* and *skin* to the rhymal complement position (GP does not posit a separate coda constituent), preceded by an empty nucleus which lacks phonological

features. The stop then forms the onset of the following onset–nucleus pair, as shown below in (4).

(4) *s*C sequence in GP



This structure successfully explains why *s*C sequences behave differently from genuine branching onsets (e.g. /pl, tr, kw, θr/), which consist of an obstruent followed by a sonorant – i.e. a sequence displaying rising sonority – whereas *s*C sequences may display not only rising sonority (e.g. /sm, sn, sl, sw/) but also falling sonority (e.g. /sp, st, sk/).

To substantiate this structure, Kaye (1992) draws on evidence from Italian, Ancient Greek, European Portuguese and British English. The present study, however, focuses exclusively on the representation of *s*C sequences in English and therefore limits the discussion to English data, rather than providing a comparative analysis across a wider range of languages.

In Kaye (1992), the English evidence centres on the behaviour of /s/ + consonant sequences in dialects that permit the postconsonantal glide /j/ (the "yod") after coronals, as in *tune* [tjuːn] and *lute* [ljuːt]. These dialects – particularly Southern British English and certain Canadian and Australian varieties – exhibit a crucial restriction: /j/ never occurs after a genuine branching onset, that is, one in which two consonants are jointly dominated by a single onset node (e.g. /pl, bl, kl, gl/). For instance, words like *pluvial*, *plumage*, *fluid*, *occlusion*, or *glue* cannot have a yod after the two onset consonants: **pl*[j]*uvial* or **gl*[j]*ue* are ill-formed. The distribution therefore shows that /j/ is systematically banned after branching onsets.

In contrast, /j/ freely appears after *s* + consonant sequences, as in *st*[j]*upid*, *sp*[j]*ew* and *sk*[j]*ew*. If *s*C sequences were branching onsets, they should behave like /pl/ or /kl/ and prohibit /j/; yet they do not. The most straightforward explanation is structural: /s/ + C does not form a single onset but spans two syllable domains – /s/ occupies the rhymal complement of an empty-nucleus syllable, and the following consonant serves as the onset of the next syllable. This pattern accords precisely with the GP view that *s*C sequences are heterosyllabic as previously given in (1).

Kaye (1992) argues that the glide /j/ occurs only after non-branching onsets: for example, *tune* /tuːn/ → /t**j**uːn/ and *spew* /s.puː/ → /s.p**j**uː/—but is excluded after branching onsets, as in **pl*[j]*uvial* or **gl*[j]*ue*. Since *s*C sequences permit the occurrence of /j/, they are taken to belong to the former type. The behaviour of postconsonantal /j/ in English has therefore been interpreted as providing empirical support for the GP claim that word-initial /s/ + C sequences do not constitute branching onsets.

However, this argument crucially depends on the assumption that the glide in question is part of the onset. As noted by Davis and Hammond (1995) and Polgárdi (2015; 2022), the phonotactic restrictions on /j/ with a preceding consonant are not robustly attested. Instead, /j/ appears to be restricted primarily by the following vowel, occurring only before [uː] and [ʊə]. This suggests that /j/ forms a unit with the following vowel, and may therefore be more appropriately analysed as part of the nucleus.

Even if this evidence is not decisive, *s*C sequences still display certain similarities to singleton onsets such as /t-/, /p-/ and /k-/, rather than to branching onsets such as /pl-/ and /gl-/, although the precise basis for this comparison remains open to further examination, and it is not entirely clear that *s*C sequences indeed have the structure proposed in (4). While this raises a degree of scepticism regarding the original motivation for the proposed representation, the next section turns to a separate set of issues concerning the structure in (4) within GP.

3.2. Absence of vowel-zero alternation as evidence for an empty nucleus

The empty nucleus in the GP representation of word-initial *s*C sequences remains unrealised (phonetically silent) by virtue of Magic Licensing (Kaye 1992). In GP, empty nuclei are normally required to be licensed in order to remain silent, either (i) because they occur in domain-final position or (ii) because they are properly governed by a following melodically filled nucleus. The former condition is regulated by the Domain-Final Empty Nucleus (DFEN) parameter (Harris 1994, 162; 179–191), while the latter follows from Proper Government, whereby a melodically filled nucleus properly governs its adjacent empty counterpart (Kaye 1990; Charette 1991; Harris 1994, 191–193). The sequence of an empty nucleus followed by an *s*C cluster, however, occurs in positions where neither of these conditions is satisfied. To account for this, Kaye (1992) proposes Magic Licensing as a parametric option whereby an empty nucleus preceding an *s*C sequence may remain silent without being properly governed. This mechanism allows the empty nucleus to remain phonetically unrealised. While classic GP provides motivation for positing several licensing mechanisms that allow empty nuclei to remain phonetically unrealised, it does not clarify why three distinct types of licensing are invoked for what appears to be the same phonetic outcome, nor does it consider whether analogous mechanisms operate in other components of grammar, such as morphosyntax.

A further issue concerns the empirical basis for positing a word-initial empty nucleus preceding /s/. Unlike word-final empty nuclei – whose existence is supported by vowel-zero alternations (e.g. *kiss* [kʰɪs] vs *kisses* [kʰɪs**ɪ**z] or [kʰɪs**ə**z]) – no comparable alternations in English indicate the presence of a word-initial empty nucleus, although similar alternations are found in other languages such as Portuguese (Kaye 1992; Polgárdi 2022). No such alternation is observed in *s*C sequences: forms like *spin* or *school* never alternate with [ə]*spin* or [ə]*school*.[1] The absence of

any overt reflex of the empty nucleus within English itself makes the analysis less directly supported by phonetic evidence. And this lack of overt evidence potentially weakens the GP account of *s*C sequences, since it depends on a structural element whose existence cannot be directly observed in surface alternations.

4. The representation of *s*C sequences in the Strict CV framework

4.1. A Strict CV account of *s*C sequences

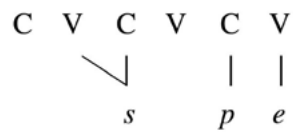
Polgárdi (2022) argues that the behaviour of *s*C sequences (e.g. /sp, st, sk/) is best explained not by traditional GP-type analyses, but within a Strict CV framework, where phonological structure consists only of alternating C and V positions. More specifically, the paper argues that *s*C sequences are not genuine onset sequences; rather, they involve the spreading of the melodic content of *s* from a C position to an adjacent V position, thereby removing the need to posit either empty nuclei or Magic Licensing.

In this analysis, the position of /s/ is parametrically determined across languages: in languages like English, the spreading of the melodic content of /s/ from a C position to the following V position, as in (5a); in Italian and Portuguese, on the other hand, the spreading of that of /s/ to the preceding V position as in (5b).

(5) Strict CV representation of word-initial *s*C-sequences

a. English

b. Italian, Portuguese

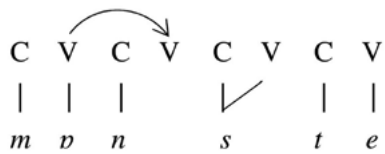


The proposal accounts for a wide range of phonological facts in a unified way. First, it explains the atypical phonotactics of *s*C sequences. Unlike ordinary branching onsets (e.g. /pl, tr, kw/), *s*C sequences permit structures that violate general sonority constraints, such as falling sonority (/sp, st, sk/) and combinations that are otherwise disallowed. Under the Strict CV analysis, this follows naturally because /s/ is not part of a true onset cluster but occupies a structurally distinct position, which removes the need to treat these sequences as exceptional.

Second, the proposal captures their behaviour in processes like vowel lengthening and gemination. In languages such as Italian, *s*C sequences pattern like coda-onset sequences rather than simple onsets: for example, they block tonic vowel lengthening and fail to trigger gemination processes that normally apply before word-initial onsets. In the Strict CV framework, this behaviour follows from the internal structure of *s*C sequences, where the presence and distribution of empty positions (properly governed) determine whether lengthening or gemination can occur.

Third, the analysis in (5a) is theoretically appealing in that it permits a uniform structural representation of English *s*C sequences, both in word-initial position and in other environments, as exemplified in (6).

(6) Strict CV representation of word-internal *s*C sequences



Finally, the analysis explains their similarities to syllabic consonants. The paper argues that the branching structure proposed for *s* is formally identical to that of syllabic sonorants: in both cases, a consonantal melody is associated with a V position. Moreover, their distributional restrictions are parallel, in that such "syllabic" structures are only permitted under specific structural conditions (e.g. when required to avoid ill-formed representations). This parallelism provides independent support for the proposed representation.

Taken together, these points show that the Strict CV analysis not only avoids ad hoc mechanisms like Magic Licensing, but also derives several otherwise unrelated properties of *s*C sequences from a single representational principle.

4.2. Melodic sharing in *s*C sequences

Within the Strict CV framework, branching structure is assumed to play a central role. In its more familiar usage in prosodic or morphosyntactic representations, a branching structure typically refers to a configuration in which constituents combine to form a higher-order set. However, the notion of branching invoked in the structure under consideration differs fundamentally from this standard interpretation. In the present case, C and V do not constitute parts of a larger constituent corresponding to /s/. Rather, what is described as a branching structure is a configuration in which the segmental properties associated with *s* are shared across both C and V positions. This raises an important question: what does it mean for the segmental properties of /s/ to be shared between C and V?

In Polgárdi (2022), the internal segmental composition of /s/ is not explicitly addressed. If we approach the issue within Element Theory, /s/, as a coronal fricative, it may be assumed to consist minimally of an element |A| (or an equivalent feature) associated with coronality, and an element |H| associated with frication. Let us suppose that these elements are linked to a C position and, simultaneously, to a V position. A central tenet of Element Theory is that elements are not inherently restricted to either consonantal or vocalic positions; rather, their phonetic interpretation depends on positional context. Thus, |A| is interpreted as coronality in

a C position, but as an open vowel quality in a V position. Similarly, |H| is realised as voicelessness or aspiration in a C position, but as high tone in a V position. Under this assumption, a C position bearing |A| and |H| would be phonetically realised as [s], whereas the same melodic configuration linked to a V position would yield a high-pitched open vowel. In combination, this would predict a phonetic outcome resembling [sá], rather than a purely consonantal realisation.

However, Polgárdi (2022) assumes that a structure in which /s/ is associated simultaneously with C and V positions is realised not as [sá], but as syllabic [s̩]. If this interpretation is to be maintained, it becomes necessary to clarify the representational mechanism involved. In particular, one must assume that the elements |A| and |H| are not linked to the V position, which must instead remain melodically empty. In such a case, the V position would not surface as schwa, yet its syllabicity would still be realised. At a minimum, this requires an explicit account. That is, rather than analysing the structure as one in which C and V share a single melodic expression, it may be more appropriate to assume that the noise component of [s] phonologically masks the vocalic quality associated with the empty nucleus, while leaving its syllabic contribution intact. Without an explicit formulation of such a mechanism, the claim that the /s/ in *s*C sequences constitutes a syllabic [s̩] spanning both C and V positions remains theoretically under-specified.

A further issue, discussed in Section 3.2, arises if one assumes that an underlying V position follows *s* in *s*C sequences and that *s* is associated with both C and V. Under such an analysis, one would expect that in cases where *s* is associated only with the C position, leaving the V position unassociated, that V should be phonetically realised, yielding outputs of the type [sə]. Indeed, such patterns are attested in English, for example in forms like *kiss* [kɪs] plus the plural suffix *-es* [-z], realised as [kɪsəz] (cf. Harris 1994, among others). While it is possible, from a historical perspective, to appeal to related languages such as Italian as evidence for similar structural configurations, such arguments are not sufficient for a synchronic analysis of present-day English. From a synchronic standpoint, unless there is independent evidence for vowel-zero alternations involving the hypothesised V position, it may be difficult to justify the presence of an underlying vowel between the members of an *s*C cluster.

In fact, this issue is closely related to a key theoretical assumption underlying the Strict CV framework. Within this approach, phonological representations are assumed to consist of a linear sequence of alternating C and V positions, irrespective of whether a given language exhibits overt vowel-zero alternations. If this assumption is taken to be fundamental, then it must be maintained even in the absence of direct synchronic evidence for empty nuclei in the language under investigation. Under such a view, the absence of phonetic realisation corresponding to empty V positions in a particular language does not, in itself, constitute sufficient grounds for rejecting their presence. Rather, if comparable phenomena are attested in historically related languages, this may be taken as providing indirect support

for positing empty V positions. From a theoretical standpoint, therefore, this line of reasoning is not without merit within the Strict CV framework.

5. The representation of *s*C sequences in Precedence-free Phonology

5.1. /s/+stop sequences as singleton segments

The preceding sections have examined the representation of /s/+stop sequences within classical GP and the Strict CV framework. The present section turns to their representation within Precedence-free Phonology (PfP). Before developing a detailed formal account, it is useful to outline a central working hypothesis: within PfP, /s/+stop sequences are taken to correspond to a single phonological domain that is phonetically externalised as a consonantal unit. If such a domain is systematically realised in this way, it is natural to interpret it as corresponding to a single segment.

This assumption naturally leads to the proposal that /s/+stop sequences do not consist of two independent segments, but rather form a single segment, i.e. a singleton segment. The following discussion briefly motivates this interpretation by reconsidering the distributional patterns reviewed in (2), which are reproduced below for convenience.

(7) Empirical patterns of /s/+stop sequences = (2)

a. Word-initial three-consonant sequences (#$C_1C_2C_3$-)
   When three consonants occur word-initially, only /sp, st, sk/ may appear in the C_1C_2 position, as in *split* /splɪt/, *strike* /stɹaɪk/ and *squeeze* /skwiːz/. Other combinations are unattested.
b. Word-initial two-consonant sequences (#C_1C_2-)
   Only when C_1 = /s/ do we observe violations of the Sonority Sequencing Principle – according to which, in syllable-initial position, sonority is expected to rise from less sonorous to more sonorous segments – as in /sp, st, sk/ (*spoil* /spɔɪl/, *stout* /staʊt/, *school* /skuːl/).
c. Word-final two-consonant sequences (-C_1C_2#)
   In non-inflected, non-derived words, final consonant sequences are overwhelmingly restricted to /sp, st, sk/, as in *lisp* /lɪsp/, *mist* /mɪst/ and *ask* /ɑːsk/. By contrast, sequences such as /bl, tr, kw/, which are permitted word-initially (*blue* /bluː/, *tree* /triː/, *quick* /kwɪk/), do not occur word-finally.
d. Word-medial position (-C_1C_2- in non-inflected/non-derived words)
   When attention is restricted to non-inflected, non-derived forms, medial *s*C sequences are also heavily skewed towards /sp, st, sk/. Although combinations such as *sm* or *sf* are attested (e.g. *Asmara* /æsˈmɑːrə/, *asphalt* /ˈæsfælt/), they are rare and lexically restricted. By contrast, /sp, st, sk/ are robustly attested (e.g. *whisper* /ˈwɪspə/, *aster* /ˈæstə/, *basket* /ˈbɑːskɪt/).

In fact, these patterns follow naturally if /sp, st, sk/ are reanalysed as singleton segments. First, if /sp, st, sk/ are single segments, then apparent three-consonant sequences such as *split* /splɪt/, *strike* /stɹaɪk/ and *squeeze* /skwiːz/ are in fact sequences of only two segments. This eliminates the need to explain why only these specific combinations occur in the C_1C_2 position of apparent three-consonant sequences (#$C_1C_2C_3$-).

Second, the apparent violation of the Sonority Sequencing Principle disappears. Under the traditional view, sequences such as /sp, st, sk/ (e.g. *spoil* /spɔɪl/, *stout* /staʊt/, *school* /skuːl/) violate the expectation that sonority should rise in syllable-initial position. However, if these sequences are treated as single segments, no such violation arises, since there is no sequence of two independent consonants. Instead, they are comparable to other stop segments such as /p^h, t^h, k^h/ and /p, t, k/.

Third, the restriction on word-final sequences follows straightforwardly. In non-inflected, non-derived words, final sequences are overwhelmingly restricted to /sp, st, sk/, as in *lisp* /lɪsp/, *mist* /mɪst/ and *ask* /ɑːsk/. If these are analysed as singleton segments, their occurrence in word-final position is no different from that of other stops /p, t, k/. The absence of sequences such as /bl, tr, kw/ in this position (cf. *blue* /bluː/, *tree* /triː/, *quick* /kwɪk/) can then be attributed to an independent restriction against word-final consonant sequences in non-derived forms.

Finally, in the word-medial position, the apparent diversity of *s*C sequences largely disappears once morphological complexity is controlled for. While forms such as *asphalt* /ˈæsfælt/ or *Asmara* /æsˈmɑːrə/ contain /sf/ and /sm/, these are relatively rare and lexically restricted. By contrast, /sp, st, sk/ are robustly attested in monomorphemic vocabulary, as in *whisper* /ˈwɪspə/, *aster* /ˈæstə/ and *basket* /ˈbɑːskɪt/. These patterns align closely with the behaviour expected of singleton segments.

A separate line of research offers a contrasting perspective on /s/+stop sequences, arguing that these are not true sequences at all, but rather, that they constitute complex-singleton segments comparable to affricates like [t͡ʃ] and [d͡ʒ]. This view has been advanced by a number of scholars – among them van de Weijer (1996), who presents cross-linguistic evidence, and earlier works on English by Fudge (1969), Ewen (1982), Fujimura and Lovins (1978) and Selkirk (1982). Related observations are also found in Broselow (1983) for L2 English and Barlow (1997) in studies of phonologically delayed English-acquiring children.

Given these considerations, various segmental theories could, in principle, be employed to model the representation of /s/+stop sequences. In what follows, the analysis is developed within the framework of PfP, which provides a model for capturing the properties of these structures. Specifically, PfP allows complex phonological units to be represented without recourse to linear ordering relations, making it particularly suitable for analysing /s/+stop sequences as unified segments whose internal structure is phonetically realised in a temporally extended manner.

5.2. Overview of Precedence-free Phonology

Before discussing the representation of /s/+stop sequences in PfP, this section briefly outlines how PfP differs from other models of phonological representation.

PfP is a representational framework that redefines the way phonological structure is conceived and organised. It aims to minimise redundancy in representations and to tighten the constraints on expressive capacity, thereby achieving theoretical economy while maintaining a close correspondence with the computational principles that underlie syntactic structure. Unlike traditional models that rely on prosodic hierarchies and temporal sequencing to describe surface patterns, PfP assumes that phonological organisation can be captured independently of linear ordering or prosodic categories. In doing so, it provides a unified perspective in which phonology and syntax are governed by a shared structural mechanism (Nasukawa 2017a; 2017b; 2020; in press; Backley and Nasukawa 2020).

At the core of this approach lies the assumption that phonological representations are constructed by recursive applications of **Merge**, the same combinatorial operation that plays a role in syntax. During lexicalisation, morpheme-internal representations arise through the successive merger of phonological primitives known as **elements –** minimal, privative units such as |A|, |I| and |U| that encode contrastive properties. Complex segments are formed when these primitives combine hierarchically, creating structured configurations that express the internal composition of sounds. This stands in sharp contrast to traditional models, which posit multiple prosodic or metrical constituents (e.g. syllable, rhyme, nucleus, foot) as independent organising domains.

A central tenet of PfP is its **rejection of temporal precedence**. Phonological structures do not encode sequencing or ordering relations. Instead, surface order is interpreted as a consequence of hierarchical configuration – analogous to how word order in syntax emerges from structural relations rather than by being specified directly. This precedence-free architecture streamlines the representational system and highlights the structural parallels between phonology and syntax, suggesting a unified model of grammatical computation.

PfP further departs from conventional prosodic theories by dispensing with categories such as onset and nucleus, which are central to Government Phonology (GP) and metrical phonology. In PfP, elements serve simultaneously as segmental and prosodic exponents, capturing both melodic and hierarchical properties within a unified representational dimension (Backley and Nasukawa 2020; Nasukawa 2020). This integration allows prosodic behaviour – such as syllabicity, prominence and weight – to be derived from the relational properties of the elements themselves, without recourse to prosodic templates.

By eliminating both prosodic constituents and temporal ordering, PfP offers a model of phonological structure that is minimal in its assumptions yet rich in explanatory power. It retains the empirical advantages of earlier frameworks while

establishing a more transparent correspondence between phonological and syntactic computation.

5.3. Structure-building in PfP

Within PfP, phonological representations are constructed exclusively through the **recursive merger of elements**, without reference to prosodic constituents or precedence relations between units. In this model, structure is purely hierarchical and relational, and thus makes no reference to empty nuclei or other abstract timing units that are central to the representational architecture of GP.

Building on the proposals of Nasukawa and Backley (2008; 2015) and Backley (2011), PfP adopts a system of six privative elements – |A I U ʔ H L| – as the sole building blocks of phonological structure (cf. Kaye et al. 1985; 1990; Harris 1994; Scheer 2004). As illustrated in (8), each element has the capacity to express either a **vocalic** or **consonantal** property, depending on how it is hierarchically combined with other elements in the structure.

(8) Elements (Backley 2011; Nasukawa 2017a; 2017b)

resonance elements	*vowel realisation*	*consonant realisation*
|A| (|mass|)	non-high	coronal, uvular, pharyngeal place
|I| (|dip|)	front	dental, palatal place
|U| (|rump|)	rounded, back	labial, velar place

non-resonance elements	*consonant realisation*	*vocalic realisation*
|ʔ| (|edge|)	occlusion	creaky voice (laryngealised)
|H| (|noise|)	noise, voicelessness, aspiration	high tone
|L| (|murmur|)	nasality, obstruent voicing	nasality, low tone

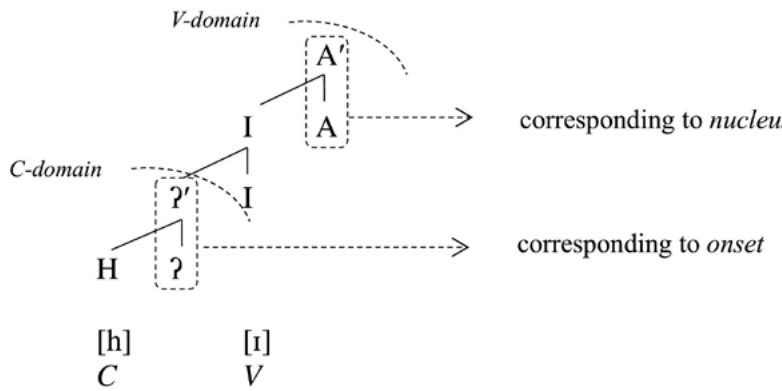
In PfP, the function traditionally associated with the syllable *nucleus* is, in languages such as English, performed by the resonance element |A|. When |A| is realised on its own, it surfaces phonetically as [ə]. A representation consisting solely of |A| thus corresponds to what GP terms the *empty nucleus* (see Kaye et al. 1990; Charette 1991; Harris 1994; Scheer 2004). When |A| combines with other resonance elements and assumes the role of head, the resulting configuration gives rise to one of the *corner vowels*. For example, when |A| takes |I| as its dependent, the resulting structure is interpreted phonetically as [i]. This outcome follows from the general principle of head-dependent realisation in (9), according to which the acoustic properties of a dependent element may override or mask those of the head element – in this case, |A|.

(9) **Principle of phonetic realisation for head-dependent structures** (Nasukawa 2014; 2015; 2016; 2017a; 2017b; 2020; Nasukawa and Backley 2015; 2017; Backley and Nasukawa 2020)
Dependents, which are not necessary for structural well-formedness, are phonetically more salient in terms of their modulated carrier signal than heads, which are important for building structure.

A parallel mechanism operates in the consonantal domain. Here, the structural base is the element |ʔ|, which functions as the consonantal counterpart to the vocalic element |A| (Backley 2011; Nasukawa and Backley 2008). When realised in isolation, |ʔ| surfaces phonetically as [ʔ]. However, when |ʔ| enters into a head–dependent relation and takes |H| as its dependent, the acoustic properties of |H| dominate those of the head element, yielding the surface realisation [h]. This pattern, like that observed for vowels, reflects the general principle outlined in (9): the dependent element may determine the perceptible features of the structure, even when hierarchically subordinate.

The elements |A| and |ʔ| – serving as the heads of the vowel and consonant domains, respectively – can also enter into a head-dependent configuration (for readers unfamiliar with PfP, Nasukawa (2020) provides a useful overview). In such a configuration, the vowel-domain head |A| functions as the dominant element over the consonantal head |ʔ|, producing a composite structure that effectively corresponds to a syllable-like unit, as illustrated in (10).

(10) Structure corresponding to a ‘syllable’ (Nasukawa et al. 2019): /hɪ/ ‘he’



Once a structure equivalent in size to a syllable is established, the two domains involved – the ‘consonantal’ domain headed by |ʔ| and the ‘vocalic’ domain headed by |A| – are prevented from being phonetically expressed at the same time. Their realisation instead proceeds in sequence, following the ordering principle presented in (11).

(11) **Principle of linearisation** (Nasukawa et al. 2019, 634–689)
A domain located at a lower level (dependent) is phonetically realised before a domain located at a higher level (head).

According to this principle, the domain governed by |ʔ| surfaces first, followed by the domain governed by |A|, ensuring that the dependent element precedes its head in phonetic realisation.

(10) illustrates a structure corresponding to a 'syllable.' The same type of hierarchical structure can be recursively combined through Merge, forming larger units under head-dependency relations, such as a structure corresponding to a 'foot.' These units may in turn combine further to yield more complex hierarchical configurations, ultimately giving rise to longer morphemes and words. All such phonological structures are assumed to be phonetically externalised in accordance with the principles set out in (9) and (11).

5.4. /sp, st, sk/ as phonetic realisation of elements within the |ʔ| domain

In PfP, as outlined in the previous section, precedence relations are entirely absent from phonological representations; instead, phonological structure is defined exclusively in terms of asymmetric dependency relations among features, referred to as elements. This property applies not only to relations between segments or between larger prosodic domains, but also to the internal structure of units that are externalised as segments. In other words, multiple melodic elements may be phonetically realised in a temporally extended manner without requiring linear sequencing in the underlying representation.

This approach extends naturally to other so-called contour segments, such as affricates and prenasalised segments. In many segmental theories, these are represented as ordered sequences of features or nodes. In PfP, by contrast, no such ordering is assumed; rather, affricates and prenasalised segments are analysed as configurations of dependent features whose phonetic interpretation yields a contour-like realisation (cf. Lombardi 1990; Schafer 1995; Scheer 2003; Nasukawa 2005).

Within this framework, the behaviour of /s/+stop sequences can be understood as an instance of the same general mechanism. Accordingly, /sp, st, sk/ may be treated in the same way, namely as 'single' segments whose internal structure gives rise to their characteristic phonetic and distributional properties.

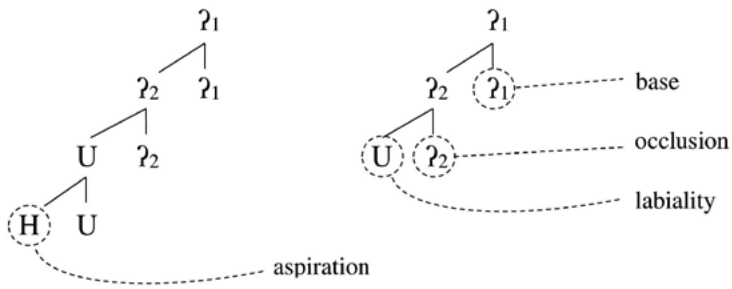
Following the line of argument developed in Nasukawa and Backley (2008) and related work, it is assumed that when multiple place elements are present within a single segment, their phonetic realisation involves a temporal staggering of the acoustic cues associated with each element. In strong positions, these cues are fully externalised, resulting in a contour-like realisation. Consequently, what is traditionally described as an *s*C sequence (e.g. /sp/) is reinterpreted as a single

expression whose initial portion is realised with the acoustic properties of a coronal fricative [s], followed by the stop component.

In order to examine this structure in greater detail, it is first necessary to consider the internal structure of English stops, which are assumed to constitute part of /s/+stop sequences. English stops are commonly analysed as exhibiting a fortis–lenis laryngeal contrast (Harris 1994; Jessen and Ringen 2002; Honeybone 2005; Iverson and Salmons 2006, among others). When stops occur in strong positions, the former (fortis) are realised as aspirated voiceless stops, whereas the latter (lenis) surface as unaspirated voiceless stops. In what follows, this contrast is illustrated by means of the representative pair /p^{h}/ and /p/.

(12) The structures of fortis and lenis stops

a. fortis stop /p^{h}/ b. lenis stop /p/



As illustrated above, both (12a) and (12b) share the same base configuration, in which the topmost position in the structure is occupied by |ʔ|. A hierarchical representation composed of elements within the domain of this structurally highest head |ʔ| is phonetically externalised as a single consonantal unit. In the present case, both (12a) and (12b) involve a base |ʔ| (represented above as $|ʔ_{1}|$), which takes $|U\ ʔ|_{ʔ}$ (represented above as $|U\ ʔ_{2}|_{ʔ2}$) as its dependent. When this configuration is externalised without further modification, it is realised phonetically as the lenis stop /p/.

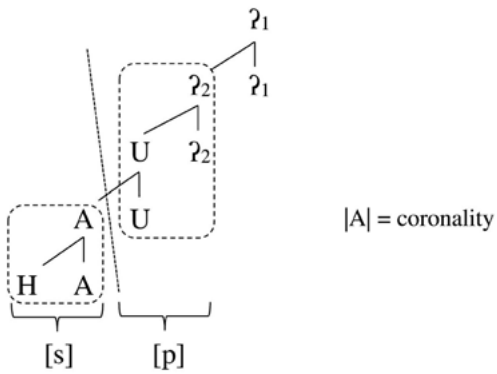
When an additional element |H| is introduced into the most deeply embedded part of the structure, the resulting configuration corresponds to (12a), which is phonetically realised as the fortis stop /p^{h}/.

Given this analysis, the question then arises as to how the phonological structure of /sp/ should be represented. The present paper proposes that /sp/ is characterised by the following structure.

The structure in (13) differs from that in (12b) in that the most deeply embedded part of the former, namely, the dependent of |U|, is occupied by the complex element |H A|. A further point of contrast with (12a) is that the element |A|, which is absent in (12a), is present in (13) as a dependent of |U|. As a result, the structure in (13) contains two resonance elements, |U| and |A|, both of which are associated with the place of articulation. It is proposed that a configuration in which two resonance

elements occur within a single |ʔ|-headed domain is phonetically externalised as a contour segment.

(13) The structure of /sp/ (/ˢp/)



This interpretation is consistent with the analysis of affricates proposed by Nasukawa and Backley (2008), who argue that affrication is entirely a matter of phonetic manifestation and does not involve any segment-internal ordering of elements. On this account, affrication is a performance strategy that enhances the perceptibility of complex place cues (e.g. |I| and |A| in /ʧ/ and /ʤ/), rendering segments with complex resonance more accessible to listeners. This is achieved by enhancing the portion of the acoustic signal that contains aperiodic noise energy, which is relatively rich in place information.

Returning to (13), the presence of two resonance elements within a single domain is therefore expected to give rise to a temporally extended phonetic realisation, in which the relevant cues are staggered so as to improve the perceptibility of complex place elements. In the case of (13), the resonance elements involved are |U| and |A|. When these are phonetically realised, the cues associated with |A| – together with its dependent |H| – are externalised prior to those associated with |U|. This ordering follows from the Principle of Linearisation stated in (11), according to which a domain located at a lower (dependent) level is phonetically realised before a domain located at a higher (head) level.

Although the discussion above has focused on /sp/, analogous configurations are posited for /st/ and /sk/, each involving a |ʔ|-headed domain containing multiple resonance elements. As in the case of /sp/, the co-occurrence of these elements within a single domain is assumed to yield a contour-like phonetic outcome, characterised by the temporal staggering of their associated cues.

A final issue to be addressed concerns the status of *s*C sequences other than /s/+stop (i.e. /sp, st, sk/), such as /sl, sm, sn, …/. The question is whether these sequences share the same structural representation as /s/+stop, or whether they

require a different analysis. The position adopted here is that, unlike /sp, st, sk/, such sequences are best analysed as combinations of two independently realised segments.

This conclusion is supported by their distributional behaviour. As shown in (2), /s/+stop sequences exhibit a number of distinctive phonotactic properties, including their restricted distribution across word-initial, word-medial and word-final positions, as well as their exceptional behaviour with respect to the Sonority Sequencing Principle. Crucially, these peculiarities are not systematically observed in other types of *s*C sequences. In most cases, sequences such as /sl, sm, sn, …/ display distributional patterns that are consistent with those of ordinary consonant sequences.

On this basis, it is proposed that /sp, st, sk/ differ fundamentally from other *s*C sequences in their structural organisation. Whereas the former are analysed as single segments with an internally complex structure, the latter are more appropriately treated as sequences of two independent segments, each corresponding to a separate phonological domain.

6. Conclusion

This paper has argued that English /s/+stop sequences (/sp, st, sk/) are best analysed not as consonant sequences, but as singleton segments with internally complex structure. The argument was based on a range of distributional asymmetries, which distinguish these sequences from other *s*C combinations and are difficult to capture under traditional cluster-based analyses.

A comparison of existing approaches within Government Phonology and the Strict CV framework has shown that, although these models provide valuable insights, they rely on representational devices such as empty nuclei or melodic spreading whose empirical motivation remains limited in the case of English. In contrast, the PfP framework offers a more constrained and principled alternative, in which phonological structure is defined in terms of hierarchical dependency relations without reference to linear ordering.

Within this framework, /s/+stop sequences are analysed as single |ʔ|-headed domains containing multiple resonance elements. Their characteristic phonetic properties are derived from the temporal staggering of these elements, in accordance with general principles of phonetic interpretation. This account not only eliminates the need for additional structural mechanisms but also provides a unified explanation for the phonotactic behaviour of these sequences.

Finally, the analysis distinguishes /s/+stop sequences from other *s*C combinations, which are argued to consist of two independent segments. This contrast further supports the view that the exceptional behaviour of /sp, st, sk/ reflects a fundamental difference in their internal structure. More broadly, the findings

of this study lend support to the PfP approach as a viable and explanatory model of phonological representation, capable of capturing both structural and phonetic generalisations in a unified manner.

Acknowledgements

I gratefully acknowledge constructive comments on an earlier draft by Phillip Backley and two anonymous reviewers. This work was supported by the following MEXT/JSPS KAKENHI grants: Grant-in-Aid for Scientific Research (S) Grant Number JP19H05589; Grant-in-Aid for Scientific Research (A) Grant Numbers JP19H00532 and JP24H00085; Grant-in-Aid for Scientific Research (B) Grant Numbers JP20H01269 and JP24K00059; and Grant-in-Aid for Scientific Research (C) Grant Number 22K00513.

Notes

1 It may appear that forms such as *stride* /straɪd/ and *astride* /əˈstraɪd/ provide evidence for the optional presence of a weak, schwa-like vowel preceding *s*C sequences. However, the initial vowel in *astride* is not lexically empty but is part of the lexical representation. Historically, this initial *a-* derives from a reduced form of a preposition (*on* or *in*) in Old English, and synchronically it forms part of a limited set of lexicalised items (e.g. *asleep*, *awake*, *aside*). Importantly, this element is not a productive prefix in present-day English and cannot be freely attached to arbitrary bases. Therefore, such cases do not constitute evidence for a word-initial empty nucleus preceding /s/.

References

Backley, Philip. 2011. *An Introduction to Element Theory.* Edinburgh: Edinburgh University Press.

Backley, Philip, and Kuniya Nasukawa. 2020. "Recursion in Melodic-prosodic Structure." *Morpheme-internal Recursion in Phonology*. Ed. Kuniya Nasukawa. Boston & Berlin: Mouton de Gruyter, 11–35.

Barlow, Jessica. 1997. "A Constraint-based Account of Syllable Onset Patterns in Phonological Development." PhD diss, Indiana University.

Broselow, Ellen. 1983. "Nonobvious Transfer: On Predicting Epenthesis Errors." *Language Transfer in Language Learning*. Ed. Susan Gass, and Larry Selinker. Rowley, MA: Newbury House. 269–280.

Charette, Monik. 1991. *Conditions on Phonological Government.* Cambridge: Cambridge University Press.

Ewen, Colin. J. 1982. "The Internal Structure of Complex Segments." *The Structure of Phonological Representations, Part I*. Ed. Harry van der Hulst, and Norval Smith. Dordrecht: Foris. 27–67.

Fudge, Erik. 1969. “Syllables.” *Journal of Linguistics* 5: 253–286.
Fujimura, Osamu, and Julie B. Lovins. 1978. “Syllables as Concatenative Phonetic Units.” *Syllables and Segments*. Ed. Alan Bell, and Joan B. Hooper. Amsterdam: North-Holland. 107–120.
Davis, Stuart, and Michael Hammond. 1995. “On the Status of Onglides in American English.” *Phonology* 12: 159–182.
Harris, John. 1994. *English Sound Structure.* Oxford: Blackwell.
Honeybone, Patrick. 2005. “Diachronic Evidence in Segmental Phonology: The Case of Obstruent Laryngeal Specifications.” *The Internal Organization of Phonological Segments*. Ed. Marc van Oostendorp, and Jeroen van de Weijer. Berlin: Mouton de Gruyter. 319–354.
Iverson, Gregory K., and Joseph C. Salmons. 2006. “On the Typology of Final Laryngeal Neutralization: Evolutionary Phonology and Laryngeal Realism.” *Theoretical Linguistics* 32: 205–216.
Jessen, Michael, and Catherine Ringen. 2002. “Laryngeal Features in German.” *Phonology* 19: 189–218.
Kaye, Jonathan. 1990. “‘Coda’ Licensing.” *Phonology* 7: 301–330.
Kaye, Jonathan. 1992. “Do You Believe in Magic? The Story of s+C Sequences.” *SOAS Working Papers in Linguistics and Phonetics* 2: 293–313. Reprinted 1996. *A Festschrift for Edmund Gussmann*. Ed. Henryk Kardela, and Bogdan Szymanek. Lublin: University Press of the Catholic University of Lublin. 155–176.
Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1985. “The Internal Structure of Phonological Elements: A Theory of Charm and Government.” *Phonology Yearbook* 2: 305–328.
Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1990. “Constituent Structure and Government in Phonology.” *Phonology* 7: 193–231.
Lombardi, Linda. 1990. “The Nonlinear Organization of the Affricates.” *Natural Language and Linguistic Theory* 8: 375–425.
Nasukawa, Kuniya. 2005. *A Unified Approach to Nasality and Voicing*. Berlin & New York: Mouton de Gruyter.
Nasukawa, Kuniya. 2014. “Features and Recursive Structure.” *Nordlyd* 41.1, *Special Issue on Features*: 1–19.
Nasukawa, Kuniya. 2015. “Why the Palatal Glide is Not a Consonantal Segment in Japanese: An Analysis in a Dependency-based Model of Phonological Primes.” *The Segment in Phonetics and Phonology*. Ed. Eric Raimy, and Charles Cairns. Malden, MA: Wiley-Blackwell, 180–198.
Nasukawa, Kuniya. 2016. “A Precedence-free Approach to (De-)Palatalisation in Japanese.” *Glossa: A Journal of General Linguistics* 1.1:9. DOI: 10.5334/gjgl.26
Nasukawa, Kuniya. 2017a. “Extending the Application of Merge to Elements in Phonological Representations.” *Journal of the Phonetic Society of Japan* 21: 59–70.

Nasukawa, Kuniya. 2017b. "The Phonetic Salience of Phonological Head-dependent Structure in a Modulated-carrier Model of Speech." *Beyond Markedness in Formal Phonology*. Ed. Bridget D. Samuels. Amsterdam: John Benjamins. 121–152.

Nasukawa, Kuniya. 2020. "Linearisation and Stress Assignment in Precedence-free Phonology: The Case of English." *Radical: A Journal of Phonology* 1: 239–291.

Nasukawa, Kuniya. In press. "A Precedence-free Approach to Phonology." *The Wiley Blackwell Companion to Phonology*. Ed. Kuniya Nasukawa, Bridget Samuels, Geoffrey Schwartz, and Miklós Törkenczy. 2nd ed. Oxford: Wiley-Blackwell.

Nasukawa, Kuniya, and Philip Backley. 2008. "Affrication as a Performance Device." *Phonological Studies* 11: 35–46.

Nasukawa, Kuniya, and Philip Backley. 2015. "Heads and Complements in Phonology: A Case of Role Reversal?" *Phonological Studies* 18: 67–74.

Nasukawa, Kuniya, and Philip Backley. 2017. "Representing Moraicity in Precedence-free Phonology." *Phonological Studies* 20: 55–62.

Nasukawa, Kuniya, Philip Backley, Yukari Yasugi, and Masatoshi Koizumi. 2019. "Challenging Universal Typology: Right-edge Consonantal Prominence in Kaqchikel." *Journal of Linguistics* 55.3: 611–641.

Polgárdi, Krisztina. 2015. "Vowels, Glides, Off-glides and On-glides in English: A Loose CV Analysis." *Lingua* 158: 9–34.

Polgárdi, Krisztina. 2022. "The Representation of *s*C-clusters: A Strict CV Analysis." *Glossa* 7: 1–29.

Rialland, Annie. 1994. "The Phonology and Phonetics of Extrasyllabicity in French." *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Ed. Patricia A. Keating. Cambridge: Cambridge University Press. 136–159.

Schafer, Robin. 1995. "Headedness in the Representation of Affricates." *The Linguistic Review* 12: 61–87.

Scheer, Tobias. 2003. "On Spirantisation and Affricates." *Living on the Edge*, *28 Papers in Honour of Jonathan Kaye*. Ed. Stefan Ploch. Berlin: Mouton de Gruyter. 283–301.

Scheer, Tobias. 2004. *A Lateral Theory of Phonology: What Is CVCV, and Why Should It Be?* Berlin: Mouton de Gruyter.

Selkirk, Elisabeth. 1982. "The Syllable." *The Structure of Phonological Representations, Part II*. Ed. Harry van der Hulst, and Norval Smith. Dordrecht: Foris. 337–383.

Weijer, Jeroen van de. 1996. *Segmental Structure and Complex Segments.* Tübingen: Niemeyer.

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Sławomir Zdziebko, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.04

Sławomir Zdziebko
https://orcid.org/0000-0001-8303-8773
The John Paul II Catholic University of Lublin

A Substance-free Autosegmental Account of the Coronal Mutation in English

Abstract: This paper puts forward a restrictive approach to subsegmental architecture, which assumes that sub-skeletal representations are composed of obligatorily binary branching nodes and substance-free emergent features which serve as pointers to phonetic interpretation. We put the theory to a test by applying its principles to the set of alternations of coronal consonants attested in English before affixes *-ion*, *-ian* and *-ious* in words such as *infectious*, *suggestion*, *revision*, *expansion*, etc. It is demonstrated that the requirement for the structure to be binary branching as well as the requirement for English consonants to contain at most three features derive the correct outputs of the coronal mutation. It is also argued that the postulated approach accounts for the behavior of the items which resist the mutation, such as *rebellion*, *Arthurian*, *oblivious* through the last resort operation of node epenthesis.

Keywords: substance-free phonology, autosegmental phonology, consonant mutation, features, phonology–phonetics interface

1. Introduction

This paper appeals to the hierarchical and autosegmental nature of phonological representations and proposes a model in which phonological structures are composed of strictly hierarchical binary branching nodes, which at the same time may be associated with other nodes in a manner known from the classic works in autosegmental phonology.

At the same time, the approach employed below assumes a strict separation of phonology and phonetics. The approaches which postulate that phonology is a formal system independent of phonetic substance (see Hale and Reiss 2008; Odden 2022; Scheer 2022, among others) are obliged to put forward a coherent approach to externalization or the phonology–phonetics interface. A natural way of conceiving of such an interface is that it appeals to the operation of intermodular translation in which phonological structure is associated with phonetic information and is re-written as such at the interface (see e.g. Scheer 2022).

This work postulates that the hierarchical nature of autosegmental phonological representations plays a crucial role for the translation operation. To be precise, we postulate that the phonetic interpretation of the phonological structure proceeds in a bottom-up fashion, commencing at the bottom-most branching node.

The aim of the paper is to apply the general principles of such a framework to the data from English. In particular, we put forward a substance-free autosegmental analysis of the consonant mutation that affects coronal obstruents in the environment of affixes *-ion*, *-ian* and *-ious*. It is assumed that those affixes are lexically equipped with a floating node, which is expected to attach to the structure of the stem-final consonants. The observed mutation is the consequence of the structural modifications that the affected consonants undergo when the floating node is integrated into their representation.

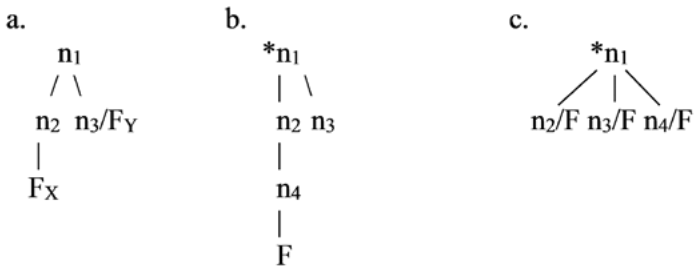
In general, the framework outlined in this paper subscribes to the classic assumption of autosegmental phonology whereby nodes and features may act independently of other features and nodes which constitute the make-up of a given segment. To be more precise, nodes and features may be lexically unassociated, i.e. floating, and are assumed to spread independently of other nodes and features. At the same time, the current proposal differs from most autosegmental approaches to feature geometry (see Clements 1985; Sagey 1986; Clements and Hume 1995, among others) in that it postulates that nodes and features are devoid of phonetic substance. As a consequence, the nodes postulated in this work do not carry the traditional labels such as ‘Place,’ ‘Manner’ or ‘Laryngeal,’ etc., which constitute a direct appeal to the phonetic, as opposed to the phonological, aspects of sound systems.

Section 2 introduces the nodes and features as the building blocks of the sub-skeletal phonological structure and discusses the basic well-formedness conditions that a grammatical representation must fulfil. Section 2.1 discusses the workings of phonetic interpretation, while Section 2.2 briefly addresses the proposed sources of cross-linguistic differences between sound systems. Section 3 begins with the discussion of the data and introduces the sub-classes of stems affected by coronal mutation. Sections 3.1–3.4 present the analyses of the stems which undergo the mutation. Section 3.5 provides the analysis of the consonants which systematically resist the mutation. Section 4 concludes the article.

2. A substance-free approach to sub-skeletal architecture: the framework

The framework employed in this article rests on the assumption that the sub-skeletal portion of phonological segments is composed of nodes (n) and features (F). Following Nasukawa (2017; 2018), we assume that the lexical phonological representations are built through the operation Merge, known from the syntactic literature (Chomsky 1995; 2000). As a consequence of Merge, nodes form a binary branching spine. All nodes must be binary branching, save the terminal nodes. Features may be attached to terminal or non-terminal nodes. In the latter case, they constitute separate branches on their own. The legitimate structure constructed in this fashion is illustrated in (1a). Structures in (1b, 1c) are impossible.

(1)



As illustrated in (1a), node n_1 may branch into two nodes (n_2 and n_3) or a node (n_2) and a feature (F_Y). Structure (1b) is ungrammatical as node n_2 is a non-branching non-terminal node. Structure (1c) is ungrammatical as it contains a ternary branching node n_1.

Nodes and features differ in their status in that the features, but not the nodes, constitute the instructions for the phonetic realization of the structure (see below for details). While the associations between nodes and features may be removed during the derivation, the associations between nodes are fixed and cannot be removed. At the same time, new associations between nodes may be added to the structure in the course of the derivation.

In the current approach, features are emergent in the sense of Mielke (2008), i.e. there is no universal set of distinctive features inherent in the grammar. Instead, features are postulated at the stage of language acquisition in order to account for the two kinds of phonological behavior of segments. Firstly, the features account for the existence of lexical contrasts between segments, i.e. segments such as /t/ and /s/ in words such as /kɪt/ and /kɪs/ differ in their featural content as the two words have distinct lexical representations. Secondly, the features are postulated to account for the existence of natural classes of segments which behave in a uniform way with respect to phonological processes or phonotactic generalizations observed in a language. Thus, the fact that both /s/ and /t/ give rise to /ʃ/, when subject to the coronal mutation discussed in this paper, points to certain common aspects of

the featural make-up of /t/ and /s/. At the same time, the English /s/ and /t/ differ in their featural representation as only the former may be found at the beginning of words before the lateral, cf. *slow* vs **tlow*.

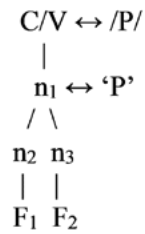
Features employed in this study are substance-free, unary and privative. The privative nature of phonological features, in which the absence of a feature is opposed to its presence, as well as the assumption that representations are constructed on the basis of their phonological behavior, results in a radical non-specification of the phonological representation. The approach postulated here shares this property with approaches such as Element Theory (Kaye et al. 1985; Harris and Lindsey 1995; Backley 2011, among many others), Government Phonology 2.0 (Pöchtrager 2015), Radical CV Phonology (van der Hulst 2020) or Particle Phonology (Schane 1984). The phonetic detail is not represented phonologically but may be read off phonological representation by the phonetic interpretation procedure, which converts the hierarchical phonological representations into phonetic labels understood by the articulatory-perceptual system.

2.1. Phonetic interpretation of the structure

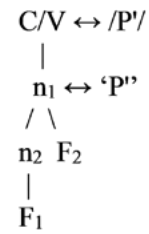
Apart from the formal requirement that phonological representation be composed of binary branching nodes, a no less important well-formedness condition is that it be phonetically interpretable. Phonetic interpretation is an operation which converts the phonological representation into phonetic labels. Phonetic interpretation (PI) operates in a segment-to-segment fashion, commencing from the right margin of the interpreted linguistic unit (word/morpheme/stem). PI proceeds in a strictly bottom-up order, starting with the bottom-most branching node of the representation and supplies phonetic (articulatory and acoustic) labels computable by the phonetic subcomponent of the grammar. The phonetic labels are lexically stored constellations of phonetic information which correspond to phonological nodes and are selected on the basis of the featural content of the relevant nodes. Phonetic interpretation may be sensitive to the presence of a given combination of features as well as the asymmetrical relation of dominance between the features. This means that representation (2a) is realized as a phonetic label P, while the representation (2b), despite being composed of the same set of features, may be realized as a distinct label P'.

(2)

a.



b.



The representations in (2) are supplied with different phonetic content on the basis of two lexically stored interpretational conventions (3a) and (3b).

(3)
a. $\{F_1;F_2\} \leftrightarrow P$
b. $\{F_2 > F_1\} \leftrightarrow P'$

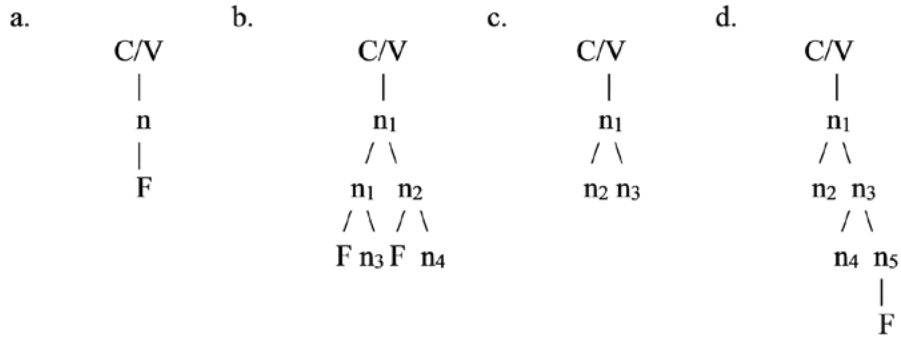
The interpretational convention (3a) says that a node under which features F_1 and F_2 co-occur (;), is to be supplied with a phonetic label P. The convention (3b) says that a node under which feature F_2 dominates (>) features F_1 is supplied with the node P'. A feature F_X dominates another feature F_Y if the first node which dominates F_X dominates F_Y. The two conventions are not exactly conflicting as the dominance relation is considered to make (3b) more specific and causes it to be selected over (3a) as a realization of the representation (2b). At the same time, convention (3b) cannot be selected to supply the phonetic label for the representation (2a), which does not contain the dominance relation. Convention (3a), on the other hand, may realize structure (2b), but only in the absence in the lexicon of a convention (3b).

The higher nodes inherit the phonetic labels supplied for the lower nodes and contribute additional phonetic information which is combined with the former and sent to phonetic execution at the skeletal level represented by nodes C (consonant) or V (vowel).

A phonological structure is phonetically uninterpretable if it does not contain a branching node (4a), if it does not contain an unambiguously lowest branching node (4b), or if the branching node at which the interpretation takes place does not contribute a feature for interpretation (4c, 4d). To be precise, the interpretational cycles at nodes n_1 in (4c) and (4d) do not contribute new phonetic substance and are, therefore, spurious. Intuitively speaking, a general appeal to the economy of the derivation speaks against spurious interpretational cycles. In more formal, graph-geometric terms, the condition that rules out structures such as (4c) and (4d) is a restriction on the distribution of empty nodes within structure. Such a condition says that an empty node may not dominate another empty node, where a node n_X dominates node n_Y if the first node which dominates node n_X also dominates n_Y. Such a formal condition rules out representation in (4c) as node n_1 dominates the empty nodes n_2 and n_3. It also rules out configuration (4d), where node n_1 immediately dominates the empty node n_2 and at the same time dominates the empty node n_4.

Furthermore, a structure is deemed uninterpretable if the lexicon of a given language does not contain an interpretational convention that could supply the phonetic label for the relevant node. In such a case, the derivation crashes.

(4) Phonetically uninterpretable structures



The structures presented in (4) are universally uninterpretable. The case in which a well-formed structure cannot be interpreted due to the absence of an appropriate interpretational convention is a language-specific situation. Languages may, therefore, differ with respect to what structure they are capable of externalizing. The following section discusses further reasons for the observable cross-linguistic variation in phonology.

2.2. Why do sound systems differ?

The first answer to the above question has been suggested in the previous section: languages differ because they have different interpretational conventions, which supply different nodes with different phonetic labels. In other words, the first and major source of cross-linguistic variation in sound systems is that languages have different lexicons. Ideally, the lexicon should be the only source of such a variation. If such a situation were obtained, the cross-linguistic variation would boil down to languages differing in their superficial phonetic detail as a result of the existence of the phonology-phonetics interface. The phonological systems of languages would be completely uniform. Unfortunately, such a situation is a naïve idealization. An overwhelming body of cross-linguistic evidence suggests that languages differ in their phonological systems. The aim of phonological theory is to discover and formalize the limits of such a variation in a concise and empirically contentful way.

This paper postulates that much of the cross-linguistic variation observed in the phonological behavior of segments boils down to what will be referred to as the Featural Complexity Parameter (FCP) setting assumed by a given system. Let us assume that languages select a maximal number of features which a segment in a given language may be composed of. The cut-off points at a Featural Complexity hierarchy that a language selects directly affect the possible outputs of phonological processes attested in such a language and, less directly, the number of segments such a language may contain in its inventory. Let us further assume that English

allows its consonants to be composed of at most three features. Hawaiian allows its consonants to contain two features at most, while Polish allows its consonants to contain up to four features. A partial hierarchy of this kind is illustrated in (5).

(5) Featural Complexity Parameters (consonants)

maximal number of features per segment

Hawaiian (2) English (3) Polish (4) 5?

A language which would allow one feature per segment most likely does not exist. Such a language would have trouble forming natural classes of segments. Given the existence of languages which utilize many more consonantal contrasts than Polish, we take it that languages may use more than four features per segment. The exact empirical consequences of the parametric variation of the postulated form cannot be discussed here. The claim that English does not allow more than three features per one consonant has paramount consequences for the application of coronal mutation in this language.

To conclude the exposition, let us note that the assumptions that constitute the framework presented in this section are not new. The binary branching structures in phonology have been argued for in Pöchtrager (2015), Schwartz (2010), Nasukawa (2017; 2018), van der Hulst (2020), among others. The privative nature of phonological primes has been the cornerstone of many approaches to phonological representations, Element Theory being probably the best-established one. The emergent nature of phonological features has been around at least since Mielke (2008). The claims that phonological representations are free from the phonetic substance have given rise to a growing body of works (see Hale and Reiss 2008; Odden 2022). The claim that the interpretational subcomponents of the grammar operate in a bottom-up fashion, starting from the lowest branching node, is the established view on semantic interpretation (see Heim and Kratzer 1998). In sum, the novelty of the framework presented in this section lies primarily in that it is a new combination of well-known and often well-established claims about how the phonological grammar works. The following sections focus on how this combination of assumptions may be fruitfully employed to derive a set of relatively complex morphophonological alternations attested in English, i.e. the coronal mutation.

3. The analysis of the coronal mutation in English

Although English is much less abundant in morphophonological consonant mutations than e.g. the system of Polish, some such changes are on the record. The most common and best-analyzed mutations attested in English are velar softening (6), t-spirantization (7), coronal palatalization (8), and coronal mutation (9). They have been analyzed in a variety of frameworks in Chomsky and Halle (1968),

Rubach (1984), Halle and Mohanan (1985), Kang (2000), Halle (2005), Radomski (2010), among many others.

(6) Velar softening

a. *elektri*/k/ – *elektri*/s/*-ity*
b. *fanati*/k/ – *fanati*/s/*-ism*
c. *publi*/k/ – *publi*/s/*-ize*
d. *analo*/g/*-ous* – *analo*/ʤ/*-y*
e. *ri*/g/*-our* – *ri*/ʤ/*-id*
f. *pedago*/g/*ue* – *pedago*/ʤ/*-y*

(7) t-spirantization

a. *pira*/t/e – *pira*/s/*-y*
b. *intima*/t/*e* – *intima*/s/*-y*
c. *innocen*/t/ – *innocen*/s/
d. *pruden*/t/ – *pruden*/s/

(8) Coronal palatalization

a. *concep*/t/ – *concep*/ʧ/*-ual*
b. *gra*/d/e – *gra*/ʤ/*-ual*
c. *u*/z/*e* – *u*/ʒ/*-ual*
d. sek/s/ – *sek*/ʃ/*-ual*

(9) Coronal mutation

a. *elek*/t/ – *elek*/ʃ/*-ion*
   celebra/t/*e* – *celebra*/ʃ/*-ion*
   infec/t/ – *infec*/ʃ/*-ious*
   educa/t/*e* – *educa*/ʃ/*-ion*
b. *sugges*/t/ – *sugges*/ʧ/*-ion*
   conges/t/ – *conges*/ʧ/*-ion*
   diges/t/ – *diges*/ʧ/*-ion*
c. *impre*/s/ – *impre*/ʃ/*-ion*
   aggre/s/*-ive* – *aggre*/ʃ/*-ion*
   preten/s/*e* – *preten*/ʃ/*-ious*
   Thra/s/*e* – *Thra*/ʃ/*-ian*
d. *ero*/d/*e* – *ero*/s/*-ive* – *ero*/ʒ/*-ion*
   inva/d/*e* – *inva*/s/*-ive* – *inva*/ʒ/*-ion*
d'. *conclu*/d/*e* – *conclu*/s/*-ive* – *conclu*/ʒ/*-ion*
   comprehen/d/ – *comprehen*/s/*-ive* – *comprehen*/ʃ/*-ion*

e. *condescen*/d/ – *condescen*/ʃ/-*ion*
conten/d/ – *conten*/ʃ/-*ious*
f. *rev*/aɪz/*e* – *rev*/ɪʒ/-*ion*
dec/aɪd/*e* – *dec*/ɪʒ/-*ion*
coll/aɪd/*e* – *coll*/ɪʒ/-*ion*
g. *rebe*/l/ – *rebe*/l(i)j/-*on* – *rebe*/l(i)j/-*ous*
victor/r/-*y* – *victo*/r(i)j/-*ous*
compa/n/-*y* – *compa*/nj/-*on*
en/v/-*y* – *en*/v(i)j/-*ous*

The status of velar softening (6) as a phonological change is questionable due to its notorious underapplication (cf. *anar*[k]-*ism, cate*[k]-*ize*, *monolo*[g]-*uist*) and possible sensitivity to spelling (only the /k/ spelled as <c> is affected) and will not be analyzed here.

While t-spirantization (7) is more regular in that it takes place in numerous abstract nouns, especially those ending in -*y* /i/, it admits a large number of exceptions (cf. *calami*/t/-*ous* – *calami*/t/-*y*, *warran*/t/ – *warran*/t/-*y*). Due to the existence of those exceptional cases, we will leave the status of t-spirantization for further research.

Coronal palatalization applies regularly before the affix -*ual* /(ju)əl/. In the case of some items, it is optional and register-dependent (see *sek*/sjuəl/ vs sek/ʃuəl/, *gra*/djuəl/ vs *gra*/dʒuəl/).[1]

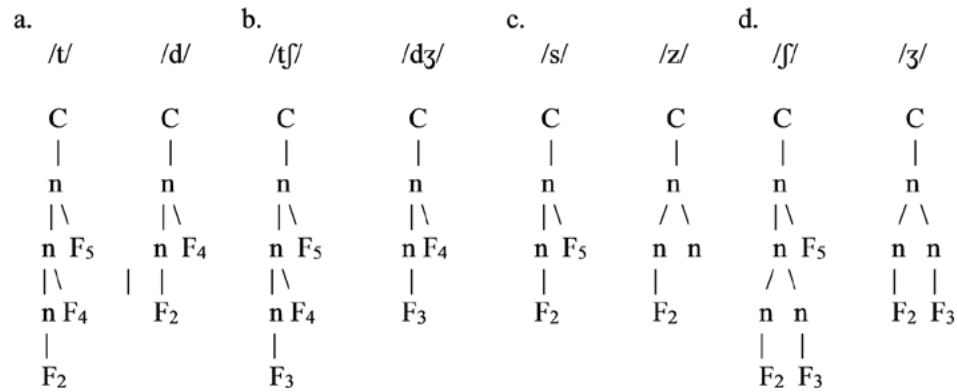
Given the space limitations, this paper focuses only on the descendants of Early Modern English palatalization illustrated in (9). These changes are jointly referred to as *coronal mutation*. The relevant changes affect alveolar obstruents obligatorily before the affixes -*ion,* /(ij)ən/ and -*ious* /(ij)əs/ and often variably before affix -*ian* /(ij)ən/. They cause /t/ to surface as post-alveolar /ʃ/ (9a) unless it is preceded by /s/, in which case the post-alveolar affricate /tʃ/ surfaces (9b). The items in (9c) illustrate the cases in which the underlying /s/ surfaces as /ʃ/. The stems in (9d) and (9e) surface with the lenis post-alveolar spirant /ʒ/ or the fortis spirant /ʃ/, if preceded by the nasal. The items in (9d), surface with stem-final /d/ if they are verbs, and with /s/ in front of the affix -*ive* /ɪv/.

In the items such as (9f) the consonant mutation is accompanied by a vowel mutation. Since a reasonably constrained model of phonology cannot derive a pattern in which the diphthong /aɪ/ is shortened to /ɪ/ before alternating /z/, we will disregard vowel alternations and focus only on the consonantal portion of the alternation stems, i.e. the derivation of /ʒ/ from /z/.

Segments other than coronal obstruents resist the mutation (9g) and the stem-final segments in relevant words are followed by a vowel /i/ or a semi-vowel /j/.

To begin with, let us present preliminary lexical representations of the segments which constitute the input and the output of the mutation, i.e. the coronal obstruents found in English (see (10)).

(10) Coronal obstruents in English



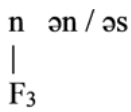
The representations found in (10) are to be treated as the default configurations which render the relevant phonetic objects. The discussion in the sections to follow will reveal that certain representations may be richer, i.e. contain more nodes and features, than what (10) illustrates and that certain nodes may be rearranged with respect to other nodes. All such manipulations are legitimate so long as the representations conform to the general principles discussed in Section 2: the binary-branching requirement, Featural Complexity Parameter and the need for the representations to be phonetically interpretable.

The interpretational conventions, which convert the representations of English coronal obstruents into their phonetic realizations, are presented in (11).

(11)
a. $\{F_5\} \leftrightarrow$ 'fortis obstruent'
b. $\{F_4\} \leftrightarrow$ 'occlusion/non-continuant'
c. $\{F_4;F_3\} \leftrightarrow$ 'post-alveolar non-continuant'
d. $\{F_2;F_3\} \leftrightarrow$ 'post-alveolar'
e. $\{F_2;F_4\} \leftrightarrow$ 'alveolar non-continuant'
f. $\{F_2;F_5\} \leftrightarrow$ 'fortis alveolar obstruent'

Let us further assume that the affixes before which coronal mutation takes place are lexically represented with a floating node hosting feature F_3, which anchors onto the preceding linearized material, triggering the relevant mutations (see (12)).

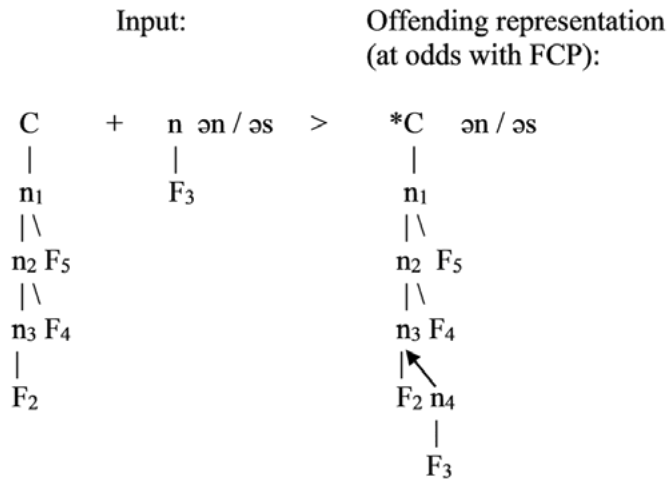
(12) The representation of affixes *-ion*, *-ian* /ən/, *-ious* /əs/



3.1. The derivation of stems in /t/ and /s/ (*obligation*, *suggestion*, *impression*)

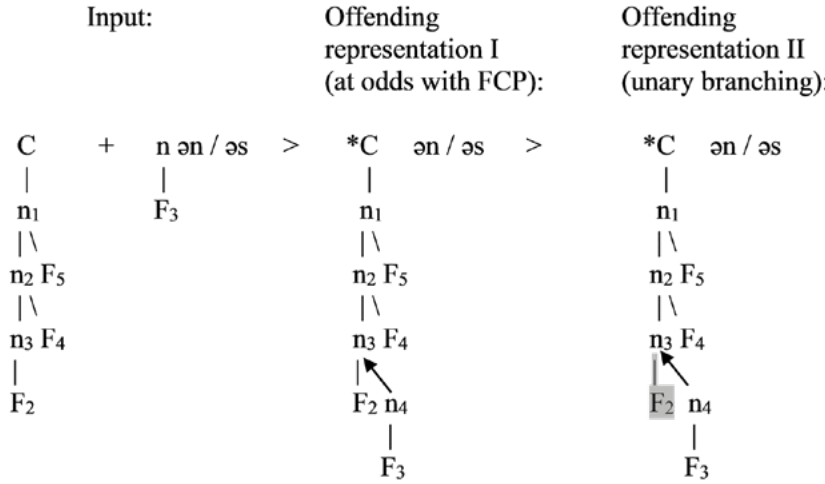
Let us propose that the floating node anchors onto the stem-final consonant[2] as low as possible without violating the universal and language-specific constraints discussed in Section 2. When one of the affixes found in (12) is concatenated with stems which lexically terminate in the fortis alveolar plosive /t/ (10a) the floating node attempts to anchor as the sister to feature F_2. The resulting structure, however, is at odds with the Featural Complexity Parameter (FCP) setting for English, according to which English does not tolerate consonants specified for more than three features (see (13)).

(13) Offending structure I



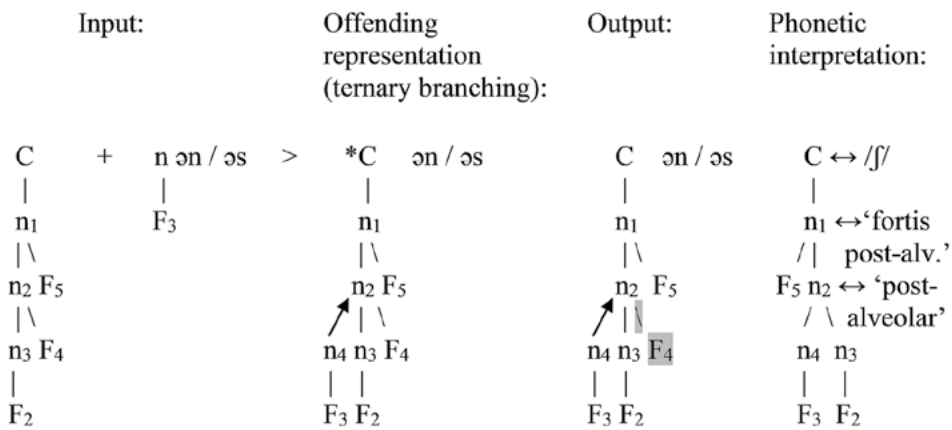
Note that the repair operation, which results in the delinking and deletion of feature F_2, is not an option as it results in the derivation of a structure in which node n_3 is unary branching and non-terminal (14).[3]

(14) Offending structure II (derivation of a unary branching representation)



Since it is not possible for the floating node to anchor as the sister of feature F_2, it attempts to anchor as the sister of feature F_4. This results in the illegal ternary branching structure, which may, however, be repaired by the deletion of feature F_4. This is illustrated in (15).

(15) Legitimate derivation /t/ > /ʃ/



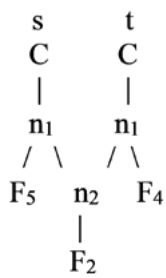
The phonetic interpretation of the output structure in (15) commences by establishing the exponent of the bottom-most branching node, i.e. n_2. The convention which realizes the features dominated by this node is ((11d), $\{F_2;F_3\}$ ↔ 'post-alveolar'). The relevant node is interpreted as 'post-alveolar'. The next branching node in the structure inherits the label 'post-alveolar' and contributes the label

'fortis obstruent', by convention (11a). The representation is realized as the fortis post-alveolar fricative /ʃ/.

In sum, the derivation described above causes the stems which lexically terminate in the alveolar plosive /t/ (*celebra*/t/, *infec*/t/, etc.) to surface with the post-alveolar spirants in words such as *celebra*/ʃən/ and *infec*/ʃəs/, i.e. when concatenated with affixes *-ion* and *-ious*.

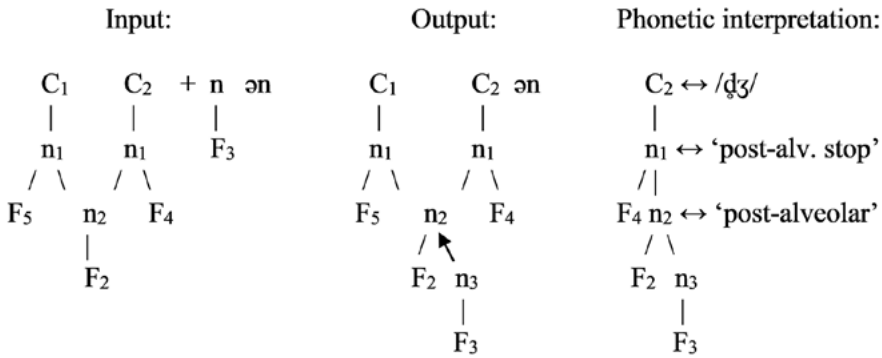
The derivation proceeds in a minimally different way if the alveolar plosive is preceded by the fricative /s/ (10c). It is a well-established fact about English that the contrast between the lenis and fortis plosives is neutralized after the consonant /s/. In the current framework, this means that the plosives which follow /s/ are represented without the structural layer containing feature F_5. The /st/ sequence found in stems such as *sugge*/st/ is represented as in (16), i.e. as a partial geminate in which the fricative and the stop share node n_2 hosting feature F_2.

(16) The representation of the sequence /st/



The concatenation of the relevant stems with the affixes equipped with the floating node is illustrated in (17).

(17) The concatenation of stems in /st/ + *ion* (e.g. *suggest-ion*)

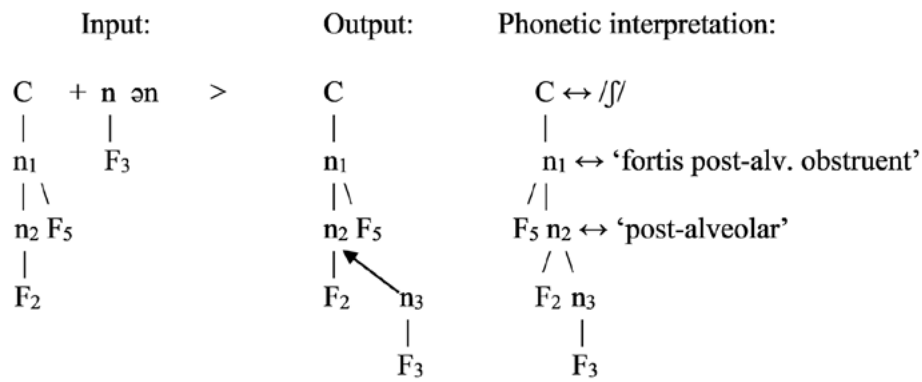


In this case, the bottom-most branching node is again interpreted by convention ((11d), $\{F_2;F_3\} \leftrightarrow$ 'post-alveolar'). The feature attached to the higher branching

node further contributes to the label 'occlusion' and the output is realized as the post-alveolar non-continuant, i.e. a non-fortis /tʃ/ or a voiceless version of the /dʒ/.[4]

The concatenation of the stems which terminate in the spirant /s/ (10c) in the lexical representation with the relevant affixes is illustrated in (18).

(18) The concatenation of stems in /s/ + *ion* (e.g. *agress-ion*)



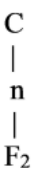
The output is realized as /ʃ/ by virtue of conventions ((11d), $\{F_2;F_3\}$ ↔ 'post-alveolar') and ((11a), $\{F_5\}$ ↔ 'fortis obstruent').[5]

3.2. The triplets in /d/, /s/ and /ʒ/ (*conclude – conclusive – conclusion*)

As noted above, the stems found in (9d) give rise to triplets terminating in /d/ if they are verbs, /s/ if they are followed by the affix *-ive* and the post-alveolar spirant in the presence of the affix *-ion*. The *comprehend/expand* subclass (9d') constitutes a somewhat different set of cases, as they surface with /ʃ/ before the affix *-ion*. This subclass will be analyzed in Subsection 3.3.

Let us postulate that stems such as *conclude* terminate in an abstract, i.e. phonetically uninterpretable, 'coronal' consonant specified only for feature F_2 (see (19)).

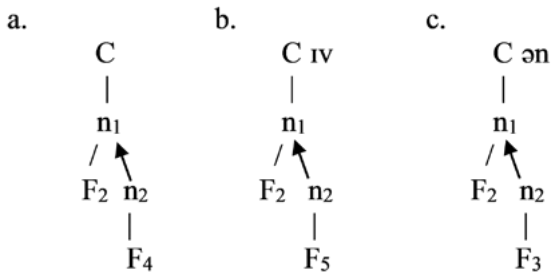
(19) The final segment in stems such as *conclude – conclusive – conclusion*



Let us further assume that morphosyntactic head v, which categorizes the class of stems including *conclude*, *comprehend, expand*, etc. and contributes to their interpretation as verbs, has the floating node hosting feature F_4 ('occlusion')

as its sole phonological exponent. At the same time, the affix *-ive*, categorizing the relevant stems as adjectives, is lexically represented with a floating node and feature F_5 ('fortis obstruent').[6] When concatenated with the stems which terminate in the consonant represented as (19), the floating nodes anchor as sisters to feature F_2 (see (20)).

(20) The derivation of items such as *clonclu*/d/*e*, *conclui*/s/*-ive* and *conclu*/ʒ/*-ion*



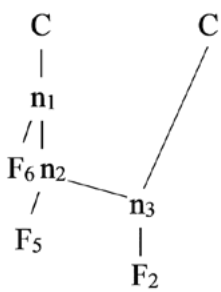
The branching nodes in (20a) and (20b) are interpreted by virtue of specialized conventions ((11e), $\{F_2;F_4\} \leftrightarrow$ 'alveolar non-continuant') and ((11f), $\{F_2;F_4\} \leftrightarrow$ 'fortis alveolar obstruent'). Ultimately, they are realized as /d/ and /s/, respectively. The representation in (20c) is translated into phonetic labels by convention ((11d), $\{F_2;F_3\} \leftrightarrow$ 'post-alveolar') and pronounced as /ʒ/.

3.3. The stems in -nC (*comprehend/expand*, *condescend*)

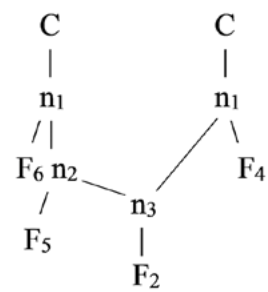
This section discusses the derivation of the stems which terminate in a cluster of a nasal consonant and the lenis alveolar obstruent, i.e. classes (9d') and (9e). In both those classes, the cluster is obligatorily homorganic. This fact is represented structurally as the nasal consonant and the following obstruent sharing the node which hosts feature F_2 (see (21)).

(21)

a. *comprehend/expand*-class



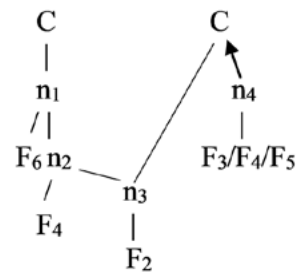
b. *condescend*-class



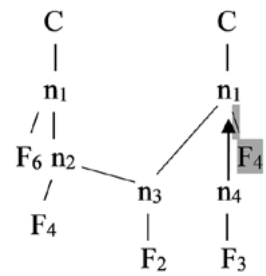
Feature F_6 contributes the property of nasality by virtue of the interpretational convention $\{F_6\} \leftrightarrow$ 'nasal.'[7] Importantly, the Featural Complexity Parameter setting for English prevents floating nodes from docking as the sister to feature F_2, as that would result in the nasals hosting more than 3 features. Thus, the floating nodes are forced to anchor onto the higher nodes. In the *comprehend/expand* class (21a), this is node C. The anchoring results in the representation (22a) and the phonetic interpretation takes place at node C. In the *condescend* class, the floating node docks onto the node hosting feature F_4. The phonological grammar avoids the derivation of the ternary branching structure by removing feature F_4 (see (22b)).

(22)

a. *comprehend/expand*-class



b. *condescend*-class

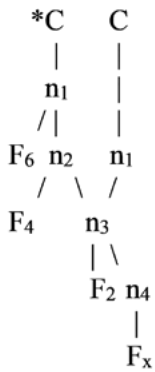


Given the interpretational conventions we have been working with so far, the structures featuring F_3 in (22) should be realized as /ʒ/. This is the case, as the only way for a structure to be interpreted as a 'fortis obstruent' is for it to contain feature F_5. However, such a result is incorrect. As a matter of fact, the contrast between the lenis /ʒ/ and fortis /ʃ/ is neutralized in favor of the latter after /n/ in English. English does not feature clusters such as /nʒ/.[8]

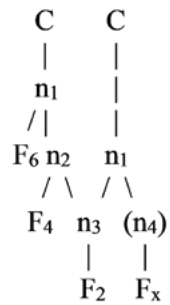
Let us propose that the absence of such partially homorganic nasal-lenis fricative clusters is a consequence of two factors. The first of them is the Featural Complexity Parameter setting, which says that English sets a limit to the number of features in consonants: a consonant cannot contain more than three features. As a consequence, if a following consonant has a complex featural specification, i.e. if there is more than one feature which occupies the bottom-most branching node, the features which define the nasal cannot dominate two such features. There are two series of fricatives in English which have complex place specification. These are dental fricatives /θ/ and /ð/ and the post-alveolars. Thus, a cluster composed of a nasal and post-alveolar/dental fricative is not represented as in (23a). Instead, it must be represented as in (23b).

(23)

a. illegal n+C cluster



b. legitimate n+C cluster



In sum, in the case of complex specification, a nasal and a following segment may only form a partial geminate, never a 'full' geminate.[9]

Let us further propose that the presence of the fortis realization of the relevant fricatives after nasals is a consequence of the availability of two interpretational conventions illustrated in (24).

(24)

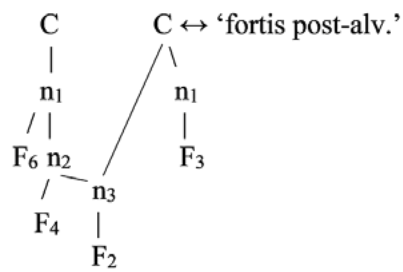
a. $\{F_2 > F_3\} \leftrightarrow$ 'post-alveolar'

b. $\{F_2;F_3\} \leftrightarrow$ 'fortis post-alveolar'

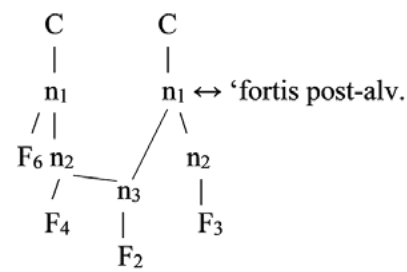
Convention (24a) says that a node which hosts a dominance relation between features F_2 and F_3 contributes the label 'post-alveolar' to the phonetic interpretation of the structure. According to (24b), the node under which features F_2 and F_3 simply co-occur, contributes the label 'fortis post-alveolar.'[10] Since the structures in (22) do not contain the asymmetrical relation of dominance between features F_2 and F_3, it is the convention (24b) which supplies the phonetic labels for the relevant nodes. This is illustrated for the two subclasses in (25).

(25)

a. *comprehens-ion/expans-ion*-class



b. *condescens-ion*-class

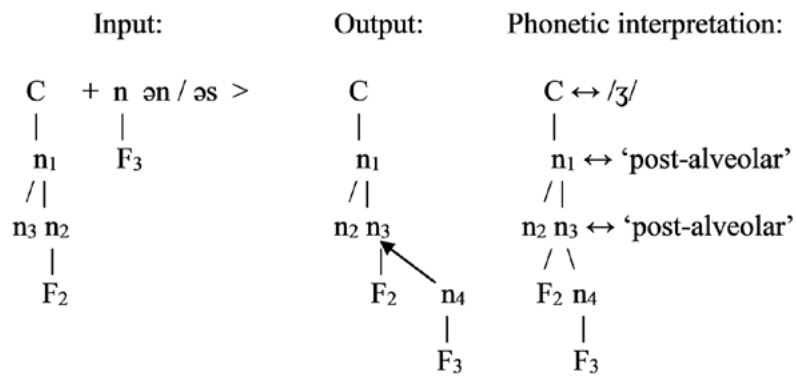


In sum, in items such as *comprehend/expand* and *condescend*, the final consonant mutates into the fortis fricative because the floating node cannot anchor in a manner that would contribute an additional feature to the structure of the nasal, as such a structure would contain four features. Due to the availability of conventions (24), the resulting representations are interpreted as fortis.

3.4. The *revise – revision* class

On the assumption that the stems from the *revision*-class terminate in an underlying /z/, the integration of the floating node proceeds as presented in (26).

(26)



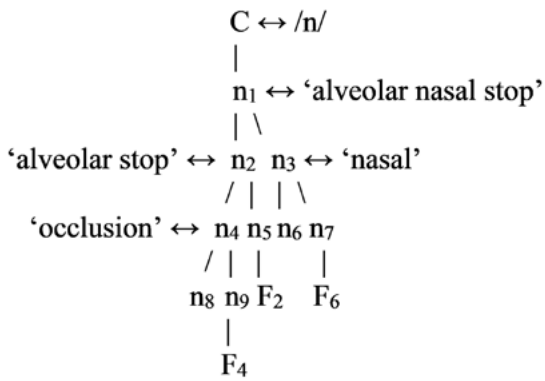
The floating node anchors as the sister to feature F_2. The resulting branching node is interpreted as post-alveolar by convention (24a). The relevant consonants are realized as the spirant /ʒ/.

3.5. Segments that resist the mutation

The stems that resist mutation in the context of affixes *-ion*, *-ian* and *-ious* do it for several reasons. However, all those reasons follow from the general conditions that define a well-formed subsegmental structure discussed in Section 2 as well as the requirement that a phonological representation be phonetically interpretable.

The nasal /n/ attested in the item *compan-ion* is structurally too large to be enriched with another feature. Let us postulate that the structure of the nasal is as presented in (27).

(27) The representation of /n/ in English

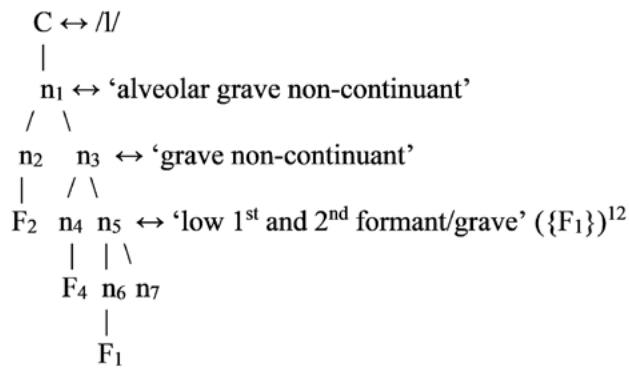


The augmentation of any of the layers of the structure in (27) with an additional feature is at odds with the Featural Complexity Parameter setting for English. At the same time, a removal of the lexically available features is precluded as it leads to the derivation of an ill-formed or uninterpretable structure. For example, if the floating node were to anchor onto node n_9 in (27) with simultaneous removal of feature F_4, the resulting structure would contain a non-branching non-terminal node n_9. The removal of features F_2 and F_4 would result in the derivation of structures of the kind presented in (4c) and (4d), i.e. structures in which empty nodes are locally dominated by other empty nodes. In sum, a node carrying feature F_3 cannot be added to structure (27).

The lateral (28) is also structurally complex and resists enrichment with an autosegmental node. The featural make-up of the lateral is meant to account for the processes resulting in the weakening of /l/ attested in different accents of English: the darkening of the clear /l/ to /ɬ/ and the vocalization of /ɬ/ to /w/. The former process is analyzed as the detachment of feature F_2, while the latter is expressed as a removal of feature F_4. The representation in (28) illustrates a clear /l/ which may undergo a positional darkening to /ɬ/.[11]

As in the case of the nasals, a floating node cannot be integrated with the structure (28) without the violation of the general well-formedness conditions or the derivation of an uninterpretable structure.

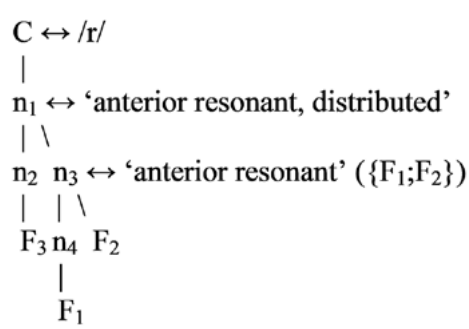
(28)



For instance, the anchoring of a floating node onto node n_4 with a simultaneous removal of feature F_2 would result in the absence of the unambiguously bottom-most branching node (as in (4b) above) and the phonetic interpretation algorithm could not commence. The anchoring of the floating node onto node n_2 would preclude the removal of feature F_2, as that would produce a unary branching non-terminal node n_2. Importantly, the features F_4 and F_1 cannot be removed from the representation (28) as that would result in configurations in which empty nodes (here: n_4 and n_6 or n_6 and n_7) find themselves in an illegal configuration defined in Section 2.1 (see (4c), (4d) above). Furthermore, the anchoring of the floating node onto node n_6 would result in node n_6 becoming a branching node dominating features F_1 and F_3. Crucially, English does not have an interpretational convention which would supply phonetic content for the combination $\{F_1; F_3\}$. Thus, the direct combination of features F_1 and F_3 is uninterpretable in English, and its presence would lead to the crash at the phonology–phonetics interface.[12]

At the same time, features F_1 and F_3 may be found in a single segment in English, provided that they are not interpreted phonetically at the same cycle.[13] Here we postulate that the case at hand is the rhotic, whose postulated representation is found in (29).

(29)

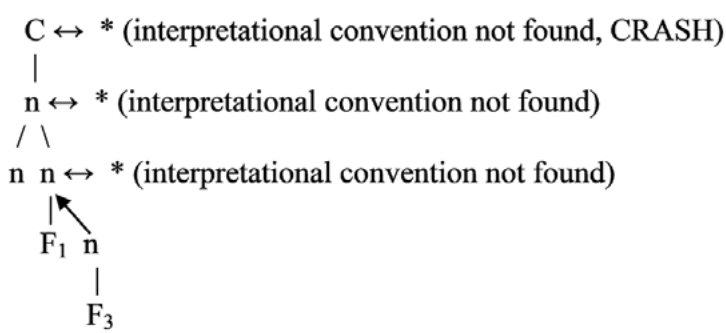


The presence of F_3 in the rhotic is justified given that, in a great many varieties of English, coronal obstruents tend to be pronounced with an alveo-palatal place of articulation, when followed by /r/, cf. *misery* rendered as *mi*/ʒr/*y*, *grocery* as *gro*/ʃr/*y* (see Wells 2008). Feature F_1 contributes to the 'rounded/labial' quality in the phonetic interpretation of the English rhotic.[14]

The anchoring of the floating node carrying feature F_3 onto the representation in (29) must necessarily involve the delinking of the lexically present feature F_3, as a segment cannot carry two manifestations of the same feature. Consequently, the floating node cannot anchor onto node n_2, which would end up as a unary non-terminal node. The attachment of the floating node to the nodes n_3 and n_4 would result in the derivation of the uninterpretable combination of features F_1 and F_3.[15] In sum, given the general assumptions, the floating node carrying feature F_3 cannot be integrated into representation (29).

The stems, such as *envious* /ɛnv(i)jəs/ and *oblivion* /ɒblɪv(i)jən/, resist mutation for a similar reason. The anchoring of the floating node hosting feature F_3 gives rise to a structure which cannot be interpreted phonetically due to the fact that the English lexicon does not contain an interpretational convention that would realize the output.

(30) The anchoring of the floating node onto /v/

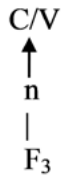


The possibility of the derivation crashing due to the unavailability of the interpretational convention is a consequence of the minimalist assumption that the (phonological) grammar exists as an optimal means of meeting the interface condition of interpretability.

If the floating node cannot anchor onto the preceding stem-final obstruent, it is realized as a vowel /i/ or an approximant /j/, giving rise to forms such as *rebe*/liən/~/ljən/, *Arthu*/riən/~/rjən/, *ceremo*/niəs/~/njəs/, etc. We propose to account for these realizations as instantiating the last resort strategy, which involves an epenthesis of a node. Note that if a floating node carrying a feature simply anchors onto a skeletal position (31), the resulting structure is not phonetically interpretable, due to the unavailability of a branching node. The operation of node epenthesis

allows the floating structure to form a branching node, which may be subject to phonetic interpretation.

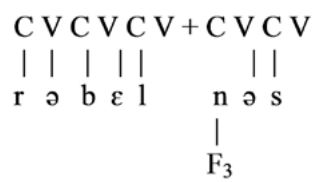
(31) A phonetically uninterpretable structure



The current analysis follows the core assumptions of the approach to syllabic structure formulated in Lowenstamm (1996), Scheer (2004) and Cyran (2010), among others. The framework known as Strict CV Phonology assumes that CV is the only syllable type and all languages represent their morphemes using the sequences of C positions followed by V positions, which give rise to consonants and vowels, respectively. Consequently, words which terminate in consonants do so only superficially, and at the syllabic or skeletal level they terminate in empty V positions, i.e. V positions devoid of subsegmental structure. Similarly, vowel-initial items in fact begin in empty C positions.

Given these general assumptions, the stem *rebel* terminates in an empty V position and the affix *-ious*, begins with an empty C. The concatenation (+) of these two morphemes is represented as in (32).

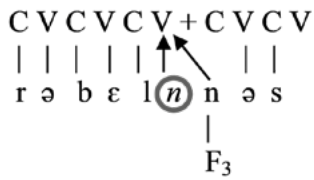
(32)



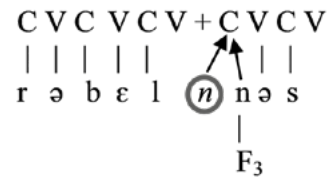
Since the floating node cannot anchor onto the /l/, as the latter is structurally too complex, the grammar invokes a last resort operation of node epenthesis, which allows the floating structure to be realized on the empty V, which is pronounced as /i/. Alternatively, the floating node may anchor onto the affix-initial C position, which, in such a case, is realized as /j/. The two representational options are illustrated in (33). The epenthetic nodes are italicized and encircled.

(33) Node epenthesis

a. *rebe*/liəs/



b. *rebe*/ljəs/



In sum, stems which terminate in the nasal, the lateral, the rhotic and labials resist mutation for two reasons. Either their structures cannot be augmented with a floating node without violating general conditions constraining the shape of the subsegmental phonological representation or the anchoring of the floating node results in the derivation of the uninterpretable structure.

Whenever the floating node cannot dock onto the target consonant, the grammar employs the last resort strategy of node epenthesis, which allows the floating node to dock onto the empty syllabic positions, deriving the representations with vowel /i/ or consonant /j/.

4. Conclusion

This paper demonstrated that a substance-free autosegmental framework, which employs privative features, can successfully account for the complex and opaque pattern of coronal mutation in English. The attested outputs of the mutation are a function of several factors. The first of them is the size of the floating material that anchors onto the stem-final segments to provoke the mutation. In the case of coronal mutation, this is a node and a feature. The second factor is the lexical representation of the stem-final consonants. If the target consonant is structurally too complex, the anchoring may require adjustments, as in many cases of the underlying /t/. If the lexical representation makes no available place for the anchoring of a node (see (27), (28)), or if they already contain an instance of the anchoring feature, the docking may be blocked. Finally, the docking may be blocked if it gives rise to a phonetically uninterpretable consonant. In English, the combination of the features F_1, which typically contributes 'grave/labial/rounded/anterior non-coronal/low 1st and 2nd formant' phonetic label and F_3 ('post-alveolar/palatal/distributed/high 2nd formant') does not correspond to any interpretational convention. In other languages, however, the combination of similarly interpreted features may be realized as a palato-labial consonant (Polish) or a front-rounded vocoid (French).

According to the approach outlined in this contribution, sound systems differ in the availability of particular interpretational conventions and the settings of the Featural Complexity Parameter. The exact limits that these two factors set on the possible shapes of sound systems seem to be an exciting and promising direction of research.

Notes

1 A reviewer inquires about the reasons for treating pairs such as *revise – revision* and *impress – impression* as cases of 'mutation,' while the pairs such as *sex – sexual* are analyzed as 'palatalization.' The reason for analyzing the effects of affix *-ual* and *-ion* as triggering a different type of process is the effect they have on coronal plosives. While the former appears to turn alveolar plosives into post-alveolar affricates (e.g. *concep*/t/ – *concep*/tʃ/*-ual*), the latter produces post-alveolar spirants (e.g. *concep*/t/ – *concep*/ʃ/*-ion*). In this work, the terms 'palatalization' and 'mutation' are used simply as labels that distinguish between the two processes and carry no information about their nature.

2 See van Oostendorp (2006) and Trommer (2011) for proposals as to why floating autosegments favor heteromorphemic over tautomorphemic anchoring.

3 Features F_4 and F_5 in (14) cannot be deleted for similar reasons.

4 In careful speech, the initial sound of the cluster, i.e. /s/, may resist assimilation and be pronounced with the alveolar place of articulation. In such cases, an alternative representation is proposed in which /s/ and /t/ do not share the node hosting the feature F_2 ('alveolar').

5 The derivation of the pair *confu*/z/*e – confu*/ʒ/*-ion* proceeds in an analogous fashion with the output interpreted as lenis due to the absence in the structure of the feature $\{F_5\}$ ('fortis obstruent').

6 In fact, we have found only one stem which features a lenis obstruent and the affix *-ive*. This is the stem *gerund*, which gives rise to the word *gerundive.* This word clearly differs semantically from most other words featuring the affix *-ive* and its status as an adjective is questionable. Given these facts, one might be inclined to assume that the affix *-ive* attested in *gerundive* is a different affix than the one attested in items such as *conclusive*, *extensive, investigative*, etc.

7 In (21), the representation of nasals is structurally simplified for clarity of exposition. We present the complete structure of nasals in (27) in Section 3.5.

8 This generalization holds for the established English vocabulary, including fully assimilated loans. As pointed out by one of the reviewers, certain recent borrowings may not conform to it. Thus, the French loanword *mélange* ends in /nʒ/. It is possible that such words as well as other lexical exceptions are simply not represented as partial geminates at all.

9 As pointed out by one of the reviewers, English shows a general dispreference for geminates. Why this is the case is beyond the scope of this work. Still, for the framework outlined here, the reason why English avoids geminates must be extra-grammatical. A hypothetical language which differs from English only in that it has geminates is not an impossible language from the point of view of the phonological grammar.

10 Since dental fricatives do not participate in the relevant mutations, we postulate a convention $\{F_1 > F_2\}$ to contribute the label 'dental' in both lenis /ð/ and fortis /θ/. The latter additionally contains feature F_5, interpreted as 'fortis obstruent.'

11 The dark /ɬ/ in accents which feature vocalization has the representation similar to (28). The only difference is that features F_2 and F_4 switch their respective places. As a consequence, the phonetic interpretation of the relevant structure proceeds in a different way with the label 'alveolar grave' supplied at node n_3.

12 A reviewer points out that the role of the interpretational conventions is simply to make sure that certain combinations of features are impossible in a given language. While it is true that the effect of the (absence of) interpretational conventions in a given system is the same as the effect of feature co-occurrence restrictions known from other frameworks, the difference between the two devices is that interpretational conventions are not part of phonology but rather part of the lexical knowledge of the speaker. In fact, this work postulates that feature co-occurrence restrictions do not exist.

13 In other words, F_1 and F_3 may be part of the same representation so long as the interpretation does not have to invoke an interpretational convention $\{F_1;F_3\} \leftrightarrow$ '…,' which does not exist in English.

14 Different degrees of lip protrusion are commonly observed to accompany the production of /r/ in both the rhotic and non-rhotic varieties of English (see King and Ferragne 2020 for a recent instrumental investigation and references). The 'labial' quality of the English rhotic is most vividly manifested in the varieties of South-East London English, which realize the rhotic as the labio-dental approximant /ʋ/ (see Tollfree 1999, 174).

15 Note that the attachment of the floater to the node n_3 would have to be accompanied by the delinking of the feature F_2 to avoid the formation of a ternary branching node.

References

Backley, Phillip. 2011. *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge: MIT Press.

Chomsky, Noam. 2000. "Minimalist Inquiries: The Framework." *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Ed. Roger Martin, David Michaels, and Juan Uriagereka. Cambridge: MIT Press. 89–155.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Clements, George. N. 1985. "The Geometry of Phonological Features." *Phonology Yearbook* 2: 225–252.

Clements, George. N., and Elizabeth Hume. 1995. "The Internal Organization of Speech Sounds." *Handbook of Phonological Theory*. Ed. John Goldsmith. Oxford: Basil Blackwell, Oxford. 245–306.

Cyran, Eugeniusz. 2010. *Complexity Scales and Licensing in Phonology.* Berlin & New York: Mouton de Gruyter.

Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise.* Oxford: Oxford University Press.

Halle, Morris. 2005. "Palatalization/Velar Softening: What It Is and What It Tells Us about the Nature of Language." *Linguistic Inquiry* 36.1: 23–41.

Halle, Morris, and K. P. Mohanan. 1985. "Segmental Phonology of Modern English." *Linguistic Inquiry* 16.1: 57–116.

Harris, John, and Geoff Lindsey. 1995. "The Elements of Phonological Representation." *Frontiers of Phonology: Atoms, Structures, Derivations*. Ed. Jacques Durand, and Francis Katamba. London & New York: Longman. 34–79.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar.* Oxford: Wiley-Blackwell.

Hulst, Harry van der. 2020. *Principles of Radical CV Phonology: A Theory of Segmental and Syllabic Structure.* Edinburgh: Edinburgh University Press.

Kang, Seouk-Keun. 2000. "An Optimality Theoretic Account of Phonological Opacity in English." *Studies in Phonetics, Phonology and Morphology* 6.2: 307–332.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1985. "The Internal Structure of Phonological Elements: A Theory of Charm and Government." *Phonology Yearbook* 2: 305–328.

King, Hannah, and Emmanuel Ferragne. 2020. "Loose Lips and Tongue Tips: The Central Role of the /r/-typical Labial Gesture in Anglo-English." *Journal of Phonetics* 80: 100978. DOI: 10.1016/j.wocn.2020.100978

Lowenstamm, Jean. 1996. "CV as the Only Syllable Type." *Current Trends in Phonology. Models and Methods.* Ed. Jacques Durand, and Bernard Laks. Salford, Manchester: European Studies Research Institute, University of Salford. 419–441.

Mielke, Jeff. 2008. *The Emergence of Distinctive Features.* Oxford: Oxford University Press.

Nasukawa, Kuniya. 2017. "Extending the Application of Merge to Elements in Phonological Representations." *Journal of the Phonetic Society of Japan* 21.1: 59–70.

Nasukawa, Kuniya. 2018. "The Phonological Shape of Morphemes Constructed by Merge." *Phonological Externalization*. Ed. Hisao Tokizaki. Volume 3. Sapporo: Sapporo University. 1–8. DOI: 10.24467/onseikenkyu.21.1_59

Odden, David. 2022. "Radical Substance-free Phonology and Feature Learning." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4: 500–551. DOI: 10.1017/cnj.2022.10

Oostendorp, Marc van. 2006. *A Theory of Morphosyntactic Colours*. Amsterdam: Meertens Institute.

Pöchtrager, Markus A. 2015. "Binding in Phonology." *Representing Structure in Phonology and Syntax*. Ed. Marc van Oostendorp, and Henk van Riemsdijk. Berlin: Mouton de Gruyter. 255–275.

Radomski, Marek. 2010. "An Optimality Theory Analysis of Velar Softening in English." *Lubelskie Materiały Neofilologiczne* 34: 65–80.

Rubach, Jerzy. 1984. "Segmental Rules of English and Cyclic Phonology." *Language* 60.1: 21–54.

Sagey, Elizabeth Caroline. 1986. "The Representation of Features and Relations in Non-linear Phonology." PhD diss, Massachusetts Institute of Technology.

Schane, Stanford, A. 1984. “The Fundamentals of Particle Phonology.” *Phonology Yearbook* 1: 129–156.

Scheer, Tobias. 2004. *A Lateral Theory of Phonology. What is CVCV, and Why Should It Be?* Berlin: Mouton de Gruyter.

Scheer, Tobias. 2022. “3 x Phonology.” *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4: 444–499. DOI: 10.1017/cnj.2022.22

Schwartz, Geoffrey. 2010. “Phonology in the Speech Signal – Unifying Cue and Prosodic Licensing.” *Poznań Studies in Contemporary Linguistics* 46.4: 499−518.

Tollfree, Laura. 1999. “South East London English: Discrete versus Continuous Modelling of Consonantal Reduction.” *Urban Voices. Accent Studies in the British Isles.* Ed. Paul Foulkes, and Gerard J. Docherty. London: Arnold. 163–184.

Trommer, Jochen. 2011. “Phonological Aspects of Western Nilotic Mutation Morphology.” Habilitation thesis, University of Leipzig.

Wells, John. 2008. *Longman Pronunciation Dictionary*. 3rd ed. Harlow: Pearson/Longman.

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by the Authors, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.05

Bert Botma
https://orcid.org/0000-0002-6764-9782
Leiden University

Hamza Khwaja
https://orcid.org/0009-0000-8516-2567
Leiden University

Joseph Smith
https://orcid.org/0009-0006-6378-7009
Leiden University

A Stratal Approach to English Affricates

Abstract: There is a general consensus in the phonological literature that the English affricates /tʃ/ and /dʒ/ are unitary segments rather than clusters of a stop and a fricative. However, for /tʃ/, not all the evidence is consistent with this approach. While /tʃ/ patterns as a single segment in linguistic performance, some authors on British pronunciation have pointed out that glottalization targets /tʃ/ as if it were a cluster (Higginbottom 1964; Gimson and Cruttenden 2001). This paper proposes that the conflicting evidence can be resolved under a stratal model (Bermúdez-Otero 2018a). We argue that /tʃ/, in the relevant dialects, is a cluster at the stem level but surfaces as a monosegmental affricate in the course of the derivation. The behaviour of /dʒ/, on the other hand, suggests that it is a single segment at all levels of representation.

Keywords: affricates, phonological theory, Stratal Optimality Theory, gemination, glottal reinforcement

1. Introduction

While the internal structure of the English affricates /tʃ/ and /dʒ/ has been a subject of debate (see Lin 2011; van de Weijer 2014), it is generally assumed that they

are unitary segments rather than clusters of a stop and a fricative. This assumption is consistent with native speakers' intuitions and is supported by evidence from linguistic performance (Section 3). However, not all evidence supports the single-segment approach. As some authors on British pronunciation have pointed out (Higginbottom 1964; Gimson and Cruttenden 2001), glottalization targets /tʃ/ as if it were a cluster, applying even in intervocalic contexts, where /p/, /t/, and /k/ are never glottalized in standard accents (Section 4). This behaviour has not received a satisfactory explanation under the single-segment approach; Gussenhoven (1986), in an attempt to extend Kahn's (1976) framework to British English, resorts to an ad hoc rule of affricate gemination to account for the glottalization of intervocalic /tʃ/. We believe that the apparent conflict between the behaviour of /tʃ/ as a single segment at the level of performance but as a cluster in some phonological patterns can be resolved under a stratal model. Adopting a Stratal Optimality Theory framework (Bermúdez-Otero 2012; 2018a), we argue that /tʃ/ is a cluster at the stem level – the level of representation at which forms are stored in the mental lexicon – but coalesces to a monosegmental affricate in the phonological output (Section 5). This analysis requires /tʃ/ to undergo coalescence in the course of the derivation, paralleling the more familiar coalescence of /t, d, s, z/ + /j/ clusters (yod coalescence). Of course, this conclusion is applicable only to those varieties of English with the glottalization pattern under discussion. Moreover, we do not extend our analysis of /tʃ/ to its lenis counterpart /dʒ/, as we find additional evidence for a monosegmental /dʒ/ at the stem level (Section 6).

2. English affricates and their distribution

Phonetically, affricates are stops followed by homorganic frication. At least six such sounds can be identified in English (Jones 1969, 160). They are listed in (1) along with their distribution in monomorphemic words.

(1) *Phonetic affricates in English*

	initial	medial	final
/tʃ/	*chess*	*kitchen*	*match*
/dʒ/	*joke*	*badger*	*edge*
/tr/	*train*	*petrol*	
/dr/	*drum*	*hydra*	
/ts/		*pizza*	*quartz*
/dz/		*Pudsey*	*adze*

To this list we might add the dental affricate [t̪θ] of *eighth* and, for some speakers, its voiced counterpart [d̪θ] or [d̪ð], as in *width*, but these are limited to final position in words with a *-th* suffix (Gimson 1962, 166; Ward 1945, 138).

The phonotactic distribution of /tʃ/ and /dʒ/ presents mixed signals as to their segmental status. It is immediately apparent that /tʃ/ and /dʒ/ have a freer distribution than the other phonetic affricates. Like other stop-liquid and stop-glide clusters, /tr/ and /dr/ must be followed by a vowel, while /ts/ and /dz/, like other stop-sibilant clusters, are banned from word-initial position.[1] This distributional freedom seems to count in favour of a single-segment analysis of /tʃ/ and /dʒ/. Yet their distribution is still not as free as that of other single obstruents. Van de Weijer (2014) observes that while both stops and fricatives can combine with a following liquid or glide in onset clusters (*pure*, *brew*, *twice*, *drip*, *quick*, *glow*, *feud*, *three*, *slow*, *shrewd*, *Vladimir*, *Zwicky*), no such possibilities are available to /tʃ/ and /dʒ/. If these affricates are analyzed as clusters, this restriction would follow from a general restriction limiting complex onsets to two segments, a restriction that can also be observed, for example, in the history of yod dropping: traditional RP retains /j/ after #l, as in *lewd* and *lurid*, but has lost it after #Cl, as in *blue* and *clue* (Szigetvári 2016).[2] However, van de Weijer argues that the absence of onset clusters containing /tʃ/ and /dʒ/ can be accounted for by their complexity at the sub-segmental level, and need not lead to the conclusion that they are bisegmental.

It should be noted that this restriction no longer applies, at least at the phonological output level, for innovative speakers who have reanalyzed /tr/ and /dr/ as /tʃr/ and /dʒr/. This reanalysis is apparent in children's spellings such as ⟨CHRAY⟩ for *tray* and ⟨JRAGIN⟩ for *dragon* (Read 1971; Treiman 1993). Evidence for this change in the adult grammar can be found in a common pattern of /s/-retraction where /s/ has shifted to /ʃ/ before /tʃ/. This change is not just triggered by etymological /tʃ/, as in *mischief*, but is also fed by yod coalescence, as in *stupid* |(#stj > #ʃtʃ), and by pre-/r/ affrication, as in *strong* (#str > #ʃtʃr). The lack of (categorical) retraction in the clusters /spr/ and /skr/ shows that retraction in /str/ cannot be triggered directly by the /r/ across the intervening consonant, and hence must be triggered by the adjacent affricate (Bailey et al. 2022). English speakers tend to affricate /tr/ and /dr/ automatically in loanwords and when speaking foreign languages, suggesting that pre-/r/ affrication remains an active phonological process. It is therefore conceivable that speakers with [tʃr] and [dʒr] in the output still have /tr/ and /dr/ in stem-level representations, in which case van de Weijer's generalization would still hold at the stem level.

The affricate /tʃ/ is phonetically distinct from the cluster of /t/ + /ʃ/ that arises due to morpheme concatenation. This distinction, which can be observed by comparing e.g. *scorching* with *courtship*, has been taken by some authors as evidence for the monosegmental status of the affricate /tʃ/ (e.g. Jensen 1993, 30; van de Weijer 2014). However, this /t/ + /ʃ/ cluster arises only in compounds and in derivations with the *-ship* suffix, which likely have the prosodic structure of compounds.[3] These

/t/ + /ʃ/ sequences therefore always have a prosodic-word boundary between the two consonants. Note that a similar phonetic distinction can be observed between /tr/ and /t/ + /r/ in *fortress* vs. *courtroom*, and between /dr/ and /d/ + /r/ in *kindred* vs. *mind reader* (Gimson 1962, 167; Wells 1990). In the speech of one of the authors (JS), there is also a distinction between /t/ + /s/, as in *hot sauce*, which is realized with obligatory glottal replacement of the /t/ ([ˈhɔʔ͡sɔːs]), and /ts/, as in *pizza*, where glottal replacement is optional ([ˈpɪjʔsə] ~ [ˈpɪjʔtsə]). Since prosodic boundaries can condition the application of phonological rules, we do not think that these distinctions are informative regarding the segmental status of affricates.

3. Evidence from linguistic performance

Fromkin (1971) observes that /tʃ/ and /dʒ/ behave as indivisible units in speech errors. She records several cases where speech errors target the constituent consonants of onset clusters individually, e.g. *fish grotto* → *frish gotto*, *brake fluid* → *blake fruid*. Yet she finds no case where the stop or fricative component of /tʃ/ or /dʒ/ was deleted or transposed independently. While Cruttenden (in Gimson and Cruttenden 2001) and van de Weijer (2014) cite Fromkin's findings in favour of the single-segment analysis without further qualification, Fromkin herself draws the following conclusion:

> One may assume that the old phonemic controversy, as to whether such affricates should be considered one segment or two, is solved *for linguistic performance*, and that affricates should be considered single segments *in the production of speech*, for speakers of English. (Fromkin 1971, 33, emphasis added)

Since speech errors canonically involve transposition of segments belonging to different words, we can conclude, given fairly standard assumptions about modularity (see e.g. Bermúdez-Otero 2012), that they occur within or after the phrase-level module. Though we are not sure whether speech errors manipulate the word-level output or the phrase-level output, we can be confident that they do not manipulate stem-level representations.

Cowan et al. (1985) study the productions of "fluent backward talkers" – speakers who can reliably reverse the order of segments in words at or near a normal speech rate. They find that these speakers keep affricates, as well as diphthongs, in their original order, e.g. *join* [dʒɔɪn] → [nɔɪdʒ], whereas they usually reverse the order of consonants in clusters, e.g. *ghost* [go^{u}st] → [tso^{u}g].[4] Since reversal operates on fully inflected word-forms, e.g. *aching* [ekɪŋ] → [ŋɪke] ~ [g(ə)nɪke], it seems most likely that the input to reversal is the output of the word-level module, perhaps with some influence from orthography. Cross-linguistic surveys of language games (see especially Bagemihl 1988, 413–488) suggest that the word-level output is a typologically common input to language games.

Textbooks often appeal to native speakers' intuitions as an argument for the monosegmental status of affricates. Gimson (1962, 168) writes that native speakers "consider that *chip*, *catch*, consist of three parts in the same way as *tip*, *ship*, or *cat*, *cash*." Dividing words into segments is in essence a simple language game, so like most other language games, it likely has access to word-level representations. We conclude that evidence from language games and native speaker intuitions, along with evidence from speech errors, gives us good reason to believe that /tʃ/ and /dʒ/ are single segments in the word-level output.

4. Glottalization of /tʃ/ in British English

Support for the bisegmental status of /tʃ/ comes from a pattern of glottal reinforcement found in some varieties of British English, including RP.[5] Based on the descriptions in, among others, O'Connor (1952), Higginbottom (1964), Roach (1973), Wells (1982), and Gussenhoven (1986), the conditions under which reinforcement of /tʃ/ occurs can be summarized as in (2):

(2) *Glottal reinforcement of* /tʃ/
a. /tʃ/ is reinforced word-internally following a stressed syllable and between two unstressed syllables, unless preceded by an obstruent: *orchard* [ˈɔːʔtʃəd], *furniture* [ˈfɜːnɪʔtʃə], *pilchard* [ˈpʰɪlʔtʃəd], but *capture* [ˈkʰaʔptʃə].
b. /tʃ/ is not reinforced when followed by a stressed vowel, whether it bears primary or secondary stress: *achieve* [əˈtʃiːjv], *kerchief* [ˈkʰɜːˌtʃiːjf].
c. /tʃ/ is always reinforced word-finally, regardless of stress, while word-initial /tʃ/ is never reinforced, even if the preceding word ends in a vowel: *teach Anne* [ˈtʰiːjʔtʃ ˈan], but *the Chilean* [ðə tʃiˈlɛjən].

Glottal reinforcement of /tʃ/ applies in a wider range of contexts than that of /p, t, k/: while /tʃ/ can be glottalized intervocalically, glottalization of word-medial /p, t, k/ requires the presence of a following consonant, e.g. *accident* [ˈaʔksɪdənt], *applicant* [ˈaʔplɪkənt]. This leads Cruttenden and Gimson (2001, 173) to note that "the [ʃ] element of /tʃ/ is in some sense the 'following consonant' which allows glottalization of the preceding [t]."

The distribution of glottalized /tʃ/ is strikingly similar to that of the flapped allophone of /t/ in American English. Compare the examples in (2) with the following American English examples, which have [ɾ] in the same set of environments: (a) *city* [ˈsɪɾi], *obesity* [oˈbisəɾi], but *victor* [ˈvɪktɚ]; (b) *attack* [əˈtʰæk], *latex* [ˈleˌtʰɛks]; (c) *let Ann* [ˌlɛˈɾæn], but *the tarantula* [ðə tʰəˈɾæntʃələ]. According to an analysis first proposed by Kahn (1976), flapping occurs whenever /t/ is ambisyllabic, provided certain segmental conditions are met.[6] Gussenhoven (1986) applies this analysis to British English /tʃ/. Following Kahn, he assumes that the syllabification rules of English make consonants ambisyllabic in certain contexts.

In addition to this, Gussenhoven proposes a rule of affricate gemination. This rule supplies /tʃ/ with an extra position on the timing tier when it is linked to the preceding syllable, that is, when it is ambisyllabic or a pure coda. According to this analysis, therefore, the glottalized affricates in (2) are geminates in the phonological surface form, i.e. *orchard* [ˈɔːʔttʃəd], *furniture* [ˈfɜːnɪʔttʃə], etc. Below, we follow this aspect of Gussenhoven's analysis, but we offer a different account of why affricate gemination occurs.

Gussenhoven's analysis predicts that /tʃ/ surfaces either as a singleton onset or as a geminate, but never as a singleton coda. A word-final /tʃ/ is geminated regardless of what follows, not just when the next word begins with a vowel.[7] The absence of simple coda /tʃ/ is consistent with the idea that the phonetic interpretation of fortis stops depends strictly on their position within a syllable. Coda /p, t, k/ in British English are characteristically glottalized and unreleased; /tʃ/, on the other hand, always has a pulmonic (fricative) release, even word-finally before a consonant (cf. e.g. Gimson 1962, 171). We conclude from this that the most natural account is one in which all instances of glottalized /tʃ/ are geminates.

The notion of ambisyllabicity has been challenged by several authors (see e.g. Jensen 2000; Harris 2004; 2013; Bermúdez-Otero 2011). Bermúdez-Otero argues that ambisyllabicity is an unnecessary representational device in a theory that incorporates the prosodic hierarchy in combination with a stratal architecture. An early analysis of /t/-flapping in this type of approach appears in Kiparsky (1979). In Kiparsky's analysis, a word-level rule lenites /t/ when it is not foot-initial and not preceded by a consonant. At the phrase level, word-final consonants are resyllabified as onsets when the following word starts with a vowel; lenited /t/ is realized as [ɾ] in onsets and as [ʔt] in codas. The distribution of geminate /tʃ/ can similarly be reformulated in terms of foot structure: at the word level, /tʃ/ is geminated whenever it is not foot-initial and not preceded by an obstruent.

As noted, Gussenhoven's analysis (along with its foot-based alternative) posits a rule of affricate gemination. However, this analysis does not explain why such a rule would exist, and why it targets affricates but not simple plosives. A cluster analysis, on the other hand, provides a straightforward motivation for why /tʃ/ is geminated, given that gemination also affects other cluster-initial stops. To our knowledge, Higginbottom (1964) is the first to argue explicitly for such an analysis, based on the parallel between glottal reinforcement of /t/ preceding /ʃ/ and of stops preceding a liquid or a glide:

> I have favoured treating /tʃ/ as a sequence of plosive plus fricative, i.e. as two phonemes, on account of its possibilities of reinforcement. This would remove the anomaly of the reinforcement of /tʃ/ before vowels, in a position where the plosives are not reinforced, since the reinforced plosive element would then be regarded as preceding a fricative; i.e. *teaching* would contain a plosive-fricative cluster, where reinforcement

> would be expected. The objection to this is the behaviour of /tʃ/ within the word with regard to stress; in this respect, it follows the plosive-liquid and -semivowel sequences in contrast to clusters of plosive-fricative. Also, for most speakers of RP, /tʃ/ would be the only plosive-fricative cluster in syllable initial position. (Higginbottom 1964, 141)

In our view, Higginbottom's second objection – that /tʃ/ is a possible syllable onset, while other plosive-fricative clusters are not – can serve as an answer to her first objection. She observes that the distribution of /tʃ/-glottalization resembles that of clusters consisting of a fortis stop plus a liquid or glide. In both cases, glottalization is sensitive to stress; we find glottal reinforcement before an unstressed syllable, e.g. *butcher* [ˈbʊʔtʃə], *petrol* [ˈpɛʔtrəl], *accurate* [ˈaʔkjʊrət], but not before a stressed syllable, e.g. *achieve* [əˈtʃijv], *attract* [əˈtrakt], *accuse* [əˈkjʉwz]. Under a bisegmental analysis of /tʃ/, the stress-sensitive pattern applies to just those clusters that form legal branching onsets, i.e. /pl, kl, pr, tr, kr, pj, tj, kj, tw[8], kw/ and /tʃ/. The cluster /tʃ/ just happens to be the only member of this set with a fricative as its second element. By contrast, a stress-insensitive pattern applies to all clusters that are not legal onsets, including the stop-liquid cluster /tl/: glottalization occurs in both *atlas* [ˈaʔtləs] and *Atlantic* [əʔtˈlantɪk], and in both *accident* [ˈaʔksɪdənt] and *accept* [əʔkˈsɛpt] (see Andrésen 1968, 107).

The affricate /tʃ/ parallels stop-sonorant sequences not only in the stress-sensitive conditioning of glottalization but also in the phonetic properties of the glottalized output forms. As first documented by Iles (1960) as characteristic of London speech, the fortis stop has a noticeably lengthened closure in these contexts, accompanied by phased laryngeal gestures – constricted at the approach (glottal reinforcement) and spread at the release, where /tʃ/ has a voiceless fricative portion and the stop-sonorant sequences have a devoiced liquid or glide. Iles's transcriptions include the following: *spatula* [ˈspætʔ͡tʃələ], *batchelor* [ˈbætʔ͡tʃələ], *teacher* [ˈtiːtʔ͡tʃə], *sacrifice* [ˈsækʔ͡kr̥ifais], *pantry* [ˈpæntʔ͡tr̥i], *culprit* [ˈkʌlp͡ʔpr̥it͡ʔ]. These are precisely the phonetic properties we would expect if these stops are geminates in the phonological output form – coda properties followed by onset properties, and increased duration.

While "fake geminates" can be found in some heteromorphemic contexts, e.g. *unnamed*, *guileless*, *bookcase*, English lacks a gemination contrast within the stem domain. Gemination of /tʃ/ and of fortis stops before a liquid or glide occurs in fully predictable contexts. We therefore assume that there are no geminates in English stem-level representations, and that gemination in these sequences is derived on-line by the word-level phonology. Below, we use the term "affricate gemination" to refer to the process that derives geminate /tʃ/, and "pre-sonorant gemination" to refer to gemination of /p, t, k/ before a liquid or glide. A formal account of these processes in Stratal Optimality Theory is presented in Section 5.

5. A stratal approach to English /tʃ/

Our analysis of affricate gemination builds on the analysis of pre-sonorant gemination in Khwaja et al. (2025; in press). Khwaja et al. observe that in varieties with pre-sonorant gemination, the medial stops in the words in (3a) have a phased realization: they are first glottalized, as is typical of codas, and then released with weak aspiration, as is typical of unstressed onsets (the weak aspiration is heard as devoicing of the following sonorant). In the words in (3b), on the other hand, the stops are initial in a foot and no gemination takes place.

(3)

a. *petrol*	[ˈpɛʔt.tr̥əl]	b. *attract*	[əˈtr̥akt]
accurate	[ˈaʔk.kr̥ʊrət]	*accuse*	[əˈkr̥ɯwz]
chocolate	[ˈtʃɔʔk.kl̥ət]	*acclaim*	[əˈkl̥ejm]
awkward	[ˈɔːʔk.kw̥əd]	*request*	[rəˈkw̥ɛst]

Drawing on earlier analyses of gemination in Indic (Vaux 1992) and West Germanic (Bermúdez-Otero 1999), Khwaja et al. derive the clusters in (3a) from heterosyllabic sequences of a stop and a sonorant. At the stem level, these stops are syllabified as codas due to the effect of the alignment constraint AlignL ([$_{\sigma}$CC, Σ]), which penalizes branching onsets that are not aligned with the left edge of a foot. AlignL([$_{\sigma}$CC, Σ]) dominates the constraint CodaCond(ition), which penalizes coda consonants whose place of articulation is not shared with the following onset (see, among others, Itô 1986; Yip 1991).[9] Importantly, codas that form the first half of a geminate satisfy CodaCond, since they share their place specification with the following onset (which contains the second half of the geminate). Gemination can therefore be viewed as a strategy to avoid a CodaCond violation. The stop-sonorant clusters in (3b) are foot-initial and therefore form branching onsets; this syllabification satisfies both AlignL([$_{\sigma}$CC, Σ]) and CodaCond. The different syllabifications in (3a) and (3b) reflect the tendency for prosodically strong positions, such as the left edge of a foot, to support more complexity than prosodically weak positions.

The constraint tableaux in (4ab) show the interaction between AlignL ([$_{\sigma}$CC, Σ]) and CodaCond at the stem level for the forms *poplar* and *apply*. We assume that AlignL([$_{\sigma}$CC, Σ]) and CodaCond are both ranked above *[$_{\sigma}$CC, a syllable markedness constraint that prohibits complex onsets. This ranking ensures that the syllabification of stop-sonorant clusters is stress-sensitive; these clusters receive a heterosyllabic parse in foot-medial position, as in (4a), and a tautosyllabic parse in foot-initial position, as in (4b).

(4) *Stress-sensitive syllabification of medial stop-sonorant clusters (stem level)*

a.

/pɔplə/	AlignL([$_{\sigma}$CC, Σ)	CodaCond	*[$_{\sigma}$CC
☛ ˈpɔp.lə		*	
ˈpɔ.plə	*!		*

b.

/əplaj/	AlignL([$_{\sigma}$CC, Σ)	CodaCond	*[$_{\sigma}$CC
əp.ˈlaj		*!	
☛ ə.ˈplaj			*

The ranking order of the two constraints is reversed at the word level. We assume that gemination is enforced by the faithfulness constraint MaxLink-μ/C, which requires a consonant that is moraic in the input to retain its mora in the output. The combined effect of MaxLink-μ and CodaCond produces an optimal form with a geminate stop, as shown in (5).

(5) *Resyllabification of medial stop-sonorant clusters (word level)*

ˈpɔp.lə	MaxLink- μ/C	CodaCond	AlignL([$_{\sigma}$CC, Σ)	*[$_{\sigma}$CC
ˈpɔp.lə		*!		
ˈpɔ.plə	*!		*	*
☛ ˈpɔp.plə			*	*

We now extend this analysis to affricate gemination. Under the assumption that /tʃ/ is a cluster at the stem level, which we formalize as undominated ranking of the inventory constraint *t͡ʃ, the constraint ranking in (4) predicts that /tʃ/ is subject to the same stress-sensitive syllabification pattern as stop-sonorant clusters.

(6) *Stress-sensitive syllabification of medial affricates (stem level)*

a.

/bʊtʃə/	*t͡ʃ	AlignL([$_{\sigma}$CC, Σ)	CodaCond	*[$_{\sigma}$CC
☛ ˈbʊt.ʃə			*	
ˈbʊ.tʃə		*!		*
ˈbʊ.t͡ʃə	*!			

b.

/ətʃɪjv/	*t͡ʃ	AlignL([$_{\sigma}$CC, Σ)	CodaCond	*[$_{\sigma}$CC
ət.ˈʃɪj.v			*!	
☛ ˈə.ˈtʃɪj.v				*
ə.ˈt͡ʃɪj.v	*!			

Given the evidence from linguistic performance presented in Section 3, we assume that /tʃ/ must coalesce to a single segment at the word level. This requires a change in the ranking of the inventory constraint *t͡ʃ from undominated at the stem level to low-ranked at the word level. Coalescence of /tʃ/ also violates the anti-coalescence constraint Unif(ormity). To drive coalescence, we propose a constraint *[$_{\sigma}$TT, which bans obstruent clusters in onsets. Such clusters are cross-linguistically marked because their members have an insufficient sonority distance (Parker 2012). We can assume this constraint is undominated at the word-level, since English has no obstruent-obstruent onsets on the surface (assuming the /s/ of #sp, #st, #sk is not part of the onset). Paralleling the case of pre-sonorant gemination, the affricate surfaces as a geminate due to the effect of MaxLink-μ/C.

(7) *Coalescence of* /tʃ/ *(word level)*

a.

ˈbʊt.ʃə	*[$_{\sigma}$TT	MaxLink-μ/C	CodaCond	Unif	*t͡ʃ
ˈbʊt.ʃə			*!		
ˈbʊt.tʃə	*!				
☛ ˈbʊt.t͡ʃə				*	*
ˈbʊ.t͡ʃə		*!		*	*

b.

ə.ˈtʃɪj.v	*[$_{\sigma}$TT	MaxLink-μ/C	CodaCond	Unif	*t͡ʃ
ə.ˈtʃɪj.v	*!				
☛ ə.ˈt͡ʃɪj.v				*	*

It remains to discuss gemination of word-final /tʃ/. Following Harris and Gussmann (2002), we regard all word-final consonants as onsets of an empty-headed syllable.[10] Under this assumption, the analysis of word-final /tʃ/ in *teach* is the same as for foot-medial /tʃ/ in *butcher*. At the stem level, /tʃ/ receives a heterosyllabic parse, while at the word level it undergoes gemination and coalescence, just like the /tʃ/ in *butcher*. This is shown in (8).

(8) *Coalescence of word-final* /tʃ/ *(word level)*

ˈtɪjt.ʃ	*[$_{\sigma}$TT	MaxLink-μ/C	CodaCond	Unif	*t͡ʃ
ˈtɪjt.ʃ			*!		
ˈtɪjt.tʃ	*!				
☛ ˈtɪjt.t͡ʃ				*	*
ˈtɪj.t͡ʃ		*!		*	*

In phrasal contexts, the second half of the geminate may be resyllabified with a following word-initial vowel. This phrase-level resyllabification is insensitive to stress, allowing the geminate affricate to appear before a stressed vowel in surface forms, e.g. *catch up* [ˈkʰaʔtˈt͡ʃʌ.p], *teach Ann* [ˈtʰɪjʔtˈt͡ʃa.n]. Compare the geminate [ʔt.t͡ʃ] in these forms with the non-glottalized onset /p/ in *keep up* [ˈkʰɪj.ˈpʌ.p], *stop Ann* [ˈstɔ.ˈpa.n]. Since the word-final /p/ of *keep* [ˈkʰɪj.p] and *stop* [ˈstɔ.p] is an onset in the word-level output, it is not associated with a mora; it therefore remains a simple onset after phrase-level resyllabification without incurring a violation of MaxLink-μ/C.

As noted in Section 4, the overapplication of affricate gemination before word-initial stressed vowels parallels the overapplication of flapping in the same context in American English, as in *get up* [ˈgɛ.ˈɾʌ.p], *let Ann* [ˈlɛ.ˈɾa.n]. Under Kiparsky's (1979) analysis of flapping, /t/ is lenited at the word level unless preceded by an obstruent or the left edge of a foot; under our analysis of affricate gemination, /tʃ/ receives a heterosyllabic parse at the stem level under the very same conditions. Both processes are counterbled by phrase-level resyllabification, resulting in the same pattern of overapplication.

6. Affricate and pre-sonorant gemination: two independent patterns

The analysis presented in Section 5 applies to varieties of English with both pre-sonorant gemination and affricate gemination – common among younger speakers of Southern Standard British English. However, affricate gemination and pre-sonorant gemination are also attested independently of each other. O'Connor (1952), Higginbottom (1964) and Roach (1973) all report gemination of /tʃ/ for their own idiolects, but not but do not have pre-sonorant gemination. Higginbottom even distinguishes, in her own speech, between underlying /tʃ/ and [tʃ] derived from underlying /tj/ by yod coalescence, geminating the former but not the latter:

> There is no contrast in my idiolect between /tʃ/ and /tj/; *tune* has the same initial as *chew*. Likewise the affricate in *virtue*, *situation* is identical with that in *feature*, except that in the two former words there is no reinforcement, where from the stress

> placement a reinforced /tʃ/ would be expected. This lack of reinforcement is found only before the vowels /u uː uə/, i.e. where the sequence /tj/ occurs in other idiolects. Since I do not reinforce plosive-semivowel sequences within the morpheme, it is possible to regard these non-reinforced sequences as examples of /tj/, rather than /tʃ/. (Higginbottom 1964, 141)

Roach (1973) also observed this distribution in some of his informants, who produced *matches* [ˈmaʔt.tʃɪ.z] with gemination but *saturated* [ˈsa.tʃə.ˌrɛj.tɪ.d] without. Interestingly, he also found the exact inverse of this distribution from other informants – i.e. simple onset [.tʃ] in *matches* [ˈma.tʃɪ.z] but gemination in *saturated* [ˈsaʔt.tʃə.ˌrɛj.tɪ.d]. This latter group of speakers had pre-sonorant gemination across the board but geminate [ʔt.tʃ] only when derived from underlying /tj/. We present a formal interpretation of these two patterns below.

We have seen in Section 5 that affricate gemination requires /t͡ʃ/ to be absent from the stem-level inventory. If the grammar allows /t͡ʃ/ at the stem level, no gemination is predicted; rather, /t͡ʃ/ is syllabified just like the plosives /p, t, k/. By demoting the inventory constraint *t͡ʃ to a low ranking at the stem level, we predict pre-sonorant gemination without affricate gemination. The distinction between *matches* [ˈma.tʃɪ.z] without gemination but *saturated* [ˈsaʔt.tʃə.ˌrɛj.tɪ.d] with gemination, as documented by Roach (1973), follows if the stem-level representation of *saturated* contains a /tj/ cluster which undergoes coalescence to [t͡ʃ] at the word-level. (Yod coalescence is avoided at the stem-level due to high-ranked UNIFORMITY.) The gemination of /t͡ʃ/ in this case is in fact pre-sonorant gemination, with an added step of yod coalescence.

(9) *Stress-sensitive syllabification of* /tj/ *(stem level)*

/satjərɛjt/	UNIF	ALIGNL([$_{\sigma}$CC, Σ)	CODACOND	*[$_{\sigma}$CC	*t͡ʃ
☛ ˈsat.jə.rɛj.t			*		
ˈsa.tjə.rɛj.t		*!		*	
ˈsa.t͡ʃə.rɛj.t	*!				*

(10) *Yod coalescence with gemination (word level)*

ˈsat.jə.rɛj.t + d	MAXLINK-μ/C	CODACOND	ALIGNL([$_{\sigma}$CC, Σ)	UNIF	*t͡ʃ
ˈsat.jə.rɛj.tɪ.d		*!			
ˈsat.tjə.rɛj.tɪ.d			*!		
☛ ˈsat.t͡ʃə.rɛj.tɪ.d				*	*
ˈsa.t͡ʃə.rɛj.tɪ.d	*!			*	*

In (10), the constraint driving yod coalescence is AlignL([$_{\sigma}$CC, Σ) – the same constraint we used to motivate stress-sensitive syllabification at the stem level. This ranking generates the conservative distribution of yod coalescence in which /tj/ and /dj/ remain intact in foot-initial positions but coalesce foot-internally (see Szigetvári 2016). For more innovative speakers who apply yod coalescence across the board, Uniformity would need to be ranked below the general markedness constraint on complex onsets *[$_{\sigma}$CC.

We now turn to the varieties of English with affricate gemination but no pre-sonorant gemination. In these varieties, the stem-level phonology must apply the stress-sensitive pattern only to /tʃ/, and not to stop-sonorant clusters. To analyze this pattern, we return to the observation that obstruent-obstruent clusters are cross-linguistically marked relative to obstruent-sonorant clusters. We define a constraint AlignL([$_{\sigma}$TT, φ) which penalizes obstruent-obstruent clusters that are not aligned with the left-edge of a foot. The correct outcomes are generated if this constraint is ranked above CodaCond at the stem level, while AlignL([$_{\sigma}$CC, Σ) is ranked below CodaCond.

(11) *Stress-sensitive syllabification of affricates vs. onset maximization for* /pl/ *(stem level)*

/bʊtʃə/	*t͡ʃ	AlignL([$_{\sigma}$TT, φ)	CodaCond	*[$_{\sigma}$TT	AlignL([$_{\sigma}$CC, Σ)
☛ ˈbʊt.ʃə			*		
ˈbʊ.tʃə		*!		*	*
ˈbʊ.t͡ʃə	*!				

/pɔplə/	*t͡ʃ	AlignL([$_{\sigma}$TT, φ)	CodaCond	*[$_{\sigma}$TT	AlignL([$_{\sigma}$CC, Σ)
ˈpɔp.lə			*!		
☛ ˈpɔ.plə					*

7. Evidence against a cluster analysis of /dʒ/

While the arguments for treating English affricates as single segments apply to both /tʃ/ and /dʒ/, the evidence for a bisegmental analysis pertains to /tʃ/ only, as /dʒ/ is never affected by glottalization. Indeed, as far as /dʒ/ is concerned, we see no reason to depart from the standard view that it patterns as a single segment.

Support for this comes from an examination of the default stress pattern in English polysyllabic verbs (Harris 1994, 40). Disregarding a single final consonant (which is the onset of an empty-headed syllable in our view, and is regarded as extrametrical in other approaches), stress falls on the final syllable if it is heavy,

otherwise on the penultimate syllable. Harris gives the examples in (12). The final syllables in (12ab) are heavy and therefore attract stress; the final syllables in (12c) are light, so stress falls on the penultimate syllable; the forms in (12d) show that polysyllabic verbs ending in /dʒ/ pattern with those in (12c) – i.e. that /dʒ/ counts as a single consonant for the purpose of stress assignment.

(12)

a.	b.	c.	d.
torménˈt	*cajóle*	*édit*	*mánage*
laménˈt	*maintáin*	*astónish*	*fórage*
collápse	*caróuse*	*cáncel*	*encóurage*

These data provide strong evidence for the monosegmental status of /dʒ/ at the stem level, where prosodic structure is built (see Bermúdez-Otero 2012). It is striking, however, that there do not appear to be any examples of polysyllabic verbs with a final unstressed syllable ending in /tʃ/.[11] The stress facts do not, therefore, undermine our analysis of /tʃ/ as a cluster at the stem level; rather, they suggest that at this level /tʃ/ and /dʒ/ may have a different segmental status.

Additionally, there is some historical evidence that /tʃ/, but not /dʒ/, is dispreferred word-finally after an unstressed vowel. Kenyon (1969), in his discussion of the origins of /dʒ/ in English, notes the following:

> Another source of **dʒ** in English is from the voicing of **tʃ** by lack of stress in unaccented syllables, as in *Harwich* **hærɪdʒ**, *Dulwich* **dʌlɪdʒ**, *Woolwich* **wulɪdʒ**, *cabbage* (ME cabache), *knowledge* (ME knowleche), *spinach* (OF spinache, -age). (Kenyon 1969, 150)

If /dʒ/ counts as a single segment at the stem level while /tʃ/ counts as two, then this historical change can be viewed as a cluster simplification. Cluster simplification is expected in this context, since English allows only a highly restricted set of consonant clusters word-finally after an unstressed vowel (see Hammond 1999, 198–200).

8. Conclusion

The stratal analysis proposed in this paper resolves two ambiguities concerning the English fortis affricate /tʃ/. First, at the phonetic level, /tʃ/ has both onset-like and coda-like properties in foot-medial positions after a vowel or sonorant, with glottalization preceding the closure and a degree of aspiration accompanying the fricative release. We have argued that this phonetic realization corresponds to an output phonological form containing a geminate affricate [t.tʃ].

Second, while /tʃ/ is consistently treated as a single segment in speech errors, language games, and the intuitions of native speakers, it appears to pattern like

a cluster in syllabification for many speakers of British English. We resolve this paradox by proposing that, for these speakers, /tʃ/ is a cluster at the stem level, where syllable and foot structure are built, but coalesces to a single segment [t͡ʃ] at the word level – much as /tj/ may also coalesce to [t͡ʃ] at the word-level. Where the input to the word level contains a complex onset [.tʃ] or [.tj], coalescence yields a simple onset affricate [.t͡ʃ] (e.g. *achieve* [ə.ˈtʃɪj.v] → [ə.ˈt͡ʃɪj.v]; *attune* [ə.ˈtjʉw.n] → [ə.ˈt͡ʃʉw.n]); where the input contains a heterosyllabic cluster [t.ʃ] or [t.j], coalescence yields a geminate affricate [t.t͡ʃ] as a result of faithfulness to mora associations (e.g. *butcher* [ˈbʊt.ʃə] → [ˈbʊt.t͡ʃə]; *saturate* [ˈsat.jə.rɛj.t] → [ˈsat.t͡ʃə.rɛj.t]).

Since /tʃ/ and /dʒ/ are a fortis-lenis pair, it is often taken for granted that they must be alike in their internal structure and segmental status. However, we have not found any phonological phenomena in which /dʒ/ behaves like a cluster. On the contrary, we find evidence from stress assignment that /dʒ/ is a single segment even at the stem level. We therefore maintain that monosegmental /d͡ʒ/ is present in the stem-level inventory even in varieties of English where its fortis counterpart /t͡ʃ/ is absent – an asymmetry that finds some support in diachrony.

Notes

1 There are some loanwords with initial /ts/ and /dz/, e.g. *tsunami*, *dzo*. The online version of the Merriam-Webster dictionary notes that these have variant pronunciations without the initial stop.

2 This would imply that the /s/ in sCC clusters is not part of the onset. This is supported by the observation that traditional RP (and many other accents) have no yod dropping in words like *student* and *skewer*.

3 One type of evidence for this is that derivatives with *-ship* admit coordinated structure deletion, e.g. *guardian- and custodianship*. This has been used as a diagnostic for prosodic-word status in Dutch and German (Hall 1999); it seems reasonable to apply this diagnostic to English as well.

4 The fluent backward talkers in Cowan et al.'s experiment were all speakers of American English. It is an open question whether fluent backward talkers with the glottalization patterns described in Section 4 would have the same reversals. We do not know, but we expect they would.

5 This is not the only argument for treating /tʃ/ as a cluster. Cruttenden and Gimson (2001, 173) note that /tʃ/ in RP is "generally subject to … glottal replacement of the [t] element alone as in [kəʊʔʃ]." Kenyon (1969, 149) notes that "in America the **tʃ** sound is usually preserved in words like *wrench* **rɛntʃ**, *inch* **ɪntʃ**, *gulch* **gʌltʃ**, which in England are sometimes, but not invariably, pronounced as **rɛnʃ**, **ɪnʃ**, **gʌlʃ**." We restrict our attention to the behaviour of /tʃ/ in glottal reinforcement.

6 For details of this approach, see Kahn (1976) and Gussenhoven (1986).

7 Gussenhoven does not comment on the syllabic affiliation of a word-final /tʃ/ before a pause or a word-initial consonant, as in *teach* ‖, *teach me*. In Section 5, we suggest that the second half of the geminate forms the onset of a final empty-headed syllable.

8 The cluster /tw/ has a limited distribution, however. We find no examples of foot-medial /tw/. This cluster occurs between a primary and secondary stress in *patois* and *Antwerp*; see Khwaja et al. (in press) for a discussion of possibilities of glottalization in this context. /tw/ occurs word-medially before an unstressed vowel in *Antoinette* (we thank an anonymous reviewer for bringing this example to our attention), but we believe the cluster is in fact foot-initial in this word. Following the approach of Bermúdez-Otero (2018b), *Antoinette* should have the following foot structure: ($_{\Sigma}$ ˌan) ($_{\Sigma'}$ tʰwə ($_{\Sigma\circ}$ ˈnɛt)).

9 In his analysis of West Germanic gemination, Bermúdez-Otero (1999) assumes that gemination is driven by the constraint (Syllable)Contact, which penalizes coda–onset sequences with rising sonority. Khwaja et al. (in press) replace Contact with CodaCond. This makes it possible to view gemination as one of a number of strategies to avoid place-specified codas.

10 With a possible exception for word-final syllabic sonorants, as in *buckle* [ˈbʌkl̩], *button* [ˈbʌʔn̩], which may occupy a syllable nucleus.

11 There are a few verbs with a final stressed syllable ending in short vowel + /tʃ/, e.g. *attách*, *detách*, *dispátch*, but Harris (1994, 276, note 16) disregards these examples because they contain potentially segmentable prefixes. He considers these forms to be analogous to *abút*, *permít*, *compél*, where stress falls on a light final syllable because the penultimate syllable is an unstressable prefix.

References

Andrésen, Bjørn Stålhane. 1968. *Pre-glottalization in English Standard Pronunciation*. Oslo: Norwegian Universities Press.

Bailey, George, Stephen Nichols, Danielle Turton, and Maciej Baranowski. 2022. "Affrication as the Cause of /s/-retraction: Evidence from Manchester English." *Glossa* 7.1: 1–30. DOI: 10.16995/glossa.8026

Bagemihl, Bruce. 1988. "Alternate Phonologies and Morphologies." PhD diss, University of British Columbia.

Bermúdez-Otero, Ricardo. 1999. "Constraint Interaction in Language Change: Quantity in English and Germanic." PhD diss, University of Manchester.

Bermúdez-Otero, Ricardo. 2011. "Cyclicity." *The Blackwell Companion to* Phonology. Ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice. Volume 4. Malden, MA: Wiley-Blackwell. 2019–2048.

Bermúdez-Otero, Ricardo. 2012. "The Architecture of Grammar and the Division of Labor in Exponence." *The Morphology and Phonology of Exponence*. Ed. Jochen Trommer. Oxford: Oxford University Press. 8–83.

Bermúdez-Otero, Ricardo. 2018a. "Stratal Phonology." *The Routledge Handbook of Phonological Theory*. Ed. S. J. Hannahs, and Anna R. K. Bosch. Abingdon: Routledge. 100–134.

Bermúdez-Otero, Ricardo. 2018b. "Diagnostics of the Moraic Trochee from Proto-Germanic to Present-day English." Handout presented at the Workshop

on the Foot in the Phonological History of English, ICEHL XX, Edinburgh, August 27.
Cowan, Nelson, Martin D.S. Brain, and Lewis A. Leavitt. 1985. "The Phonological and Metaphonological Representation of Speech: Evidence from Fluent Backward Talkers." *Journal of Memory and Language* 24: 679–698.
Fromkin, Victoria A. 1971. "The Non-anomalous Nature of Anomalous Utterances." *Language* 47.1: 27–52.
Gimson, A. C. 1962. *An Introduction to the Pronunciation of English*. London: Edward Arnold.
Gimson, A. C., and Alan Cruttenden. 2001. *Gimson's Pronunciation of English*. London: Edward Arnold.
Gussenhoven, Carlos. 1986. "English Plosive Allophones and Ambisyllabicity." *Gramma* 10.2: 119–141.
Hall, T. Alan. 1999. "The Phonological Word: A Review." *Studies on the Phonological Word*. Ed. T. Alan Hall, and Ursula Kleinhenz. Amsterdam & Philadelphia: John Benjamins. 1–22.
Hammond, Michael. 1999. *The Phonology of English: A Prosodic Optimality-theoretic Approach*. Oxford: Oxford University Press.
Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.
Harris, John. 2004. "Release the Captive Coda: The Foot as a Domain of Phonetic Interpretation." *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Ed. John Local, Richard Ogden, and Rosalind Temple. Cambridge: Cambridge University Press. 103–129.
Harris, John, and Edmund Gussmann. 2002. "Word-final Onsets." *UCL Working Papers in Linguistics* 14: 1–42.
Higginbottom, Eleanor. 1964. "Glottal Reinforcement in English." *Transactions of the Philological Society* 63.1: 129–142.
Iles, Laurence A. 1960. "ðə ɡlotḷaizeiʃn əv vɔislis plousivz in *lʌndən spiːtʃ." *Le Maître Phonétique* 113: 14–15.
Itô, Junko. 1986. "Syllable Theory in Prosodic Phonology." PhD diss, University of Massachusetts, Amherst.
Jensen, John T. 1993. *English Phonology*. Amsterdam: John Benjamins.
Jensen, John T. 2000. "Against Ambisyllabicity." *Phonology* 17: 187–235.
Jones, Daniel. 1969. *An Outline of English Phonetics*. 9th ed. Cambridge: Cambridge University Press.
Kahn, Daniel. 1976. "Syllable-based Generalizations in English Phonology." PhD diss, Massachusetts Institute of Technology.
Khwaja, Hamza, Bert Botma, Joseph Smith, and Janet Grijzenhout. 2025. "The Prosodic Conditioning of Glottalization in British English." Poster presented at The 31st Manchester Phonology Meeting, Manchester, May 29.
Khwaja, Hamza, Bert Botma, Joseph Smith, and Janet Grijzenhout. In press. "Gemination and the Allophony of English Fortis Stops." *Theoretical Pho-*

nology and the Phonology of English. Ed. Patrick Honeybone, and Martin Krämer. Cambridge: Cambridge University Press.

Kenyon, John Samuel. 1969. *American Pronunciation*. 10th ed. Ann Arbor: George Wahr.

Kiparsky, Paul. 1979. "Metrical Structure Assignment is Cyclic." *Linguistic Inquiry* 10.3: 421–441.

Lin, Yen-Hwei. 2011. "Affricates." *The Blackwell Companion to Phonology*. Ed. Marc van Oostendorp, Colin J. Ewen, Elisabeth Hume, and Keren Rice. Volume 1. Malden, MA: Wiley-Blackwell. 367–390.

O'Connor, Joseph D. 1952. "RP and the Reinforcing Glottal Stop." *English Studies* 33: 214–219.

Parker, Steve. 2012. "Sonority Distance vs. Sonority Dispersion – A Typological Survey." *The Sonority Controversy*. Ed. Steve Parker. Berlin & Boston: Mouton De Gruyter. 101–166.

Read, Charles. 1971. "Pre-school Children's Knowledge of English Phonology." *Harvard Educational Review* 41.1: 1–34.

Roach, Peter J. 1973. "Glottalization of English /p/, /t/, /k/ and /tʃ/ – A Re-examination." *Journal of the International Phonetic Association* 3.1: 10-21.

Szigetvári, Péter. 2016. "The Curious Case of Cj Clusters in English." *The Even Yearbook* 12: 93–105.

Treiman, Rebecca. 1993. *Beginning to Spell: A Study of First-grade Children*. New York: Oxford University Press.

Vaux, Bert. 1992. "Gemination and Syllabic Integrity in Sanskrit." *Journal of Indo-European Studies* 20.3–4: 283–303.

Ward, Ida C. 1945. *The Phonetics of English*. 4th ed. Cambridge: W. Heffer & Sons.

van de Weijer, Jeroen. 2014. "Affricates in English as a Natural Class." *Above and Beyond the Segment: Experimental Linguistics and Phonetics*. Ed. Johanneke Caspers, Jos Pacilly, Willemijn Heeren, and Yiya Chen. Amsterdam: John Benjamins. 350–358.

Wells, John C. 1982. *Accents of English*. 3 vols. Cambridge: Cambridge University Press.

Wells, John C. 1990. "Syllabification and Allophony." *Studies in the Pronunciation of English: A Commemorative Volume in Honor of A.C. Gimson*. Ed. Susan Ramsaran. London: Routledge. 76–86.

Yip, Moira. 1991. "Coronals, Consonant Clusters, and the Coda Condition." *The Special Status of Coronals: Internal and External Evidence*. Ed. Carole Paradis and Jean-François Prunet. San Diego: Academic Press. 61–78.

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Silke Hamann and Radosław Święciński, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.06

Silke Hamann
https://orcid.org/0000-0002-6588-0892
University of Amsterdam

Radosław Święciński
https://orcid.org/0009-0009-9454-7456
Amsterdam University of Applied Sciences

How Language-specific Perceptual Cues Can Explain Differences in Onset Clusters: The Case of /tr/ in Polish and English

Abstract: /tr/ onset clusters in Polish and English show different phonetic realizations. Contra previous accounts locating the cause for this in a different gestural timing or a different phonological representation, we propose the two languages employ different perceptual cues necessary for the perception of consonants in language-specific clusters: in English, voiceless plosives are followed only by sonorants in onsets, and the sonorant release sufficiently cues the plosive, while in Polish, with its much larger cluster inventory, an audible release of a plosive followed by another plosive is necessary to identify the place of articulation of the first plosive. A formalization of the mapping between /tr/ and its language-specific cues is provided and employed for native and non-native perception. The resulting predictions are compared to existing perceptual studies.

Keywords: phonetics–phonology interface, onset clusters, perceptual cues, Polish, English, phonological perception

1. Introduction

Languages with complex onsets and codas differ greatly in the type of consonant clusters they allow. On the one hand, there are languages like English that have

many clusters but only a few violate the Sonority Sequencing Principle (SSP; Selkirk 1984), such as onset /st/, /sf/ or /spl/.[1] On the other hand, there are languages like Polish that have a large cluster system, many of these clusters with a sonority plateau, such as /pt/, /fx/, /sf/, or even a sonority reversal, such as onset /wk/, /rd͡z/ or /mʂ/.[2] Unsurprisingly, these two language types share onset clusters of rising sonority, such as /kw/ and /pl/, though these clusters are not necessarily realized in the same way.

In this article, we examine one such onset cluster, namely /t/ plus rhotic in English and Polish, inspired by an article by Schwartz with the same focus (Schwartz 2023). This cluster is transcribed phonologically as /tr/ henceforth. Though phonologically not identical in the two languages, as the phonemes constituting the cluster are contrasting with different sets of phonemes in Polish and English and are potentially undergoing different phonological processes, we are not further concerned with these phonological differences. The point of the present paper is that /tr/ in the two languages has different phonetic realizations, and that these phonetic differences do not imply or require differences in phonological representation, thus arguing for a strict distinction between phonetics and phonology.

In English, the /t/ is phonetically realized as alveolar, and the rhotic most commonly as an approximant transcribed as [ɹ] (Cruttenden 2014; Ogden 2017). The two consonants of the cluster /tr/[3] have a shorter duration than corresponding consonants in a non-cluster (e.g. O'Shaughnessy 1974). Furthermore, the plosive is rhotically released (the plosive closure is post-alveolar with the tongue shape anticipating the following rhotic; Ogden 2017), and the rhotic is also influenced by the plosive: it is at least partially devoiced (Tsuchida et al. 2000; Ogden 2017). This also holds for other plosive-liquid and (for) plosive-glide clusters such as /pr, kr, pl, kl, tj, kj, pj, tw, kw, pw/. English devoiced /tr/ often shows affrication (cf. Cruttenden 2014, 189ff.). In the following, we do not distinguish between a partially devoiced sonorant and a homorganic voiceless fricative because the difference is very subtle, and to our knowledge no clear acoustic criteria have been proposed for its differentiation.

In Polish, the plosive is typically dental (Wierzchowska 1980; Dukiewicz and Sawicka 1995, 150) and the rhotic is mainly realized as a trill [r] or a tap, though accounts differ as to which variant is predominant (see Wierzchowska 1980; Stolarski 2013; 2015). Approximant or fricative articulations are also documented (Jaworski and Gillian 2011). The two consonants in the cluster have, impressionistically, the same duration as in isolation[4] and are usually clearly acoustically separated (see e.g. Cieślak 2015). Furthermore, a short vowel-like element is often observed in between (ibid.), which may be considered an *intrusive* vowel following the definition by Hall (2006), as the sound is optional, does not participate in phonological processes (e.g. it does not form a syllable nucleus), and the speakers are often not aware of its presence, in contrast to *epenthetic* vowels. Dłuska (1983) remarks that for Polish /r/, a vocalic element is always present initially and finally

whenever /r/ is adjacent to a consonant or silence and is a crucial perceptual cue that helps to distinguish /r/ from /d/, e.g. [*V*rak] 'crayfish' or [d*V*rɔga] 'dear', where we use *V* to denote this vocalic element. Dłuska further maintains that it also appears between closures in a trill articulation. According to the same author, the auditory quality of this vocoid is difficult to identify, but it resembles the Polish vowels /ɛ/ or /ɨ/, or assumes the quality of the adjacent vowel. In contrast to English, almost no phonetic influence of /t/ on /r/ (such as devoicing) can be observed in the syllable onset position in Polish (Bruni 2011, 81–94 found a minimal devoicing effect of a preceding voiceless obstruent on a following onset /r/ and /l/ in Polish). Devoicing may appear word finally, though, e.g. [mɛtr̥] 'meter' or when a voiceless consonant follows, e.g. [pjotr̥ka] 'Piotrek, gen. sg.' (see e.g. Bruni 2011, 94 for complete devoicing in the first context; in both contexts, a trill realization seems common).

There are also clear articulatory differences between the two languages, which are covered in Section 3.1 below, as they are sometimes considered the source of the above-described acoustic-phonetic discrepancies.

In the present article, we propose that the differences in acoustics and in articulation are caused by different perceptual cues that the two languages require for /tr/ but also for other clusters with a plosive followed by a consonant: Polish listeners need explicit cues for the plosive, namely a salient burst with formant transitions, especially in a cluster where it is followed by another plosive (e.g. /pt/), as otherwise the place of articulation of the first plosive would not be retrievable. This cue requirement is generalized to all plosives in clusters, thus also to /tr/, where in principle a closure silence followed by an alveolar rhotic release would be sufficient to cue /t/. English listeners, on the other hand, do not need a burst to cue the place of the plosive in a cluster, as these plosives are only ever followed by sonorants (/r/, /l/, /w/ or /j/), where the formant transitions in the following sonorant provide sufficient cues to the place of articulation of the plosive. Hence a temporal overlap of cues is possible in English, and perceptually more efficient, resulting in a (partially) devoiced sonorant that cues at the same time the voicelessness of the plosive and the type of sonorant (as English does not contrast voiced with voiceless sonorants).

We provide an explicit formalization of these language-specific cues for onset /tr/ and their perception and production in Polish and English in the model of Bidirectional Phonetics and Phonology, which not only accounts for the phonetic difference in /tr/ cluster between the two languages, but also makes predictions on native and cross-linguistic perception that are then compared to the results of perception experiments by Aperliński and Schwartz (2015) and Schwartz (2023). Our formalization also shows that with the postulation of language-specific cues, a difference in articulation (and articulatory timing) follows automatically, and that a postulation of phonologically different structures to account for perceptual and articulatory differences is not necessary.

This article is structured as follows. Section 2 briefly elaborates two types of accounts that have been proposed previously for the difference in the clusters of these two languages: the first referring to articulatory timing differences, the second to phonological representational differences. Section 3 then introduces the theoretical framework in which our formalization is given, and compares it to other theories on the phonetics–phonology interface. Section 4 provides a formalization of the native perception and production of English and Polish /tr/ clusters, and tests this by applying it to cross-language input and comparing the results to existing perception data. Section 5 discusses the assumptions and implications of the presented formalization, supported by findings on Polish rearticulated geminates. Section 6 concludes.

2. Existing accounts

2.1. Articulatory account

Articulatory studies have shown that there is a difference in /tr/ onset clusters between the two languages in terms of gestural timing: there is more gestural overlap between the two component consonants in English (e.g. Marin and Pouplier 2010) compared to a minimal gestural overlap between the consonants in Polish (e.g. Hermes et al. 2017),[5] see the schematic representation in Figure 1.

Fig. 1. Simplified representations of a two-consonant onset cluster with overlapping gestures (on the left), representing English /tr/, and non-overlapping gestures (on the right), representing Polish /tr/. Rising lines represent the gestural onset, flat lines the gestural steady-state, and falling lines the gestural offset of each consonant.

This difference in gestural timing between the two languages holds not only for onset /tr/ clusters, though we will not elaborate on other clusters.

An EMA study by Schwartz et al. (2024) directly compared the articulation of clusters in both languages by looking at Polish-English bilinguals, and found larger gestural lags in Polish than English clusters for the same speakers. This gestural timing difference has been considered the cause of the above-mentioned acoustic differences in /tr/ clusters between the two languages: the lag required in the gestural timing of onset clusters in Polish has been argued to cause a difference in the acoustics of the cluster (see e.g. Bruni 2011, 111).

Interestingly, in the literature on gestural timing, perceptual motivations such as recoverability have been acknowledged, see e.g. Kwon and Chitoran (2024,

175) on Georgian: "[s]peakers seem to organize consonant gestures with more temporal distance, when tighter timing could result in unfavorable perceptual consequences … . On the contrary, tighter timing seems to be tolerated when it would not work against the perceptual recoverability of the consonants."

2.2. Phonological account

In his detailed study of /tr/ clusters in English and Polish, which was the incentive for the present article, Schwartz (2023) correctly states that traditional phonological approaches that refer to sonority cannot account for the observed differences between this cluster in the two languages: The cluster is of rising sonority in both, and involves comparable phonemes in both[6] (despite the different phonetic realization of the rhotic, see Chabot 2019). Schwartz proposes an alternative phonological account, couched in the framework of Onset Prominence (Schwartz 2010; 2016; 2023), which includes both phonetic realizations and phonological representations. The building block in this theory is a CV sequence (following Scheer 2004), from which all types of clusters are derived. The two languages differ in the representation of the /tr/ onset-cluster (and all rising sonority clusters) in this theory, in that English has a non-branching structure in the top node of its representational tree, while Polish has a branching structure, cf. Figure 2.

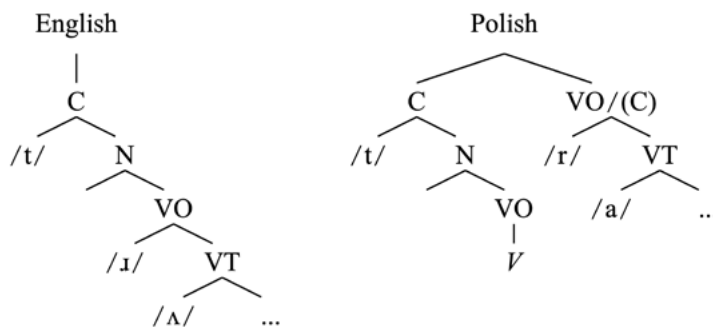


Fig. 2. Representation of English /trʌ/ on the left and Polish /tra/ on the right in the framework of Onset Prominence (based on Schwartz 2023, 647 and 648). C = Closure, N = Noise, VO = Vocalic Onset, VT = Vocalic Target. *V* indicates the position of the potential intrusive vowel in Polish.

In Figure 2, the structure for the /tr/ cluster in Polish, on the right, differs from that in English, on the left, as the former has a branched-off VO node at the top, which creates space for an intrusive vowel in the left-branch (under the VO node there). This structure is again different from the structure of the English sequence /tər/ as in the word *terrain*, given in Figure 3, where the schwa is not intrusive but a full vowel that forms its own syllable (and complete tree).

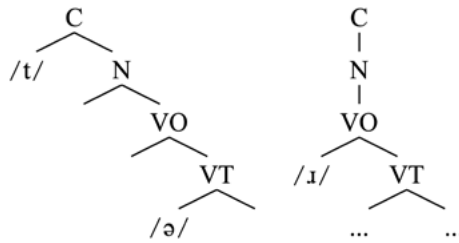


Fig. 3. Representation of English /tər/ in *terrain* (based on Schwartz 2023, 648)

According to Schwartz, the difference in phonological structure in English vs Polish /tr/ in Figure 2 directly reflects a difference in phonetic realization: it predicts gestural overlap for the cluster in English (with resulting devoicing), referred to as "synchronous articulation" by Schwartz (2023, 624), and non-overlap or "asynchronous articulation" for Polish.

However, the Onset Prominence representation of /tr/ captures not only phonetic implementation but also information of prosodic word structure, according to Schwartz (2023). While in English only the rhyme is of relevance for minimal prosodic words (as is common cross-linguistically), in Polish, Schwartz claims, the onset contributes to prosodic weight. Inflected nouns are usually minimally either VC, CVC, or CCV, which in Schwartz' view shows that the minimal prosodic word in Polish consists either of **C**CV, i.e. a syllable with a branching structure as in Figure 2 right, or (C)V**C**, a syllable with a coda. Words consisting only of CV are "sub-minimal for the *purposes of morphology*, and often pronounced as enclitics" (Schwartz 2023, 632, italics ours).[7] The correlation between prosodic word and morphology that Schwartz draws is unusual, as the prosodic (or phonological) word is defined as a domain for phonological processes and is not identical to morphological constituents (e.g. Trask 1996, 275; but see also Cetnarowska 2000 for a discussion of the (mis-)alignment of prosodic and morphological constituents in Polish).

Both Kwon and Chitoran (2024, 175), as representatives of the articulatory account discussed in Section 2.1, and Schwartz et al. (2024), as representatives of the phonological account in Section 2.2, refer to a potential perceptual explanation for language-specific differences in clusters. Schwartz et al. (2024, 214) elaborate that Polish might have less gestural overlap of the second consonants in a three-consonant cluster "to increase perceptibility of the second consonant … . Since preconsonantal consonants … lack of CV transition cues …, sufficient discriminability may demand that speakers lengthen C2 in three member onsets as a way of making C2 easier for listeners to hear."

The following two sections take this point that perceptibility of a contrast in consonant clusters might play an important role in the difference between the two languages further: we present a theory that formalizes the perception process with

language-specific perceptual cues (Section 3), and formalize therein the perception and production of onset /tr/ clusters in English and Polish (Section 4).

3. A view on the phonetics–phonology interface

Our account is couched in the framework of Bidirectional Phonetics and Phonology (BiPhon; Boersma 2007; 2011). BiPhon is a generative model in assuming that linguistic production and comprehension involve the generation of phonological surface representations that are not stored, but are necessary to either access the underlying lexical representation in comprehension, or to further generate a phonetic form in production. In its generativeness and its strict division between phonetics and phonology it is radically different from Exemplar-theoretic approaches (e.g. Bybee 1988; 2001; Pierrehumbert 2001), but falls within the group of substance-free phonological approaches (such as Hale and Kissock 2007; Blaho 2008; Hale and Reiss 2008; Iosad 2012) that do not postulate any phonetic information in the phonological representations. Instead, BiPhon treats phonetic substance in the separate phonetic representations of auditory and articulatory form. Differing from the substance-free approaches by Hale and Kissock (2007), Hale and Reiss (2008) and Volenec and Reiss (in press), and from early generative approaches such as Chomsky and Halle (1968), BiPhon assumes that the mapping between phonology and phonetics is not universal but belongs to the realm of linguistic knowledge, following the long tradition of studies showing the language-specificity of phonetics (e.g. Fromkin 1975; Keating 1988; Pierrehumbert 1990). BiPhon includes a language-specific mapping between phonology and phonetics, which has to be learned, just as phonological, morphological, semantic and syntactic knowledge and their mappings have to be learned. Phonetics is thus part of grammar, i.e. part of the knowledge that speakers and listeners have of their language(s).

With respect to processing directions, BiPhon formalizes both the production and the comprehension of speech, departing in this point from other phonological theories, which restrict their formalization to the production process, assigning perception a supplementary role at most (for the latter, see e.g. Steriade 2004; Flemming 2013).

The mapping from auditory form onto phonological surface form and vice versa constitutes the actual phonetics–phonology interface in BiPhon (for a discussion of alternatives in other generative approaches, see e.g. Hamann 2011), and the articulatory form only plays a role when trying to achieve a specific auditory form, as elaborated below.

The mappings in BiPhon have been formalized in Optimality Theory (BiPhon-OT, see e.g. Boersma and Hamann 2009; Boersma and van Leussen 2017), which requires negatively formulated constraints and a large set of constraints, in Harmonic Grammar (BiPhon-HG, see e.g. Zhou and Hamann 2024; 2026), which

allows for positive constraints and therefore drastically reduces the number of constraints, and in Neural Networks (BiPhon-NN, see e.g. Seinhorst et al. 2019; Boersma et al. 2020; 2022). In this article, we will employ BiPhon-HG with constraint satisfaction.

Figure 4 provides an overview of the BiPhon representations necessary for the present formalization, namely the phonological surface form, given in slashes, and the two phonetic representations of auditory and articulatory form, both given in square brackets (they are further distinguished with subscripts Aud and Art, respectively, where necessary).

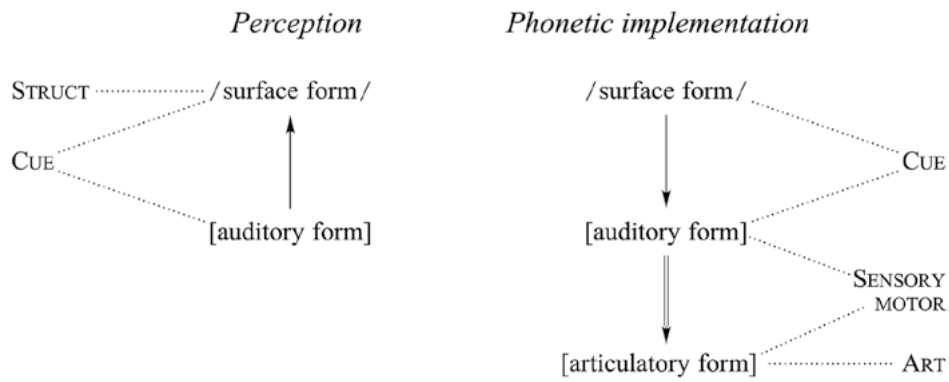


Fig. 4. Part of Bidirectional Phonetics and Phonology relevant for the present study: process of perception on the left, and phonetic implementation on the right. The double arrow indicates an automatic mapping. Constraints are in small capitals.

Figure 4 also shows the two processing directions: on the left is perception, which involves the mapping of the auditory form onto the surface form by so-called cue constraints (Cue), but which is also influenced by language-specific structural constraints (Struct) regulating phonotactic restrictions. Speech perception is thus not independent of phonology: the auditory form is only mapped onto phonological categories and sequences that are licit in the listener's language.

On the right-hand side of Figure 4 is the process of phonetic implementation. This consists of two parallel steps, the mapping of the surface form onto an auditory target, performed by the same cue constraints as employed in perception, and the mapping of the auditory form onto an articulation that can cause the associated auditory form. The latter mapping is regulated by sensory-motor constraints, and is considered automatic for the present purposes: differences in muscles movements, etc. are only of relevance in our modelling if they result in a different auditory form because only then are they observable for speakers and listeners. The choice of an articulatory form can be restricted by articulatory constraints (Art) that evaluate

gestural co-ordination and complexity of gestures, amongst other things. In a parallel evaluation of auditory and articulatory forms, articulatory constraints can have an influence on the choice of the auditory form (see e.g. Boersma and Hamann 2008 on the impact of articulatory constraints in the evolution of dispersed sibilant inventories).

Figure 4 presents only the part of BiPhon used in the formalization; the larger model includes higher-level representations like underlying, morphological and semantic form (see e.g. Boersma and van Leussen 2017 for modelling five levels of representation).

Concerning the different phonetic realizations between Polish and English /tr/ clusters, BiPhon does not assume that they are caused by distinct phonological structures for the two, as opposed to Schwartz's phonological account (Section 2.2): the difference in clusters between the two languages is located in the phonetic module and cannot be read off phonological representations (due to language-specific phonetics combined with substance-free phonology). However, instead of postulating that language-specific gestural timing is the source of the difference, as proposed in the articulatory account described in Section 2.1, BiPhon assumes that the cause of the difference is language-specific perceptual cues: Polish has several tautosyllabic plosive-plosive clusters in all potential combinations, which can only be properly perceived if there is a sufficient auditory cue for the place of articulation of the first plosive. This is provided by an audible release burst with formant transitions. This audible release burst is then generalized to positions where it is perceptually not necessary as place cue, as in /tr/ clusters (how this generalizes also to other positions is taken up in Section 5.1). English, on the other hand, does not require such an audible burst, as it does not have tautosyllabic plosive-plosive clusters, and therefore /tr/ is cued very efficiently by a devoiced rhotic release, signaling at the same time the plosive and the following sonorant. The articulation of the clusters is then dictated by these cues: the gestural timing is chosen to achieve the required perceptual cues in each language.

The following section formalizes this proposal in detail.

4. Formalization

The onset cluster /.tr/, where "." stands for the syllable boundary, is realized differently in the two languages considered here. To make these differences explicit, we use IPA symbols and additional annotations (cf. also Boersma and Hamann 2009). In terms of auditory form, the cluster is realized as [_$^{t\text{ɔ}}$r] in Polish and as [_^{ɻ}] in English, where the underscore [_] stands for the silence during plosive closure, [$^{t\text{ɔ}}$] stands for an audible burst with formant transitions, [r] stands for the low formants

of a (fully voiced) alveolar rhotic trill or tap, and [tr̥] for a (partially) devoiced, potentially affricated, rhotically-released plosive typical in English.

With these auditory events, we can now formulate cue constraints that map the auditory information onto phonological structure and vice versa. The following mappings are given in the production direction, from surface form onto cues, relevant for the production of the /.tr/ clusters in the two languages in Section 4.1. The reverse interpretation is introduced and used for the perception in Section 4.2. All cue constraints are formalized positively, and the satisfaction of these constraints is rewarded (shown in the sections below).

(1) Cue constraints:

a) /t(C)/ [tə] "A voiceless [CORONAL, +anterior] plosive before a consonant in a cluster needs to be phonetically realized with an audible burst with formant transitions."

b) /.tr/ [tr̥] "A voiceless [CORONAL, +anterior] plosive plus rhotic in an onset cluster needs to be phonetically realized with a devoiced rhotic release."

c) /r/ [ɹ] "A rhotic /r/ needs to be realized as a voiced approximant with continuous formants."

d) /r/ [r] "A rhotic /r/ needs to be realized as a tap or sequence of taps (a trill)."[8]

e) /t/ [_] "A plosive needs to be realized with a closure silence."

The burst "[tə]" in constraint (1a) can include an audible vowel-like component, i.e. the intrusive vowel that has been reported for Polish, which we transcribe here for simplicity sake as schwa.[9] The C in the phonological surface form in (1a) specifies the context, namely that of a following consonant, and therefore is in brackets.[10] As pointed out at the end of the previous section, this constraint is of importance for Polish plosive-plosive clusters (i.e. when C is a plosive), where the audible burst is the only cue to the place of articulation of /t/, but the constraint is generalized in Polish to all possible consonants. In contrast to the constraint relevant for the English onset cluster in (1b), the Polish constraint in (1a) is not restricted to the onset position: it also holds for coda clusters and hetero-syllabic clusters, while the English realization of the rhotic as devoiced rhotic release only holds for the onset position after a voiceless obstruent. Constraint (1c) specifies the realization of English rhotic /r/ in all other contexts, and constraint (1d) that of the Polish rhotic in most contexts (in this study, we do not deal with devoicing of Polish /r/ after voiceless obstruents when a word boundary or a voiceless obstruent is following).

Constraint (1e), specifying the realization of a voiceless plosive with a silent closure, is not included in the following evaluations for reasons of space: they all have silent closures in the auditory form and plosives in the surface form. The other four constraints are given in all evaluations but have different weights in the two languages.

4.1. Native production

Tableaux (2a) and (2b) formalize the production of /.tr/ by English and Polish speakers, respectively. There are three possible candidates in each evaluation, the first with an auditory form containing the cues that we would expect for a Polish speaker [_$^{t\ni}$r]$_{Aud}$ (audible burst and a trill), the second a form similar to the Polish but with an English rhotic [_$^{t\ni}$ɹ]$_{Aud}$ (audible burst and an approximant rhotic), and the third with an auditory form with cues to be expected for English [_^{ɻ}]$_{Aud}$ (a voiceless rhotic release). These three are paired with their corresponding articulations, given here as schematic representations of non-overlapping gestures [ ]$_{Art}$ for the first and second, and overlapping gestures [ ]$_{Art}$ for the third.

Columns 2–5 give the cue constraints defined in (1a)–(d). The row above them gives their language-specific weights. For instance, the constraint "/t(C)/ [$^{t\ni}$]" has a low weight of 0.1 in English (2a), as /tC/ clusters in this language are not realized with an audible release, whereas in Polish (2b) they need to be realized with an audible release, hence the high weight of 1.0. The constraints can be satisfied by the candidates, as indicated by a "1." The first candidate in (2a), e.g. satisfies constraint "/t(C)/ [$^{t\ni}$]," because the evaluated form "/.tr/" fulfills the phonological structure "/t(C)/" and the first candidate contains the auditory form "[$^{t\ni}$]." This candidate also satisfies the third constraint, the realization of the rhotic /r/ by a trill [r]. Its harmony score *H* in the last row is therefore the weight of the first constraint x its satisfaction plus the weight of the third constraint x its satisfaction, which is 0.1 plus 0.1, thus an *H* of 0.2. The third candidate without the audible release only satisfies the constraint "/.tr/ [^{ɻ}]" with the weight 1, and thus has a harmony score 1.0, which is higher than that of the other two candidates, therefore the third candidate wins (as indicated by the pointing finger).

(2a) Phonetic implementation of /.tr/ in English:

	0.1	1.0	0.1	0.5	*H*
/.tr/	/t(C)/ [$^{t\ni}$]	/.tr/ [^{ɻ}]	/r/ [r]	/r/ [ɹ]	
[_$^{t\ni}$r]$_{Aud}$ [ ]$_{Art}$	1		1		0.2
[_$^{t\ni}$ɹ]$_{Aud}$ [ ]$_{Art}$	1			1	0.6
☞ [_^{ɻ}]$_{Aud}$ [ ]$_{Art}$		1			1.0

(2b) Phonetic implementation of /.tr/ in Polish:

	1.0	0.1	0.5	0.1	*H*
/.tr/	/t(C)/ [$^{t\ni}$]	/.tr/ [$^{\downarrow}$]	/r/ [r]	/r/ [ɹ]	
☞ [_$^{t\ni}$r]$_{Aud}$ [⌒]$_{Art}$	1		1		1.5
[_$^{t\ni}$ɹ]$_{Aud}$ [⌒]$_{Art}$	1			1	1.1
[_$^{\downarrow}$]$_{Aud}$ [⌒]$_{Art}$		1			0.1

In Polish (2b), the first candidate, with an audible release, wins (with an *H* of 1.5), as expected.

Tableau (4) formalizes the production of the English /.tər/ sequence, which requires an additional cue constraint on the short realization of the schwa, given in (3). As the schwa is usually very short in English (see e.g. Schwartz 2023, 634), this constraint has a high weight in English.

(3) /ə/ [ᵊ] "The vowel schwa needs to be realized with short, schwa-like formant transitions."

(4) Phonetic implementation of /.tər/ in English:

	0.1	1.0	0.1	0.5	0.8	*H*
/.tər/	/t(C)/ [$^{t\ni}$]	/.tr/ [$^{\downarrow}$]	/r/ [r]	/r/ [ɹ]	/ə/ [ᵊ]	
[_$^{t\ni}$r]$_{Aud}$ [⌒]$_{Art}$	1		1		1	1.0
☞ [_$^{t\ni}$ɹ]$_{Aud}$ [⌒]$_{Art}$	1			1	1	1.4
[_$^{\downarrow}$]$_{Aud}$ [⌒]$_{Art}$		1				1.0

The first and second candidates, with schwa formants in the auditory form, are more likely to correspond to an articulation with an active tongue gesture for the schwa, represented by the red contour in the following form: [⌒]$_{Art}$. We assume that this articulatory form does not differ in its auditory result from [⌒]$_{Art}$ and therefore does not have to be considered as separate candidate pair.

Polish has no phonemic schwa; therefore, /.tər/ would never be input to native production; hence, we do not provide a formalization of its phonetic implementation.

4.2. Native perception

The same constraints and their language-specific rankings that we established in Section 4.1 for production can now be used to model perception. For this, we use the cue constraints with a reverse interpretation, e.g. constraint (1b) "/.tr/ [ᵗ̪ɾ]" now reads as "A devoiced rhotic release in the auditory form is mapped onto a plosive followed by a rhotic in the phonological surface form." Tableaux (5a) and (5b) formalize the perception of the native cues for /.tr/ in English and Polish, respectively. The two tableaux thus differ in the auditory form in the input. Both have as first candidate the surface form /.tr/, because this form exists in both languages. English has as second candidate the form /.tər/, the most likely alternative. In Polish, which lacks a phonemic schwa, we consider /.tɛr/ as the most likely alternative,[11] given as second candidate in (5b). We therefore also need a cue constraint that maps the short schwa sound that occurs in Polish clusters with a plosive as the first member onto this /ɛ/, "/ɛ/ [ᵊ]". This constraint has a low weight as the very short, schwa-like formants are not in line with the realization of the full, front vowel.

In both tableaux, the first candidate, /.tr/, is the winning candidate, they thus model the native perceptions correctly. In each evaluation, those constraints are satisfied which correspond in their auditory cues to (part of) the input auditory form.

(5a) Perception of [_ᵗ̪ɾ] in English:

	0.1	1.0	0.1	0.5	0.8	*H*
[_ᵗ̪ɾ]	/t(C)/ [ᵗᵊ]	/.tr/ [ᵗ̪ɾ]	/r/ [r]	/r/ [ɹ]	/ə/ [ᵊ]	
☞ /.tr/		1				1.0
/.tər/						0

(5b) Perception of [_ᵗᵊr] in Polish:

	1.0	0.1	0.5	0.1	0.2	*H*
[_ᵗᵊr]	/t(C)/ [ᵗᵊ]	/.tr/ [ᵗ̪ɾ]	/r/ [r]	/r/ [ɹ]	/ɛ/ [ᵊ]	
☞ /.tr/	1		1			1.5
/.tɛr/			1		1	0.7

Indirect support for the difference between Polish and English in the weight that we assume for the burst cue, i.e. constraint "/t(C)/ [ᵗᵊ]," comes from a perception experiment by Aperliński and Schwartz (2015). They cross-spliced plosive burst and transitional cues and found that Polish native listeners paid far more

attention to the burst cue than English native listeners when identifying the place of articulation for these plosives.

4.3. Cross-language perception

The tableaux in the previous sections formalized (part of) the native grammars for English and Polish and could correctly handle the production input /.tr/ and the corresponding native perception inputs. It is interesting to test these grammars on perceptual input from the other language. In (6a) we feed the English realization [_ʰ] to a Polish listener, and in (6b) Polish [_ᵗˢr] to an English listener:

(6a) Perception of English [_ʰ] by Polish listeners:

	1.0	0.1	0.5	0.1	0.2	*H*
[_ʰ]	/t(C)/ [ᵗˢ]	/.tr/ [ʰ]	/r/ [r]	/r/ [ɹ]	/ɛ/ [ᵊ]	
☞ /.tr/		1				0.1
/.tɛr/						0

(6b) Perception of Polish [_ᵗˢr] by English listeners:

	0.1	1.0	0.1	0.5	0.8	*H*
[_ᵗˢr]	/t(C)/ [ᵗˢ]	/.tr/ [ʰ]	/r/ [r]	/r/ [ɹ]	/ə/ [ᵊ]	
/.tr/	1		1			0.2
☞ /.tər/			1		1	0.9

Tableau (6a) predicts that English [_ʰ] will be perceived as /.tr/ by Polish listeners, even though it is not their native realization of this cluster (indicated by the low harmony score). Potential alternative candidates not included in this tableau for reasons of space are /.tʂr/ and /.tɐr/, where Polish listeners perceive the friction noise often present in devoiced English [_ʰ] as a separate sibilant. Such forms would violate the Polish structural constraints */.tʂr/ and */.tɐr/ (also not included for reasons of space), as these onsets do not occur in native words.

Tableau (6b) predicts that English listeners perceive Polish [_ᵗˢr] as their native /.tər/. This is not surprising, as [_ᵗˢɹ] was the production output of the English sequence /.tər/ as formalized in (4), and the Polish form [_ᵗˢr] differs from this mostly in the unusual trill realization of the rhotic but also in slightly different place cues for the plosive; all perceived as a non-native accent. But if English /.tər/ and Polish /.tr/ have similar auditory forms, then one would expect that Polish native

listeners perceive English [_ᵗˢɹ] as their /.tr/ (though as produced with a non-native accent due to different plosive place and rhotic cues). This is what the following tableau (7) predicts:

(7) Perception of English [_ᵗˢɹ] by Polish listeners:

	1.0	0.1	0.5	0.1	0.2	*H*
[_ᵗˢɹ]	/t(C)/ [ᵗˢ]	/.tr/ [ᵗɻ]	/r/ [r]	/r/ [ɹ]	/ɛ/ [ᵊ]	
☞ /.tr/	1			1		1.1
/.tɛr/				1	1	0.3

Our cross-language modelling thus predicts that Polish native listeners have problems distinguishing English minimal pairs like *train* [_ᵗɻ] – *terrain* [_ᵗˢɹ], as both map onto their native /.tr/, while English native listeners are predicted to misperceive Polish /.tr/ words (realized as [_ᵗˢr]) as their native /.tər/ (realized as [_ᵗˢɹ]).

A perception experiment performed by Schwartz (2023) with Polish native listeners that were first-year students of English showed that they were less accurate than native English listeners in identifying English *terrain* (and *derive*), although this difference was not significant. In addition, the Polish native listeners showed significantly longer reaction times for these words than the English native listeners.[12] This provides some support for our predictions. Unfortunately, there do not seem to exist any perception experiments that have tested the perception of Polish /Cr/ clusters by English listeners, or the discrimination of English minimal pairs like *train – terrain* by naïve Polish listeners to further support or refute the predictions that emerged from our formal modelling.

5. Remaining issues

Three topics remain to be addressed in this section, namely potential evidence for the generalization of cues that are relevant for /tr/ clusters to other contexts and other languages (Section 5.1), implications of our formalization for the division between phonetic and phonological processes in English and Polish (Section 5.2), and a short note on previously used terminology (Section 5.3).

5.1. Generalizations: Polish rearticulated geminates and second languages

One important assumption we made in the preceding formalization is that the requirement for an audible burst for Polish /t/ is generalized from a pre-plosive position, where it is absolutely necessary, to another position, namely before /r/,

where it is in principle not necessary for the perception of place of articulation.[13] Is there any evidence for such a generalization? We think that this type of evidence comes from geminates, which exist in Polish both underlyingly and due to morphophonological processes. Examples for plosive geminates are given in (8) (from Polakowska 2024, 8).

(8) /lɛkki/ [lɛkki] 'light'
/ɔt+tak/ [ɔttak] 'just like that'
/w+fɔtɛlu/ [ffɔtɛlu] 'in an armchair'

These geminates are produced with a lengthened constriction, as usual for geminates in the languages of the world, providing sufficient perceptual cues to the place of articulation (in the release) and their geminate nature (in the duration of closure). However, Polish is typologically unusual as it allows geminates to also be realized with a "perceptible release of the first consonant, especially observable in the case of stops and affricates" (Rojczyk 2022, 118), called "rearticulated" geminates. This rearticulation has not been reported for English derived geminates. Rojczyk (2022) found no significant differences in the perception accuracy for both "single-articulated" (i.e. normally articulated geminates) and rearticulated geminates for intervocalic Polish geminates.[14] The release of the first part is thus not necessary in Polish plosive geminates. Why are they nevertheless realized so often with a release? We propose that this is due to the cue constraint "/t(C)/ [ᵗˢ]" proposed in Section 4.1, which also applies to the first part of a geminate: the constraint did not specify the context "(C)", which can be identical to /t/. The existence of re-articulated geminates therefore provides evidence for our assumption that Polish listeners generalize "/t(C)/ [ᵗˢ]" from C being a heterorganic plosive to being a sonorant (where the audible release is not necessary), as they also generalize it to a homorganic plosive (where an audible release is also not necessary).

Further evidence for a generalization comes from foreign language pronunciation: Polish learners are reported to have "noticeably more frequent and stronger releases in English stop sequences than English native speakers" (Rojczyk et al. 2013, 6), indicating that Polish speakers also seem to generalize their cue constraint from their first to their non-native language(s), contributing to a non-native accent.

5.2. Implications for the phonetics–phonology divide

The BiPhon model presented here makes a strict distinction between phonetic and phonological representations and between phonetic implementation and phonological processes. While phonetic implementation involves several cues, which can differ gradually depending on factors like speech rate and phonetic context, phonological processes tackle abstract categories that either change or do not. For the case of /tr/ clusters in Polish and English, our modelling implies that English devoicing of post-obstruent rhotics (and liquids in general) is a phonetic process,

not a phonological one: it is part of the phonetic implementation of obstruent-liquid clusters captured in cue constraints, and is gradual, as it can range from slight devoicing to full affrication (a distinction we did not make in this paper, due to a lack of clear acoustic criteria). This differs from phonological treatments of this process, see e.g. Iverson and Salmons (1995, 372–374) who argue that onset clusters of obstruent and sonorant in English share the feature [spread glottis]. Similarly, our formalization considered the "intrusive vowel" in Polish clusters as a phonetic cue to a burst release, and not a phonological category that is inserted via a phonological process. This is in line with the distinction that Hall (2006) makes between "intrusive" and "epenthetic vowels" (see Section 1).

5.3. Terminology

Schwartz (2023, 633) calls the Polish perception of the English word *terrain* as *train* a "perceptual schwa deletion," in analogy to the "illusory vowel" or "perceptual epenthesis" effect (e.g. Dupoux et al. 1999; Dehaene-Lambertz et al. 2000; Dupoux et al. 2011). Kwon and Chitoran (2024) introduce the term "illusory cluster perception" for the comparable case of Georgian speakers who perceive two consonants with a salient burst and vowel transitions in between as a cluster. In our view, these terms are problematic, as they imply that there is imperfect perception on the side of the L2 listeners, as opposed to perfect perception on the native listeners' side, whereas what we observe is the application of an optimized native perception grammar to a non-native auditory input. Linguists need to be aware of the fact that there is no perfect or flawless perception of auditory input, that we all perceive any speech input with our native ears (see Hamann and Miatto 2025 for a similar argument), and that we need to acquire the knowledge of how to optimally perceive auditory cues in a new, foreign language.

6. Conclusion

In the previous sections, we provided arguments that /.tr/ clusters in English and Polish differ in their phonetic implementation, even though they are phonologically the same (or comparable). We showed how this language-specific implementation can be formalized in the model of Bidirectional Phonetics and Phonology, where the same knowledge (cue constraints and their weights) is used for speech perception, too, entailing that the interpretation of auditory-acoustic information is likewise different across languages. This was shown with the language-specific perception of short formant transitions that follow a plosive release, which are interpreted as a phonological vowel in English but not in Polish. Emilian dialects of Italian are similar to Polish in this respect, as they also treat an audible burst release as important perceptual cue for the plosive, and not as a phonological vowel (Cavirani and

Hamann 2023). Interestingly, these Emilian dialects, like Polish, have heterogenic plosive-plosive clusters, where an audible release is necessary to identify the place of articulation of the first plosive.

In contrast to previous accounts, which simply postulate a difference between English and Polish /.tr/ (such as "the two languages employ different gesture alignments for clusters," or "they have a different underlying phonological structure"), our account can motivate this difference: the two languages have different cluster inventories, which require different perceptual cues. This different cue requirement then results in different gestural alignments in the two languages. We hope to have shown that it is not necessary to assume a difference in phonological representations between the cluster /tr/ in the two languages only because there is a phonetic difference.

Acknowledgments

We are thankful to the audience at the 31st Manchester Phonology meeting, and the constructive comments by two anonymous reviewers. All errors of course remain our own.

Notes

1 For this reason, the /s/ in these clusters is often considered a syllable adjunct (or appendix) or extrametrical, i.e. not contributing to the sonority profile of the syllable (see e.g. Selkirk 1982).

2 In Polish, several clusters are only possible through affixation. For the purpose of this article, it is irrelevant whether clusters are monomorphemic or bimorphemic.

3 They can also be part of the larger onset cluster /str/, which is not further dealt with in this article.

4 A direct comparison of the segmental duration of /t/ and /r/ in isolation and in the cluster /tr/ does not seem to exist. Wierzchowska (1980) reports that the intrinsic duration of sounds in Polish is rather constant. The acoustic measurements by Jassem (1973) and Igras, Ziółko and Ziółko (2013) only show that /r/ is noticeably shorter than /t/.

5 Several measures of gestural overlap or gestural alignment can be found in the literature, such as the c-center effect and the related consonant coupling (see e.g. Browman and Goldstein 1988; Byrd 1995; Hermes et al. 2008), relative overlap (e.g. Chitoran et al. 2002; Gafos et al. 2010) and constriction duration overlap (e.g. Hoole et al. 2013; Crouch et al. 2023). We refrain from a detailed discussion and comparison as it is not relevant for our argumentation.

6 It is common in phonology and especially phonological typology to equate phonemes that are transcribed with the same symbol across languages, even though they usually engage in different contrasts and/or phonological processes. What is more, phonological typology would not be possible if a comparison across languages were

only allowed if the sounds in question would show exactly the same contrasts and phonological behavior.

7 Schwartz (2016, 58) elaborates that words of the form CV, such as *ra* or *ga*, are largely restricted to function words, while onset cluster words like *gra* are possible content words in Polish. He refers for this difference to Garrett (1999), who writes that "the minimal word in Polish is CVX" (p.18), making reference neither to the prosodic word nor to onset clusters.

8. The realization of a tap or trill requires short sequences of non-closure, i.e. sonorant material (see e.g. Ladefoged and Maddieson 1996, 218), cf. Dłuska's (1983) report of a short vocalic element before, after and in-between Polish /r/ (mentioned in Section 1). We do not treat these tap/trill-specific sonorant periods as separate cues to the Polish rhotic in this article.

9 Note that the perceptual interpretation of a short duration of formant transitions as a vowel or not is language-dependent, see Hamann and Miatto (2025) for a formalization of these language-dependent perceptions of simple burst noise and burst with formant transitions for Korean, English and Italian.

10 In a more complete account, additional cue constraints that further specify the realization of this consonant would be required.

11 Similar to schwa in English, /ɛ/ is the typical epenthetic vowel in Polish, as in clitics after *z* 'with/from' or *w* 'in/on/at' before some fricatives, e.g. *z*/ɛ/ *skałą* 'with a rock' vs *z matką* 'with the mother,' or *w*/ɛ/ *wtorek* 'on Tuesday' vs *w poniedziałek* 'on Monday.' The vowel /ɛ/ is also the vocalic realization in Polish alternating with zero in some noun stems, also referred to as *yer*, as in /sfɛtɐr/ – /sfɛtrɨ/ 'sweater sg. – pl.' (e.g. Bethin 2011), as a reviewer pointed out to us.

12 As the participants were first-year students of English, they are expected to have had considerable knowledge of English and thus an L2 perception grammar of English that departs from their native Polish perception grammar. Interestingly, *terrain* received more correct responses than *derive* because of the Polish cognate *teren*, causing L1 interference.

13 A further, less controversial generalization can be made, namely that the requirement for an audible release is not restricted to the coronal plosive /t/, but holds for all plosives in Polish. This could be formalized with the more general cue constraint "/plosive/ [*audible release*]."

14 Though there was a difference in reaction time: rearticulated geminates were perceived faster. This is not surprising as the audible release of the first part followed by cues indicating the same place and manner in the second part are explicit cues that this is a (re-articulated) geminate, while for a "single-articulated" geminate, the listener needs to listen to the whole duration of the geminate and to relativize this to the duration of the preceding vowel or the whole word before they can deduct that this is a geminate.

References

Aperliński, Grzegorz, and Geoffrey Schwartz. 2015. "Release Bursts vs. Formant Transitions in Polish Stop Place Perception." *Proceedings of the 18th Inter-*

national Congress of Phonetic Sciences. Ed. The Scottish Consortium for ICPhS 18. Glasgow: University of Glasgow.

Bethin, Christina Y. 2011. "Polish Syllable Structure." *The Blackwell Companion to Phonology*. Ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Karen Rice. Volume V. Malden, MA: Wiley-Blackwell. 2609–2630.

Blaho, Sylvia. 2008. "The Syntax of Phonology: A Radically Substance-free Approach." PhD diss, University of Tromsø.

Boersma, Paul. 2007. "Some Listener-oriented Accounts of h-aspiré in French." *Lingua* 117.12: 1989–2054. DOI: 10.1016/j.lingua.2006.11.004

Boersma, Paul. 2011. "A Programme for Bidirectional Phonology and Phonetics and Their Acquisition and Evolution." *Bidirectional Optimality Theory*. Ed. Anton Benz, and Jason Mattausch. Amsterdam & Philadelphia: John Benjamins. 33–72.

Boersma, Paul, Titia Benders, and Klaas Seinhorst. 2020. "Neural Network Models for Phonology and Phonetics." *Journal of Language Modelling* 8.1: 103–177. DOI: 10.15398/jlm.v8i1.224

Boersma, Paul, Kateřina Chládková, and Titia Benders. 2022. "Phonological Features Emerge Substance-freely from the Phonetics and the Morphology." *Canadian Journal of Linguistics* 67: 611–669.

Boersma, Paul, and Silke Hamann. 2008. "The Evolution of Auditory Dispersion in Bidirectional Constraint Grammars." *Phonology* 25: 217–270.

Boersma, Paul, and Silke Hamann. 2009. "Loanword Adaptation as First-language Phonological Perception." *Loanword Phonology*. Ed. Andrea Calabrese, and W. Leo Wetzels. Amsterdam & Philadelphia: John Benjamins. 11–58.

Boersma, Paul, and Jan-Willem van Leussen. 2017. "Efficient Evaluation and Learning in Multi-level Parallel Constraint Grammars." *Linguistic Inquiry* 48: 349–388.

Browman, Catherine P., and Louis Goldstein. 1988. "Some Notes on Syllable Structure in Articulatory Phonology." *Phonetica* 45: 140–155.

Bruni, Jagoda. 2011. "Sonorant Voicing Specification in Phonetic, Phonological and Articulatory Context." PhD diss, University of Stuttgart.

Bybee, Joan. 1988. "Morphology as Lexical Organization." *Theoretical Morphology*. Ed. Michael Hammond, and Michael Noonan. San Diego: Academic Press. 119–141.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Byrd, Dani. 1995. "C-centers Revisited." *Phonetica* 52: 285–306.

Cavirani, Eduardo, and Silke Hamann. 2023. "Formalising Phonological Perception: The Role of Voicing Assimilation in Consonant Cluster Perception in Emilian Dialects." *Journal of Linguistics* 60.2: 253–284.

Cetnarowska, Bożena. 2000. "On the (Non-)Recursivity of the Prosodic Word in Polish." *ZAS Papers in Linguistics* 19: 1–21.

Chabot, Alex. 2019. "What's Wrong with Being a Rhotic?" *Glossa* 4: 38. DOI: 10.5334/gjgl.618

Chitoran, Ioana, Louis Goldstein, and Dani Byrd. 2002. "Gestural Overlap and Recoverability: Articulatory Evidence from Georgian." *Papers in Laboratory Phonology 7*. Ed. Carlos Gussenhoven, and Natasha Warner. Berlin: De Gruyter. 419–448.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper & Row.

Cieślak, Joanna. 2015. "English-Polish Differences in the Organization of Consonant Clusters: A Study in Phonetic Code-switching." MA thesis, Adam Mickiewicz University in Poznań.

Crouch, Caroline, Argyro Katsika, and Ioana Chitoran. 2023. "Sonority Sequencing and Its Relationship to Articulatory Timing in Georgian." *Journal of the International Phonetic Association* 53.3: 1049–1072.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*. 8th ed. London & New York: Routledge.

Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux, and Ariel Gout. 2000. "Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study." *Journal of Cognitive Neuroscience* 12.4: 635–647.

Dłuska, Maria. 1983. *Fonetyka polska. Artykulacje głosek polskich.* 2nd ed. Warszawa: Państwowe Wydawnictwo Naukowe.

Dukiewicz, Leokadia, and Irena Sawicka. 1995. *Gramatyka współczesnego języka polskiego: fonetyka i fonologia*. Kraków: Wydawnictwo Instytutu Języka Polskiego PAN.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. "Epenthetic Vowels in Japanese: A Perceptual Illusion." *Journal of Experimental Psychology: Human Perception and Performance* 25: 1568–1578.

Dupoux, Emmanuel, Erika Parlato, Sonia Frota, Yuki Hirose, and Sharon Peperkamp. 2011. "Where Do Illusory Vowels Come from?" *Journal of Memory and Language* 64.3: 199–210.

Flemming, Edward S. 2013. *Auditory Representations in Phonology*. London & New York: Routledge.

Fromkin, Victoria. 1975. "The Interface between Phonetics and Phonology." *Proceedings of the 8th International Congress of Phonetic Sciences*. Leeds.

Gafos, Adamantios I., Philip Hoole, Kevin Roon, Chakir Zeroual, Cécile Fougeron, Barbara Kühnert, Marioapaola D'Imperio, and Nathalie Vallée. 2010. "Variation in Overlap and Phonological Grammar in Moroccan Arabic Clusters." *Laboratory Phonology* 10: 657–698.

Garrett, Edward. 1999. "Minimal Words Aren't Minimal Feet." *UCLA Working Papers in Linguistics* 111: 92–122.

Hale, Mark, and Madelyn Kissock. 2007. "The Phonetics–Phonology Interface and the Acquisition of Perseverant Underspecification." *The Oxford Handbook of Linguistic Interfaces*. Ed. Gillian Ramchand, and Charles Reiss. Oxford: Oxford University Press. 81–102.

Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise.* Oxford: Oxford University Press.

Hall, Nancy. 2006. "Cross-linguistic Patterns of Vowel Intrusion." *Phonology* 23: 387–429.

Hamann, Silke. 2011. "The Phonetics–Phonology Interface." *Continuum Companion to Phonology*. Ed. Nancy C. Kula, Bert Botma, and Kuniya Nasukawa. London: Continuum. 202–224.

Hamann, Silke, and Veronica Miatto. 2025. "Three Language-specific Phonological Interpretations of Release Bursts and Short Vowel-like Formants." *Epenthesis and Beyond: Recent Approaches to Insertion in Phonology and Its Interfaces*. Ed. Ji Yea Kim, Veronica Miatto, Andrija Petrović, and Lori Repetti. Berlin: Language Science Press. 199–223.

Hermes, Anne, Martine Grice, Doris Mücke, and Henrik Niemann. 2008. "Articulatory Indicators of Syllable Affiliation in Word Initial Consonant Clusters in Italian." *Proceedings of the 8th International Seminar on Speech Production*. Ed. Rudolph Sock, Susanne Fuchs, and Yves Laprie. Strasbourg: LORIA. 433–436.

Hermes, Anne, Doris Mücke, and Bastian Auris. 2017. "The Variability of Syllable Patterns in Tashlhiyt Berber and Polish." *Journal of Phonetics* 64: 127–144.

Hoole, Philip, Marianne Pouplier, Stefan Benus, and Lasse Bombien. 2013. "Articulatory Coordination in Obstruent–Sonorant Clusters and Syllabic Consonants: Data and Modelling." *Rhotics: New Data and Perspectives*. Ed. Lorenzo Spreafico, and Alessandro Vietti. Bolzano: Bozen-Bolzano University Press. 81–98.

Igras, Magdalena, Bartosz Ziółko, and Mariusz Ziółko. 2013. "Length of Phonemes in a Context of their Positions in Polish Sentences." *Proceedings of the 10th International Conference on Signal Processing and Multimedia Applications*. 59–64.

Iosad, Pavel. 2012. "Representation and Variation in Substance-free Phonology: A Case Study in Celtic." PhD diss, University of Tromsø.

Iverson, Gregory K., and Joseph C. Salmons. 1995. "Aspiration and Laryngeal Representation in Germanic." *Phonology* 12.3: 369–396.

Jassem, Wiktor. 1973. *Podstawy fonetyki akustycznej*. Warszawa: Państwowe Wydawnictwo Naukowe.

Jaworski, Sylwester, and Edward Gillian. 2011. "On the Phonetic Instability of Polish /r/." *Poznań Studies in Contemporary Linguistics* 47.2: 380–398.

Keating, Patricia A. 1988. "The Phonology–Phonetics Interface." *Linguistics: The Cambridge Survey. Volume I: Linguistic Theory: Foundations*. Ed. Frederick Newmeyer. Cambridge: Cambridge University Press. 281–302.

Kwon, Harim, and Ioana Chitoran. 2024. "Perception of Illusory Clusters: The Role of Native Timing." *Phonetica* 81.2: 153–184.

Marin, Stefania, and Marianne Pouplier. 2010. "Temporal Organization of Complex Onsets and Codas in American English: Testing the Predictions of a Gestural Coupling Model." *Motor Control* 14: 380–407.

O'Shaughnessy, Douglas. 1974. "Consonant Durations in Clusters." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22.4: 282–295.

Ogden, Richard. 2017. *An Introduction to English Phonetics*. 2nd ed. Edinburgh: Edinburgh University Press.

Pierrehumbert, Janet B. 1990. "Phonological and Phonetic Representation." *Journal of Phonetics* 28: 375–394.

Pierrehumbert, Janet B. 2001. "Exemplar Dynamics: Word Frequency, Lenition, and Contrast." *Frequency Effects and the Emergence of Linguistic Structure*. Ed. Joan Bybee, and Paul J. Hopper. Amsterdam & Philadelphia: Benjamins. 137–157.

Polakowska, Weronika. 2024. "The Perception of Polish Geminates: A Study on Geminate Articulation and Social Meaning." BA thesis, University of Amsterdam. https://www.fon.hum.uva.nl/archive/2024/2024-BA-WeronikaPolakowska.pdf

Rojczyk, Arkadiusz. 2022. "The Perception of Rearticulated and Single-articulated Geminates in Polish." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.1–2: 118–126. DOI:10.1017/cnj.2022.5

Rojczyk, Arkadiusz, Andrzej Porzuczek, and Marcin Bergier. 2013. "Immediate and Distracted Imitation in Second-language Speech: Unreleased Plosives in English." *Research in Language* 11.1: 3–18.

Schwartz, Geoffrey. 2010. "Phonology in the Speech Signal – Unifying Cue and Prosodic Licensing." *Poznań Studies in Contemporary Linguistics* 46.4: 499–518.

Schwartz, Geoffrey. 2016. "On the Evolution of Prosodic Boundaries: Parameter Settings for Polish and English." *Lingua* 171: 37–73.

Schwartz, Geoffrey. 2023. "All TRs Are not Created Equal: L1 and L2 Perception of English Cluster Affrication." *Journal of Linguistics* 59.3: 623–654.

Schwartz, Geoffrey, Katarzyna Dziubalska-Kołaczyk, and Radosław Święciński. 2024. "Phonotactic and Morphonotactic Influences on the (A)Synchronicity of Consonant Clusters in Polish." *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 141.3: 205–217.

Selkirk, Elizabeth. 1982. "The Syllable." *The Structure of Phonological Representations*. Ed. Harry van der Hulst, and Norval Smith. Dordrecht: Foris. 337–382.

Selkirk, Elizabeth. 1984. “On the Major Class Features and Syllable Theory.” *Language Sound Structure: Studies in Phonology*. Ed. Mark Aronoff, and Richard T. Oehrle. Cambridge MA: MIT Press. 107–136.

Seinhorst, Klaas, Paul Boersma, and Silke Hamann. 2019. “Iterated Distributional and Lexicon-driven Learning in a Symmetric Neural Network Explains the Emergence of Features and Dispersion.” *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne. 1134–1138.

Steriade, Donca. 2004. “Directional Asymmetries in Place Assimilation: A Perceptual Account.” *Phonetically Based Phonology*. Ed. Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge: Cambridge University Press. 219–250.

Stolarski, Łukasz. 2013. “Articulation of the Polish /r/ in the Intervocalic Position.” *Language, Thought and Education: Across Systems*. Ed. Marek Kuczyński, and Leszek Szymański. Zielona Góra: Oficyna Wydawnicza Uniwersytetu Zielonogórskiego. 21–35.

Stolarski, Łukasz. 2015. “Further Analysis of the Articulation of /r/ in Polish – The Postconsonantal Position.” *SKY Journal of Linguistics* 28: 349–379.

Trask, Robert Lawrence. 1996. *A Dictionary of Phonetics and Phonology*. London & New York: Routledge.

Tsuchida, Ayako, Abigail C. Cohn, and Masanobu Kumada. 2000. “Sonorant Devoicing and the Phonetic Realization of [Spread Glottis] in English.” *Working Papers of the Cornell Phonetics Laboratory* 13: 167–181.

Volenec, Veno, and Charles Reiss. In press. “Cognitive Phonetics: The Universal Phonology–Phonetics Interface.” *Cambridge Handbook of Linguistic Interfaces*. Ed. Antonio Fábregas, Laia Mayol, Clàudia Pons-Moll, and Yanina Prystauka. Cambridge: Cambridge University Press.

Wierzchowska, Bożena. 1980. *Fonetyka i fonologia języka polskiego*. Wrocław: Ossolineum.

Zhou, Chao, and Silke Hamann. 2024. “Modelling the Acquisition of the Portuguese Tap by L1-Mandarin Learners: A BiPhon-HG Account for Individual Differences, Syllable-position Effects and Orthographic Influences in L2 Speech.” *Glossa* 9.1. DOI: 10.16995/glossa.9692

Zhou, Chao, and Silke Hamann. 2026. “A Revised Multi-modal Formal Account for Orthography-induced Sound Patterns.” *Linguistic Inquiry*. Advance publication. DOI: 10.1162/LING.a.585

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Karolina Broś, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.07

Karolina Broś
https://orcid.org/0000-0001-6701-9257
University of Warsaw

Lip Tracking and Acoustics at the Phonetics–Phonology Interface: Coronal Stop Deletion in American English

Abstract: American English is well-known for optional /t d/ deletion in clusters, as in *best* [bes] *man*, *kept* [kʰep] *going* or *whipped* [wɪp] *butter*. However, research shows that even in the absence of an acoustic trace of /t d/, articulatory gestures related to the alveolar constriction are still there, although they may be of a smaller magnitude, which questions the completeness of deletion from a phonological standpoint. This study uses motion capture to look at /t d/ trapped in between consonants vs a neutral context and explore the interplay of acoustics with articulation. The results show significant differences between sequences with deleted and retained coronals in both lips and acoustic measurements. Some gradient effects and weakening strategies are also discussed, pointing to the need for a more wholesome approach to the study of (apparent) sound deletion in the context of overall articulatory strategies.

Keywords: coronal deletion, articulation, motion capture, lip tracking

1. Introduction and literature review

Final /t d/ deletion is a well-known phenomenon of English phonology whereby consonants /t/ and /d/ are deleted when preceded by another consonant in connected speech. Some examples include *best man* pronounced as [bes mæn] or *kept going* [kʰep goʊɪŋ]. The process has been reported for many varieties of English, from

American, through Chicano English to British dialects (Coetzee 2004; Baranowski and Turton 2020). As an example of a variable process that is optionally applied by speakers, it has been studied for many years now (Labov and Cohen 1967; Labov et al. 1968; Tagliamonte and Temple 2005; Walker 2012; Tamminga 2016). Unfortunately, different studies have led to contrary findings and varying conclusions concerning such matters as the phonological vs morphological conditioning of the process, as well as its putative categoricity. Importantly, the completeness of deletion is still subject to debate (Purse 2019).

The discrepancies between the findings of previous studies can be seen as due to methodological limitations. On the one hand, /t d/ deletion is a connected speech phenomenon, which means that it should be studied based on natural, spontaneous productions. Only then can generalizations about its frequency and conditioning be reliable. This assumption has led to an upsurge in corpus-based studies of the phenomenon (e.g. Baranowski and Turton 2020; Renwick et al. 2014; Walker 2012). However, while uncontrolled samples of this kind can provide more tokens of /t d/, they can also introduce several types of bias related to the greater frequency of certain morphological forms or phonetic contexts. For instance, it has been suggested that /t d/ deletion is more prevalent in monomorphemic words such as *mist* compared to complex words in which /t d/ serve as grammatical markers (e.g. *missed*, Guy 1980; Wolfram 1969). Not all studies confirm this effect, however, and it has been proposed that the quantitative difference in the frequency of deletion can be an artefact resulting from the greater frequency of monomorphemic forms in the language or sample (Sonderegger et al. 2011). The same applies to the reported effect of the preceding consonant. It has often been claimed that preceding /s/ and /n/ are the most common drivers of deletion, but it is precisely /st/ and /nt/ clusters that are the most prevalent (Temple 2009). Some researchers tried to control for these confounds in the data and still found evidence for phonetic and morphological effects in certain dialects (Baranowski and Turton 2020).

Another important question related to the process of /t d/ deletion concerns the phonological status of the consonant itself. Palatography and electromagnetic articulography (EMA) studies have shown that complete deletion of the /t/ or /d/ gesture happens only rarely (Purse and Turk 2016). In other words, although the consonant is not audible, the tip of the tongue does move upwards to produce it (i.e. to touch the alveolar ridge) but this gesture is either incomplete or masked by neighboring gestures (Browman and Goldstein 1990). Purse (2019), for instance, registered only up to 6% of true /t d/ deletions in a sample of American English speakers. The problem is, however, that such articulatory studies require a controlled procedure, with specifically designed tokens, typically a word list or minimal pairs to read out, and only allow for testing a small number of stimuli and participants. Thus, it may be that the lack of complete deletion observed in these studies is a result of eliciting unnatural speech and not capturing the phenomenon in its full form. Some indication of this interpretation is the result found by Purse for three of the five

tested speakers in which there was a bimodal distribution of /t d/ pronunciations despite the lack of complete deletion. Although most tokens were produced with some degree of tongue tip raising, there was a clear difference between audible and inaudible tokens. Those speakers seemed to produce two types of stops: a full stop and a weakened (i.e. lenited) stop type in which the tongue gesture was much smaller (i.e. incomplete). It would be interesting to know whether this tendency is confirmed on a larger sample of participants and enhanced when the experimental procedure does not include short word lists. This is precisely the aim of this project.

1.1. Using the motion capture technique

As mentioned above, studying /t d/ deletion using acoustic data is not necessarily reliable because the consonant that is apparently deleted and cannot be heard can nevertheless be covertly produced to some extent. This can only be confirmed or disconfirmed using advanced articulatory techniques, such as EMA, which consists in putting sensors on the participant's tongue to track its exact position. However, such methods are both time-consuming and resource intensive.[1] There is, nonetheless, an alternative that is both easier to implement and better at getting more data from a larger sample: *motion capture*. It has been used to study orofacial movements accompanying speech for some time now (Sangari 2002; Esteve-Gibert and Prieto 2013; Żygis et al. 2017). These studies focused on speech in prominent positions, and on different magnitudes of mouth opening or brow raising depending on the type of utterance or speech modality. Recently, an implementation of motion capture that uses inexpensive video recording equipment has been found to allow for optical tracking of the mouth and lips when studying individual sound productions (Holbrook et al. 2019; Krause et al. 2020). As a result, recordings can be made in the lab or in the field using an internet camera. With a specially designed algorithm based on OpenFace face-tracking, the method allows for extracting lip measurements from video material. These include not only lip closure vs lip opening, but also lip compression, which allows for distinguishing different degrees of lip closure. The method was tested empirically by Broś and Krause (2024), who collected video data testing the production of Spanish stop weakening in postvocalic and postconsonantal position. The extracted lip movement data was used to test relative lip opening in the productions of native speakers' /p/ and /b/ sounds. The method proved useful not only in showing significantly different degrees of lip aperture for /p b/, but also in showing differences in lip trajectories depending on whether a preceding coronal /s/ was pronounced or not. As a result, Broś and Krause were able to show phonological differences between stops in underlying and derived onset positions. The method was also fine-grained enough to show prosodic prominence effects. In this project, I use a similar technique and comparable measurements to look at a different segmental process. More specifically, I compare the degrees of lip closure and lip compression in anticipation of and

following /t d/ pronunciation (or deletion) in sequences of /t d/ flanked by labial sounds in American English.

1.2. Studying lips when looking at coronal sounds

While motion capture looks promising when looking at labial sounds or vowels, in which case the lips are actively involved, as demonstrated by Krause et al. (2020) and Broś and Krause (2024), this logic does not directly apply to /t d/. However, as I will show in this paper, with an appropriate context and approach, coronal (non) deletion can be studied as well. More precisely, the design of the project focuses our attention on labial sounds and their production next to, and in conjunction with, coronal /t d/. In this case, not individual sounds or sequences of vowels and consonants, but sequences of sounds starting with a bilabial closure are compared. Several conditions are taken into account and lip closure and compression are measured and compared with acoustic data in order to evaluate relative /t d/ weakening. The general idea behind this design is that in a sequence of words containing a final /t d/ in which the target sound is preceded and followed by a bilabial sound, e.g. *sto**pped p**edaling* or *sto**pped m**oving* (sequences /ptp/ or /ptm/) we necessarily have lip closure, then release (opening) for bilabial stop release and [t] production (i.e. tongue tip raising) in the middle of the sequence and then lip closure again for [p] or [m]. If the [t] is not pronounced, however, no release and hence no lip opening is expected between the two bilabials – one longer closure is produced. If the [t] is inaudible but tongue raising occurs, as suggested by earlier EMA-based literature, this suggests gestural overlap or masking. Consequently, I hypothesize that there might be a change in lip compression even if lip closure is not fully released. To the best of my knowledge, such predictions have not been studied to date. Thus, the design used in the present study may help capture intermediate effects (e.g. weakened segments) by allowing a reliable comparison of minimal pairs in which the stop is pronounced with those with (apparently) deleted stops. We can also compare such cases with 'neutral' sequences, such as *ste**pped o**ut*, where there is no gestural plan for the lips to close after the [t]. As a result, the motion capture technique can provide some evidence for gestural organization and planning.[2]

Furthermore, apart from measuring lip compression and the degree of opening, we can also look at the duration of the sequence. If there is a difference in the duration of closure for [pt] depending on whether the [t] was pronounced (audible) or not, this will inform us of incomplete deletion. In other words, if the [t] is completely deleted and there is no tongue tip raising gesture, then we would expect the [pt] sequence to be shorter than when the [t] is present and the same as in a control condition that does not contain an underlying /t/ at all (e.g. *ste**p o**ut, sto**p m**oving*). If the duration of such a sequence with an inaudible [t] is longer, this means that it includes the inaudible [t] at least in some part and hence deletion is not complete.

We have gestural weakening. Thus, with the above design we may be able to get information on the presence or absence of tongue tip raising without ever looking at the tongue and having to engage in EMA. Phonetic annotation of the audio files and extraction of acoustic parameters, such as sound and sequence duration, intensity and voicing can serve as additional parameters in the determination of surface productions.

With this in mind, the aim of the present paper is to determine whether we can reliably distinguish between the different variants of /t d/ being produced in a more categorical manner, i.e. pronounced stop, (inaudible) weakened stop, (inaudible) deleted stop or the observed differences are gradient in nature. To do this, I look at specific sound sequences embedded in longer sentences produced both when reading and when retrieving them from memory to elicit a greater number of more natural productions. A large number of sentences is used to collect a larger data sample from more speakers. The answer to the above research question will help disentangle the phonetics from the phonology of /t d/ weakening in American English. At the same time, the study allows me to test an alternative, more user-friendly data collection method focused on speech articulation, hopefully advancing our knowledge of articulatory coordination. It is worth noting that the motion capture technique has not been used as a proxy in the study of lingual gestures to date and the present work is the first to make such an attempt.

2. Theoretical framework

In their seminal work on Articulatory Phonology, Browman and Goldstein (1990) propose that deletion is always incomplete because the articulatory gestures are not canceled in fast speech or weakening contexts. Neither do they change their articulatory plan. Instead, a reduction in the magnitude of gestures takes place, causing gestural overlap. This, in turn, changes the acoustic output. When gestures do not reach their targets, reach them faster or later than planned, some sounds may appear to be deleted as they are rendered inaudible. The same can happen if gestural masking occurs, i.e. one gesture is hidden or actively co-occurs with another. This is the most probable scenario in timing changes of gestures occurring at different tiers (e.g. tongue vs lip movements). Thus, the apparent deletion of a segment is not a categorical process but a surface consequence of gradient changes in articulatory coordination. This, in turn, is of interest to researchers studying the phonetics and phonology interface.

In the case of coronal stop deletion and its putative categoricity, the existence of different rates of process application depending on morphological structure, for instance, has led to the proposal of semi-categorical changes whereby weakened, incomplete gestures leading to non-canonical, perhaps inaudible pronunciations

of /t d/ might be interpreted as allophonic variants of full stops in the course of phonological analysis (e.g. Guy 1991; Bermúdez-Otero 2010; see also Purse and Turk 2016). In his investigation of the issue at hand, Purse (2019) shows that the situation is more nuanced than previously assumed as inaudible /t d/ segments can be produced with or without tongue tip raising, although the latter situation is much less common. In this study, I take the insights of Articulatory Phonology as the basis for the exploration of the relative influence of lip movements on speech coordination over sequences of sounds containing possible 'deletion' (or masking) targets, i.e. coronal stops. More specifically, I look at differences in lip configuration prior to and during the expected /t d/ pronunciation. I also look at acoustic markers that accompany either audible or inaudible /t d/ productions to see the interplay of acoustic and articulatory factors.

The main acoustic marker explored in this study is sound and sequence duration, as compared between analogous sound structures in target and control conditions. Furthermore, given previous interpretations of coronal stop deletion as weakening, I use two more lenition markers. One of them is intensity, a common parameter in consonant lenition studies, typically used to measure relative constriction degree that aligns well with articulatory data (Parrell 2010). Intensity has been used by numerous researchers (Carrasco et al. 2012; Hualde et al. 2010; Hualde and Nadeu 2011; Hualde et al. 2011). It can be calculated as minimum intensity of a sound subtracted from the maximum intensity of the preceding vowel (e.g. Broś et al. 2021; Broś and Krause 2024). The lower the resulting score, the less constricted and hence the more weakened the tested consonant. In this study, I follow the same method with one subtle change. Since I am analyzing both individual first sounds and sequences of two consonants (bilabial + coronal) together as key sounds/clusters, the score corresponds to the difference between the minimum intensity of the bilabial or bilabial+coronal and the preceding vowel.

The second measurement employed in this study is harmonics-to-noise ratio which was used e.g. by Broś et al. (2021) as a marker of relative stop weakening in postvocalic position. This measurement has been found reliable in studies on partly harmonic sounds in which periodicity can be measured to explore the degree of weakening assuming that the process leads to greater aperture, and hence sonority and relative vowel-likeness (Bárkányi and Kiss 2010). Any gradient effects in the form of sound weakening and (partial) deletion should lead to changes in this parameter.

Finally, I look at lip trajectories over target sequences that show changes in lip aperture and lip compression, which potentially mark different articulatory strategies and weakening degrees. The aim of the study is to test the following hypotheses:

1) Sequence and cluster durations will differ depending on the deletion status (i.e. between coronal retention, apparent deletion and controls). They will be the shortest in controls and the greatest in retention cases.

2) Intensity difference will differ depending on the deletion status (i.e. between coronal retention, apparent deletion and controls). The highest values are expected for retention, the lowest for controls.
3) HNR will differ depending on the deletion status (i.e. between coronal retention, apparent deletion and controls). The lowest values are expected for retention, the highest for controls.
4) Lip aperture will differ depending on the deletion status (i.e. between coronal retention, apparent deletion and controls). The highest values are expected for retention, the lowest for controls.
5) Lip thickness will differ depending on the deletion status (i.e. between coronal retention, apparent deletion and controls). The highest values are expected for retention, the lowest for controls.

In all cases, the following interpretation applies. If apparent deletions (i.e. inaudible /t d/) pattern with retention, this would suggest complete gestural masking or overlap. If apparent deletions pattern with controls, this would suggest complete deletion. If they have intermediate values, this is consistent with a weakening / partial overlap interpretation.

3. Methodology

The method used in the experiment was motion capture combined with acoustic measurements. The study was conducted using a Micro ultrasound system for speech research from Articulate Instruments Ltd (Edinburgh, UK) in the California State University Channel Islands' Psychology lab in February and March 2025. Here, only lip tracking and acoustic data are reported. The experiment lasted around 60 minutes and tested the following conditions:

- /t d/ trapped between two bilabials – bilabial context, e.g. *stopped making*
- analogous bilabial-bilabial sequences without an intervening /t d/ – bilabial control, e.g. *stop making*
- /t d/ trapped between a bilabial and a vowel – vowel context, e.g. *stepped out*
- analogous bilabial-vowel sequences without an intervening /t d/ – vowel control, e.g. *step out*

Based on previous literature, it was assumed that /t d/ deletion in target sequences would be elicited around 50% of the time, although this rate could be lower since morphologically complex words were used to ensure comparability between conditions. According to previous studies on coronal stop deletion in English regular verbs are less prone to weakening compared to irregular verbs and morphologically simple words. The rate of /t d/ deletion before vowels is generally

low; hence, it was assumed that this would occur only rarely, the vowel context playing the role of a second control condition. Thanks to the above design, both audible and inaudible /t d/ sounds can be reliably compared.

3.1. Materials

98 sentences were constructed in which a regular verb ending in a bilabial sound was the first key word in the target phrase. The verb was presented either in the present or in the past tense and followed by a word beginning with a bilabial sound or a vowel. All bilabials were used, i.e. /p b m/. Each sentence was relatively short to enable memorization. The same verbs were used across the four conditions giving quadruplets such as the following:

(1) He scrubbed marks off the wall.
(2) I had to scrub my hands with alcohol after work.
(3) She scrubbed an old pan clean.
(4) She decided to scrub in for the surgery.

This gave us the following sequences: /ptm/, /ptb/, /mdp/ and /bdm/ in the bilabial context; /pm/, /pb/, /mp/ and /bm/ in bilabial controls; /ptV/, /bdV/ and /mdV/ in the vowel context (the vowels varied across sentences); and /pV/, /bV/ and /mV/ in vowel controls. The verbs used in the study are listed in Table 1.

Table 1. Verbs used in the experiment

Voiceless /t/	Verbs ending in -p	stop, chop, crop, clap, nap, dip, drip, whip, slip, trip, ship, skip, group, prep
Voiced /d/	Verbs ending in -b	sob, rob, grab, dub, scrub, snub, rub
	Verbs ending in -m	cram, scam, trim, dim, sum, numb, hum

The full list of sentences used in the experiment is presented in the Appendix. Prior to conducting the experiment, an earlier version of the list was consulted with native speakers from California to ensure lexical consistency with the local variety of English and language usage, as well as grammatical correctness, simplicity and contextual adequacy. Four psychology students from the host institution gave their feedback, which led to the amendment of the list.

3.2. Participants and procedure

The data under analysis comes from 13 participants (2 males) aged 20–28.[3] Nine of them were of Hispanic or mixed Hispanic/Caucasian ethnicity, one reported to be Asian, and one of mixed American Indian and African American ethnicity. The

remaining two participants identified as Caucasian. Six of the Hispanic participants speak Spanish on an everyday basis. All of the participants except one were born in California.

The participants were prepared for data collection by the experimenter: a 3D-printed nonmetallic UltraFit headset produced by Articulate Instruments was put on the participant's head that contained a microphone and a small camera focused on the lip area. Articulate Assistant Advanced software (henceforth, AAA; Articulate Instruments Ltd 2019) was used to present and record data. After calibration, trial items were presented on the screen for the participant to read out. These included a few individual words testing different vowel sounds (as a baseline), followed by one practice sentence. The headset and instructions were adjusted during this period. After the practice session, the experiment was put in automated mode whereby each sentence was presented on the screen for 5 seconds. The participant had to read the sentence out loud after an initial auditory prompt. After the predefined time passed, the next screen appeared. This time, the participant only saw the number of the sentence and had to repeat the previously read sentence from memory. At the third trial, the procedure was repeated. Next, a different sentence appeared on the screen and the experiment proceeded in a similar fashion. A total of 98 sentences were presented in a pseudo-randomized order, each repeated three times. There were four randomized lists. Each new participant was assigned a different list (1,2,3,4,1,2,3,4 and so on). After finishing the data collection part, the participants were asked to fill in a short demographic questionnaire.

3.3. Data extraction and analysis

After data collection, video and audio signals were synched and data concerning the positioning of the lips was extracted. AAA's internal machine learning algorithm was used to mark lip contours frame by frame for each participant. Audio files were extracted and annotated in Praat (Boersma and Weenink 2025) by two junior annotators, followed by corrections made by the author. High-fidelity headphones were used to ensure precise auditory analysis (BeyerDynamic DT 770 Pro 80 Ohm). Annotations included several tiers marking the three-segment sequence, the key cluster (first bilabial + /t/ or /d/ when applicable), as well as preceding vowel and following sound (bilabial or vowel). Apart from this, repetition number and target verb were annotated, as well as /t d/ status: deleted/retained, and preceding sound status: weakened/not weakened.

To explain the annotation procedure, whenever the /t d/ was not audible and the visual inspection of the spectrogram did not confirm the presence of a burst, aspiration or energy after the initial bilabial, or an additional period of silence after the preceding bilabial release (if applicable), the sound was annotated as apparently deleted. The preceding sound (i.e. the initial bilabial) could also be weakened as

it stands in a position rendering lack of plosion according to standard English allophony rules. Thus, it was also determined whether this sound was fully released or not. Additional notes specified whether weakening occurred via weak release, lack of release (burst), fricativization, nasal release (if applicable) or apparent complete deletion.

The resultant TextGrids were imported in AAA, which allowed for the extraction of lip data for each tested sound sequence. The data were extracted at 11 timepoints calculated proportionally over the sequence and included lip aperture (i.e. the difference between the positions of the two inner lips at their midpoints) and lip thickness (i.e. the difference between the inner and outer lower lip at its midpoint). The aim of using lip thickness was to measure lip compression, i.e. tenseness of lip closure, which may be relevant if lips do not open completely during articulation. On the acoustic part, sound and sequence durations were extracted, alongside the voice report and intensity measurements. This was performed using a dedicated Praat script.

After extraction, both acoustic and motion capture data were preprocessed in R (R Core Team 2020) and analyzed using generalized linear mixed-effects models using the *lme4* package (Bates et al. 2018). The *emmeans* package was used to calculate simple effects and perform post-hoc tests (Lenth 2019). For all models, marginal means were estimated to explore interactions. The resulting simple effects were tested via *t*-tests, using the Kenward–Roger method of estimating the degrees of freedom. Familywise error was contained using Tukey's method. Descriptive plots and effects plots were generated using *ggplot2* (Wickham 2016). The exact model structure is reported for each of the tested variables in Section 4.

4. Findings

The results of the study are first presented descriptively. This is followed by the report of statistical models testing the hypotheses for each variable of interest. In all cases, I refer to coronal deletion / deletion status in the sense of apparent deletion, as judged based on the acoustic and auditory analysis of the recorded data. In other words, I refer to inaudible stops. After reporting the results from both acoustics and articulation, an interpretation of these apparent deletions is then given in Section 5.

The data encompasses a total number of 3607 observations. The number of analyzed sequences differed depending on the participant given the exclusion of mispronunciations and acoustically or articulatorily bad data. Some of the sentences were omitted or not repeated correctly from memory, leading to further exclusions. Additionally, two participants had fewer trials than expected given technical problems with the software. All in all, the mean number of analyzed sequences out of the target 297 was 277. The table of data by sequence type and participant is presented in Table 2.

Table 2. Datapoints by sequence type and participant[4]

Condition	B01	B05	B09	B17	B21	B23	M03	M05	M12	M13	M18	M21	M23
Bilabial context	81	70	84	88	87	79	82	86	84	84	83	86	83
Bilabial control	67	63	72	76	80	82	83	76	76	80	78	82	86
Vowel context	57	45	55	60	61	60	62	61	61	63	63	58	62
Vowel control	61	46	59	61	63	59	57	56	59	62	59	62	57

There were a total of 1762 control items in the data and 1845 target sequences with possible deletions. Deletion took place 399 times in the bilabial context (37%), and only 27 times in the vowel context, i.e. only 3.5% of the time. The rare occurrence of deletion before vowels is in line with the literature while the bilabial context shows a slightly lower rate than expected. The deletion rates varied from speaker to speaker, ranging from 6 to 67 deleted items. The mean number of deleted items was 35, i.e. 24.5%, with some speakers reaching almost 50% and one speaker having none.

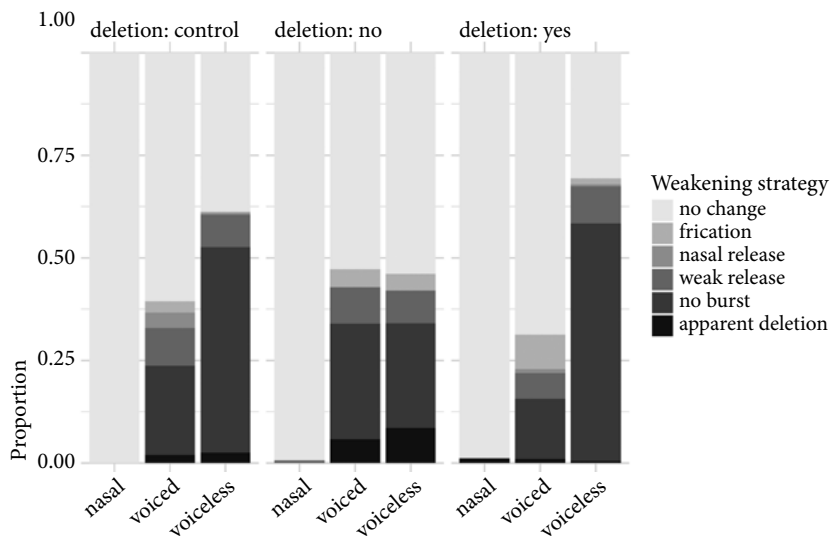


Fig. 1. Bilabial weakening strategies depending on deletion status and type of sound starting the sequence. 'No change' means that no sound weakening occurred. *Deletion: control* includes control sentences constructed with the present tense, *Deletion: no* shows examples of past tense with /t d/ retention, and *Deletion: yes* shows cases of past tense with inaudible /t d/.

Apart from /t d/ deletion, the participants also used a range of other weakening strategies in the key cluster that concerns the first sound (i.e. bilabial [p], [b] or [m], as in *shi**pp**ed boxes, du**bb**ed movies, su**mm**ed prices*). These included weak release, lack of burst, and nasal plosion, as well as fricativization or apparent deletion of /p b/ preceding the coronal. All in all, weakening of the preceding sound was applied in around 36% of the cases in the bilabial context and 28% in the vowel context. This also happened in the control conditions (46% and 15%, respectively).

The weakening strategy depended both on deletion status and on the type of segment beginning the sequence, i.e. /m/, /b/ or /p/, with nasals being the least likely to weaken at all and voiceless segments being the most prone to lack of burst accompanying coronal deletion (see Figure 1).

Importantly, while weakening probability differs depending on deletion status, it seems that there is a contrast in cluster simplification depending on target sound. When /t/ is deleted, the proportion of weakened preceding sounds is higher (around 60–70%). When /d/ is deleted, the proportion of weakened preceding sounds is lower (around 20%), which shows a contrary tendency (see Figure 2).

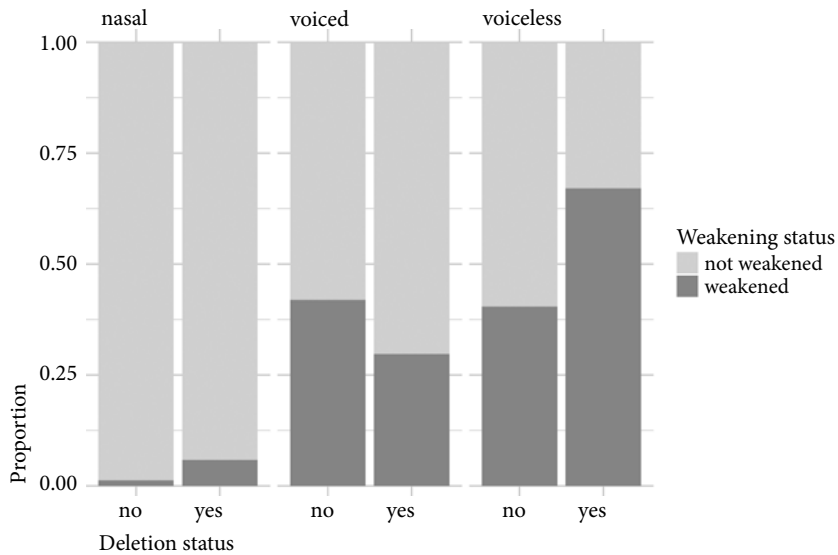


Fig. 2. Probability of weakening depending on deletion status and target sound. Control cases are omitted.

Given the overlap of /t d/ deletion and preceding sound weakening in the data, the different types of sequence productions can be descriptively labeled in a hierarchical order as: a fully produced cluster with plosion on both stops (27.6% of the data), a lenited cluster with preceding bilabial weakening (11.7%), a sequence with coronal deletion and no other weakening (6.7%), a doubly weak

cluster with both coronal deletion and preceding bilabial weakening (5.1%), a fully produced control condition (33%), and a lenited control with bilabial weakening (15.9%).

4.1. Duration

The first parameter taken into account was the duration of the key sound/cluster calculated for either the first two sounds (bilabial+coronal) in past tense targets or just the first sound (bilabial) in present tense controls. Such calculations were necessary to compare targets with controls. The mean duration of such key sounds/clusters in the data was 111 ms (SD=46 ms). This differed slightly depending on the right-hand context and showed a bigger difference depending on target/control status. In the bilabial context, the mean duration was 129 ms (99 ms in bilabial controls), and in the vowel context, the mean duration was 132 ms (83 ms in vowel controls). The duration did not differ significantly by type of sequence, except for /pb/ sequences which were particularly long (around the duration of /ptb/, i.e. circa 139 ms). This is most likely due to the fact that key sound/cluster duration was calculated for the sequence without the right-hand sound – just /p/ in the case of /pb/. However, delimiting the /p/ in /pb/ was not possible in most cases given the lack of /p/ burst and hence key sound/cluster duration often corresponds to the duration of the whole sequence here. In this case, sequence duration containing the whole sequence of two or three sounds, depending on target/control status is perhaps a more reliable parameter. The general mean of sequence duration is 194 ms (SD=106 ms), with marked differences between target sequences and controls (217–232 ms vs 148–186 ms).

If we take deletion into account, the data shows that sequences with coronal deletion are markedly shorter than those without it (159 ms, *SD*=50ms vs 238 ms, *SD*=70 ms), and close to the controls (165 ms, *SD*=59ms). Similarly, sequences produced with weakened preceding bilabials are shorter than those with fully produced preceding stops (175 ms vs 201 ms on average). All in all, the data shows a gradient hierarchy of sequences depending on the discussed strategies of cluster production, with an interaction between coronal deletion and preceding sound weakening, as illustrated in Figure 3.

Similar effects were observed for key sound/cluster durations. However, preliminary data analysis revealed some discrepancies between values obtained for the whole sequence versus the duration of the key sound/cluster in deletion vs retention contexts. Since both overall sequence duration and sound/cluster duration may depend on the particular constitution of the sound triplets or doubles, it was further explored whether sequences beginning with a nasal differed from those beginning in a voiced sound (/b/) and/or those beginning in a voiceless sound (/p/). This was then tested in the mixed-effects linear models built for the data.

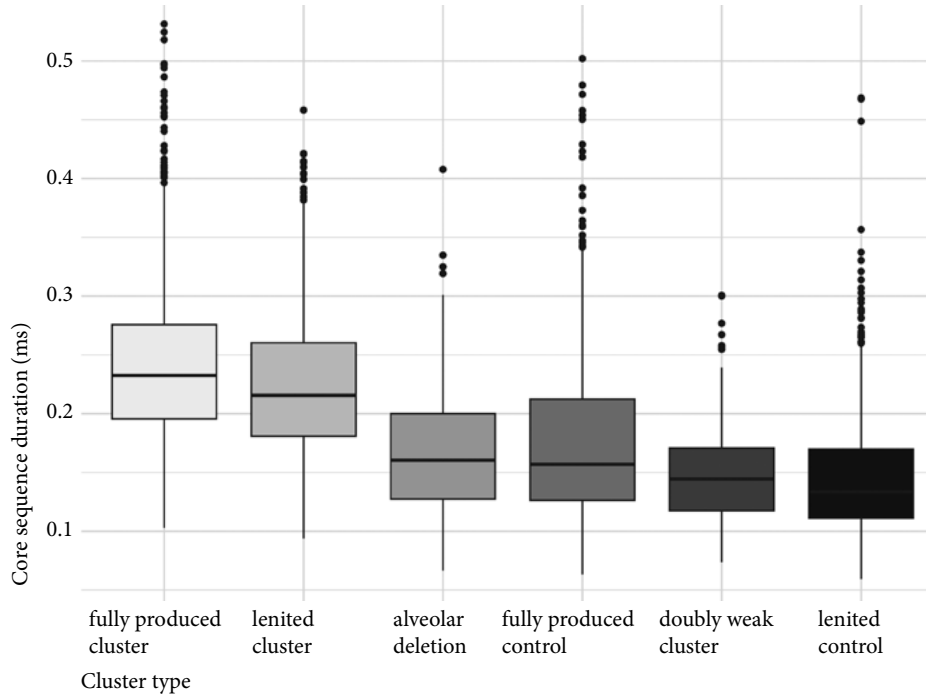


Fig. 3. Sequence duration in milliseconds depending on cluster type

The first model had sequence duration as the dependent variable, deletion status and preceding sound weakening, as well as sequence beginning as fixed factors, two interactions and a maximal random structure including speaker and item variables that allowed for model convergence. Adding the third interaction (weakening * sequence beginning) did not improve model fit.

*sequence duration ~ weakening * deletion + deletion * sequence beginning + (1+weakening|speaker) + (1+deletion+weakening|item)*

The results of the model show a significant effect of weakening and deletion, and a significant interaction between deletion and sequence beginning (see Table 3). The effect size of the model, measured using the r2beta function was R^2=.729 (upper.CL=.829, lower.CL=.638), with deletion as a major contributor, followed by weakening.

Table 3. Table of coefficients from the sequence duration model

	Estimate	**Std. Error**	**Df**	***t* value**	***p* value**
(Intercept)	0.173	0.014	35.592	12.721	<.001
Weakening: weakened	-0.026	0.005	29.392	-5.221	<.001

	Estimate	**Std. Error**	**Df**	***t* value**	***p* value**
Deletion: yes	-0.001	0.01	26.435	-0.144	.887
Deletion: no	0.077	0.012	31.584	6.429	<.001
Sequence beginning: voiced	-0.001	0.013	32.303	-0.06	.952
Sequence beginning: voiceless	0.004	0.012	30.682	0.355	.725
Weakening: weakened*deletion: yes	0.012	0.007	1653.808	1.663	.097
Weakening: weakened*deletion: no	0.019	0.005	2109.417	4.055	<.001
Deletion: yes*sequence beginning: voiced	0.004	0.014	25.103	0.319	.753
Deletion: no*sequence beginning: voiced	-0.036	0.015	26.243	-2.396	<.05
Deletion: yes*sequence beginning: voiceless	-0.004	0.013	32.327	-0.323	.749
Deletion: no*sequence beginning: voiceless	-0.005	0.014	28.143	-0.386	.702

The interaction plot (Figure 4) shows that sequences with deletion pattern with controls, while sequences with coronal stop retention are significantly longer. This was confirmed by pairwise comparisons estimated based on the marginal means from the model. No other contrasts were reliable. As for the weakening variable, it points to shorter duration overall in weakened sequences but pairwise comparisons only show a significant difference in the control condition (t=5.14, p<.001).

The second model was built for key sound/cluster duration as a dependent variable. Here, data points in which the initial sound/cluster could not be delimited and hence the annotated interval was equal to the whole sequence were excluded from the analysis, which gave a smaller number of observations (3262). The maximal model that converged was the following.

*sound/cluster duration ~ weakening * deletion + deletion * sequence beginning + (1+deletion+sequence beginning|speaker) + (1+deletion|item)*

In this model, there was a significant difference between clusters with both retained and deleted coronals on the one hand, and controls on the other. Similarly, weakening was statistically significant, as well as sequence beginning, along with one of the interactions: deletion and sequence beginning. The effect size of the model was R^2=.831 (upper.CL=.890, lower.CL=.771), with deletion and sequence beginning as major contributors.

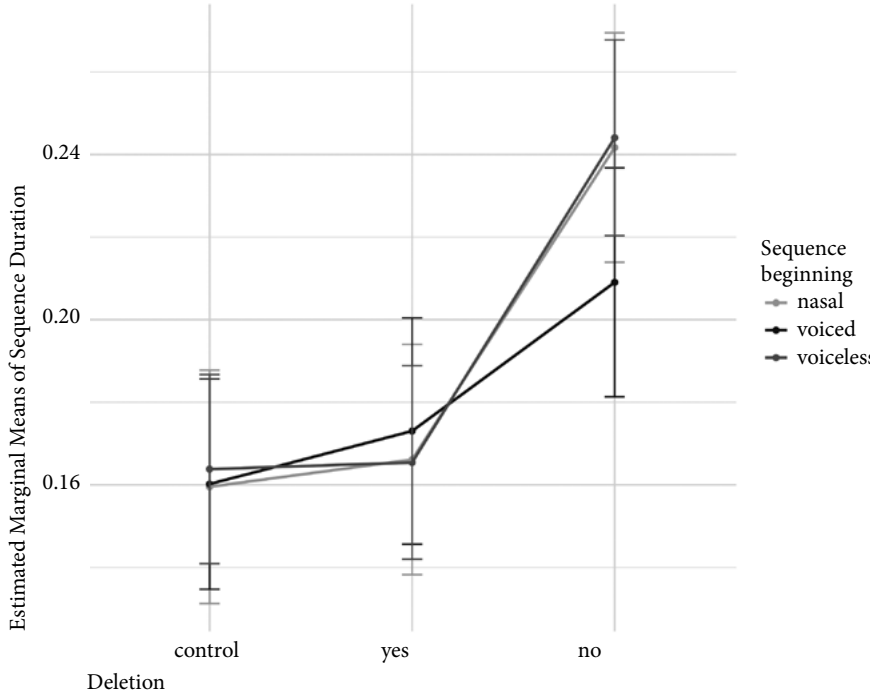


Fig. 4. Interaction between deletion and sequence beginning. The plot shows marginal means for sequence duration estimated from the model.

Table 4. Table of coefficients from the cluster duration model

	Estimate	Std. Error	Df	*t* value	*p* value
(Intercept)	0.084	0.005	25.952	16.652	<.001
Weakening: weakened	-0.009	0.002	1956.676	-4.47	<.001
Deletion: yes	0.013	0.006	20.885	2.142	<.05
Deletion: no	0.065	0.005	28.331	12.999	<.001
Sequence beginning: voiced	-0.008	0.005	29.227	-1.532	.136
Sequence beginning: voiceless	0.012	0.006	31.517	2.063	<.05
Weakening: weakened*deletion: yes	0	0.004	1004.598	-0.113	.91
Weakening: weakened*deletion: no	0.003	0.003	1906.435	1.191	.234
Deletion: yes*sequence beginning: voiced	0.006	0.008	18.118	0.731	.474
Deletion: no*sequence beginning: voiced	-0.013	0.006	20.536	-2.383	<.05
Deletion: yes*sequence beginning: voiceless	-0.007	0.008	23.023	-0.941	.356
Deletion: no*sequence beginning: voiceless	-0.017	0.005	24.279	-3.422	<.01

As in the first model, marginal means were calculated using the *emmeans* package, with SE, *df,* and 95% CIs. Pairwise comparisons show that sequence beginning does modulate cluster durations, as suspected. While overall durations seem to be different between controls and sequences with coronal deletion, once the first sound in the sequence is taken into account, this difference disappears. The data show that the differences between sequence beginnings are too small to be significant within any of the levels of deletion, however, except for the difference in the duration of controls between those beginning with a voiceless and a voiced sound *(t*=-4.665*, p*<.01). Figure 5 shows the relevant interaction.

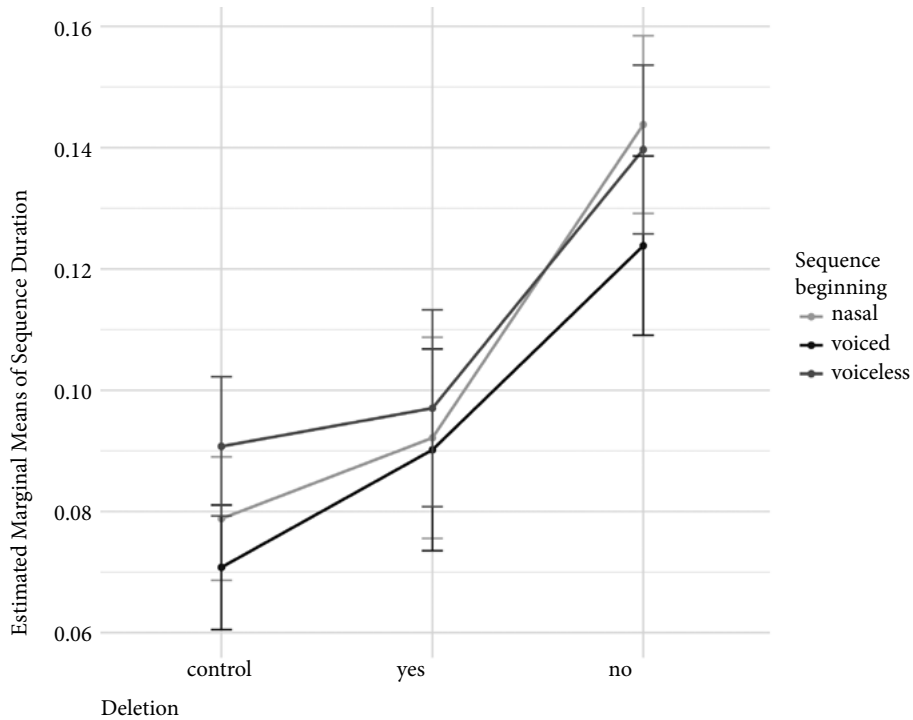


Fig. 5. Interaction between deletion and sequence beginning in the key sound/cluster duration model

4.2. Intensity difference

Intensity difference in the key sound/cluster was greater when no deletion took place, suggesting greater constriction. Additionally, these values were higher than in control sequences in which intensity was driven by the first segment, greater in the case of the nasal (i.e. smaller intensity difference). Deletion cases seem to be placed somewhere in between. This was true for sequences beginning with /m b/ but not in the case of /p/, in which there were no differences between any of the

conditions. Figure 6 illustrates this graphically. There were no markable differences depending on cluster weakening.

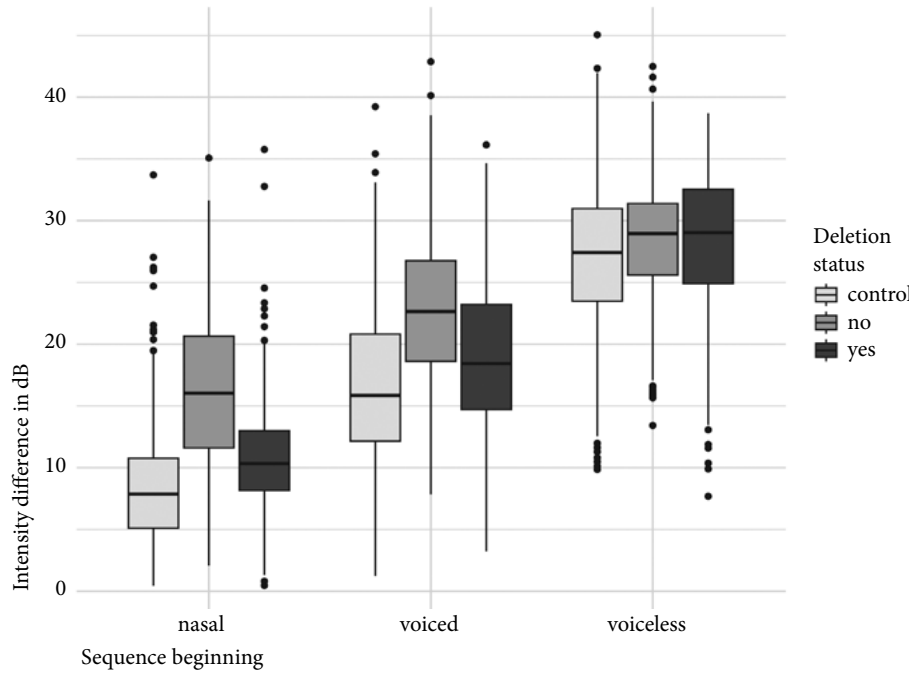


Fig. 6. Intensity difference calculated for the key sound/cluster with respect to the preceding vowel depending on deletion and type of sound at sequence beginning

A linear mixed effects model built for intensity difference as a dependent variable is as follows.

*intensity difference ~ weakening * deletion + deletion*sequence beginning + (1+deletion+weakening|speaker) + (1+deletion|item)*

The results showed a significant effect of deletion and sequence beginning, and a significant interaction between deletion and sequence beginning, which demonstrates that sequences with apparently deleted coronals have intermediate values between sequences with retained coronals and controls, albeit not in the case of sequences beginning with /p/ (see Table 5). In the latter case, there is no difference even between retention and deletion or control. Nonetheless, pairwise comparisons do not confirm an intermediate variant as the differences between deletion and controls were not significant. Additionally, the difference between deletion and retention was not significant in sequences beginning with /b/ (t=0.24, p=.25). Figure 6 shows the interaction plot.

Table 5. Table of coefficients from the intensity difference model

	Estimate	Std. Error	Df	*t* value	*p* value
(Intercept)	8.222	1.004	31.648	8.19	<.001
Weakening: weakened	0.13	0.529	19.688	0.245	.809
Deletion: no	7.89	0.939	29.166	8.4	<.001
Deletion: yes	3.092	0.971	27.193	3.185	<.01
Sequence beginning: voiced	8.533	1.258	22.93	6.784	<.001
Sequence beginning: voiceless	18.645	1.097	24.342	16.994	<.001
Weakening: weakened*deletion: no	0.046	0.463	2772.274	0.099	.921
Weakening: weakened*deletion: yes	-1.102	0.709	1446.631	-1.554	.12
Deletion: no*sequence beginning: voiced	-1.939	1.229	23.464	-1.578	.128
Deletion: yes*sequence beginning: voiced	-0.024	1.252	25.015	-0.019	.985
Deletion: no*sequence beginning: voiceless	-6.488	1.083	25.909	-5.992	<.001
Deletion: yes*sequence beginning: voiceless	-0.088	1.162	33.268	-0.076	.94

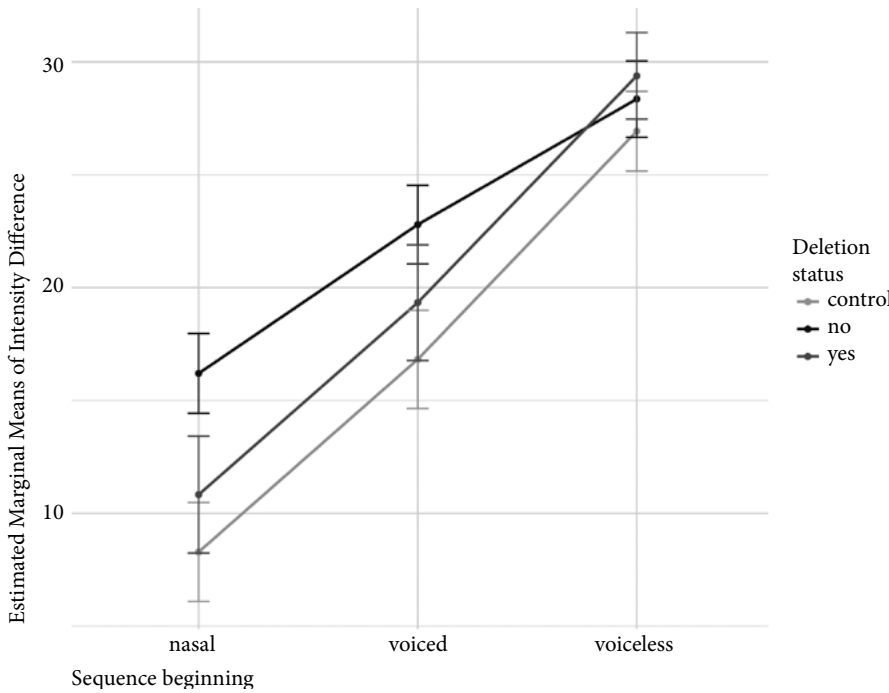


Fig. 7. Interaction between deletion status and sequence beginning type based on marginal means of intensity difference in the key cluster from the model

4.3. Harmonics-to-noise ratio

The harmonics-to-noise ratio calculated for the key sound/cluster did not differ depending on deletion status or weakening status. There were differences in the overall values of this parameter depending on which sound was the first in the sequence. This is the expected result as nasal sounds are more harmonic and voiceless stops are the least harmonic and therefore drive the values down. Table 6 presents a summary.

Table 6. Summary of HNR values per sequence beginning and deletion status

sequence beginning	deletion	mean	Sd	Min	Max	N
Nasal	control	15.506	4.837	2.477	31.076	504
Nasal	no	15.911	3.611	5.791	26.97	400
Nasal	yes	15.887	5.044	2.646	28.632	103
Voiced	control	8.165	4.737	0.542	28.174	496
Voiced	no	7.693	3.472	0.714	19.181	420
Voiced	yes	8.325	4.335	1.349	22.623	101
Voiceless	control	4.814	2.862	-0.221	16.371	762
Voiceless	no	4.862	2.441	0.012	13.792	599
Voiceless	yes	4.655	2.719	-0.026	15.093	222

The model built for HNR had the following structure.

*mean HNR ~ weakening * deletion + deletion*sequence beginning + weakening sequence beginning + (1+weakening|speaker) + (1+weakening|item)*

The results show a significant effect of all the fixed factors, and two interactions (see Table 7). The effect size was R^2=.739.

Table 7. Table of coefficients from the HNR model

	Estimate	Std. Error	Df	*t* value	*p* value
(Intercept)	15.616	0.667	19.8	23.396	<.001
Weakening: weakened	-3.033	0.988	160.675	-3.069	<.01
Deletion: no	0.178	0.206	3294.296	0.867	.386
Deletion: yes	0.969	0.338	3303.266	2.867	<.01
Sequence beginning: voiced	-8.094	0.479	29.325	-16.893	<.001

	Estimate	Std. Error	Df	*t* value	*p* value
Sequence beginning: voiceless	-11.5	0.432	35.208	-26.62	<.001
Weakening weakened*deletion: no	-1.288	0.281	2981.577	-4.577	<.001
Weakening: weakened*deletion: yes	-0.213	0.443	2940.277	-0.481	.63
Deletion: no*sequence beginning: voiced	-0.323	0.309	3309.861	-1.045	.296
Deletion: yes*sequence beginning: voiced	-0.418	0.493	3315.334	-0.849	.396
Deletion: no*sequence beginning: voiceless	0.521	0.309	3212.336	1.687	.092
Deletion: yes*sequence beginning: voiceless	-0.668	0.512	3212.159	-1.304	.192
Weakening: weakened*sequence beginning: voiced	4.724	0.928	127.891	5.09	<.001
Weakening: weakened*sequence beginning: voiceless	3.837	0.898	188.279	4.271	<.001

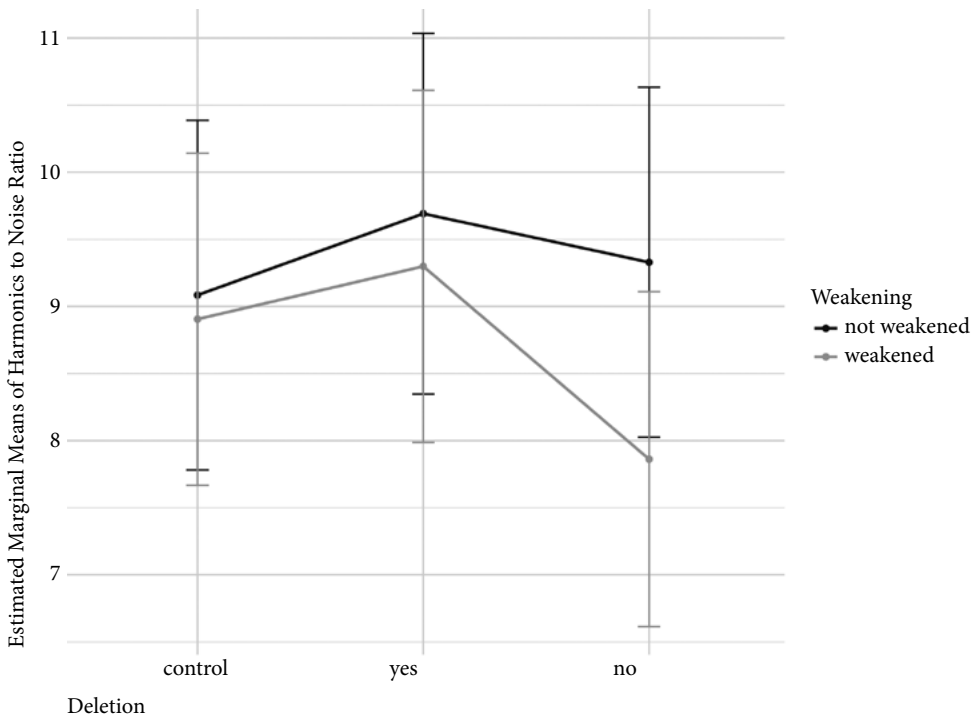


Fig. 8. Interaction between deletion status and weakening based on marginal means of HNR in the key cluster from the model

Pairwise comparisons based on the marginal means from the model show significant differences between controls and retention (t=4.48, p<.001), and between deletion and retention (t=3.96, p<.001) in the weakened condition but not in the unweakened cases. The relevant interaction plot is shown in Figure 8. Additionally, when looking at sequence beginning, the differences between controls and retention were reliable in the case of the nasal and /b/, and the difference between deletions and retentions was significant in /b/ only (t=3.42, p<.05). None of the contrasts were significant in the case of /p/ sequences. HNR differences within each level of deletion status and between the different sequence beginnings were confirmed to be reliable, as expected.

4.4. Lip tracking

Lip parameters calculated for the midpoint of each tested sequence show the same tendencies as the acoustics, with a greater lip aperture in non-deletion cases, in line with the expectation that the lips open or decompress for the tongue gesture (see Figure 9). In this case, differences between sequences beginning with nasals vs voiced vs voiceless sounds are negligible, although sequences beginning with /b/ do show slightly smaller values in the non-deletion condition.

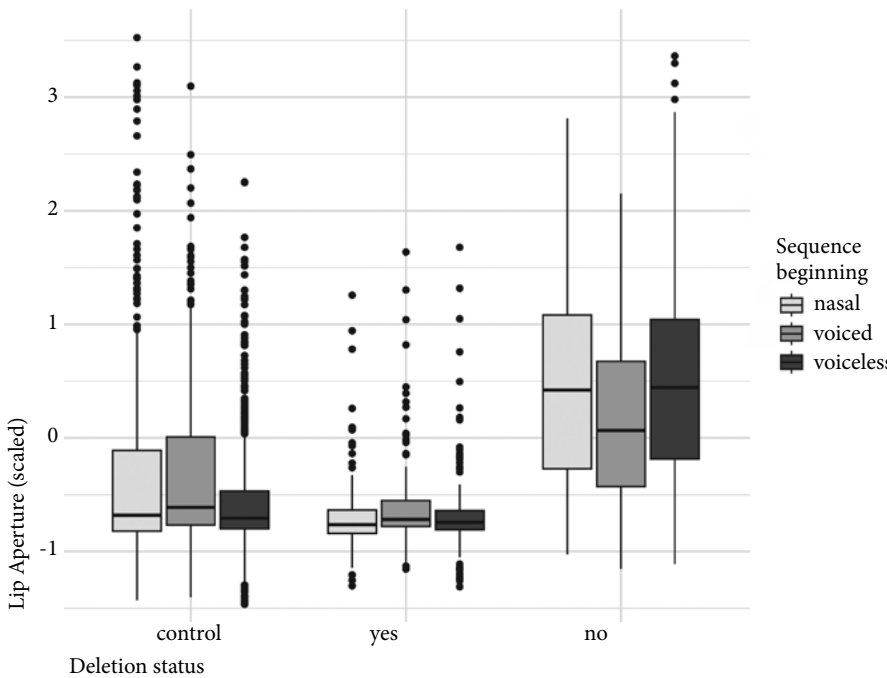


Fig. 9. Scaled lip aperture values at sequence midpoint by deletion status and sequence beginning. Vowel context and bilabial context values are taken together.

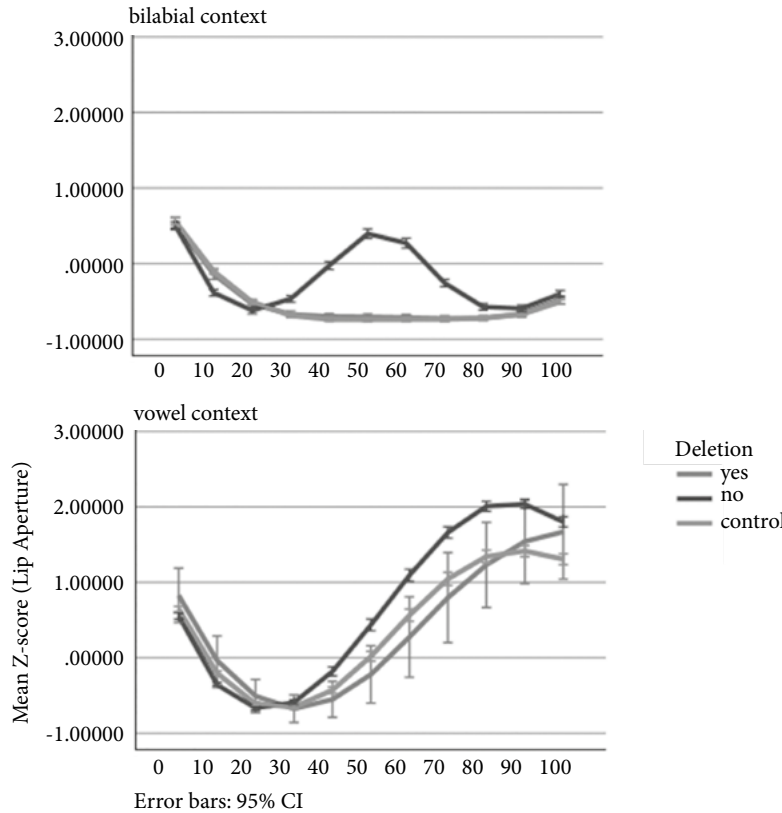


Fig. 10. Lip aperture trajectories depending on deletion status and context. The *x* axis shows percentage point of the trajectory over the sequence. Note that in the bilabial context, the trajectories showing *deletion: yes* and *controls* overlap almost completely and it is difficult to distinguish between the two lines.

As for weakening, it seems to affect only controls and not target sequences. The right-hand context plays a greater role instead, given that the lips either close (again) to produce a bilabial sound or open further for the following vowel, which significantly affects the overall trajectories. This is shown in Figure 10 where the whole trajectories are plotted, indicating that midpoint values represent maximum aperture in bilabial contexts, and inflection point in the vowel context. A marked change from the baseline lip closure in bilabial productions can be observed, and then either a comeback to closure at the end of the sequence or continued opening (in vowel contexts). Similar tendencies take place in lip thickness trajectories (not shown on the plot).

In the case of lip thickness, the situation is more nuanced. Controls pattern with sequences containing apparent deletions but only in the bilabial context. In vowel contexts the overall values of lip thickness are greater and controls show

slightly higher values than both target sequences (with and without deletions). This suggests greater aperture at midpoint, probably in relation to the start of lip opening for the following vowel which can be delayed in the case of clusters (see Figure 11).

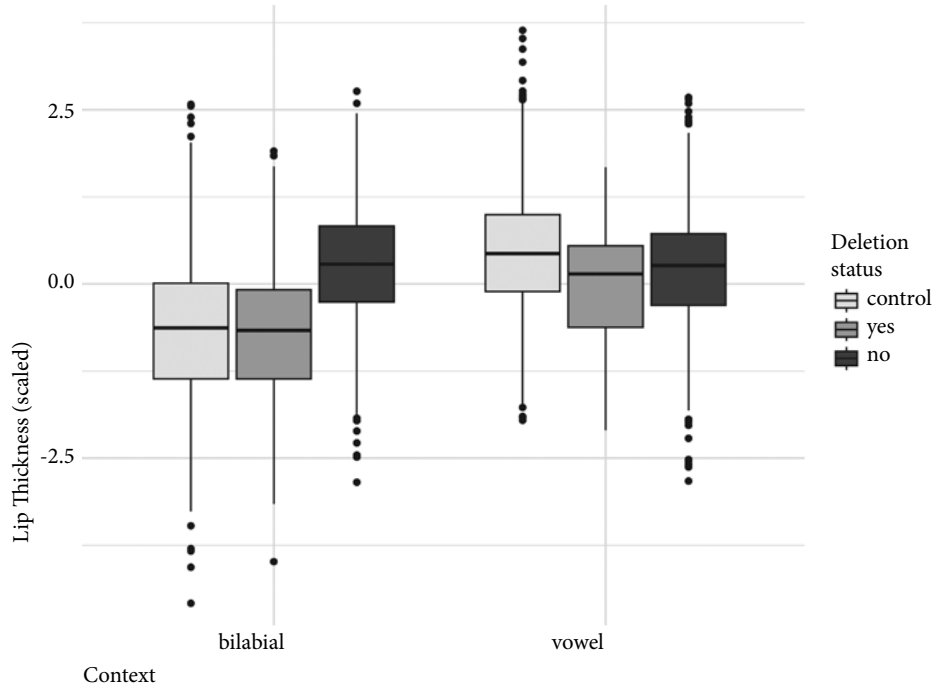


Fig. 11. Scaled values of lip thickness at sequence midpoint by context and deletion status

Table 8 shows a summary of lip aperture and lip thickness values depending on the context and deletion status in the data.

Table 8. Lip aperture and lip thickness – summary of scaled values

Lip aperture		**Mean**	**Std dev.**	**Min**	**Max**
Bilabial context	control	-0.761	0.285	-1.465	2.126
Bilabial context	deletion: yes	-0.699	0.312	-1.31	1.303
Bilabial context	deletion: no	0.327	0.681	-1.11	3.299
Vowel context	control	-0.006	0.816	-1.231	3.524
Vowel context	deletion: yes	-0.002	0.858	-1.057	1.678
Vowel context	deletion: no	0.441	0.915	-1.153	3.363

Lip thickness		Mean	Std dev.	Min	Max
Bilabial context	control	-0.653	1.045	-4.583	2.578
Bilabial context	deletion: yes	-0.742	1.092	-10.507	1.905
Bilabial context	deletion: no	0.272	0.824	-2.846	2.765
Vowel context	control	0.468	0.89	-1.956	3.64
Vowel context	deletion: yes	0.006	0.951	-2.098	1.675
Vowel context	deletion: no	0.171	0.916	-10.507	2.677

The two lip parameters were used as dependent variables in linear mixed effects models, with deletion status, weakening and context as fixed factors. The first model structure was:

*scaled lip aperture ~ weakening*deletion + deletion*context + (1+context+weakening|speaker) + (1+weakening|item)*

The results show a significant effect of deletion and context but not weakening. Both interactions were significant (see Table 9). The effect size was 0.8 (R^2=.809, upper.CL=.847, lower.CL=.771).

Table 9. Table of coefficients from the lip aperture model

	Estimate	Std. Error	Df	*t* value	*p* value
(Intercept)	-0.712	0.071	120.465	-10.036	<.001
Weakened	-0.073	0.048	52.44	-1.512	.137
Deletion: yes	-0.065	0.092	242.352	-0.703	.482
Deletion: no	1.059	0.086	196.051	12.334	<.001
Context: vowel	0.743	0.119	64.594	6.249	<.001
Weakened*deletion: yes	0.267	0.08	164.704	3.336	<.01
Weakened*deletion: no	-0.015	0.063	71.991	-0.241	.81
Deletion: yes*vowel context	0.078	0.171	425.685	0.457	.648
Deletion: no*vowel context	-0.649	0.128	143.851	-5.074	<.001

While the interaction between deletion status and weakening seems to be showing a possible three-way contrast in the case of weakened tokens (see Figure 12), Tukey-adjusted pairwise comparisons based on the marginal means from the model do not confirm this effect, only showing a significant difference in lip aperture between either controls or deletion cases and non-deletion cases in both weakened and non-weakened sequences.

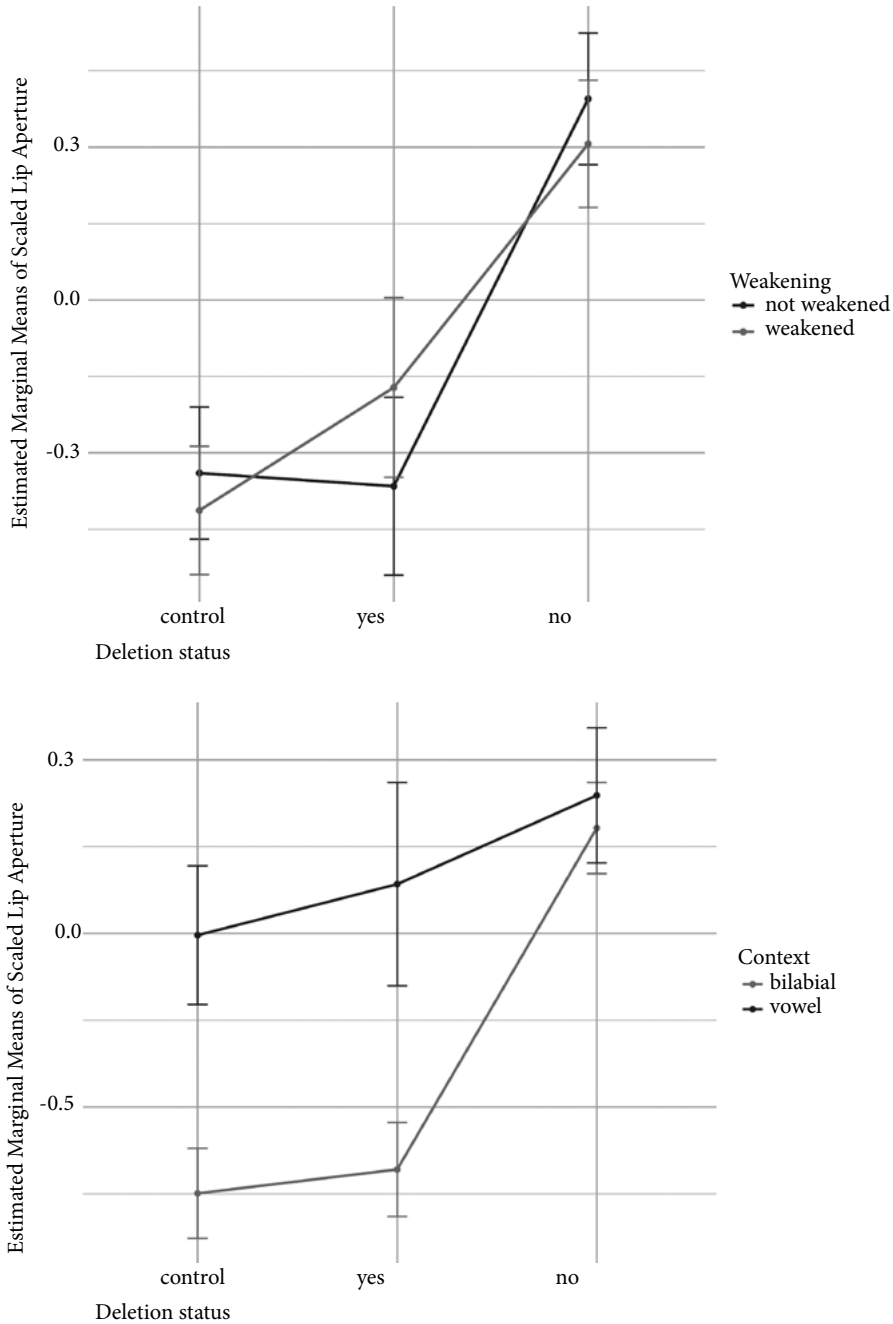


Fig. 12. Interaction plots from the lip aperture model

The interaction between context and deletion shows overall higher values in vowel contexts, with smaller differences between controls and deletions compared to non-deletions. Pairwise comparisons based on the marginal means indicate that controls are significantly different from non-deletions in both vowel and bilabial contexts, while apparent deletions are only significantly different from non-deletions in the bilabial context.

The second lip tracking model was built for lip thickness as follows:

*scaled lip thickness ~ weakening*deletion + deletion*context +*
(1+context+deletion|speaker) + (1+context+weakening|item)

The results show a significant effect of weakening, deletion and context, as well as significant interactions (see Table 10). The effect size of the model was R^2=.71 (upper.CL=.805, lower.CL=.623).

Table 10. Table of coefficients from the lip thickness model

	Estimate	Std. Error	Df	*t* value	*p* value
(Intercept)	-0.482	0.147	16.837	-3.29	<.01
Weakening: weakened	-0.277	0.069	107.027	-4.045	<.001
Deletion: yes	-0.091	0.108	36.565	-0.84	.406
Deletion: no	0.763	0.179	18.892	4.265	<.001
Context: vowel	1.007	0.159	18.287	6.332	<.001
Weakening: weakened*deletion: yes	0.283	0.118	227.432	2.394	<.05
Weakening: weakened*deletion: no	0.238	0.095	113.443	2.5	<.05
Deletion: yes*context: vowel	-0.438	0.197	535.302	-2.22	<.05
Deletion: no*context: vowel	-1.087	0.105	122.599	-10.314	<.001

The bilabial context data pattern with the results obtained for lip aperture, as shown in Figure 13. In the case of vowel contexts, apparent deletion cases show lower values of lip thickness, suggesting more compression compared to both controls and coronal retention. However, these differences resulted insignificant. By contrast, both controls and deletions are significantly different from retention but not from each other in the bilabial context (t=-4.92, p<.01 for the control-retention contrast, t=-4.86, p<.01 for the deletion-retention contrast, t=-0.48, p=.996 for the control-deletion contrast).

As for the weakening status: unweakened tokens with deletion show slightly more compression compared to both controls and cases of coronal retention.

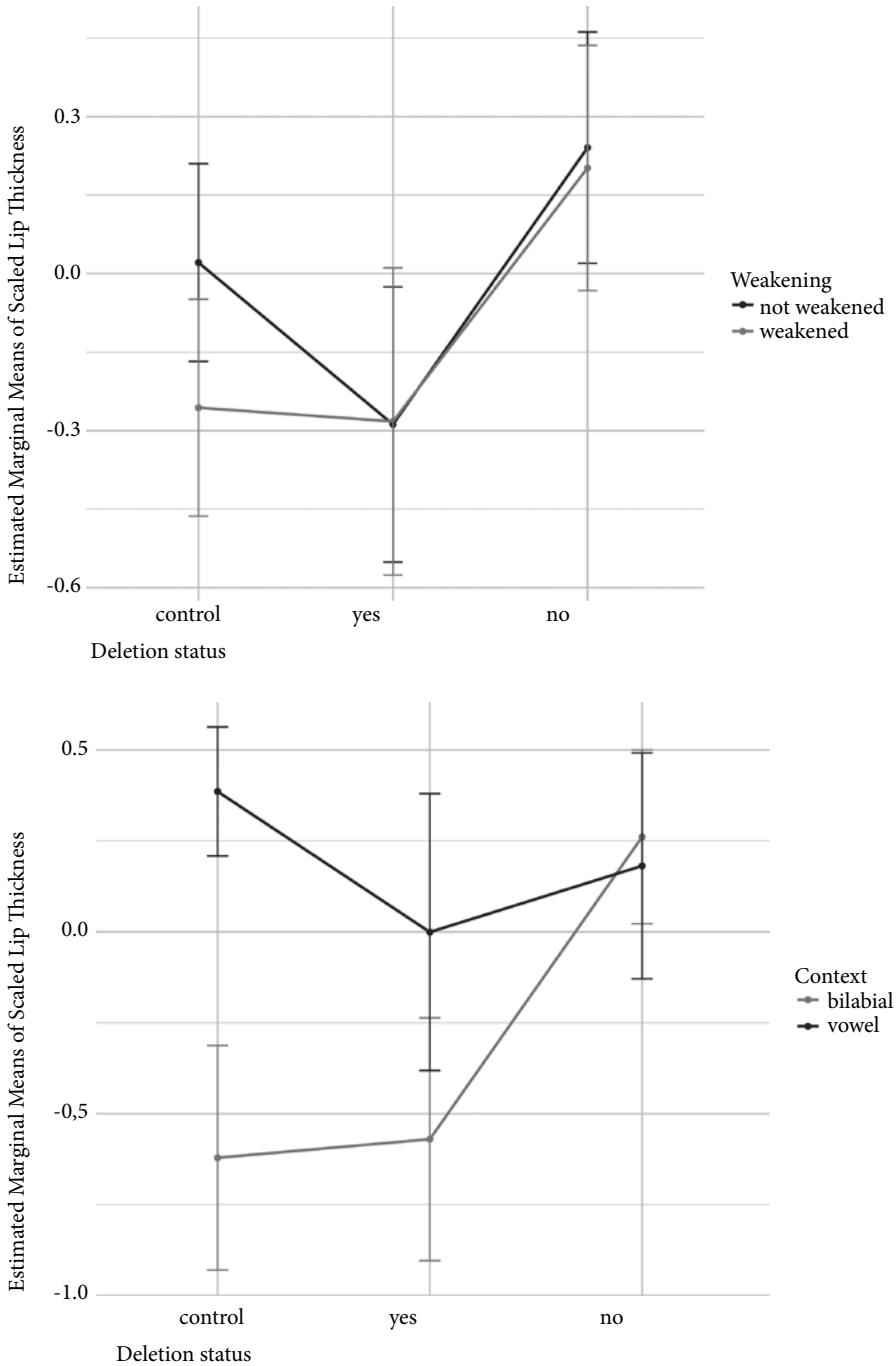


Fig. 13. Interaction plots from the lip thickness model

However, pairwise comparisons indicate that the difference is not significant between any of these conditions (t=-1.3, p=.77 between unweakened controls and retention, t=2.59, p=.111 between unweakened controls and deletion, and t=-2.81, p=.101 between unweakened retention and deletion). Similarly, none of the contrasts are significant in the weakened condition. Interestingly, there is a significant difference between weakened and unweakened controls in the direction opposite to the expectations (t=3.98, p<.01). This effect is probably driven by the overall higher values of lip thickness in vowel contexts.

5. Discussion and conclusions

The results of the study show reliable differences between apparent deletions and coronal retention, and no differences between apparent deletions and controls in several of the tested parameters. This result is consistent with the interpretation that deletion in the data, whenever applied, is complete. No reliable evidence was found for an intermediate variant between controls and retentions. In other words, whenever /t d/ are inaudible, they pattern with control sequences that did not have the coronal in the first place and therefore possibly lack a tongue gesture. This was found when measuring both sequence and key sound/cluster durations and lip aperture, and also to some extent intensity difference, HNR and lip thickness. Thus, acoustic and articulatory data are in line in predicting the recorded token types, indicating full deletion. Consequently, if we treat lip tracking data as proxies pointing to the presence or lack of tongue movements in the tested sequences, we should conclude that no relevant tongue movement was present in apparent deletion cases or, if there were tongue gestures, their magnitudes were significantly smaller than in the case of full /t d/ production. Before drawing this conclusion, however, we should take into account the other findings from the presented analysis.

First, preceding sound weakening was found as an important moderator of the observed differences between controls, deletions and retentions. While it causes a decrease in sound/cluster and sequence duration and HNR, its role in modulating lip aperture and compression is more limited. Interestingly, the data indicates that the probability of preceding sound weakening increases with /t/ but not /d/ deletion, suggesting different strategies that are probably more dependent on the whole cluster than just the coronal. This also shows that articulatory strategies taken by the speakers are not necessarily limited to cluster simplification (eliminating or weakening one sound or the other in a cluster, e.g. /t/ or /p/ in /ptb/) but seem to be additive. Weakening applies to the whole sequence planned and produced by the speaker.

Another important observation comes from the fact that there seem to be compensatory effects. While key sound/cluster duration is greater in /t/ sequences in apparent deletion cases, the whole sequence is of the same duration as /d/

sequences and controls, which suggests that speakers compensate for the shorter /d/ by lengthening the preceding bilabial. In this way, sequence beginning plays a role in modulating cluster durations and weakening strategies, as shown in all of the tested models. Nonetheless, neither of the sequence types observed as beginning and ending in different segments showed an overall different pattern than the rest. The general tendencies were the same although particular measurements were obviously affected. For instance, intensity difference in sequences beginning with a nasal is naturally smaller, given that /m/ is a sonorant, while HNR will be higher as /m/ is more harmonic compared to an oral stop. Perhaps the most important deviation from the general pattern can be observed in the lip data in which, naturally, sequences ending in a vowel exhibit greater lip opening and less compression from midpoint or even earlier until the end of the sequence. This is as expected.

It must be emphasized that the lip measurements used in this study are quite coarse as they only indicate the midpoint of the sequence, measured proportionally based on sequence durations. Since different combinations of stops and vowels were used in the tested sequences, lip thickness or aperture at midpoint may not necessarily indicate where the greatest opening took place (or not), i.e. where the putative /t d/ should have been produced. The different durations of the sounds constituting the sequences may have affected these measurements. A time series analysis over the whole sequence trajectories is planned in a separate paper. This, combined with the fact that there were slight differences between controls and apparent deletions at least in some measurements (e.g. intensity, HNR and lip thickness) but at the same time sequence durations consistently showed no significant effects between these conditions, as well as the fact that we do not have direct data concerning tongue movements, suggests that the results should be treated with caution. Given preceding sound weakening that often accompanies coronal deletion, it is possible that the observed deletions are not complete and some remnant of a tongue gesture is still there. It is also possible that, given the general weakening tendency over the whole sequence, there is substantial gestural overlap between the tongue and lips and what we see in the lips (as also indicated by duration and other acoustic measurements) masks the tongue gesture that is produced at the same time as lip closure or earlier than in retention cases due to weakening. Such an interpretation requires the analysis of the tongue data and would be in line with some of the previous studies on the subject suggesting incomplete deletions (e.g. Purse 2019).

All in all, the results of the study indicate that while inaudible /t d/ in the data seem to be elided, the observed effects are at least to some degree gradient. General sequence weakening takes place instead of just local elision, probably driven by gestural overlap and changes in articulatory magnitude. This is of consequence for the phonological interpretation of the data which necessarily requires a decision on the presence or absence of the target sound in /t d/ deletion contexts. More information

is needed on the exact temporal gesture alignment between the tongue and lips that might provide additional evidence for gestural undershoot or tongue movement absence in the course of /t d/ production. Whether deletion is complete or only apparent in this context requires a follow-up articulatory analysis which might inform us of the interplay between lip and tongue gestures in different timing conditions.

Acknowledgements

This research was conducted as a part of a Fulbright Senior Award. I would like to extend my thanks to my host and collaborator from California, Peter A. Krause, without whom collecting data would have been impossible. I would also like to thank my annotators, Sophie Johnson and Magdalena Sielużycka, and to the Reviewers for all their constructive remarks.

Notes

1 Other techniques allowing tongue movement measurements include electropalatography (EPG), which requires the use of a custom-made artificial palate, and ultrasound imaging. The latter is slightly less precise, given no stable reference point on the palate but is more user-friendly and much cheaper.

2 As rightly pointed out by a Reviewer, in line with the predictions of Articulatory Phonology and similar frameworks, lingual and lip gestures are executed on different tiers, so they are not necessarily interrelated. On the contrary, the prediction would be that they happen independently. However, given coarticulation and temporal overlap in connected speech, gestural masking may occur. Furthermore, it may be that the gestures are coordinated to the extent that changes in timing cause changes within them, e.g. a change in lip compression vs lip opening. For lack of evidence, we do not know whether undershoot of a coronal gesture is accompanied by changes in lip configuration. A motion capture study linked with tongue movement data extracted from ultrasound imaging might resolve this question. This goes beyond the present study, however, and I leave it for future work.

3 The data comes from a bigger project in which a total of 46 participants were tested. Half of them were heritage Spanish speakers, the other half – American English monolinguals, all raised in California. The project involved both motion capture and ultrasound imaging. Here, I only present the results concerning lip movements.

4 Participant codes include information on the speakers' language background. Monolingual speakers were coded 'M,' and heritage Spanish speakers raised in the US either since birth or since they were very young (0–3 years old) were coded 'B.' Group differences are not explored in this paper awaiting more data.

References

Baranowski, Maciej, and Danielle Turton. 2020. "TD-deletion in British English: New Evidence for the Long-lost Morphological Effect." *Language Variation and Change* 32.1: 1–23. DOI: 10.1017/S0954394520000034

Bárkányi, Zsuzsanna, and Zoltán Kiss. 2010. "A Phonetic Approach to the Phonology of v: A Case Study from Hungarian and Slovak." *Turbulent Sounds: An Interdisciplinary Guide*. Ed. Susanne Fuchs, Martine Toda, and Marzena Żygis. Berlin & New York: De Gruyter Mouton. 103–142.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2018. *lme4: Linear Mixed-Effects Models Using 'Eigen' and S4*. R package (version 1.1-17). https://cran.r-project.org/web/packages/lme4

Bermúdez-Otero, Ricardo. 2010. "Currently Available Data on English t/d-deletion Fail to Refute the Classical Modular Feedforward Architecture of Phonology." Paper presented at the 18th Manchester Phonology Meeting, Manchester, UK. http://www.bermudezotero.com/18mfm.pdf

Boersma, Paul, and David Weenink. 2025. *Praat: Doing Phonetics by Computer*. Computer program. http://www.praat.org/

Broś, Karolina, and Peter A. Krause. 2024. "Stop Lenition in Canary Islands Spanish – A Motion Capture Study." *Laboratory Phonology* 15.1. DOI: 10.16995/labphon.9934

Broś, Karolina, Marzena Żygis, Adam Sikorski, and Joanna Wołłejko. 2021. "Phonological Contrasts and Gradient Effects in Ongoing Lenition in the Spanish of Gran Canaria." *Phonology* 38.1: 1–40. DOI: 10.1017/S0952675721000038

Browman, Catherine P., and Louis Goldstein. 1990. "Tiers in Articulatory Phonology, with Some Implications for Casual Speech." *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Ed. John Kingston, and Mary Beckman. Cambridge: Cambridge University Press. 341–397.

Carrasco, Patricia, José Ignacio Hualde, and Miquel Simonet. 2012. "Dialectal Differences in Spanish Voiced Obstruent Allophony: Costa Rican versus Iberian Spanish." *Phonetica* 69: 149–179.

Coetzee, Andries. 2004. "What It Means to Be a Loser: Non-optimal Candidates in Optimality Theory." PhD diss, University of Massachusetts, Amherst.

Esteve-Gibert, Núria, and Pilar Prieto. 2013. "Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements." *Journal of Speech, Language, and Hearing Research* 56.3: 850–864. DOI: 10.1044/1092-4388(2012/12-0049)

Guy, Gregory. 1980. "Variation in the Group and the Individual: The Case of Final Stop Deletion." *Locating Language in Time and Space.* Ed. William Labov. New York: Academic Press. 1–36.

Guy, Gregory. 1991. "Explanation in Variable Phonology: An Exponential Model of Morphological Constraints." *Language Variation and Change* 1: 1–22.

Holbrook, Benjamin B., Alan H. Kawamoto, and Qian Liu. 2019. "Task Demands and Segment Priming Effects in the Naming Task." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45.5: 807–821. DOI: 10.1037/xlm0000631

Hualde, José Ignacio, and Marianna Nadeu. 2011. "Lenition and Phonemic Overlap in Rome Italian." *Phonetica* 68: 215–242.

Hualde, José Ignacio, Marianna Nadeu, and Miquel Simonet. 2010. "Lenition and Phonemic Contrast in Majorcan Catalan." *Romance Linguistics 2009: Selected Papers from the 39th Linguistic Symposium on Romance Languages.* Ed. Sonia Colina, Antxon Olarrea, and Ana Maria Carvalho. Amsterdam & Philadelphia: John Benjamins. 63–79.

Hualde, José Ignacio, Miquel Simonet, and Marianna Nadeu. 2011. "Consonant Lenition and Phonological Recategorization." *Laboratory Phonology* 2.2. DOI: 10.1515/labphon.2011.011

Krause, Peter A., Carly A. Kay, and Alan H. Kawamoto. 2020. "Automatic Motion Tracking of Lips Using Digital Video and OpenFace 2.0." *Laboratory Phonology* 11.1: 9. DOI: 10.5334/labphon.232

Labov, William, and Paul Cohen. 1967. "Systematic Relations of Standard and Non-standard Rules in the Grammars of Negro Speakers." *Project Literacy Reports* 8: 66–84.

Labov, William, Paul Cohen, Clarence Robins, and John Lewis. 1968. *A Study of the Non-standard English of Negro and Puerto Rican Speakers in New York City*. Volume 1. Philadelphia: US Regional Survey.

Lenth, Russell V. 2019. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package. https://cran.r-project.org/web/packages/emmeans/index.html

Parrell, Benjamin. 2010. "Articulation from Acoustics: Estimating Constriction Degree from the Acoustic Signal." *The Journal of the Acoustical Society of America* 128.4: 2289. DOI: 10.1121/1.3508033

Purse, Rachael. 2019. "The Articulatory Reality of Coronal Stop Deletion." *Proceedings of ICPhS*. Ed. Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren. Melbourne: Australasian Speech Science and Technology Association Inc. and International Phonetic Association. 1595–1599.

Purse, Rachael, and Alice Turk. 2016. "'/t d/ Deletion': Articulatory Gradience in Variable Phonology." Paper presented at LabPhon 2016.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Renwick, Margaret, Leila Baghai-Ravary, Rosalind Temple, and John Coleman. 2014. "Deletions in Big Data? The Phonetics of Word-final /t d/ in the Audio BNC." Paper presented at BAAP 2014.

Sangari, Sandra. 2002. “Visual Correlates to Focal Accent and Their Relation to the Fundamental Frequency Contour.” *The Journal of the Acoustical Society of America* 112.5: 2441. DOI: 10.1121/1.4780032

Sonderegger, Morgan, Andrea Beltrama, Theodora Chatzikonstantinou, Elizabeth Franklin, Brett Kirken, Jason Lee, Meghan Nelson, Katie Nicoletto, Tali Penslar, Hannah Provenza, Nicholas Rothfels, Max Bane, Peter Graff, and Jason Riggle. 2011. “Coronal Stop Deletion on Reality TV.” Paper presented at NWAV40, Washington, DC.

Tagliamonte, Sali, and Rosalind Temple. 2005. “New Perspectives on an ol’ Variable: (t, d) in British English.” *Language Variation and Change* 17: 281–302.

Tamminga, Meredith. 2016. “Persistence in Phonological and Morphological Variation.” *Language Variation and Change* 28: 335–356. DOI: 10.1017/S0954394516000119

Temple, Rosalind. 2009. “(t,d): The Variable Status of a Variable Rule.” *Papers in Phonetics and Computational Linguistics*. Ed. Oliver Parker Jones, and Elinor Payne. Oxford: Faculty of Linguistics, Philology, and Phonetics. 145–170.

Walker, James A. 2012. “Form, Function, and Frequency in Phonological Variation.” *Language Variation and Change* 24: 397–415.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer International Publishing. DOI: 10.1007/978-3-319-24277-4

Wolfram, Walt A. 1969. *A Sociolinguistic Description of Detroit Negro Speech*. Urban Language Series 5.

Żygis, Marzena, Susanne Fuchs, and Katja Stoltmann. 2017. “Orofacial Expressions in German Questions and Statements in Voiced and Whispered Speech.” *Journal of Multimodal Communication Studies* 4.1–2: 87–92.

Appendix

List of sentences used in the experiment

1. *She finally stopped making noise.*
2. *They clapped merrily at the actors.*
3. *She chopped more mushrooms for the soup.*
4. *I cropped my top a bit to look younger.*
5. *He often dipped marshmallows in chocolate.*
6. *Water dripped messily from the faucet.*
7. *They grouped members by age.*
8. *The baby napped before being taken to the park.*
9. *She whipped butter for the frosting.*
10. *They prepped bowls for the guests.*

11. I slipped because I was looking the wrong way.
12. She tripped before the finish line.
13. We shipped boxes of goods to the refugees.
14. I skipped breakfast to get to work early today.
15. I professionally dubbed movies in the past.
16. He scrubbed marks off the wall.
17. The Academy snubbed my favourite director.
18. She rubbed mud off her boots.
19. He sobbed miserably in the corner.
20. The thief robbed money from the cash register.
21. He grabbed mashed potatoes from my plate.
22. She summed prices at the register.
23. She crammed papers into her bag before rushing out.
24. The dentist numbed patients' gums before proceeding.
25. He hummed parts of the symphony.
26. She regularly scammed people out of cash.
27. The bonsai should be trimmed periodically.
28. The mist dimmed parts of the landscape.
29. I had to stop eating junk food.
30. We began to clap our hands after the last scene.
31. He made me chop a bunch of onions for the salad.
32. I like to nap after an afternoon meal.
33. I need to prep a pitch for the board meeting.
34. It's easy to trip in the woods.
35. Your heart will skip a beat when you see this.
36. I'm bilingual so I can dub in English and French.
37. She decided to scrub in for the surgery.
38. I didn't mean to snub our mentor.
39. Let me rub your back so you can relax a bit.
40. I'm just not built to sob over art or fiction.
41. It is not right to rob anyone of their precious time.
42. We grab an umbrella whenever they forecast rain.
43. I need to sum everything up before deciding.
44. I don't know how to cram all my ideas into a project.
45. A local anaesthetic is used to numb a specific body part.
46. He liked to hum as he walked the dog in the morning.
47. It's not easy to scam an entire organization.
48. I need to trim audio files for analysis.
49. She did not want her beauty to dim away so soon.
50. Stop making excuses and start working.
51. Everyone began to clap merrily.

52. Please chop more peppers and onions.
53. The photographer decided to crop most pictures.
54. I always dip my muffin in the coffee.
55. Blood started to drip more slowly from the wound.
56. We need to group members by age.
57. I plan to nap before going back to work.
58. Use a whisk to whip butter until fluffy.
59. We need to prep bags for the trip.
60. I managed to slip back into the classroom unnoticed.
61. I will definitely trip because the whole world is watching.
62. He needs to ship big items such as home appliances.
63. I think I'll skip Black Friday shopping this year.
64. As a side job I dub movies on weekends.
65. I had to scrub my hands with alcohol after work.
66. They're known to snub many talented artists.
67. It's good to rub muscles after exercise.
68. She got upset and started to sob moments later.
69. Modern times rob many people of their freedom.
70. Remember to grab my protein drink for the gym.
71. I had to sum prices for the order.
72. I had to cram pages of notes into a small backpack.
73. Some people use alcohol to numb painful emotions.
74. You have to hum pretty loud for the baby to fall asleep.
75. The company tried to scam people by selling bogus products.
76. I have to trim pictures for the album.
77. We encourage staff to dim patient room lights by 9 pm.
78. She stopped eating to raise a toast.
79. Something urgent has cropped up.
80. She dipped an apple in caramel.
81. The juice dripped off the edge of the table.
82. She whipped up an elaborate meal.
83. I prepped everything ahead of time.
84. They shipped everything to the wrong address.
85. I was dubbed a hopeless dreamer by my peers.
86. She scrubbed an old pan clean.
87. He got snubbed at the Oscars.
88. The doctor rubbed an ointment on the wound.
89. He sobbed uncontrollably at the sad news.
90. They robbed an old man of his wallet.
91. He grabbed an orange from the fruit bowl and took a bite.
92. He summed up an equation without effort.

93. They crammed into the elevator before it closed.
94. I was left numbed and shocked by the news.
95. He sat on the floor and hummed an old tune.
96. He scammed an elderly woman over the phone.
97. The gardener trimmed an overgrown hedge.
98. This definitely dimmed our hopes of an early settlement.

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Balogné Bérces & Sajtós, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.08

Katalin Balogné Bérces
https://orcid.org/0000-0002-7251-3839
Pázmány Péter Catholic University, Budapest & Catholic University in Ružomberok

Erika Sajtós
https://orcid.org/0009-0001-5660-6306
Pázmány Péter Catholic University, Budapest

Correlations and Functional Explanations in Laryngeal Phonetics and Phonology: Evidence from New Zealand English

Abstract: Varieties of English differ in (i) the amount of aspiration/voicing in plosives, traditionally measured in Voice Onset Time (VOT), and (ii) the degree of vowel clipping before fortis consonants (the Voicing Effect). The two seem to correlate with each other, with a possible functional explanation, but they also interact with the phonological status of vowel length within the dialect. To eliminate this latter effect, we investigate two varieties, Pākehā and Māori New Zealand English, which share the vowel system. We measure vowel duration and plosive VOT in Praat, conduct statistical analyses using the *lme4* package of *R* and show that correlations between the two features exist independently.

Keywords: laryngeal phonology, English dialects, aspiration, VOT, Voicing Effect, Pre-fortis Clipping, New Zealand English

1. Introduction

We seek to establish links between phonological language typology, systemic-functional motivations, and the fine-tuning of phonetic detail in acoustic realisations, taking phenomena from the laryngeal phonology of English as our examples.

The model of laryngeal typology currently widespread in generative frameworks is laryngeal realism (or "the narrow interpretation of [voice]," see Honeybone 2005; Iverson and Salmons 2008), distinguishing two VOT-based binary systems: voice and aspiration languages. In voice languages like Slavic or Romance, the contrast is based on vocal cord activity, i.e. voicing in the strict sense: lenis obstruents are (pre)voiced (with VOT lead), while their fortis peers are voiceless unaspirated (tenuis). Most English varieties, together with most other Germanic languages, however, are the aspiration type (also called spread glottis languages): the obstruents traditionally analysed as voiced vs voiceless actually contrast in negative or zero vs positive VOT (VOT lag), i.e. lenes are tenuis and fortes are aspirated. Nevertheless, several dialects of English constitute voice systems (especially traditional dialects in Northern Britain and certain contact varieties – see Section 3).

Tanner et al. (2020) show that English dialects also differ in the Voicing Effect (VE) (the degree of vowel clipping before fortes), and propose a scale in which, roughly speaking, US varieties exhibit a VE size larger than dialects in England and even larger than dialects in Scotland. Additionally, we note a relationship between the VE and prevocalic VOT: generally, aspiration is heavy and lenes are voiceless in the US, and aspiration is weaker in England; North/North-East England and parts of Scotland feature voice languages. Apparently, the heavier the aspiration, the larger the VE – more strongly aspirated fortes imply less voiced lenes; consequently, when aspiration lenites in codas, the fortis–lenis distinction becomes less salient, hence cued in preconsonantal vowel duration rather than VOT.

However, the functional load of vowel duration also depends on the status of vowel length, which, again, generates differences between the US, England and Scotland. This may also affect the VE scale. To test whether correlations between VOT and the VE exist independently, we investigate Pākehā (PE) and Māori (ME) English of New Zealand, with anticipated laryngeal differences but a shared vowel system. We proceed as follows. First, Section 2 introduces the Voicing Effect and the dialectal variation that has been identified in its terms, while Sections 3 and 4 discuss systemic-functional interactions between VE size and VOT on the one hand, and between vowel length, vowel quality and postvocalic consonant on the other, in varieties of English. Then, in Section 5, we turn to the dialects under investigation and present our data and methods of analysis. Section 6 explains our results, and Section 7 concludes.

2. The Voicing Effect

The Voicing Effect (VE) is the result of the influence of following consonants on the phonetic length of the preceding vowel. If a long vowel – monophthong (in English, /æ/, too), diphthong or triphthong (in varieties where they exist) – is immediately

followed by a fortis consonant (e.g. *fate*) or a nasal and a fortis consonant (e.g. *faint*) within the word, then the vowel will be shortened or clipped, hence the alternative name of the process, Pre-fortis Clipping. This phenomenon was observed and described at least as early as in Chen (1970), and continues to generate studies (more recent ones include Coretta 2020; Morley and Smith 2023, etc.).

It has also been shown that VE size is considerably larger in English than in other languages. One common method of expressing it calculates a fraction (i.e. a value below 1): the clipping of pre-fortis vowels compared to their non-pre-fortis peers (duration of $V_{pre\text{-}fortis}$ divided by duration of $V_{non\text{-}pre\text{-}fortis}$). This value is around 0.51–0.60–0.70 for English (according to House 1961; Chen 1970; Cho 2016, etc.): vowels may become 30–40% shorter, or even half their elsewhere phonetic length, before fortis consonants. In contrast, the corresponding value in other languages ranges between 0.80 (Norwegian, Korean, Russian – Chen 1970) and 0.90 (Spanish, German and French – Chen 1970; Italian and Polish – Coretta 2020; Hungarian – based on Kovács 2002), that is, a shortening of 10–20% only. VE size is sometimes expressed in studies with values above 1, as ratios (i.e. the effect of voicing/lenisness: duration of $V_{non\text{-}pre\text{-}fortis}$ divided by duration of $V_{pre\text{-}fortis}$). Since previous literature works with either one or the other, we specify both in our results below.

Research on the VE in English has revealed a wide variety of its phonetic and phonological aspects. For example, it has been shown that it is triggered by phonological rather than physiological "voicing" (e.g. Fox and Terbeek 1977; Walsh and Parker 1981), and that the preceding vowel duration (V/C ratio) serves as a primary cue to the perception of the obstruent's fortisness (e.g. Jones 1950; Raphael 1972; Klatt 1976). More recently, however, research has turned to VE size in spontaneous speech (rather than isolated, read words) and non-pre-pausal positions on the one hand, and its non-uniformity across English varieties on the other.

The study we take as our point of departure is that of Tanner et al. (2020), which surveys a number of varieties from the British Isles and North America (see Figure 1), using corpus data for a sociophonetic investigation. It finds that dialects of English differ substantially from each other in the size of the VE, while individual speakers vary little relative to their particular dialect.

Figure 1 shows that it is possible to arrange these varieties into a scale in which, roughly speaking, the VE is smallest (virtually non-existent) in Scottish dialects and largest in the US (including African American English – AAE), with England situated somewhere between the two.

However, what studies do not investigate (properly/fully) in connection with the VE is phonological/functional explanations, and the Tanner et al. (2020) study is – for obvious reasons – unable to cover a number of varieties: most conspicuously, Southern-Hemisphere Englishes. These are the gaps we attempt to fill in our research.

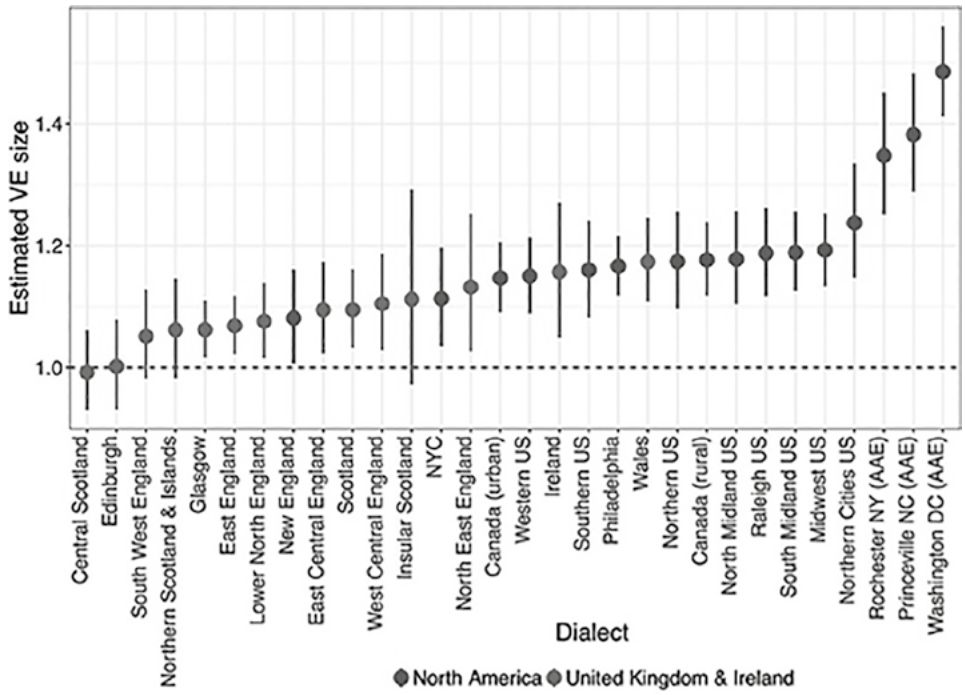


Fig. 1. VE size according to dialect in Tanner et al. (2020, 10)

3. VE size and VOT

We start with the discussion of possible functional explanations for the dialectal variation depicted in Figure 1. Why should VE size soar in the US and drop so drastically in Scotland? As Tanner et al. (2020) themselves note, the two extreme ends of the scale host varieties which feature phonological processes unrelated to but crucially impacting on the VE: the Scottish Vowel Length Rule in Scotland, and word-final devoicing/glottalling (with ensuing pre-lenis compensatory lengthening) in AAE. The former is part of traditional dialectological descriptions (Aitken 1981/2015; Wells 1982), while the latter has been the subject of more recent sociophonetic studies: Holt et al. (2016), for instance, explain that vowels in AAE lengthen before lenis consonants more than in White American English (WAE). As discussed in more detail in Section 4 below, vowel length and vowel duration interact with each other in intriguing ways, also shaping the vowel system as such, which has an effect on the workings of the VE.

We supplement these proposals with further observations. In particular, we note a systematic relationship between VE size (as in Tanner et al.'s results in Figure 1) and the average prevocalic VOT reported by prior research for fortis and lenis obstruents (see below), which, too, exhibits cross-dialectal differences in

English, and which, if displayed in the form of a diagram, would mirror the scale in Figure 1.

That is because, even though most varieties are aspiration languages (as can be recalled from Section 1), and even though quite some dialectal variation is attested in all localities, generally, aspiration is heavy in fortes and lenes are voiceless in American English (AmE), whereas generally, less heavy aspiration is typical of dialects in England. The average fortis VOT reported for AmE is 95.3 ms (Chodroff and Wilson 2017), while the corresponding value in Standard Southern British English is 56 ms (Docherty 1992). In addition, a number of voice languages with weak or no aspiration are on record in descriptions of the North/North-East of England (Durham, Yorkshire, Lancashire, the Black Country, etc.; see Wells 1982; Kerswill 1987; Clark 2004; Balogné Bérces 2022, etc.). As regards Scotland, there is a long tradition in both the dialectological and the phonetic-phonological literature of acknowledging voice languages in Scots-speaking areas (the Lowlands) (see Wells 1982; Scobbie 2006; Iverson and Salmons 2008; Sonderegger et al. 2020; Balogné Bérces 2022, etc.). These facts, when combined with the data in Figure 1, lead us to note that the heavier the overall aspiration in a system, the larger the average VE.

The explanation for this seems to be functional-systemic. In laryngeal systems, the phonetic forms of the two series of obstruents are dispersed in such a way that a convenient perceptual distance is kept between them, which appears to be constant even despite phonetic variation in the actual realisation in both aspiration and voice languages. Already in the two types of binary systems, we observe the distance maintained between aspirated and tenuis (in aspiration languages) and tenuis and pre-voiced (in voice languages): there are no languages where two sets of obstruents contrast e.g. heavily aspirated vs mildly aspirated forms, or short-lead vs long-lead forms. The distance itself varies greatly, and it is hard to define (even in ms) what this "convenient" minimum is, but what phonetic studies generally note is that even though there is substantial interspeaker variability in the overall values, each speaker makes a clear *contrast*, with the VOT higher for fortis than for lenis stops.

The dialectal variation we witness in English is restricted accordingly: if lenes are devoiced, VOT lag in fortes is pushed to enhanced degrees in order for the perceptual distance criterion to be satisfied; if lenes are pre-voiced, less aspiration in the fortes suffices. Or the reverse: if fortes decrease their VOT, the lenes have to follow suit, and assume quite some active voicing. Scobbie (2006) explicitly discusses how his investigation of Shetland English data evidences an inverse relationship between VOT lead in the lenis and VOT lag in the fortis category, whereby speakers avoid what he calls "perceptual confusion."

Therefore, we conclude that it cross-linguistically holds that a more strongly aspirated / more voiceless fortis system implies less voiced lenes, and a less aspirated/voiceless fortis system implies more voiced lenes. Interestingly, the perceptual distance between the two series tends to be optimised rather than ensured through

maximisation: there are relatively few laryngeal systems (varieties of English or other languages) "overshooting" it (see Balogné Bérces in press). Since these varieties are not within the scope of the present paper, we will ignore them in our discussion.

The issue that calls for an account next is how the correlation arises between the VOT system discussed in the previous paragraphs and the VE as shown in Figure 1. A crucial point to make is that the VOT values that are available for these phonological systems either exclusively or overwhelmingly come from measurements of word-initial/pre-vocalic (onset) plosives, while the VE affects preconsonantal vowels, i.e. it is most relevant in the relation of a vowel and a syllable-final (coda) consonant. The two phonological environments, however, significantly differ: while the onset (especially its word-initial occurrence) is a strong position in which contrasts can be fully expressed, the coda is weak in the sense that it is affected by weakening/lenition phenomena of all sorts. Specifically, the coda typically exhibits less or no aspiration in fortes (and the plosives are often unreleased, too). As a result, the fortis–lenis distinction becomes perceptually less salient: the voiceless aspirated fortis consonant has diminished aspiration and thus easily merges with its voiceless unaspirated (tenuis) lenis peer. Therefore, in that position, the fortis–lenis contrast is no more (primarily) cued in VOT but in preconsonantal vowel duration.

What follows from this, then, is the longer the onset VOT of the fortis series in a laryngeal system, the more voiceless the lenis set, and the more likely a potential coda merger that needs to be counterbalanced for the sake of perceptual discriminability. That is, the longer the onset fortis VOT, the larger the functional load of vowel duration, and therefore the larger the VE. The functional load of vowel duration, however, also depends on the status of vowel length in the language/dialect – an issue to which we turn in the following section.

4. The status of vowel length

As opposed to many other languages, which contrast short vowels and their long counterparts with little or no quality difference, the vowel system of English contains few unambiguous short–long pairs, and even pairs like KIT–FLEECE match vowels whose qualities are sharply distinct in most accents, exhibiting a lax–tense (ATR) distinction, with an additional diphthongal quality to the long/tense vowel. It is in fact generally claimed that vowel length in English is phonologically secondary to vowel quality (tenseness/laxness), especially in AmE (reflected in the absence of length indication in transcriptions – see e.g. Wells 1982, 120 and 140). This seems to apply in Southern AmE, especially southern AAE, more than elsewhere or in WAE (Holt et al. 2016). As a result, especially in AmE, modifications in phonetic duration due to the consonantal context (e.g. *heed* [hiː(j)d̥] vs *heat* [hi(j)ˀt]) will not disturb the phonological contrast (cf. *hit* [hɪˀt]).

However, the setup of the vowel inventory is able to influence the degree and status of the VE in a dialect. For example, certain varieties may have inherently shorter vowels than others, and considering the observation that short(er) vowels are less "compressible" further (Klatt 1976; also discussed in Tanner et al. 2020; Section 2), we may expect cross-dialectal differences accordingly (see the discussion in Tauberer and Evanini 2009). Tanner et al. (2020, 10) observe in their data that high vowels clip to a greater extent than non-high vowels. What this might entail is that the distribution of vowel height within a vowel system affects the amount of attestable VE. For example, a vowel shift raising several phonemes in a dialect might contribute to more overall phonetic shortening than found in a system with fewer raised or less raised vowels. Vowel mergers, for example that of the KIT and FLEECE sets, too, are expected to influence the size and vowel-specific effects of the VE.

A major Scottish innovation, the Scottish Vowel Length Rule (also called Aitken's Law), has been studied quite extensively from the perspective of its relation to the VE. It characterises the vowel systems of traditional dialects in (Lowland) Scotland primarily (hence the name), but it is also found in the far north of England. Although it varies somewhat in different dialects, according to this rule all vowels except [ɪ ʌ ɛ] are long before the voiced fricatives, before /r/ and word-finally, but short elsewhere (Aitken 1981/2015). Note how this causes a difference in e.g. the above example word *heed* between the "mainstream" VE system of most Englishes and the Scottish VLR. It is beyond doubt that this property of Scottish varieties fundamentally determines their placement in the scale in Figure 1 (also remarked by Tanner et al. 2020; Section 2). However, alternations mirroring the Scottish VLR have also been reported for the north of England (e.g. Watt and Ingham 2000), which might distort the picture about (parts of) England, away from the mainstream VE pattern and towards the Scottish one; moreover, as Llamas et al. (2011) show on the example of Tyneside English, the Scottish VLR can coexist with the VE, with some vowels being affected by one and some by the other, which complicates the matter even further.

In addition, the status of vowel length can be influenced by the presence of post-vocalic devoicing in a phonological system, too. Such cases are also attested in varieties of English, leading to further dialectal variation. Word-final /z/ devoicing, for instance, seems to be generally present in AmE, with Southern speakers applying it at higher rates (Hargrave et al. 2022). AAE and a number of regional varieties have developed final obstruent devoicing, typically as a substrate effect in regions that were historically populated by immigrant groups speaking languages characterised by that rule (German, Polish, Dutch, Yiddish, etc.). In AAE, the stops it affects may subsequently be accompanied by glottal reinforcement or (for /d/) glottal replacement, while the duration of the preceding vowel remains unchanged: e.g. [bæːt] or [bæːtʔ] or [bæːʔ] for *bad* (Wolfram 1994; Farrington 2018, etc.). In such a system, preconsonantal vowel duration will carry the burden of encoding the lenisness of the consonant, irrespective of the rest of the laryngeal

settings – an effect found to characterise AAE more than WAE (see research cited in Holt et al. 2016).

In fact, the literature seems to report two, apparently independent developments in Southern US English in general, and AAE in particular. One concerns phonetic vowel duration: as we have seen, an enhanced form of the VE has been observed in Southern WAE, and an even more increased pre-lenis lengthening in AAE, amplified by final devoicing and glottalling (also mentioned in Tanner et al. 2020; Section 2). The other characteristic feature of these varieties is the decreased contrastive function of (phonological) vowel length in tense–lax vowel pairs (Holt et al. 2016). Although the causal relation between these two (especially its direction) is unclear, it seems that vowel length, marking vowel contrasts in other varieties alongside vowel quality, is being replaced with vowel duration, whose function to cue following obstruent fortisness/lenisness is being promoted from secondary to primary (or even fundamental). This explains the excessive degree of the VE in African American varieties in Tanner et al.'s data and their location at the highest VE end of the scale in Figure 1.

It goes without saying, then, that the expected correlation, discussed in Section 3 above, between the onset VOT of plosives and the VE exhibited by preconsonantal vowels, is heavily affected by aspects of the vowel system itself, and that these factors may also contribute to the VE scale established by Tanner et al. (2020). Scottish varieties occupy one extreme end of the scale, not only because they are voice systems, but also because the Scottish Vowel Length Rule overwrites the VE. On the other hand, and at the other end of the scale, AAE exhibits what we previously dubbed an "excessive degree of VE, for the reasons explained above, even though it has been reported to have prevoicing in the lenis obstruent series (Herd 2020; for more discussion, see Balogné Bérces in press). When no such additional factors are present, presumably in (most of) the rest of the scale, the correlation we assume between the VE and average VOT values does show itself. We see it in the general difference between AmE and the English varieties mentioned above, but it is also apparent in the relatively high ranks in Figure 1 of the dialects of Ireland and Wales, known to exhibit relatively high rates of aspiration, even pre-aspiration (Wells 1982). To obtain further empirical corroboration of our proposal as well as to fill a number of research gaps, we investigated two dialects of New Zealand English (NZE), which lie beyond the scope of Tanner et al.'s study, but whose vowels, crucially, overlap: Pākehā English (PE) and Māori English (ME).

5. New Zealand English: data and analysis

As Hay et al. (2008) explain, being isolated by vast oceans from the rest of the world, New Zealand was one of the countries which were occupied the latest. About

1000 years ago, Polynesian people arrived, who called their new home Aotearoa and came to be known as the Māori. *Māori* means 'normal,' and the Māori people use this name to distinguish themselves from the European settlers called *Pākehā*, who started to arrive in large numbers only in 1840. By 1858, the non-Māori settlers outnumbered the Māori in New Zealand.

Since New Zealand is situated on a group of islands, the language of the British settlers, PE could develop without much interaction with Standard Southern British English (SSBE) as the prestigious model, and at the end of the 19th century, a distinctive and remarkably homogeneous variety evolved (Trudgill 2004). As for dialectal variation, regional differences are found only in the Southland and South Otago region, where originally, settlers were mainly of Scottish origin (Wells 1982; Hay et al. 2008). Nevertheless, three social accents can be distinguished in PE throughout New Zealand: broad, general and cultivated. This is rather a continuum, as differences lie in the extent to which typical NZE features are used. These are the most consistent and emphasised in the broad accent, with a lesser degree in the general accent, and the cultivated accent being the closest to SSBE with the NZE characteristics present. This distinction also exists between registers within the same social accent, with a higher rate of NZE features in informal speech, and a lower rate thereof in formal language.

With the arrival of the British settlers, the situation of the Māori changed dramatically. Until the 1970s, the government had a strong assimilation policy (Mulgan 1989), which brought the Maori language to the brink of extinction, and, despite several language preservation and education programmes that were applied to reverse this language shift, proficiency in Māori is still steadily declining. As a consequence, all Māori people speak English (i.e. ME), and it is the dominant language for almost all of them. Based on statistical data from 2006 (Te Puni Kōkiri 2008), only 14% of the Māori population speaks Māori well, and the majority of these speakers are 55 years old or older. Similarly to PE, ME has distinctive varieties/social dialects: standard and vernacular, the former corresponding to the cultivated, the latter to the broad accent of PE, both sharing many features with varieties of PE.

Regarding the laryngeal systems of these languages, the first point to note is a relatively low degree of aspiration in both Māori and ME. Although Māori, the substrate language for ME speakers, originally had a single series of unaspirated (tenuis) obstruents, they gradually became aspirated due to being exposed to English (Holmes and Ainsworth 1996; Bauer 1997; Bell 2000; Warren and Bauer 2004). In ME, too, the degree of aspiration has been found to be lower compared to PE (and other varieties of English), and the absence of aspiration has been documented in approximately 20% of voiceless stops. As can be seen in Figure 2 below (from Maclagan and King 2007), the frequency of aspiration has gradually increased for younger speakers in both Māori and ME, with a greater change in Māori.

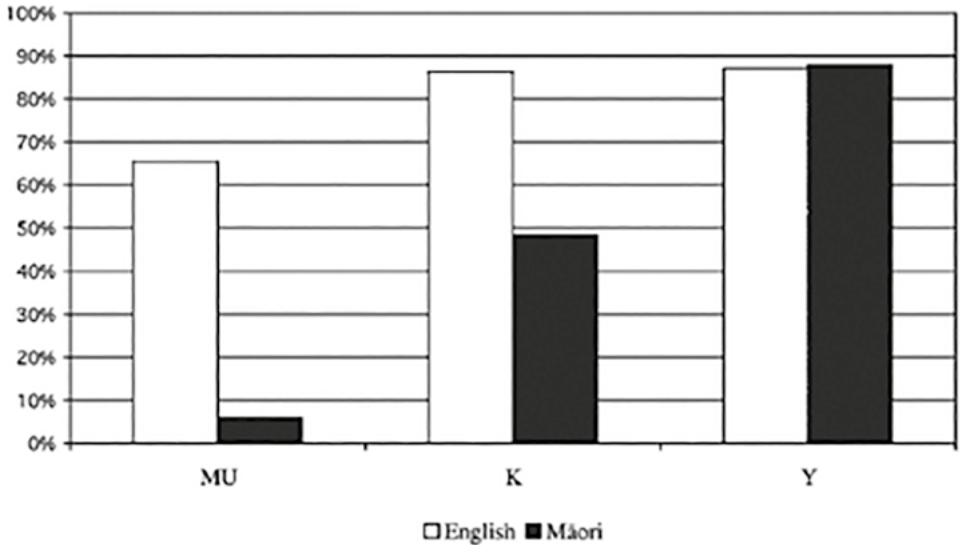


Fig. 2. Frequency of aspiration in three age groups of bilingual speakers, old (MU), middle-aged (K) and young (Y), when speaking English and Māori (Maclagan and King 2007, 88)

Maclagan and King observe similar changes in the degree of aspiration, too. Table 1 illustrates that in Māori, VOT duration more than doubled for the youngest speaker compared to the oldest speaker for all the three plosives, while in ME, a considerable increase can be observed with longer VOT duration for the youngest speaker.

Table 1. VOT duration in the three age groups in Maclagan and King (2007, 92)

	MU		K		Y	
	Māori	English	Māori	English	Māori	English
/p/	23	36	30	58	50	61
/t/	26	42	43	69	53	78
/k/	28	52	42	71	66	66
Mean	25	43	41	66	57	68
SD	10	16	18	15	22	21
n	135	98	246	98	114	101

In PE, on the other hand, fortes are aspirated (with different degrees depending on their position) and lenes have very little voicing (Bauer and Warren 2004), just like in "mainstream" varieties. In fact, traditional British prestige is gradually being

replaced with positive values attached by New Zealanders to American speech patterns (Bayard 2000), which predicts steadily high degrees of aspiration in PE.

Another phenomenon that is worthy of note here is the frequent affrication of fortis plosives, especially /t/. As analysed in detail in Hashimoto and Hume (2018), plosives exhibit a variety of phonetic variants in different phonological positions in NZE. Of these, the pre-tonic position (relevant to the aspiration data we are considering) is realised as a stopped plosive in the vast majority of cases and as a fricated plosive in a negligible proportion (for the phonetic descriptions of these variants, see Hashimoto and Hume 2018, 6–7). In addition, the stopped plosive realisation is often affricated – an observation made in a number of prior studies (e.g. Bauer and Warren 2004; Bauer et al. 2007; Hashimoto and Hume 2018) as well as our own data (see below). Hashimoto and Hume (2018) note that, contrary to previous results from AmE, /t/, rather than /k/, is the longest plosive, which they attribute to its affrication, and which we also report. Following Hashimoto and Hume (2018, endnote 7) (who in turn follow Fiasson 2015), we collapse aspirated plosives and affricated plosives into a single category.

As for VE size, some prior research (Maclagan and Hay 2007; Cho 2016) has been carried out on PE. Cho (2016), for instance, investigated the VE in minimal pairs in three varieties: PE, British English (BrE) and AmE. The greatest VE size was found in PE, with a smaller size in AmE and the smallest in BrE. However, an important factor needs to be considered when evaluating the results of previous studies: the effect of speech style. As Harris and Umeda (1974) state, most studies have used isolated words or carrier phrases when investigating VE size. In their empirical study, it is shown that there is a significant difference between the duration of vowels depending on the number of syllables in the word; that is, the more syllables the word has, the shorter the duration is. While this holds in isolated words and carrier phrases, the difference disappears in continuous speech, with approximately the same (shorter) vowel duration regardless of the number of syllables. In a more recent study, Tauberer and Evanini (2009) note that vowels with shorter duration have a smaller voiced–voiceless ratio due to the fact that they can be compressed less than vowels with longer duration (see the discussion of "compressibility" above). Furthermore, Luce and Charles-Luce (1985) highlight that speech style also affects closure duration. In other words, the closure duration of voiced and voiceless plosives shows a smaller difference and a greater variance in continuous speech than in isolated words. As a result, we expect generally lower values in data coming from spontaneous, continuous speech, in comparison to minimal pairs (Cho 2016) or word list data (Maclagan and Hay 2007).

Regarding VE size in ME, there is no previous data available, but given the direct correlation between VE size and VOT, the lower degree (or absence) of aspiration suggests smaller VE in ME.

Based on the above discussion, our research hypotheses are as follows:

1. Previous, smaller-scale studies report larger VE size in PE than in either BrE or AmE (0.56–0.60 – Maclagan and Hay 2007; Cho 2016). We hypothesise that PE is aligned with AmE, being closer to the maximal end of the scale in terms of both VOT and the VE, although with lower values than in prior research if investigated in larger samples and in a more spontaneous speech style.
2. Descriptions of ME attribute traditionally weaker/less frequent aspiration to it (as a substrate effect). Since we assume that VOT and the VE correlate, relatively low values are expected in ME for both.
3. VOT and VE correlate in both PE and ME.

To investigate whether our data support our hypotheses, the speech samples of fourteen speakers (seven Pākehā and seven Māori informants) are analysed with an approximately equal number of male and female speakers in both groups. The aim of the acoustic analysis is to investigate if there are differences in VE size and/or VOT duration between ME and PE, as well as to detect the correlation between them. Recordings of spontaneous speech were obtained from the following databases:

- For PE speakers, the audio recordings were selected from a collection available on the Tauranga City Libraries website (https://paekoroki.tauranga.govt.nz/). This digital library contains images, videos, written documents and audio recordings, and was created with the aim of sharing existing common knowledge with future generations.
- For Māori (and some PE) speakers, the digitalNZ website (https://digitalnz.org/records?tab=Audio&text=#/) was used, which consists of collections from libraries, museums, galleries and radio archives, among others.
- Our own corpus consisting of speech samples collected from contemporary New Zealanders.

Speakers were selected based on the following criteria: (1) they were born in New Zealand or moved there before the age of 10, the end of the critical period in language acquisition; (2) they were born and live outside the Southland and Otago regions as in these areas, a regional dialect is spoken which is different from the dialect spoken in the rest of the country; (3) they are educated speakers belonging to the same social class, that is, Māori speakers speak the standard, and Pākehā speakers the cultivated accent; (4) they have not spent more than 10 years/more than half of their lifetime abroad, and returned to New Zealand more than one year before the time of the recording; (5) a criterion exclusively for Māori speakers was that they should actively use the Māori language.

Because of the criterion for Māori speakers, the mean age of the informants is 55 years (standard deviation (SD): 15). As mentioned above, there is

a tendency that fewer and fewer Māori people speak the Māori language, and the majority of these speakers are over 55 years old – consequently, the mean age of our speakers is relatively high. Informants were given a code (Pākehā = "P"; Māori = "M") and were numbered consecutively, i.e. "P1," "P2," etc., so they cannot be identified by name. The speakers' background data is given in Table 2 below.

Table 2. Participants of the study

Speaker	Age	Gender	Education	Occupation
M1	77	male	tertiary	writer
M2	47	female	tertiary	university professor
M3	53	female	tertiary	musical artist
M4	71	male	tertiary	writer, politician
M5	82	male	secondary	politician
M6	35	female	tertiary	journalist
M7	50	female	tertiary	musical artist
P1	50	female	tertiary	broadcaster/writer
P2	59	male	tertiary	reporter/journalist
P3	39	female	tertiary	digital content manager
P4	51	male	tertiary	broadcaster
P5	42	male	tertiary	broadcaster
P6	41	female	tertiary	actress
P7	69	female	tertiary	writer, journalist

For each speaker, the acoustic analysis focuses on the realisation, more specifically on the duration, of the TRAP, DRESS, KIT and LOT monophthongs on the one hand, and the FLEECE, FACE and GOAT diphthongs on the other, before voiced and voiceless plosives and fricatives. The vowels were chosen in such a way that the analysis includes both diphthongs and monophthongs, and in both categories, they differ in vowel height and backness/frontness. For diphthongs, vowel height was decided according to their first element. The consonants included in the study are /p–b/, /t–d/, /k–g/, /f–v/, /s–z/, /ʃ–ʒ/ and /θ–ð/. In the case of fortis plosives, VOT was measured and compared between the two varieties. The vowels included in the analysis are given in Table 3 below, and as can be seen, the diphthongs and the monophthongs are evenly distributed.

Table 3. Vowels included in the analysis

	low	**mid**	**high**
short	LOT [ɒ]	KIT [ə]	DRESS [i]
long	FACE [æe]	TRAP [ɛ·] GOAT [ɔə]	FLEECE [ɪi]

Also, within the above categories, the number of tokens is balanced regarding both the short/long and the low/high distribution. The number of mid-vowel tokens is slightly higher, but it does not constitute a problem, as mid vowels do not contort the data in either the low or high direction. The number of tokens for each vowel is given in Table 4 below.

Table 4. The number of tokens for the different vowel qualities

	low		**mid**			**high**	
Māori	135		225			174	
	LOT	FACE	KIT	TRAP	GOAT	DRESS	FLEECE
	52	83	101	65	59	92	82
Pākehā	**low**		**mid**			**high**	
	133		227			178	
	LOT	FACE	KIT	TRAP	GOAT	DRESS	FLEECE
	69	64	111	79	37	107	71

The orthographic transcription of the recordings was made by using the *SpeakApp* (Version 2.0) software. Sentences containing the relevant tokens were cut from the audio recordings and analysed using Praat (Boersma and Weenink 2010). After annotating the extracted sentences, VOT duration was measured as well as the duration of the vowel preceding the consonants. The process of annotation was semi-automatic, as the text grid was created with the help of a Praat script, while the tokens were selected manually.

Vowel duration was measured from the first peak to the last peak in the periodic portion of the vowel, and the voiceless portion before a voiceless stop was included. In addition, the following events in the waveform and spectrogram were taken into consideration because together, they mark the end of a vowel: (1) the amplitude changes considerably; (2) a change in complexity and a decrease in the energy of the formants; (3) the beginning of aperiodicity. Figure 3 below illustrates the process of vowel duration measurements showing the acoustic events in both the waveform and the spectrogram.

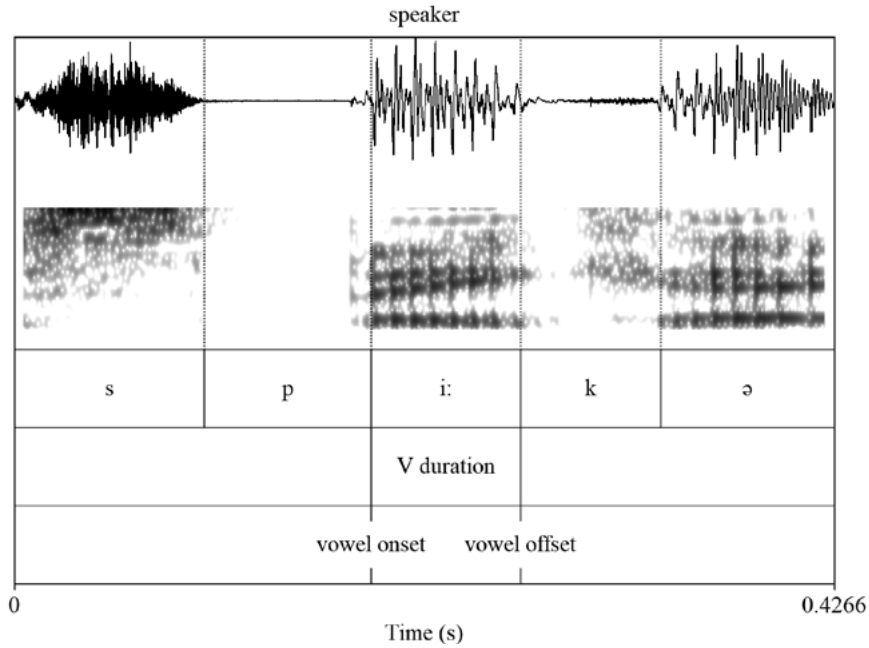


Fig. 3. Vowel onset, vowel duration and vowel offset of /i:/ in the word *speaker*

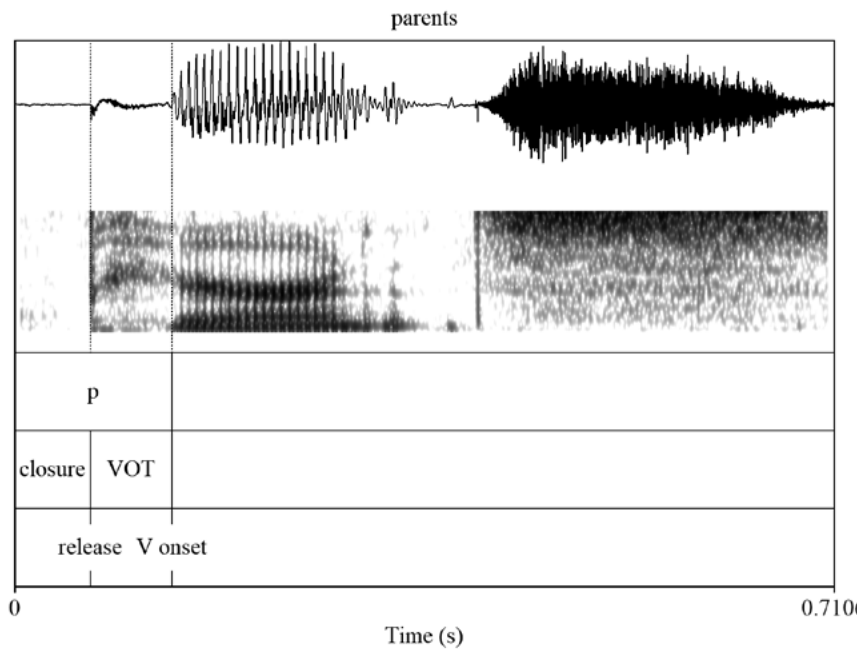


Fig. 4. Closure, release burst and VOT until vowel onset (first complete vibration) of /p/ in the word *parents*

Foulkes et al. (2010) explain VOT duration as the time period between the release burst and the onset of the following vowel, and note that it can be measured in different ways: (1) from the release burst to the onset of quasi-periodicity; (2) from the release burst to the first complete vibration; (3) from the release burst to the onset of striation in the spectrogram. In the present study, the second option was chosen, and VOT was measured for fortis plosives in word-initial position preceding a stressed vowel. In the case of multiple release bursts, VOT was measured from the final burst. Figure 4 illustrates the process of VOT measurements, showing the acoustic events in both the waveform and the spectrogram.

The number of tokens for VOT per speaker is shown in Table 5 below, and Table 6 shows the number of tokens according to the place of articulation.

Table 5. Number of tokens per speaker

Māori	**n**	**Pākehā**	**n**
M1	126	P1	112
M2	132	P2	122
M3	110	P3	118
M4	112	P4	112
M5	145	P5	148
M6	120	P6	199
M7	110	P7	135
Total	**855**	**Total**	**946**

Table 6. Number of tokens according to the place of articulation

	/p/	**/t/**	**/k/**		**/p/**	**/t/**	**/k/**
M1	12	27	20	**P1**	9	13	15
M2	10	18	19	**P2**	20	12	17
M3	13	10	15	**P3**	11	19	18
M4	9	12	17	**P4**	12	18	11
M5	18	17	18	**P5**	17	13	11
M6	7	21	13	**P6**	16	11	27
M7	11	17	14	**P7**	5	16	18
Total	**80**	**122**	**116**	**Total**	**90**	**102**	**117**

As for the place of articulation, the number of tokens of the alveolar and velar plosives shows a minimal difference, while there are slightly fewer tokens for the labial plosive. As this applies to both Māori and Pākehā speakers, it does not skew the data. The smaller number of labial plosives is due to the fact that the occurrence of words beginning with a stressed voiceless labial plosive is generally lower.

As the speech samples include spontaneous, running speech, it was not possible to control for the phonological environment and the frequency of the segments. Because of the unbalanced nature of the data, a statistical method is needed that allows the examination of the relevant factors and, at the same time, controls for the influence of other factors. Besides, the data consists of multiple measurements of VOT and vowel duration for the same participant, which means that there is a connection between these data points. Thus, there is dependency in the data, which, as Winter (2020) explains, requires the application of a mixed model as such models make appropriate estimates by incorporating non-independent clusters of data into the analysis. The primary tool to deal with non-independent data clusters is the random effect. Unlike fixed effects, random effects do not have a predictable, non-idiosyncratic influence on the data. In the present study, the *R* package, *lme4* (R Core Team 2015; Bates et al. 2015), is used with voice, ethnicity, manner of articulation and vowel height as fixed effects, and speaker as a random effect because speakers were selected randomly from the population.

6. Results and discussion

6.1. The Voicing Effect

Table 7 below shows the duration distribution of vowels in ms (with SD in brackets) before fortis and lenis consonants for Pākehā and Māori speakers, as well as the sizes of the VE. Recall from Section 2 that the VE can be expressed in two ways, as a fraction or a ratio, and that previous literature works with either one or the other; therefore, we have decided to specify both in our results.

Table 7. Vowel duration in ms (standard deviations in brackets)

	n	All vowels	Pre-fortis	Pre-lenis	Fraction/ratio
PE	538	90 (47)	80 (37)	111 (59)	0.72/1.38
ME	534	103 (44)	97 (40)	115 (51)	0.84/1.18

As is clearly shown in Table 7, vowel duration is greater for all speakers before lenis consonants than before fortis consonants. Furthermore, there is more variation in vowel duration before voiced codas than before voiceless ones, which is not

surprising given that there is a temporal limit in shortening, whereas lengthening imposes less limit. Additionally, overall vowel duration tends to be shorter in PE than in ME. Figure 5 below illustrates the differences in vowel duration before lenis vs fortis consonants in ME and PE. The greater difference in PE indicates a greater VE size.

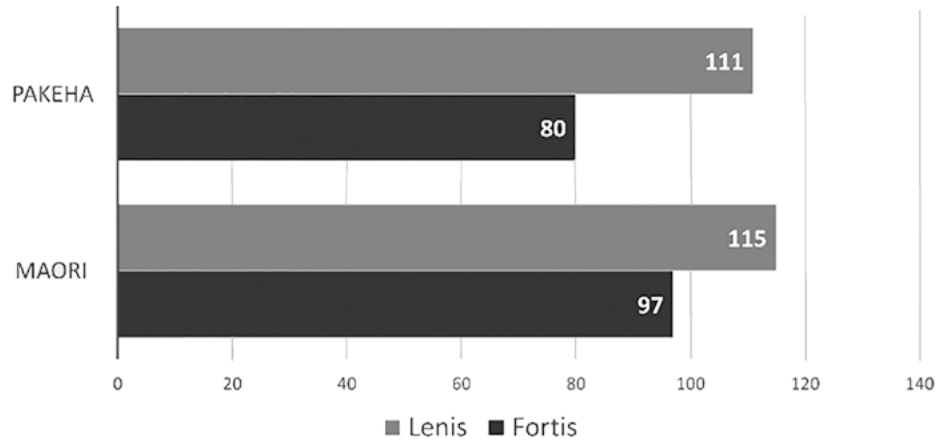


Fig. 5. Vowel duration before fortis and lenis consonants

If we compare the fraction of PE (0.72) with the results in previous studies, illustrated in Table 8, it can be seen that VE size is smaller in our data, as was expected due to the spontaneous speech, while the values obtained by Cho (2016) and Maclagan and Hay (2007) are from read material.

Table 8. Comparison of previous measurements and our data

Study	Speech style	Fraction
Maclagan and Hay (2007)	word list	0.56
Cho (2016)	minimal pairs	0.59
our data	running speech	0.72

Finally, the difference between the ratios of PE and ME is shown in Figure 6.

Fig. 6. Duration fraction by ethnicity

The graph reveals that VE size is smaller in ME compared to PE with a significant voice–ethnicity interaction (<0.0001***).

6.2. Aspiration/VOT

Regarding VOT, the dependent variable was voice onset time, and the tested predictors were ethnicity (Pākehā/Māori) and gender (male/female). For ME informants, we found weaker aspiration but no unaspirated tokens. In general, in both ME and PE, the voiceless plosives were either aspirated or affricated, with significantly longer VOT duration for female speakers (0.01668*) than for male speakers. The results also suggest that affrication is a widespread and dominant feature of these English varieties. Table 9 shows VOT duration for the labial, alveolar and velar plosives in ME and PE.

Table 9. ME and PE VOT durations in ms

	ME	**PE**
/p/	46	51
/t/	64	76
/k/	66	67
Overall mean	60	66
SD	27	27
n	318	309

The mean values clearly indicate that VOT duration is longer for Pākehā speakers for all the three plosives, with a minimal difference for the velar plosive, and greater differences for the labial and alveolar plosives.

In summary, it has been found that generally, vowel duration is shorter in PE than in ME with a greater difference in vowel duration before lenis and fortis consonants in PE. Accordingly, a greater VE size has been found in PE than in ME with a significant voice–ethnicity interaction (<0.0001***). Regarding VOT, the voiceless plosives are either aspirated or affricated in both varieties, with longer VOT duration for PE speakers, and no unaspirated tokens in ME. Even though VOT duration is longer in PE, the statistical analysis has not reached significance.

Based on the results, it can be stated that PE is aligned with AmE, being at the maximal end of the scale in terms of both VOT duration and VE size. As expected, vowel clipping due to the VE is smaller than in read speech (0.72, compared to values between 0.56–0.60 in Maclagan and Hay 2007; Cho 2016), due to our sample's more spontaneous speech style. VOT is 66 ms for Pākehā speakers, which is the average amount of aspiration in "mainstream" English varieties. Therefore, our results support our first hypothesis.

Fig. 7. ME and PE on the VE scale

The smaller VE fraction/ratio of ME (0.84/1.18) locates this variety lower on the scale than PE in terms of both VOT and VE size. However, it is towards the middle/higher, rather than the lower end of the scale – although no previous data is available to compare with, we anticipated ME to deviate from PE more massively. Since our speakers belong to the older generations, our overall mean VOT duration, 60 ms, is also higher than expected (cf. the mean value for the old age group in Maclagan and King (2007), 43 ms; see Table 1). Thus, our results only partly support the second hypothesis. Although ME is lower on the scale compared to PE, the substrate effect seems to be weaker, resulting in longer VOT values and an only slightly smaller VE size in ME. Figure 7 shows the relative positions of ME and PE on the relevant section of the VE scale established by Tanner et al. (2020) (see Figure 1 above).

The reason why our second hypothesis is not fully supported by our data is most probably stylistic. To eliminate vowel system differences, we had to choose standard ME informants: our ME speech samples come from highly educated speakers in a (radio) interview situation. Although this applies to most of the PE speakers, too, it seems to have influenced Māoris to a greater extent, leading to a less significant substrate effect than expected.

Our third hypothesis, however, that VOT and VE correlate in both PE and ME, is confirmed by the data: PE exhibits higher values, while ME has lower values, in

terms of both prevocalic plosive aspiration and variation in preconsonantal vowel duration.

7. Conclusion

The present study examined VE size and plosive VOT in the speech of Pākehā and Māori speakers in spontaneous speech. We made measurements in Praat and conducted statistical analyses using the *lme4* package of *R*, and showed that PE is at the maximal, and ME is towards the higher, end of the VE scale (i.e. both are (now) aspiration languages, even though we would expect the contact-induced structure of ME to be closer to a voice system). In harmony with prior literature, PE exhibits great VE size and long VOT values; however, as we expected, VE size is smaller than in word list data. Contrary to our initial hypotheses, the speakers in our ME sample are characterised by relatively long VOT values, but VE size is still smaller than in PE, so we observe a very weak overall substrate effect only. Besides addressing a research gap by placing PE and ME on the scale, the present study provides new empirical data in spontaneous speech for the theoretical proposal regarding the correlation between VOT and the VE.

Regarding the main theme of the paper, i.e. laryngeal phenomena showcasing how systemic, phonological motivations (expressible with the vocabulary of generative-theoretical phonology) lead to phonetic manifestations (measured with laboratory methodology), we hope to have shown that VOT and the VE correlate irrespective of the vowel system of the dialect. PE has longer average VOT and a larger VE size, ME has shorter average VOT and a smaller VE size – although our acrolect speakers use general, NZE vowels, i.e. the function of vowel length vs vowel quality is the same for both speaker groups. As explained in Section 3 above, there exists a functional-systemic explanation for the link between VE size and VOT, which seems to be supported by the data we see in both prior work and ours.

Acknowledgements

The authors are listed in alphabetical order. This study is being conducted in the frame of project no. PPKE-BTK-KUT-23-2, supported by the Faculty of Humanities and Social Sciences of Pázmány Péter Catholic University, and, for the first author, is also part of the "Laryngeal patterns in synchrony and diachrony" project sponsored by NKFI (National Research, Development and Innovation Office) grant no. 142498. We are deeply indebted to Katalin Szőke-Oláh, Bernadette Vine and Paul Warren for the help they provided in the data collection, and to László Hricisák and Katalin Mády for assisting with statistics in previous stages of the research. We are grateful to our anonymous reviewers for their insightful comments and

feedback, which have helped us improve the quality of the paper. All remaining errors are ours.

References

Aitken, Adam J. 1981/2015. "The Scottish Vowel-Length Rule." *Collected Writings on the Scots Language*. Ed. Caroline Macafee. Online. Scots Language Centre. http://medio.scotslanguage.com/library/document/aitken/The_Scottish_Vowel_length_Rule_(1981). Originally published in *So Meny People Longages and Tonges: Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh*. Ed. Michael Benskin, and M. L. Samuels. Edinburgh: The Middle English Dialect Project. 131–157.

Balogné Bérces, Katalin. 2022. "Accent Boundaries and Linguistic Continua in the Laryngeal Subsystems of English." *LingBaW* 8: 24–36.

Balogné Bérces, Katalin. In press. "Dialectal Variation in the Laryngeal Phonology of English." *Theoretical Phonology and the Phonology of English*. Ed. Patrick Honeybone, and Martin Krämer. Cambridge: Cambridge University Press.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-effects Models Using lme4." *Journal of Statistical Software* 67.1: 1–48.

Bauer, Winifred. 1997. *The Reed Reference Grammar of Māori*. With William Parker, Te Kareongawai Evans, and Te Aroha Noti Teepa. Auckland, N.Z.: Reed.

Bauer, Laurie, and Paul Warren. 2004. "New Zealand English: Phonology." *A Handbook of Varieties of English. Volume 1: Phonology*. Ed. Bernd Kortmann, and Edgar W. Schneider, in collaboration with Kate Burridge, Rajend Mesthrie, and Clive Upton. Berlin & Boston: Mouton de Gruyter. 580–602.

Bauer, Laurie, Paul Warren, Dianne Bardsley, Marianna Kennedy, and George Major. 2007. "Illustrations of the IPA: New Zealand English." *Journal of the International Phonetic Association* 37: 97–102.

Bayard, Donn. 2000. "New Zealand English: Origins, Relationships, and Prospects." *Moderna Språk* 1: 8–14.

Bell, Allan. 2000. "Maori and Pakeha English: A Case Study." *New Zealand English (Varieties of English Around the World)*. Ed. Allan Bell, and Koenraad Kuiper. Wellington: Victoria University Press. 221–248.

Boersma, Paul, and David Weenink. 2010. *Praat: Doing Phonetics by Computer*. Version 5.2. https://praat.org

Chen, Matthew. 1970. "Vowel Length Variation as a Function of the Voicing of the Consonant Environment." *Phonetica* 22: 129–159.

Cho, Hyesun. 2016. "Variation in Vowel Duration Depending on Voicing in American, British, and New Zealand English." *Phonetics and Speech Sciences* 8.3: 11–20.

Chodroff, Eleanor, and Colin Wilson. 2017. "Structure in Talker-specific Phonetic Realization: Covariation of Stop Consonant VOT in American English." *Journal of Phonetics* 61: 30–47.

Clark, Urszula. 2004. "The English West Midlands: Phonology." *A Handbook of Varieties of English. Volume 1: Phonology*. Ed. Bernd Kortmann, and Edgar W. Schneider, in collaboration with Kate Burridge, Rajend Mesthrie, and Clive Upton. Berlin & Boston: Mouton de Gruyter. 134–162.

Coretta, Stefano. 2020. "Vowel Duration and Consonant Voicing: A Production Study." PhD diss, University of Manchester.

Docherty, Gerard. 1992. *The Timing of Voicing in British English Obstruents*. Berlin & New York: Foris.

Farrington, Charlie. 2018. "Incomplete Neutralization in African American English: The Case of Final Consonant Voicing." *Language Variation and Change* 30: 361–383.

Fiasson, Romain. 2015. "Allophonic Imitation within and across Word Positions." PhD diss, University of Canterbury and Université d'Aix-Marseille.

Foulkes, Paul, Gerard J. Docherty, and Mark J. Jones. 2010. "Analysing Stops." *Sociophonetics: A Student's Guide*. Ed. Marianna Di Paolo, and Malcah Yaeger-Dror. New York: Routledge. 58–71.

Fox, Robert A., and Dale Terbeek. 1977. "Dental Flaps, Vowel Duration and Rule Ordering in American English." *Journal of Phonetics* 5.1: 27–34.

Hargrave, Rachel, Amy Southall, and Abby Walker. 2022. "Differences in Final /z/ Realization in Southwest and Northern Virginia." *American Speech* 97.3: 311–344.

Harris, Martin, and Noriko Umeda. 1974. "Effect of Speaking Mode on Temporal Factors in Speech: Vowel Duration." *The Journal of the Acoustical Society of America* 56.3: 1016–1018.

Hashimoto, Daiki, and Elizabeth Hume. 2018. "Effect of Stress on the Realization of Plosives in New Zealand English." *Te Reo – The Journal of the Linguistic Society of New Zealand* 61.1: 1–19.

Hay, Jennifer, Margaret Maclagan, and Elizabeth Gordon. 2008. *Dialects of English. New Zealand English*. Edinburgh: Edinburgh University Press.

Herd, Wendy. 2020. "Sociophonetic Voice Onset Time Variation in Mississippi English." *Journal of the Acoustical Society of America* 147.1: 596–605.

Holmes, Janet, and Helen Ainsworth. 1996. "Syllable-timing and Māori English." *Te Reo* 39: 75–84.

Holt, Yolanda Feimster, Ewa Jacewicz, and Robert A. Fox. 2016. "Temporal Variation in African American English: The Distinctive Use of Vowel Duration." *Journal of Phonetics & Audiology* 2: 121.

Honeybone, Patrick. 2005. "Diachronic Evidence in Segmental Phonology: The Case of Obstruent Laryngeal Specifications." *The Internal Organization of Phonological Segments*. Ed. Marc van Oostendorp, and Jeroen van de Weijer. Berlin & Boston: Mouton de Gruyter. 319–354.

House, Arthur S. 1961. "On Vowel Duration in English." *Journal of the Acoustical Society of America* 33: 1174–1178.

Iverson, Gregory K., and Joseph C. Salmons. 2008. "Germanic Aspiration: Phonetic Enhancement and Language Contact." *Sprachwissenschaft* 33: 257–278.

Jones, Daniel. 1950. *The Phoneme: Its Nature and Use*. Cambridge: W. Heffer and Sons.

Kerswill, Paul. 1987. "Levels of Linguistic Variation in Durham." *Journal of Linguistics* 23: 25–49.

Klatt, Dennis H. 1976. "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence." *Journal of the Acoustical Society of America* 59: 1209–1221.

Kovács, Magdolna. 2002. "Tendenciák és Szabályszerűségek a Magánhangzó-időtartamok Produkciójában és Percepciójában." [Tendencies and Rules in Production and Perception of Vowel Duration.] PhD diss, University of Debrecen.

Llamas, Carmen, Dominic Watt, Peter French, and Lisa Roberts. 2011. "Effects of the Scottish Vowel-length Rule on Vowel Quantity in Tyneside English." *Proceedings of the 17th International Congress of Phonetic Sciences*. 1282–1285.

Luce, Paul A., and Jan Charles-Luce. 1985. "Contextual Effects on Vowel Duration, Closure Duration, and the Consonant/Vowel Ratio in Speech Production." *The Journal of the Acoustical Society of America* 78.6: 1949–1957.

Maclagan, Margaret, and Jennifer Hay. 2007. "Getting *Fed* up with our *Feet*: Contrast Maintenance and the New Zealand English 'Short' Front Vowel Shift." *Language Variation and Change* 19: 1–25.

Maclagan, Margaret, and Jeanette King. 2007. "Aspiration of Plosives in Maori: Change Over Time." *Australian Journal of Linguistics* 27: 81–96.

Morley, Rebecca L., and Bridget J. Smith. 2023. "A Reanalysis of the Voicing Effect in English: With Implications for Featural Specification." *Language and Speech* 66.4: 935–973.

Mulgan, Richard. 1989. *Maori, Pakeha and Democracy*. Oxford: Oxford University Press.

Raphael, Lawrence J. 1972. "Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-final Consonants in American English." *Journal of the Acoustical Society of America* 51.4: 1296–1303.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org

Scobbie, James M. 2006. "Flexibility in the Face of Incompatible English VOT Systems." *Laboratory Phonology 8: Varieties of Phonological Competence*. Ed. Louis M. Goldstein, D. H. Whalen, and Catherine T. Best. Berlin & Boston: Mouton de Gruyter. 367–392.

Sonderegger, Morgan, Jane Stuart-Smith, Thea Knowles, Rachel MacDonald, and Tamara Rathcke. 2020. "Structured Heterogeneity in Scottish Stops Over the Twentieth Century." *Language* 96.1: 94–125.

SpeakApp. Version 2.0. https://speakapp.com

Tanner, James, Morgan Sonderegger, Jane Stuart-Smith, and Josef Fruehwald. 2020. "Toward 'English' Phonetics: Variability in the Pre-consonantal Voicing Effect across English Dialects and Speakers." *Frontiers in Artificial Intelligence* 3: 38. DOI: 10.3389/frai.2020.00038

Tauberer, Joshua, and Keelan Evanini. 2009. "Intrinsic Vowel Duration and the Post-vocalic Voicing Effect: Some Evidence from Dialects of North American English." *Interspeech 2009*. 2211–2214.

Te Puni Kōkiri. 2008. *The Health of the Māori Language in 2006*. Te Puni Kōkiri (Ministry of Māori Development) report of the 2006 Health of the Māori Language survey. https://www.tpk.govt.nz/en/o-matou-mohiotanga/te-reo-maori/the-health-of-the-maori-language-in-2006

Trudgill, Peter. 2004. *New-dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.

Walsh, Thomas, and Frank Parker. 1981. "Vowel Length and 'Voicing' in a Following Consonant." *Journal of Phonetics* 9.3: 305–308.

Warren, Paul, and Laurie Bauer. 2004. "Maori English: Phonology." *A Handbook of Varieties of English. Volume 1: Phonology*. Ed. Bernd Kortmann, and Edgar W. Schneider, in collaboration with Kate Burridge, Rajend Mesthrie, and Clive Upton. Berlin & Boston: Mouton de Gruyter. 614–624.

Watt, Dominic, and Catherine Ingham. 2000. "Durational Evidence of the Scottish Vowel Length Rule in Berwick English." *Leeds Working Papers in Linguistics and Phonetics* 8: 205–228.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Winter, Bodo. 2020. *Statistics for Linguists: An Introduction Using R*. New York: Routledge.

Wolfram, Walter. 1994. "The Phonology of a Sociocultural Variety: The Case of African American Vernacular English." *Child Phonology: Characteristics, Assessment, and Intervention with Special Populations*. Ed. John E. Bernthal, and Nicholas W. Bankson. New York: Thieme. 227–244.

Recordings:

https://paekoroki.tauranga.govt.nz
https://digitalnz.org/records?tab=Audio&text=#

Anglica. An International Journal of English Studies 35/2
ISSN: 0860-5734, e-ISSN: 2957-0905
Copyright © by Vencel Wutka, 2026
Creative Commons Attribution-ShareAlike 3.0 (CC BY-SA 3.0)
https://creativecommons.org/licenses/by-sa/3.0/
https://doi.org/10.7311/0860-5734.35.2.09

Vencel Wutka
https://orcid.org/0009-0006-1902-8232
Eötvös Loránd University

*FF as a Phonotactic Constraint in English: Evidence from an Acoustic Study

Abstract: The article presents an empirical study on the laryngeal pattern of English obstruents in 'voiceless' obstruent clusters (VOCs) in light of Szigetvári's (2020) ***FortisFortis* (*FF) phonotactic claim, which posits that adjacent fortis obstruents are prohibited within morphemes due to articulatory constraints. Using acoustic data from our target words: *a**bs**olute*, *rha**ps**ody*, and *kna**ps**ack* gathered via YouGlish (n.d.), we analyze the correlates of aspiration, segment durations, various vowel ratios, and voicing facts in Praat (Boersma and Weenink 2025), in order to test the validity of the *FF hypothesis. Rigorous statistical analysis reveals that the VOC of *rha**ps**ody* patterns with that of *a**bs**olute*, indicating phonological /bs/, rather than /ps/. Our findings provide support for the *FF constraint and suggest that English transcription conventions sometimes obscure phonological reality.

Keywords: aspiration, *FortisFortis, laryngeal pattern, statistics, phonotactic claim, transcription, voiceless obstruent clusters

1. Introduction

The phonetic realization of laryngeal oppositions and the phonological behavior of obstruent series have been subject to rigorous research. However, the laryngeal identity of obstruents in voiceless obstruent clusters (VOCs) is relatively under-explored and far from conclusive.[1]

Two-member obstruent clusters in Standard English[2, 3] are traditionally represented as: lenis+lenis (LL), lenis+fortis (LF), fortis+lenis (FL), and fortis+fortis

(FF). While such patterning is widely adopted by most transcription systems, it may largely reflect well-established orthographic conventions rather than true phonetic-phonological reality. Following G. Kiss and Szigetvári (2020) and Szigetvári (2020), we argue that the fourth category, FF clusters, is not permissible in English, as the language disallows two adjacent fortis obstruents within morphemes[4] due to a phonological constraint. To account for this phenomenon, Szigetvári (2020) proposes the ***FortisFortis* (*FF) phonotactic constraint, a claim that eliminates FF clusters by reanalyzing them either as LF or FL, depending on the phonetic cues that are associated with fortis vs lenis segments.

This study presents a phonetically grounded acoustic analysis of this constraint by taking a closer look at the putative 'FF' cluster in *rha**ps**ody*, one of our three target words, which we compare to the VOCs of *a**bs**olute* and *kna**ps**ack*. With the *FF constraint in mind, our working hypothesis implies that the bilabial stop + alveolar fricative clusters in *a**bs**olute* and *rha**ps**ody* pattern alike, yet differ from that of the morphologically complex test item, *kna**ps**ack* (*FF vs F#F). Thus, if these clusters exhibit comparable phonetic behavior, a reanalysis along the lines of an LF /bs/ may also be phonologically warranted. This would stand in stark contrast to what traditional analyses and transcriptions suggest: /æ**bs**əlut/, /ræ**ps**ədɪ/.

2. Literature review and theoretical framework

2.1. The phonetic correlates of fortis and lenis obstruents in English

English encodes the opposition between fortis and lenis obstruents primarily through a set of phonetic cues, some of which include aspiration. There may be as many as sixteen such acoustic properties that may cue the identification between fortes and lenes (Lisker 1986, 5). Pre- and postaspiration, segment duration, strength of articulation, preglottalization, and prefortis clipping may be the most well-known phonetic cues that help listeners in telling fortis and lenis obstruents apart.

Aspiration is a very robust feature in English. The distinction for utterance-initial fortis stops /p, t, k/ and the affricate /tʃ/ is carried by the release phase or by the glottal activity (Kohler 1984, 153). To initiate phonation in the following sonorant, the abducted vocal folds begin to come together only after the oral closure is released. However, it takes a short interval before they reach the adducted position required for voicing (Iverson and Salmons 1995, 370). During this post-release voicelessness phase, the aperiodic airflow ('puff of air') is perceived as aspiration, producing a relatively long Voice Onset Time (VOT), characteristic of fortis stops. G. Kiss (2017, n.p.), following Keating (1984) and Jansen (2004), notes that the conventional threshold of aspiration is 35 ms. English fortis stops are well above this cut-off point between zero and long-lag VOT. Lisker and Abramson (1964) report average VOT values for initial fortis stops in English: [pʰ] = 58 ms,

[t^h] = 70 ms, and [k^h] = 80 ms (cited in Beckman et al. 2013, 8). The interval between the release and the onset of voicing "carries intuitively a rather well-defined phonetic interpretation," constituting a defining reflex manifestation of a spread glottis (Nance and Stuart-Smith 2013, 130; Starčević 2024, 3). The rule of aspiration is often correlated with syllable-initial position (Kahn 1976, 73) and degrees of stress (Keating 1984, 306–308): Fortis stops in stressed syllables are generally more heavily aspirated than those in unstressed positions, where the glottal opening is smaller, leading to shorter VOT. Contrary to the standard assumption that strong aspiration occurs only in stressed, foot-initial positions, phonetic evidence shows that fortis stops may be realized with relatively long VOT in medial, post-tonic positions as well. G. Kiss (2017, n.p.) reports VOT values of 79.75 ms in *wri**t**er*, 55.25 ms in *hi**k**er*, and 35.75 ms in *pi**p**er*. In utterance-final position (or before a pause) other phonetic cues such as prefortis clipping (i.e. significantly shorter sonorant interval); or pre-glottalization (i.e. glottal reinforcement) are responsible for the contrast.

Lenis obstruents show none of the above-mentioned phonetic cues. They may never be aspirated, glottalized, and they do not shorten the preceding sonorant interval, hence their unmarked status in an aspirating system like English. In utterance-initial position, voicing may begin just before the release of the stop; these segments are thus characterized by zero or short-lag VOT: [b̥~p] = 1 ms, [d̥~t] = 5 ms, [g̥~k] = 21 ms (Lisker and Abramson 1964 cited in Beckman et al. 2013, 8). The phonetic realization of lenis obstruents in this position varies between partially voiced to fully voiceless (Ladefoged and Disner 2012, 139). They likely display a silent interval in isolated utterances, though some aperiodic noise may be emitted, followed instantly by the onset of voicing, which is why they resemble voiceless obstruents perceptually (Abramson and Whalen 2017, 79; Beňuš 2021, 125). Crucially, voicing in English is context-dependent rather than contrastive.[5] Therefore, negative VOT may only characterize lenis obstruents[6] if they are couched between sonorants. This type of passive (or perseverative) voicing is achieved by glottal pulses proceeding unbroken from the preceding sonorant into the lenis interval; this phenomenon is known as "unbroken voicing," "edge vibrations," or "bleed" (Davidson 2016, 43; Abramson and Whalen 2017, 79). The physiological conditions for this type of intersonorant voicing – such as a high degree of adduction, a low degree of vocal fold tension, lax walls, and movable parts – associated with lenis obstruents "are diametrically opposed to those necessary in fortis obstruents" (Kohler 1984, 162–163). Nonetheless, even in voice-friendly environments, such conditions may not be satisfied. Davidsen-Nielsen (1969, 58) and Docherty (1992, 120) reveal that even intersonorant lenis segments may be partially voiceless given that only about 57% of a lenis stop's closure tends to be voiced, while the rest remains voiceless. Stahlke (2003, 192) argues that this type of voicing is not essential to differentiate /b, d, g/ from /p, t, k/. Beňuš (2021, 119, 125) similarly claims that even though the vocal folds may maintain voicing for the initial part of the closure

of, for example, /b/, this segment is realized as a devoiced [b̥] which perceptually sounds similar to [p] due to the phonetic similarity of VOT, which is also corroborated by Ladefoged and Disner (2012, 139), who argue that an English lenis /b/ sounds more like a voiceless Spanish /p/. Overall, a lenis obstruent has two allophones, a passively voiced one ([b]) and a partially~fully voiceless one ([b̥~p]).

It is thus always possible to determine the laryngeal identity of a singleton obstruent as fortis or lenis in English. However, the laryngeal identity of obstruents in obstruent clusters is not always straightforward (G. Kiss and Szigetvári 2020, 136), which is what we turn to presently.

2.2. The suspension of aspiration and phonetically grounded experimentation

Syllabic position[7] is often linked to the distribution of aspiration in English: fortis plosives and the affricate are aspirated in syllable-initial position, but unaspirated in non-initial position, including /s/+stop$_{(fortis)}$ clusters, as in ***s***[p]*end*, ***s***[t]*ick*, or ***s***[k]*id*. However, aspiration patterns may not be reliably predicted on the basis of syllabic position: in *spit*, [p] is initial (it is in the onset) after all, but still unaspirated. Phonetically grounded models have therefore been proposed to account for the absence of aspiration in /s/+stop$_{(fortis)}$ clusters.

Davidsen-Nielsen's (1969, 55–56) perceptual, tape-cutting experiment investigated the putative 'FF' clusters of ***sp****ear*, ***sp****at*, ***st****eam*, ***st****y*, ***sc****old*, and ***sc****ore*. After removing [s], eight speakers of American English identified lenis[8] /b, d, g/ in 98%, and twenty-four speakers of British English in 92% of the cases. Since these segments lacked positive VOT and fell below the 35-ms threshold[9] for a segment to be identified as fortis, the fortis–lenis contrast appears perceptually neutralized (Davidsen-Nielsen 1969, 55; G. Kiss 2017, n.p.).

Catford (1977, 155) argues that a fortis stop is "produced with a relatively large glottal opening during stricture"; however, in /s/+stop$_{(fortis)}$ clusters these segments no longer behave as fortis, since they "[do] not contain any instruction for keeping the glottis spread" (Szigetvári 2020, 46). The resulting lack of aspiration is attributed to a narrowed glottis, which is insufficient for triggering the aspiration or devoicing of a following sonorant[10, 11] (Iverson and Salmons 1995, 376). Since a narrowed glottis is associated with lenis obstruents (explaining hence the lack of a voicing lag, Catford 1977, 114), phonetic clusters such as [sp, st, sk] are interpreted phonologically as /sb, sd, sg/ in this position. Further support comes from G. Kiss (2017, n.p.), who shows that the post-/f/ stop in ***fift****een* patterns similarly to post-/s/ [p, t, k].

Taken together, these findings suggest that post-fricative, 'fortis' stops lack important cues of fortisness in English, and may thus be analyzed as lenis obstruents phonologically, despite established transcription conventions. This does not, however, necessarily entail complete phonetic neutralization. As noted by an anonymous reviewer, residual differences may persist between the stop in *bill* and *spill*,

the latter features a higher f0 frequency in high-pitch environments both early and late in utterance (Hanson 2009, 435) and may have other laryngeal correlates not captured by VOT alone.

Crucially, such differences appear to be gradient and phonetically conditioned rather than phonologically contrastive. Monomorphemic /sb, sd, sg/ do not stand in opposition with /sp^h, st^h, sk^h/ in English. Even if the stop in *spill* differs phonetically from that of *bill*, it may nevertheless be interpreted by native speakers as lenis (and not [p^h]) after removing /s/, which does not undermine a phonological analysis, claiming these stops to be lenis. An analysis grounded in phonetics reflects the scalar nature of the phonetic realization of lenes, as opposed to what a phonological analysis strives to achieve: categorical phonological distinctions evident in lenis /b/ vs fortis /p/. A standardly assumed point where phonetics meets phonology is, for example, the ~35 ms threshold for aspiration, a point above which phonetic aspiration is indicative of phonological fortisness.

2.3. Laryngeal assimilation in English?

Given the lack of assimilation rules in English, it is suspicious to find obstruent clusters across morphemes (where assimilation tends to operate) that 'agree' in fortisness or lenisness. Data from the CUBE Dictionary (Lindsey and Szigetvári 2013) show a striking imbalance: FF clusters occur far more frequently than LL clusters. According to Szigetvári (2020, 43), FF clusters dominate word-initial and word-final positions and appear about ten times more often than LL in word-medial positions. The author writes that "the balance tilts strongly towards the marked pole of the contrast, fortis" (Szigetvári 2020, 43).

This prevalence may be attributed to Progressive Laryngeal Assimilation (PLA), which is traditionally claimed to be an active process in English. However, this PLA appears to be restricted to the suffixes *-ed* and *-s*, which are said to assimilate to the stem-final consonant in terms of their fortisness or lenisness, cf. *cloc**ked**, cloc**ks*** vs *dri**lled**, dri**lls***. For example, Harris (1994, 137) assumes an active feature reassociation rule, whereby a laryngeally unmarked obstruent acquires fortisness from a preceding fortis segment. Yet there is reason to doubt that any kind of assimilation or aspiration spreading occurs since this "would be the only case of laryngeal assimilation in English" (Szigetvári 2020, 44). We never find this effect in monomorphemic words. PLA seems to be a "very odd assimilation that does not operate morpheme internally," and if aspiration spreading worked progressively, we would expect lenis stops to become aspirated, which consistently fails to occur in English (Starčević 2024, 18), cf. *anecdote* /ænɪ**kd**ot/ *[kt], *Aztec* /æ**zt**ɛk/ *[st], or *Latvia* /læ**tv**iə/ *[tf].

This, in turn, suggests that no phonological process is required to derive, for example, a voiceless [t] in *clocke**d***. This view is corroborated by Iverson and Salmons (1995, 381), who argue that "voiced obstruents may not lie outside of

voiceless ones." Similarly, Stahlke (2003, 201) maintains that the devoicing of the past tense coronal suffix reflects "the predictable behavior of a lenis stop between voiceless glottal states."

To account for this behavior, Szigetvári (2020, 44) proposes that in phonetically constrained positions, where the phonetic cues of aspiration are absent, the segments in question are more appropriately analyzed as lenis. On this view, FF clusters need not be attributed to PLA, and forms such as *clocked* and *clocks* are better represented as /klɑ**kd**/ and /klɑ**kz**/ phonologically. This analysis does not yield any phonetic difference as the stem-final fortis stop has already clipped the preceding vowel, and in this unfavorable position, a lenis obstruent can only be interpreted phonetically as voiceless (G. Kiss and Szigetvári 2020, 139).

2.4. The hypothesis: **FortisFortis*

We have seen that it is not only fricative+stop clusters but also stop+stop clusters at the end of past tense forms that do not form FF clusters phonologically. To add to these, members of putative 'FF' clusters may be devoid of phonetic correlates that are necessary to categorically qualify them as fortis. Szigetvári (2020, 47) thus posits the ***FortisFortis** phonotactic claim, according to which no two fortis obstruents may be adjacent to each other within morphemes in English.

We aim to test this constraint and assume that *absolute* and *rhapsody* may be analyzed with identical LF /bs/ clusters phonologically, despite spelling and traditional analyses. The stop of *rhapsody* is predicted to show the following features: 1) it may not appear with a voicing lag as it is followed by an obstruent; 2) its duration is expected to be shorter; and 3) voicing bleed is assumed to be present due to planned voicing. The FL analysis, on the other hand, appears implausible as we assume that a local[12] fortis obstruent results in strong prefortis clipping potential, erroneously predicting that the vowel /æ/ of *absolute* and *rhapsody* are of identical length to that found in the morphologically complex F#F cluster of *knapsack*, which turns out not to be the case.

3. Methodology

3.1. Material and data collection

The test items *absolute*,[13] *rhapsody*, and *knapsack* were chosen based on the following criteria:

1) the target items have identical pre-cluster vowel qualities;
2) the target items have enough tokens uttered by native speakers on YouGlish (YG);
3) the target items show comparable stress patterns;
4) the target items are approximately matched in word length.

Sentence type (declarative: 110 tokens; interrogative: 3 tokens) and speakers' gender (≈ 1.6× more male) were not expected to affect our results. Likewise, utterance- or phrase-final positions, where segments tend to be longer, were unlikely to influence measurements, as most tokens occurred in non-prepausal environments where boundary-adjacent lengthening – the mechanical consequence of approaching a pause (Goldstein 2022, 5) – was not expected.

English stress patterns predict longer segmental durations in stressed than in unstressed syllables (Kim and Cole 2005, 2367), while polysyllabic shortening may reduce stressed-vowel duration as syllable count increases within polysyllabic feet; this is known as the temporal consequence of foot-structure on rhythm (Kim and Cole 2005, 2365). To mitigate such effects, we selected items with comparable word length[14] and preconsonantal vowel quality in mind.

The test item *absolute* has two possible stress patterns: *ábsolùte* vs *àbsolúte*. However, the majority of tokens sourced from YG were realized with initial primary stress (*ábsolùte*), making the comparison with the other initially stressed items, *rhápsody* and *knápsàck*, possible.

At the time of investigation, *absolute* returned 30,766 results, *rhapsody* 303, and *knapsack* 325 on YG, a website that integrates YouTube's search engine to provide access to authentic speech samples. All available YG accents were considered; however, "IE," "SCO," and "NZ" were unrepresented, restricting the analysis to "US," "UK," "AUS," and "CAN."

Subsequently, 50 random tokens were selected for each test item (150 tokens in total), and all timestamps and links were logged. Following quality control, 36 usable tokens remained for *rhapsody*, 39 for *absolute*, and 38 for *knapsack*. The 37 exclusions were due to incorrect tagging, overlapping speech, subpar audio, music, speaker duplication, or background noise.

Recordings were extracted at a sampling rate of 44,100 Hz (mono), using the internal HD audio controller of a Gigabyte Z690 UD DDR4 motherboard and saved as uncompressed wav files. The files were low-pass filtered in Audacity (ver. 3.6.1, Muse Group 2025), with a cut-off rate at 11,025 Hz, to reduce unwanted frequencies. No further voice reduction was necessary, and the files were ready for Praat analysis (ver. 6.3.18, Boersma and Weenink 2025).

Table 1. The distribution of accent and gender in the recordings[15]

Accent	*absolute*	*rhapsody*	*knapsack*	male	female
"US"	26	28	32	58	28
"UK"	11	5	4	9	11
"AUS"	2	3	–	2	3
"CAN"	–	–	2	–	2
∑	**39**	**36**	**38**	**69**	**44**

3.2. Procedure

The three main parameters examined included segment duration, stop voicing, and ratios, which allowed us to tell lenis from fortis obstruents apart. Table 2 summarizes the experimental parameters, with the right column listing their corresponding abbreviations. We present the statistical analyses and related visual illustrations of every item in Table 2 except for (1.c.), which is omitted here due to space limitations.

Segment duration was measured by first manually dividing the sound signal into temporal stretches and annotating each recording with Praat's TextGrid on a tier (titled: duration). The onset of the preconsonantal vowel was marked when the regular periodicity in the waveform and the first three formants in the spectrogram became visible, and its end boundary was marked in line with its end formants since vocal fold pulses may continue (well) into the stop's closure interval. The postvocalic stop began at the vowel's formant offset and ended at the small release burst (if there was any), which occasionally showed an abrupt discontinuity in the acoustic signal. The fricative onset roughly concurred with the stop release, ending at the cessation of friction. These segmentations enabled the measurement of the durations of the preC vowel, the postV stop, the interV CC cluster, and the V+CC cluster.

Stop voicing was estimated from quasi-periodic vibrations in the waveform. When the file contained some noise (inevitable with some recordings), both the overall noise level and the spectrogram were taken into account. Stops showing low-amplitude peaks and valleys at the start of the closure (left edge) were deemed voiced and labeled "vd" on a dedicated Praat tier (titled: voicing), the duration of which was measured in Praat. Stops with a relatively flat waveform without periodicity in the stop-closure interval were marked "vless."

To further support the analysis, V : CC, V : OCC, and SVR ratios were also calculated. All of these metrics combined allowed us to classify stops as either fortis or lenis. Figure 1 below shows a prototypical segmentation example considering our experimental parameters.

Table 2. Overview of experimental parameters

1) **Segment duration**
   a. Preconsonantal vowel
      preC V
   b. Postvocalic stop
      postV stop
   c. Intervocalic consonant cluster
      interV CC
   d. Vowel and consonant cluster
      V+CC

e. Total occurrence
OCC

2) **Stop voicing data**
   a. Voicing presence in the stop+closure interval
   vd / vless
   b. The duration of stop voicing
   –

3) **Ratio**
   a. Vowel to consonant cluster duration ratio
   V : CC
   b. Vowel to total occurrence duration ratio
   V : OCC
   c. Stop Voicing Ratio (proportion of voicing within the stop)
   SVR

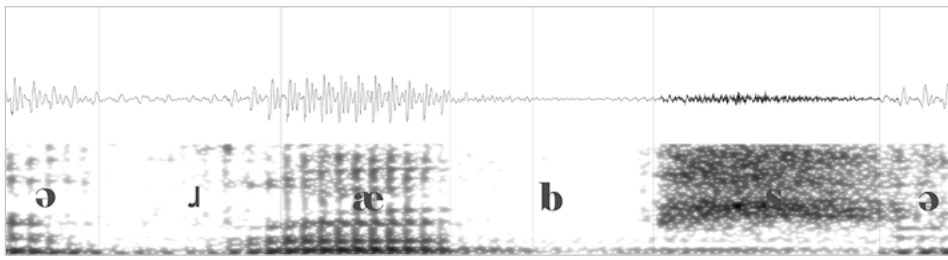


Fig. 1. A segmentation example on rhapsody

Note. This example was selected because, as pointed out by a reviewer, it is relatively hard to determine the boundary between a rhotic approximant and the following vowel, as /r/ is likely to "melt" into the vowel. In these cases, sole visual inspection of the waveform is insufficient. The formant structure and repeated auditory inspection proved helpful. What is more, low amplitude vibrations ('peaks and valleys') can be observed in the stop-closure interval (we analyze this stop as lenis, hence the transcription), the end boundary of which was marked (here with the green marker) when these waves ceased to exhibit periodic patterning.

3.3. Statistical analysis

To rigorously evaluate the similarities and differences in our data and to test the *****FF claim, descriptive statistics (standard deviation, standard error, mean, median, min., max.) were first computed in Microsoft Excel (Microsoft Corporation 2025).

Given the non-normal distribution of the data and unequal sample sizes (39–36–38), we used non-parametric methods to ensure accuracy and reliability. The Kruskal-Wallis H-test was applied to assess the means across groups, with

a $p < 0.05$ indicating statistical significance. For this test, η^2 was also calculated to determine the effect size and the proportion of variance attributable to group differences. Subsequently, pairwise comparisons were performed using the Mann-Whitney U-test, with the rank-biserial correlation (r) to interpret the size of differences.

Two custom Python 3 scripts running in Google Colaboratory (Google 2025) automated all calculations. The statistical script utilized the packages of NumPy and SciPy, while the Raincloud plots (Allen et al. 2019) and the final density and beehive plot meant for illustrations used Pandas, Matplotlib, Seaborn, and PtitPrince.

4. Findings

4.1. Introduction

This section presents the outcomes of the acoustic experiment for the three target items: *absolute*, *rhapsody*, and *knapsack*. Segment durations, stop voicing, and derived ratios are reported alongside visual representations to illustrate typical token patterns and highlight both statistically significant and non-significant differences between the items.

While absolute duration values provide an overview of temporal variation across items, examining ratios provides a much-needed complementary perspective. Ratios have been found "relatively invariant" across variations in speech rate (G. Kiss 2013, 11). Considering relative proportions may reveal a markedly different pattern, which may remain obscured when relying on 'raw' duration values alone. For this reason, we present ratio-based analyses for the duration of V+CC and V+OCC, along with our standard statistical measures.

In the following tables, we abbreviate the Kruskal-Wallis H-test as K–W and the Mann-Whitney U-tests as M–W. The short forms: A–R, A–K, and R–K stand for the pairwise comparisons between *absolute-rhapsody*, *absolute-knapsack*, as well as *rhapsody-knapsack*.

4.2. Segment duration

Besides descriptive statistics, Tables 3 and 4 present the Kruskal-Wallis and Mann-Whitney test scores for segment duration measurements, followed by their Raincloud plots in Figures 2–5.

The pairwise Mann-Whitney U-tests revealed a significant difference between *absolute* and *rhapsody*, indicating that the vowel in *rhapsody* was produced slightly slower. The stop in *knapsack* was produced with a significantly longer closure duration than in the other two items. The greater variability observed in the *rhapsody* data, however, may reflect some of the aforementioned segmentation difficulties encountered at the rhotic-vowel boundary.

Table 3. Duration of the preconsonantal vowel and postvocalic stop (in ms)

	Preconsonantal vowel			**Postvocalic stop**		
Item	***absolute***	***rhapsody***	***knapsack***	***absolute***	***rhapsody***	***knapsack***
N	39	36	38	39	36	38
Mean	114	130	118	60	57	69
Median	116	132.5	115	54	56.5	70
Min	63	58	71	28	27	35
Max	160	197	163	114	111	102
SD	24.32	35.45	20.29	21.30	19.21	17.43
SE	3.95	5.99	3.33	3.46	3.25	2.86
K-W	$H = 4.76$	$p = 0.0924$	$\eta^2 = 0.0425$	$H = 8.94$	$p = 0.0115$	$\eta^2 = 0.0798$
M-W	A-R	$p = 0.0313$	$r = 0.3551$	A-R	$p = 0.7941$	$r = 0.518$
	A-K	$p = 0.5786$	$r = 0.4629$	A-K	$p = 0.0168$	$r = 0.341$
	R-K	$p = 0.1398$	$r = 0.6001$	R-K	$p = 0.0061$	$r = 0.314$

Table 4. Duration of V+CC and the total occurrence (in ms)

	V+CC			**Total occurrence**		
Item	***absolute***	***rhapsody***	***knapsack***	***absolute***	***rhapsody***	***knapsack***
N	39	36	38	39	36	38
Mean	290	286	301	502	559	627
Median	294	274.5	301.5	491	562	624
Min	184	202	218	281	352	474
Max	393	400	366	711	785	791
SD	44.23	50.53	34.37	88.66	115.06	82.47
SE	7.17	8.54	5.65	14.38	19.45	13.56
K-W	$H = 4.01$	$p = 0.1346$	$\eta^2 = 0.0358$	$H = 25.98$	$p = 2.28e\text{-}06$	$\eta^2 = 0.232$
M-W	A-R	$p = 0.4939$	$r = 0.5463$	A-R	$p = 0.0439$	$r = 0.3643$
	A-K	$p = 0.1673$	$r = 0.5463$	A-K	$p = 3.22e\text{-}07$	$r = 0.1613$
	R-K	$p = 0.05767$	$r = 0.3713$	R-K	$p = 0.0044$	$r = 0.3070$

The V+CC data based on pairwise comparisons display a borderline non-significant difference among the target groups. Statistically significant differences were found, however, between all the groups, indicating that all the occurrences were pronounced with varying durations.

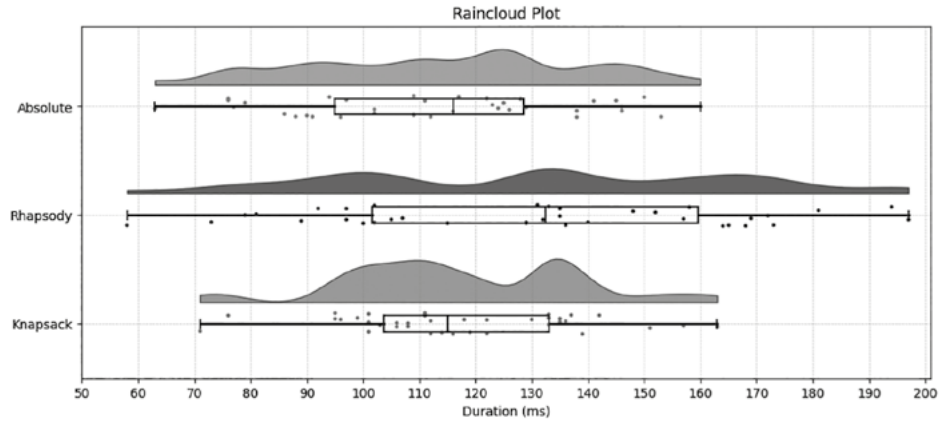


Fig. 2. Distribution and density of the preconsonantal vowel (in ms)

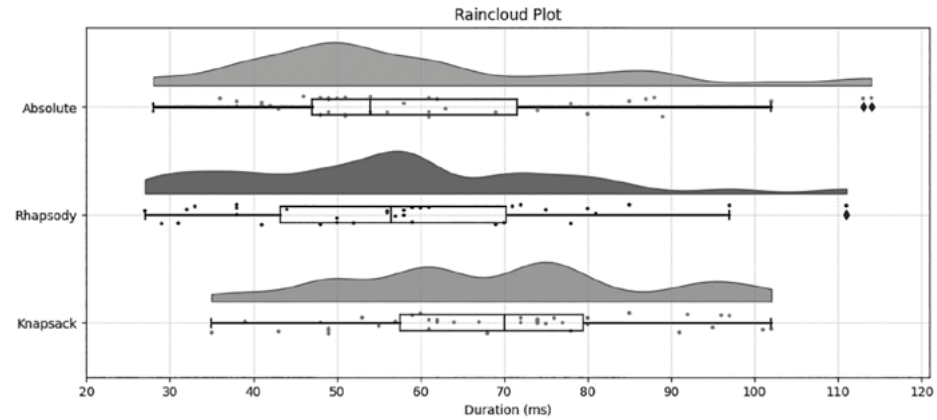


Fig. 3. Distribution and density of the postvocalic stop (in ms)

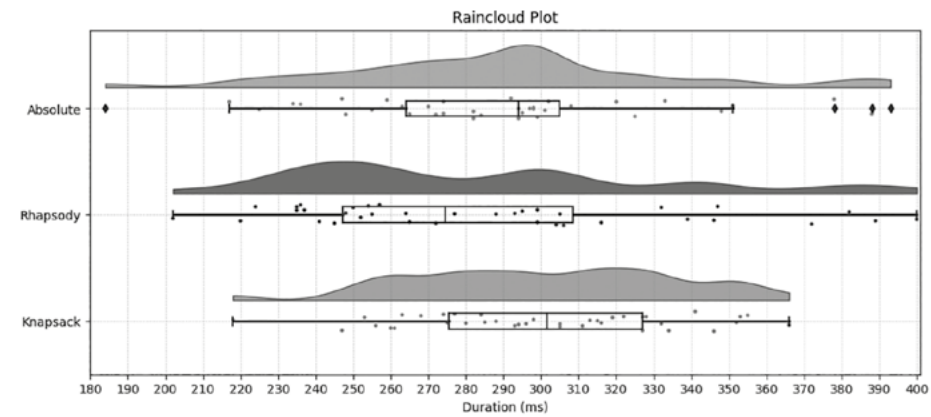


Fig. 4. Distribution and density of the vowel and consonant cluster (in ms)

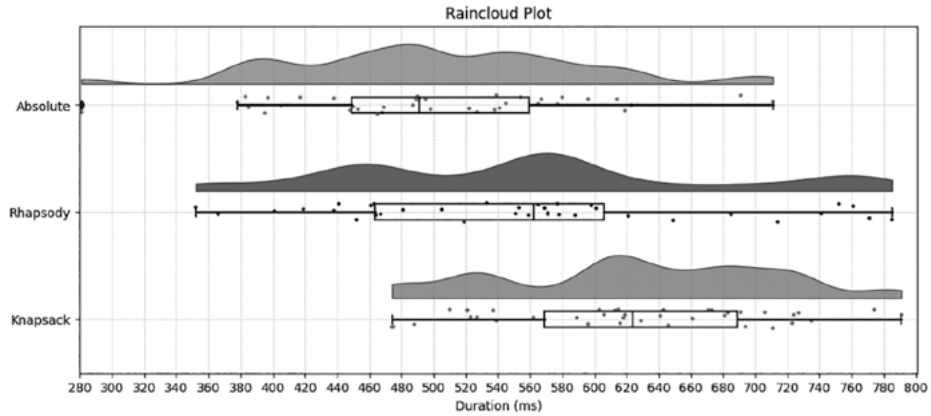


Fig. 5. Distribution and density of the total occurrence (in ms)

4.3. Duration ratios

Table 5 displays the descriptive statistics and the statistical scores for our V : CC and V : OCC duration ratios, which is similarly followed by their Raincloud plots in Figures 6–7.

Table 5. Ratio of the vowel's duration to the consonant cluster and occurrence's duration

	V : CC duration ratio			**V : OCC duration ratio**		
Item	***absolute***	***rhapsody***	***knapsack***	***absolute***	***rhapsody***	***knapsack***
N	39	36	38	39	36	38
Mean	0.67	0.85	0.39	0.23	0.24	0.19
Median	0.66	0.85	0.40	0.23	0.24	0.19
Min	0.27	0.35	0.26	0.12	0.13	0.12
Max	1.16	1.34	0.47	0.33	0.40	0.24
SD	0.19	0.23	0.05	0.05	0.06	0.03
SE	0.03	0.04	0.01	0.01	0.01	0.01
K-W	$H = 67.86$	$p=1.84\times10^{-15}$	$\eta^2 = 0.606$	$H = 18.09$	$p=0.00012$	$\eta^2 = 0.162$
M-W	A-R	$p=0.00105$	$r = 0.280$	A-R	$p=0.85$	$r = 0.487$
	A-K	$p=6.20\times10^{-11}$	$r = 0.933$	A-K	$p=0.00026$	$r = 0.741$
	R-K	$p=1.55\times10^{-12}$	$r = 0.978$	R-K	$p=0.00023$	$r = 0.749$

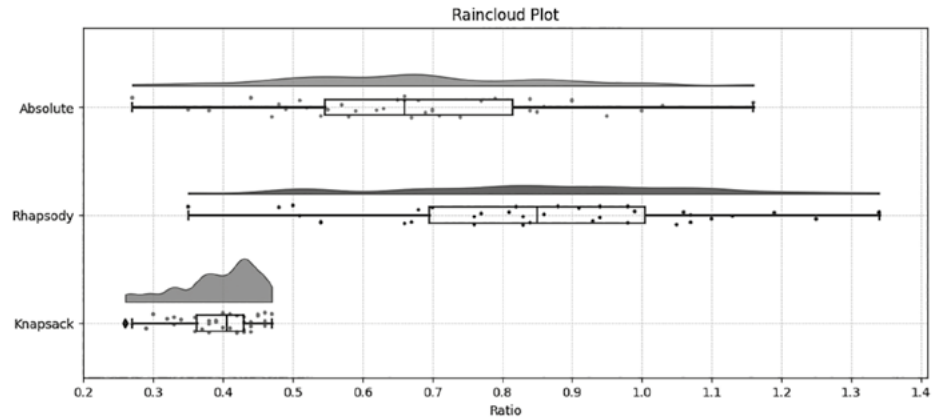


Fig. 6. Distribution and density of the V : CC duration ratio

In terms of the V : CC ratio, a highly significant difference with a very strong effect size was found between *absolute* vs *knapsack* and *rhapsody* vs *knapsack*, showing that the vowel in *knapsack* is significantly different when compared to the other two test items.

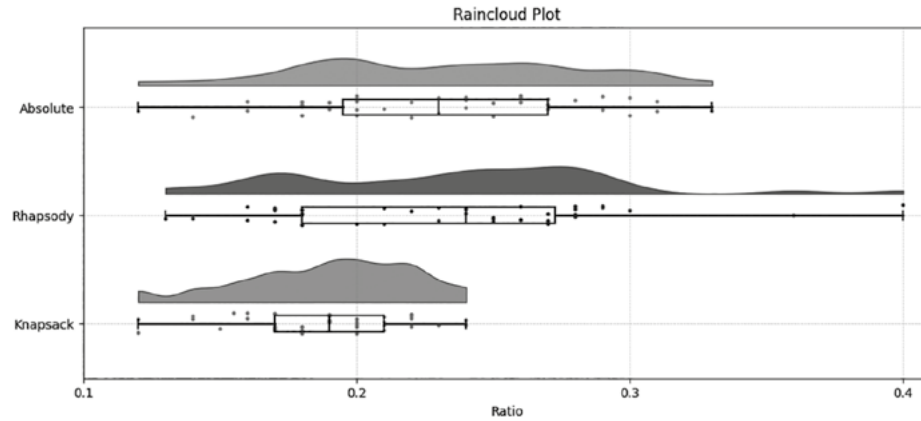


Fig. 7. Distribution and density of the V : OCC duration ratio

A statistically highly significant difference with a strong effect was also found in the V : OCC ratio data, indicating that the length of *knapsack*'s vowel was statistically shorter than that of *absolute* and *rhapsody*.

4.4. Stop voicing

A substantial part of the *absolute* (92%) and the *rhapsody* (89%) data contained some measure of voicing in the examined stop consonant, whereas in *knapsack*,

almost $^{2}/_{3}$ of the tokens (61%) were voiceless. Moving forward only those tokens are examined that have some voicing in their stop closure, the statistics of which (descriptive statistics of voicing duration and stop voicing ratio) are summarized in Table 6, followed by their Raincloud plots in Figures 8–9.

Table 6. Duration of stop voicing (in ms) and stop voicing ratio

	Duration of stop voicing			**Stop voicing ratio**		
Item	***absolute***	***rhapsody***	***knapsack***	***Absolute***	***rhapsody***	***knapsack***
N	36	32	15	36	32	15
Mean	35	28	24	0.62	0.49	0.33
Median	33	22	18	0.60	0.45	0.26
Min	9	9	11	0.18	0.16	0.12
Max	72	111	69	1.00	1.00	0.74
SD	15.56	18.95	16.92	0.24	0.23	0.19
SE	2.63	3.40	4.52	0.04	0.04	0.05
K-W	$H = 8.80$	$p = 0.012$	$\eta^2 = 0.1073$	$H = 16.31$	$p = 0.00029$	$\eta^2 = 0.1989$
M-W	A-R	$p = 0.024$	$r = 0.660$	A-R	$p = 0.021$	$r = 0.664$
	A-K	$p = 0.013$	$r = 0.724$	A-K	$p = 0.00018$	$r = 0.836$
	R-K	$p = 0.253$	$r = 0.605$	R-K	$p = 0.018$	$r = 0.717$

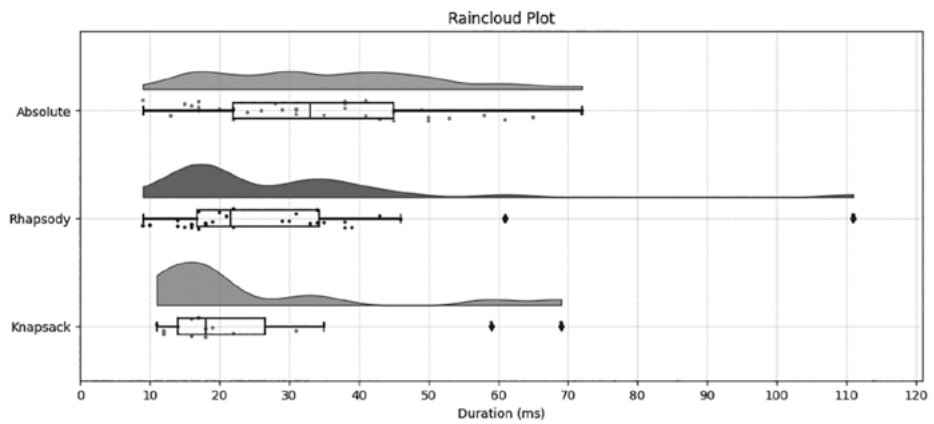


Fig. 8. Distribution and density of stop voicing (in ms)

Following from the stop voicing duration measures, the *absolute* and *rhapsody* data highlight a significantly higher degree of voicing when compared to the tokens of *knapsack*.

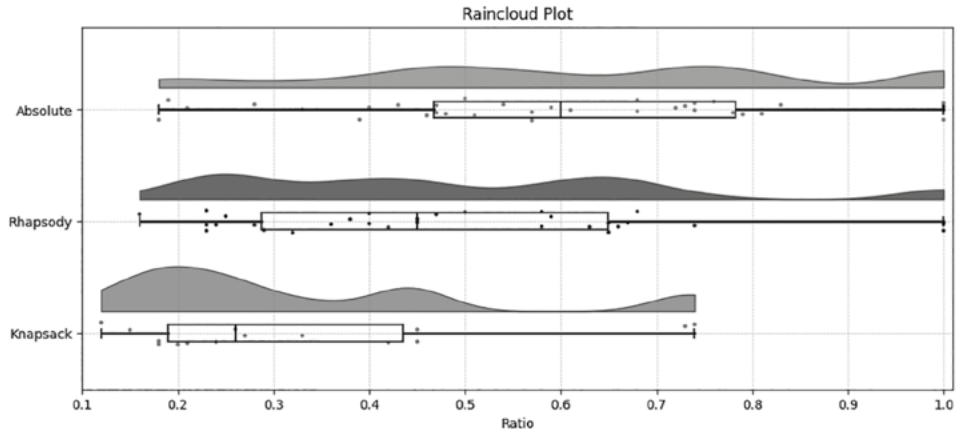


Fig. 9. Distribution and density of the Stop Voicing Ratio

The calculated stop voicing ratios correspond to the duration data. We find that the *knapsack* tokens show significantly shorter coarticulatory voicing when adjusted to *absolute* and *rhapsody*.

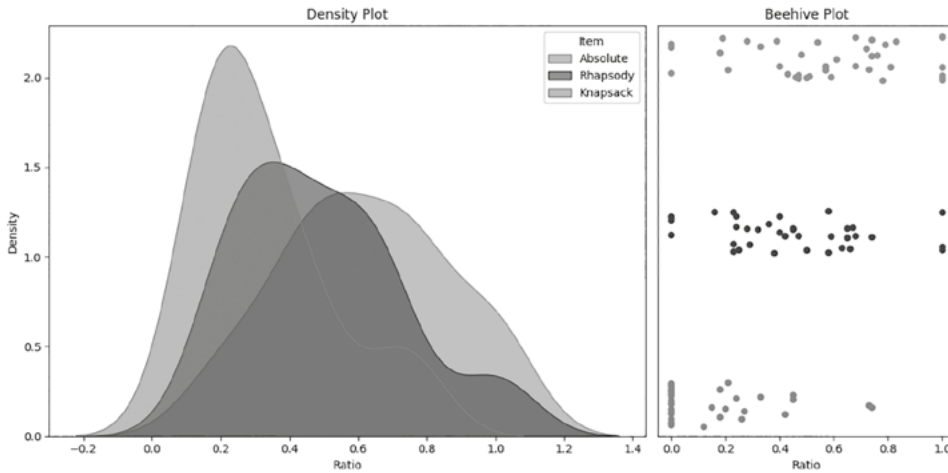


Fig. 10. Overlapping density plot and beehive plot for voicing ratio

Finally, Figure 10 aims to summarize our voicing facts. The left plot highlights that *knapsack* had the least amount of stop voicing, whereas the right plot represents that *absolute* and *rhapsody* had considerably more and longer stop voicing than *knapsack*.

5. Discussions

5.1. Interpretation of segment duration

The only statistically significant difference in segment duration was observed in the postvocalic stop, with the stop in *knapsack* averaging 9 ms longer duration than in *absolute* and 12 ms longer than in *rhapsody*, which is consistent with the phonetic fact that fortes exhibit longer consonantal durations than their lenis counterparts (Bickford 1985, 202).

Borderline non-significant differences were found in V+CC and total OCC durations, which likely reflect variable pronunciation rates in the randomly selected YG videos and may be responsible for the by-chance longer V and shorter interV CC realizations in the *rhapsody* data. Calculating ratios mitigates this complication, as ratios have been found to remain "relatively invariant" across variations in speech rates (G. Kiss 2013, 11).

The V : CC ratio in *knapsack* (0.39) is markedly lower than that of *absolute* (0.67) and *rhapsody* (0.85). This is further supported by the V : OCC ratios, which reveal that the examined Vs in *absolute* and *rhapsody* are approximately 21% and 26% longer than in *knapsack*. This finding underscores the well-known phonetic-phonological phenomenon of prefortis clipping, according to which vowels are realized shorter before fortes, which seems to operate only in *knapsack*, confirming this stop's identity as a fortis /p/. In contrast, the stops in *absolute* and *rhapsody* behave differently. The shorter realization and the consistent lack of a (prominent) clipping effect in *absolute* and *rhapsody* we interpret as the phonetic correlate of the cluster's phonological lenisness, leading us to postulate an LF /bs/, rather than the traditionally assumed 'FF' /ps/ cluster in these words.

5.2. Interpretation of stop voicing

The stop-voicing statistics also support the hypothesis that both *absolute* and *rhapsody* contain an LF cluster. In *absolute* 92% and in *rhapsody* 89% of the stops displayed some low-frequency, quasiperiodic vocal fold vibrations in the waveform, whereas *knapsack* was mostly voiceless at 61%, although 15 tokens showed some measure of voicing.

The voiced tokens in the *knapsack* data had a mean duration of 24 ms, a compact distribution, which is "usually indicative of coarticulatory voicing, i.e. voicing is just a by-product residue of the preceding vowel's final few vocal fold pulsations," which means that voice passes through the walls of the vocal tract and continues to shape the waveform (G. Kiss and Szigetvári 2020, 151; Mannell 2024, 4). By contrast, *absolute* and *rhapsody* showed longer mean voicing durations of 35 ms and 28 ms. Statistical analysis did not reveal a significant difference, but the data for *knapsack* may be right-skewed, potentially owing to two outliers (69 ms and 59 ms), which could elevate the mean toward *rhapsody*'s value and contribute

to its overall higher voicing ratio of 0.33. Such high ratios in these extreme cases may reflect planned voicing, suggesting that the stop may have been uttered as a lenis /b/ in these instances.

We find similar prompted, planned voicing, though with 1:1 ratios, in four cases of *absolute* and in three cases of *rhapsody*. However, the mean voicing ratios are substantially higher – 0.62 for *absolute* and 0.49 for *rhapsody* – indicating a passive implementation of voicing, despite the lack of an intersonorant position. This aligns with Docherty (1992, 120) and supports Iverson and Salmons' (2004, 53) claim that, even in final position, "lenis obstruents may acquire some measure of voicing from a preceding voiced segment." Further, if Bárkányi and Mády's (2012) ~30% threshold applies to English, obstruents may be perceived as (partially) voiced in languages without active voicing.

The data for *absolute* and *rhapsody*, however, are not entirely parallel. We assume that spelling may lead to a scalar distribution. Seeing <b> in *absolute* may cue some planned voicing, while <p> in *rhapsody* may not trigger the same effect. Consequently, other phonetic cues may be more responsible for the contrast, which could explain the shorter voicing duration and a smaller voicing ratio observed in the *rhapsody* data. Nevertheless, if spelling were a decisive factor, we would expect the stop to be consistently voiceless, without any vocal fold pulsations. We only identified four such tokens for *rhapsody*, which suggests that irrespective of spelling, speakers associate <p> with a lenis /b/, and usually produce it with planned voicing.

On the whole, our findings support G. Kiss and Szigetvári's claim (2020, 151) that "fortis obstruents are only voiced phonetically due to coarticulation [but] lenis obstruents may be phonetically voiced throughout allowing voicing to continue longer from the preceding vowel." The higher proportions and consistently longer stop-voicing values observed in the *absolute* and *rhapsody* align with this assumption. The facts on voicing in our analysis support the idea that *absolute* and *rhapsody* share an identical LF cluster as opposed to the F#F cluster of the morphologically complex *knapsack*.

5.3. Overall implications of the experiment

The experiment supports the proposed hypothesis that disallows FF clusters within morphemes in English. The segment/voicing durations and calculated ratios consistently support this analysis. To reiterate, *absolute* and *rhapsody* exhibit shorter stop durations, longer voiced intervals and vowel durations, as well as higher proportions of stop voicing beyond mere coarticulation, confirming an LF /**bs**/ cluster. In contrast, *knapsack* consistently displays longer stop duration, predominantly voiceless intervals, and shorter vowels indicative of prefortis clipping, supporting its F#F /**p#s**/ composition. The results presented show that orthography-driven transcriptions[16] are inaccurate. In a forthcoming large-scale study, we examine a similar cluster, namely that in *webster* vs *hipster*. Although these words are spelled – and

therefore transcribed – differently, they yield results consistent with those reported here; both appear to share an LFL /bsd/ composition. There certainly appears to be a pattern in the phonological ban on 'FF' clusters, warranting the necessity of future research.

5.4. Limitations of the study

Finally, some caveats are in order. The use of YouGlish may have its limitations. The initial selection criteria – frequency, minimal pairhood, and word length – could not be strictly maintained due to limitations in the available pronunciation data on the platform. We aimed to identify items that are spelled and therefore traditionally analyzed as containing 'FF' clusters, despite their expected similarity to LF clusters. We acknowledge that larger datasets provide greater statistical power (and the smaller scope of this investigation may limit generalizations), and that establishing a boundary between a rhotic and a vowel presents difficulties. In addition, one of the test items is one syllable shorter than the others, which may have affected the results. We plan to address these limitations in future controlled production experiments, examining native speakers' judgments on FF clusters and taking suprasegmentals into account (e.g. prosodic phrasing, rhythm, and intonation) to extend and refine the findings presented here.

6. Conclusions

This study examines selected VOCs in English and provides evidence for the ban of the proposed phonotactic constraint, *****FF. Both *absolute* /æ**bs**ələut/ and *rhapsody* /ræ**bs**ədɪ/ are consistent with an analysis involving a postvocalic lenis /b/, whereas *knapsack* /næ**p#s**æk/ contains a fortis /p/ as part of a morphologically complex F#F cluster. Longer stop duration, smaller V : CC and V : OCC ratios, and higher proportions and the ratio of planned voicing in the postvocalic stop provide statistically significant support for analyzing the VOC in *rhapsody* as /bs/, regardless of spelling or standard analyses. These findings also highlight the limitations of orthography-driven transcriptions and the importance of empirically grounded analyses coupled with phonological considerations.

Notes

1 I am grateful to Péter Szigetvári for introducing me into this constraint and for supervising this project during my MA studies. I am further grateful to Attila Starčević for his helpful guidance and careful comments. I would also like to thank the anonymous reviewers for their constructive feedback, which has greatly improved the clarity of this paper. I also acknowledge that any errors, mistakes, and inadequacies are entirely my own.

2 This study deals with Standard English only. An anonymous reviewer notes that there are varieties of English that may realize lenis obstruents with prevoicing, one such example is North Carolina English (Jacewicz et al. 2009). The reviewer also notes that if C_2 of a putative 'FF' cluster were lenis, then, under such dialectal conditions, we should expect Regressive Voicing Assimilation (RVA), something that fails to occur. This study does not deal with such varieties; however, prevoicing does not necessarily imply RVA. It may simply be the phonetic realization of the lenis series. Balogné Bérces and Huszty (2018, 170) examine potential feature spreading environments in Italian, a language known to realize lenes with prevoicing, and claim that "voice assimilation is not an integral part of the phonological system of varieties of Italian." Meccan Arabic, a dialect without aspiration, constitutes another example; the presence of voice does not trigger RVA, e.g. /yi**tb**aʕ/ > *[yi**db**aʕ] (Kenstowicz 2006). This means that certain prevoicing systems may not operate on a "true" voice pattern like Hungarian or Warsaw Polish.

3 Besides, we assume that the *FF claim holds for any variety of English which does not display 'peculiar' laryngeal phenomena. These atypical varieties include, among others, Yorkshire English, which is claimed to show aspiration spreading (Wells 1982, 366–377), and Durham English, where [voice] is proposed to function as an active feature (Kerswill 1987, 42; Cyran 2014, 201).

4 FF clusters seem to be disallowed in forms with vowelless morphs as well, e.g. 3Sg. present, nominal plural, or genitive suffixes. When referring to this constraint by "within morphemes," we mean these environments as well.

5 Recall that the phonological difference is encoded by aspiration rather than voicing in an aspirating system.

6 The reason why only unspecified lenes could 'receive' voicing in voice-friendly positions has to do with the assumption that phonetic processes cannot change an underlyingly specified phonological feature (Jessen and Ringen 2002, cited in Beckman et al. 2013, 36).

7 In a word like *abstemious* [əbs**t**émiəs], the absence of aspiration may not be simply taken to be the consequence of syllabic position. An analysis based on the Sonority Sequencing Principle of Clements (1990, 297) or Zec (1995, 86) would place the boundary between /s/ and /t/, which means that the stop is in a stressed, prevocalic syllable-initial position where aspiration is expected, yet it is suspended, cf. [əb**s**.**t**émiəs]. This analysis would however go against the Onset Maximization Principle, according to which onsets should be maximized and generally have precedence over codas (Kahn 1976, 41; Lombardi 1995, 71). Aspiration would only be expected if its morphological structure was **abs#temious*.

8 Davidsen-Nielsen's (1969) findings are corroborated by earlier studies: Lotz et al. (1960), Reeds and Wang (1961).

9 Even minimal aspiration may be insufficient to classify a stop as aspirated (Abramson and Whalen 2017, 77).

10 The devoicing of non-nasal sonorants (/l, r, j, w/) following a fortis obstruent is often attributed to aspiration spreading (cf. Harris 1994; Iverson and Salmons 1995). However, there is reason to believe that approximant devoicing only reflects a phonetic effect (just like the passive implementation of voice in intersonorant position). Spreading may not cause a class-changing process whereby an unmarked approximant

acquires a phonological feature. Rather, partial devoicing occurs since the glottis is still open during production (i.e. long-lag VOT), as in ***cl****aim* [**kl̥**eɪm], but not in *glee* [gli] (Tsuchida et al. 2000, 169; Stahlke 2003, 204; Starčević 2024, 4).

11 Tsuchida et al. (2000) report the results of a fiberscopic and acoustic investigation using transillumination. They find that in initial stressed position the glottal opening duration differs between fortis stops and fricatives, also demonstrating that /l/ is more devoiced after fortis stops than fricatives, cf. *play* vs *flee* (Tsuchida et al. 2000, 171, 175). Based on this, they argue that devoicing after stops is the result of [spread glottis], while in fricatives, it is due to aerodynamic requirements. They therefore propose to 'strip' fortis fricatives of [spread glottis] and instead contrastively mark lenis fricatives for [voice], creating a "hybrid system" (Tsuchida et al. 2000, 171, 177, 179).

12 In both words, the stressed vowel is expected to exhibit some shortening even without a local fortis obstruent (Szigetvári 2019, 102). This should not align *absolute* and *rhapsody* with *knapsack*, however.

13 A note on *absolutely*: The reason why *absolute* was preferred over *absolutely*, which has vastly more results on YG (173,708 at the time of investigation), has to do with possible frequency effects. The overuse of *absolutely* in casual speech may shorten segment durations, despite a normal speech rate.

14 We also compared clusters of equal size, since, according to Haggard's (1973) observations, we may expect about 15% shortening when comparing single, biconsonantal, and triconsonantal clusters (cited in Tsuchida et al. 2000, 174), which would have distorted the findings.

15 Please note that speakers' genders were inferred from available information and may not be their true gender identity. Assumptions were made solely for experimental purposes. These are not definitive and should be interpreted with caution.

16 It is worth noting that a hypothetical, phonologically-conscious transcription system with the *****FF constraint in mind may have disadvantages. A realized 'FL' cluster, as in e.g. *cooked* /kd/, may prompt learners of voicing-language backgrounds to apply RVA, which would reflect the phonological behavior of this obstruent series in their native languages.

References

Abramson, Arthur S., and D. H. Whalen. 2017. "Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions." *Journal of Phonetics* 63.1: 75–86. DOI: 10.1016/j.wocn.2017.05.002

Allen, Micah, Davide Poggiali, Kirstie Whitaker, Tom Rhys Marshall, Jordy van Langen, and Rogier A. Kievit. 2019. "Raincloud Plots: A Multi-platform Tool for Robust Data Visualization." *Wellcome Open Research* 4: 63. DOI: 10.12688/wellcomeopenres.15191.2

Balogné Bérces, Katalin, and Bálint Huszthy. 2018. "Laryngeal Relativism Predicts Italian." *Yearbook of the Poznań Linguistic Meeting* 4: 153–177. DOI: 10.2478/yplm-2018-0007

Bárkányi, Zsuzsanna, and Katalin Mády. 2012. "The Perception of Voicing in Fricatives." Paper presented at the 9th Old World Phonology Conference, Berlin, Germany, January 18.

Beckman, Jill, Michael Jessen, and Catherine Ringen. 2013. "Empirical Evidence for Laryngeal Features: Aspirating vs. True Voice Languages." *Journal of Linguistics* 49.2: 259–284. DOI: 10.1017/S0022226712000424

Beňuš, Štefan. 2021. *Investigating Spoken English*. Cham: Springer International Publishing.

Bickford, J. Albert. 1985. "Fortis/Lenis Consonants in Guichicovi Mixe: A Preliminary Acoustic Study." *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 29.1: 195–207. DOI: 10.31356/silwp.vol29.01

Boersma, Paul, and David Weenink. 2025. "Praat: Doing Phonetics by Computer." http://www.praat.org/

Catford, John Cunnison. 1977. *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

Clements, George N. 1990. "The Role of the Sonority Cycle in Core Syllabification." *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Ed. John Kingston, and Mary E. Beckman. Cambridge: Cambridge University Press. 283–333.

Cyran, Eugeniusz. 2014. *Between Phonology and Phonetics: Polish Voicing*. Boston & Berlin: De Gruyter Mouton.

Davidsen-Nielsen, Niels. 1969. "English Stops after Initial /s/." *English Studies* 4.1: 321–339. DOI: 10.1080/00138386908597340

Davidson, Lisa. 2016. "Variability in the Implementation of Voicing in American English Obstruents." *Journal of Phonetics* 54.1: 35–50. DOI: 10.1016/j.wocn.2015.09.003

Docherty, Gerard J. 1992. *The Timing of Voicing in British English Obstruents*. Berlin: Walter de Gruyter.

G. Kiss, Zoltán. 2013. "Measuring Acoustic Correlates of Voicing in Stops and Fricatives." *VLLXX: Papers Presented to László Varga on his 70th Birthday*. Ed. Péter Szigetvári. Budapest: Department of English Linguistics, Eötvös Loránd University & Tinta Publishing House. 289–311.

G. Kiss, Zoltán. 2017. "Aspiration of Stops after Fricatives in English: Results from a Pilot Experiment." *70 Snippets to Mark Ádám Nádasdy's 70th Birthday*. Ed. Péter Szigetvári. Budapest: Department of English Linguistics, Eötvös Loránd University.

G. Kiss, Zoltán, and Péter Szigetvári. 2020. "Telling Fortis and Lenis Apart in English Obstruent Clusters." *The Even Yearbook* 14.1: 135–156.

Goldstein, Louis. 2022. "Prosody: Temporal structure." University of South Carolina. https://sail.usc.edu/~lgoldste/General_Phonetics/Slides/Prosodic%20Phrasing%20and%20Rhythm.pdf

Google. 2025. "Google Colaboratory." https://colab.research.google.com/.

Hanson, Helen M. 2009. "Effects of Obstruent Consonants on Fundamental Frequency at Vowel Onset in English." *Journal of the Acoustical Society of America* 125.1: 425–441. DOI: 10.1121/1.3021306

Harris, John. 1994. *English Sound Structure*. Oxford & Cambridge, MA: Blackwell.

Iverson, Gregory K., and Joseph C. Salmons. 1995. "Aspiration and Laryngeal Representation in Germanic." *Phonology* 12.3: 369–396. DOI: 10.1017/S0952675700002566

Iverson, Gregory K., and Joseph C. Salmons. 2004. "Laryngeal Enhancement in Early Germanic." *Phonology* 20.1: 43–74. DOI. 10.1017/S0952675703004469

Jacewicz Ewa, Fox Robert Allen, and Lyle Samantha. 2009. "Variation in Stop Consonant Voicing in Two Regional Varieties of American English." *Journal of the International Phonetic Association* 39.3: 313–334. DOI: 10.1017/S0025100309990156

Jansen, Wouter. 2004. "Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian, and Dutch." PhD diss, Rijksuniversiteit Groningen.

Kahn, Daniel. 1976. "Syllable-based Generalisations in English Phonology." PhD diss, Massachusetts Institute of Technology.

Keating, Patricia A. 1984. "Phonetic and Phonological Representation of Stop Consonant Voicing." *Language* 60.2: 286–319. DOI: 10.2307/413642

Kenstowicz, Michael. 2006. "Phonetic Realization: Releases and Transition." https://ocw.mit.edu/courses/24-964-topics-in-phonology-phonetic-realization-fall-2006/c24d3390827b3bb7de93e24bf9d9476d_MIT24_964F06_lec10_bursts.pdf

Kerswill, Paul E. 1987. "Levels of Linguistic Variation in Durham." *Journal of Linguistics* 23.1: 25–49. DOI: 10.1017/S0022226700011026

Kim, Heejin, and Jennifer Cole. 2005. "The Stress Foot as a Unit of Planned Timing: Evidence from Shortening in the Prosodic Phrase." *Proceedings of Interspeech 2005*. 2365–2368. DOI: 10.21437/Interspeech.2005-37

Kohler, Klaus J. 1984. "Phonetic Explanation in Phonology: The Feature Fortis/Lenis." *Phonetica* 41.3: 150–174. DOI: 10.1159/000261721

Ladefoged, Peter, and Sandra F. Disner. 2012. *Vowels and Consonants*. 3rd ed. Malden, MA & Oxford: Wiley-Blackwell.

Lindsey, Geoff, and Péter Szigetvári. 2013–. "CUBE: Current British English Searchable Transcriptions." http://cubedictionary.org

Lisker, Leigh. 1986. "Voicing in English: A Catalogue of Acoustic Features Signalling /b/ versus /p/ in Trochees." *Language and Speech* 29.1: 3–11. DOI: 10.1177/002383098602900102

Lombardi, Linda. 1995. "Laryngeal Neutralization and Syllable Wellformedness." *Natural Language and Linguistic Theory* 13.1: 39–74.

Lotz, John, Arthur S. Abramson, Louis J. Gerstman, Frances Ingemann, and William J. Nemser. 1960. “The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai: A Tape-cutting Experiment.” *Language and Speech* 3.1: 71–77. DOI: 10.1177/002383096000300202

Mannell, Robert. 2024. “Speech Spectra and Spectrograms: 4. Some Detailed Views of “Heed”.” https://www.mq.edu.au/__data/assets/pdf_file/0004/911641/heed_closeups.pdf

Microsoft Corporation. 2025. “Microsoft Excel.” https://office.microsoft.com/excel

Muse Group. 2025. “Audacity.” https://www.audacityteam.org/

Nance, Claire, and Jane Stuart-Smith. 2013. “Pre-aspiration and Post-aspiration in Scottish Gaelic Stop Consonants.” *Journal of the International Phonetic Association* 43.2: 129–152. DOI: 10.1017/s0025100313000042

Reeds, James A., and William S.-Y. Wang. 1961. “The Perception of Stops after /s/.” *Phonetica* 6.1: 78–81. DOI: 10.1159/000258075

Stahlke, Herbert F. W. 2003. “Fortis and Lenis Obstruents in English.” *WORD* 54.2: 191–216. DOI: 10.1080/00437956.2003.12068832

Starčević, Attila. 2024. “Another Fortis–Lenis Language: A Reanalysis of Old English Obstruents.” *Acta Linguistica Academia* 71.3: 259–308. DOI: 10.1556/2062.2023.00705

Szigetvári, Péter. 2019. “Where Are Syllables?” *Radical: A Journal of Phonology* 1.1: 92–125.

Szigetvári, Péter. 2020. “Emancipating Lenes: A Reanalysis of English Obstruent Clusters.” *Acta Linguistica Academica* 67.1: 39–52. DOI: 10.1556/2062.2020.00004

Tsuchida, Ayako, Abigail C. Cohn, and Masanobu Kumada. 2000. “Sonorant Devoicing and the Phonetic Realization of [Spread Glottis] in English.” *Working Papers of the Cornell Phonetics Laboratory* 13: 167–181.

Wells, John C. 1982. *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

YouGlish. n.d. “YouGlish.” https://youglish.com.

Zec, Draga. 1995. “Sonority Constraints on Syllable Structure.” *Phonology* 12.1: 85–129. DOI: 10.1017/s0952675700002396

GUEST EDITOR

ARTUR KIJAK is Associate Professor of English Language and Linguistics at the Institute of Linguistics, University of Silesia in Katowice, and co-editor of *Linguistica Silesiana*, an international journal of linguistic studies. His main research interests include segmental and suprasegmental phonology, historical phonology, contrastive phonology and current phonological models. He is the author of *Polish and English Consonantal Clusters: A Contrastive Analysis within the Strict CV Framework* (2008) and *Labial-Dorsal Interactions: A Phonologically Based Approach* (2017) and co-editor of *Current Developments in English Historical Linguistics: Studies in Honour of Rafał Molencki* (2017).

AUTHORS

KATALIN BALOGNÉ BÉRCES is Associate Professor at Pázmány Péter Catholic University, Budapest, and Catholic University in Ružomberok, Slovakia. Her fields of research include phonological theory, the phonology and dialectology of English, linguistic variation and language contact, foreign-accented Englishes and pronunciation teaching. She has authored *Strict CV Phonology and the English Cross-Word Puzzle* (2008, Saarbrücken: VDM Verlag), *The Pronunciation of English* (2006, Budapest, with Szilárd Szentgyörgyi), *Beginner's English Dialectology* (2008, Budapest) and *The Structure of English* (2016/2018, Budapest). She is a member of English studies societies (ISLE, ESSE, HUSSE, SKASE) and one of the organisers of the Government Phonology Round Table.

BERT BOTMA is a university lecturer at the Leiden University Centre for Linguistics, where he teaches courses in English phonetics and phonology.

KAROLINA BROŚ is an Associate Professor at the University of Warsaw, specializing in (socio)phonetics and phonology. Her main research interests include the phonetics–phonology interface, language variation and change and the links between articulation and acoustics in speech production. To date, Dr Broś has studied lenition processes in several varieties of Spanish and is currently investigating consonantal changes in American English as a part of her Fulbright project conducted in California. In particular, she is working on optimizing research methodology that informs phonological generalizations concerning the studied languages and dialects.

SILKE HAMANN is Associate Professor at the University of Amsterdam. Her research encompasses experimental studies, formal modelling and computer simulations of L1 and L2 perception and production, all centered on the phonetics–phonology interface. She also

looks into the role of orthographic information and how this links to and interacts with phonological representations.

HAMZA KHWAJA is a PhD candidate at the Leiden University Centre for Linguistics.

KUNIYA NASUKAWA is Professor of Linguistics at Tohoku Gakuin University, Japan. He has a Ph.D. in Linguistics from University College London, and his research interests include prosody-melody interaction and precedence-free phonology. He has written many articles covering a wide range of topics in phonological theory. He is the author of *A Unified Approach to Nasality and Voicing* (De Gruyter 2005), co-editor (with Nancy C. Kula and Bert Botma) of *The Bloomsbury Companion to Phonology* (Bloomsbury 2013), co-editor (with Henk van Riemsdijk) of *Identity Relations in Grammar* (De Gruyter 2014) and editor of *Morpheme-internal Recursion in Phonology* (De Gruyter 2020).

MARKUS A. PÖCHTRAGER received his PhD from the Linguistics Department at the University of Vienna in 2006. From 2008–2017 he was Assistant Professor in Linguistics at Boğaziçi University (Istanbul), where he conducted two projects on minority languages in Turkey, and received the prize for the best teacher of the faculty twice. Since 2017 he has been a lecturer (later senior scientist) in linguistics at the University of Vienna. His research deals with phonological theory, the interaction and delimitation between phonology and morphology, structural analogy to syntax, linguistics as a cognitive science and the languages of Eurasia.

ERIKA SAJTÓS is enrolled in the PhD programme of the Doctoral School of Linguistics, Pázmány Péter Catholic University, Budapest. Her research field is sociophonetics, the phonology and phonetics of New Zealand English and other southern-hemisphere varieties. In her previous research, she investigated the history of vowel shifts in New Zealand English, their variability in style-shifting and social indexicality. Currently, she is examining aspects of the laryngeal phonology of New Zealand (Pakeha) English, Maori English and Australian English. She is a member of English studies societies (ISLE, ESSE, HUSSE). In 2023, she was awarded the LINGUIST List's Rising Star title.

JOSEPH SMITH is a Research Master student at the Leiden University Centre for Linguistics.

RADOSŁAW ŚWIĘCIŃSKI is an Associate Professor and lecturer at Amsterdam University of Applied Sciences. He is a phonetician specialising in laboratory phonology, articulatory phonetics and the phonetics–phonology interface. His research focuses on speech production, consonant cluster coordination, articulatory settings and experimental phonetics, particularly using electromagnetic articulography (EMA) and acoustic analysis of speech. He has published on Polish and English phonetics and phonology, pronunciation teaching and cross-linguistic speech phenomena. His recent work also addresses international business communication.

VENCEL WUTKA is a first-year PhD student in Linguistics at the Department of English Linguistics, Eötvös Loránd University. His main research interests include the laryngeal phonology of aspirating languages, the phonetics–phonology interface and acoustically testable phonotactic constraints. Besides teaching introductory linguistics courses at Eötvös Loránd University, he focuses on the phonetics and phonology of Germanic languages with a particular, exploratory emphasis on the application of phonotactic constraints targeting obstruents clusters in the varieties of Bavarian German and Nigerian English.

SŁAWOMIR ZDZIEBKO is Assistant Professor in the Department of English–Polish Contrastive Studies at The John Paul II Catholic University of Lublin. His main interests encompass the morphophonology of consonant mutation in Polish and English as well as the interfaces between morpho-syntax and phonology and between phonology and phonetics.

Artur Kijak
From Generative to Laboratory Phonology:
The Phonetics–Phonology Overlap

Markus A. Pöchtrager
Tales of the Unexpected (Phonotactic Dependencies)

Kuniya Nasukawa
The Representation of sC Clusters in English:
A PfP Perspective

Sławomir Zdziebko
A Substance-free Autosegmental Account
of the Coronal Mutation in English

Bert Botma, Hamza Khwaja and Joseph Smith
A Stratal Approach to English Affricates

Silke Hamann and Radosław Święciński
How Language-specific Perceptual Cues Can Explain
Differences in Onset Clusters:
The Case of /tr/ in Polish and English

Karolina Broś
Lip Tracking and Acoustics at the Phonetics–Phonology
Interface: Coronal Stop Deletion in American English

Katalin Balogné Bérces and Erika Sajtós
Correlations and Functional Explanations in Laryngeal
Phonetics and Phonology:
Evidence from New Zealand English

Vencel Wutka
Rhapsodic Revisions:
Voiceless Obstruent Clusters in English

ISSN 0860-5734 (Print)
ISSN 2957-0905 (Online)

www.anglica-journal.com