ISRG Journal of Education, Humanities and Literature (ISRGJEHL)

ISRG PUBLISHERS

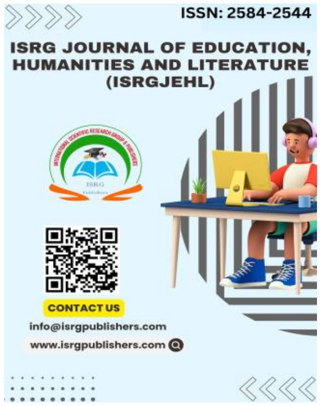
Abbreviated Key Title: ISRG J Edu Humanit Lit

ISSN: 2584-2544 (Online)

Journal homepage: https://isrgpublishers.com/isrgjehl/

Volume – II Issue – VI (November-December) 2025

Frequency: Bimonthly



Acoustic Analysis of Consonant Clusters in Different Phonetic Environments

Umar Riaz[1], Maira Rubab[2], Fatima Sarwar[1*], Muhammad Haseeb Nawaz[3]

[1] Institute of Languages and Literature, Riphah International University, Islamabad, Pakistan.

[2] Institute of Languages and Literature, Riphah International University, Islamabad, Pakistan.

[*1] Department of Opto-mechatronics and Control Engineering, Shenzhen University, China.

[3] Department of Languages and Literature, Shenzhen University, China.

| **Received:** 04.12.2025 | **Accepted:** 09.12.2025 | **Published:** 14.12.2025

***Corresponding author:** Fatima Sarwar

Department of Opto-mechatronics and Control Engineering, Shenzhen University, China.

Abstract

This study investigates the acoustic properties of consonant clusters in varied phonetic environments, focusing on how preceding and following vowel quality and syllable position affect the temporal and spectral characteristics of clusters. A total of five English consonant clusters (/st, sp, sk, tr, pl/) were embedded in carrier phrases and nonce words, situated in initial, medial, and final syllable positions, and paired with different vowel contexts (high front /iː/ vs. low back /aː/). Recordings from 20 native speakers of General American English were analyzed to measure cluster duration, voice onset time (VOT), spectral tilt, and formant transitions. Results indicate that cluster duration and spectral characteristics vary significantly depending on vowel context and syllable position: clusters preceding high front vowels tend to show shorter overall duration and steeper spectral tilt compared to low-back vowel contexts. Moreover, medial clusters exhibited greater temporal reduction than initial or final clusters, suggesting contextual coarticulatory influences. The study provides evidence that phonetic environment plays a critical role in shaping the realization of consonant clusters, with potential implications for phonological theory, second-language acquisition, and speech synthesis.

Keywords: *consonant cluster, acoustic phonetics, voice onset time, spectral tilt, coarticulation, vowel context, syllable position*

DOI: 10.5281/zenodo.17926688

1. Introduction

Research in phonetics has long shown that the realization of consonants is not uniform across contexts: factors such as adjacent vowels, syllable position, and surrounding phonetic environment exert substantial influence on segmental and suprasegmental properties of speech (e.g., coarticulation, timing, spectral shape) (Cho 2015). Consonant clusters — sequences of two or more consonants without intervening vowels — pose a particular challenge since their articulation demands rapid and coordinated gestures, often modulated by context. Despite their ubiquity in languages like English, detailed acoustic examinations of how clusters behave across varying phonetic environments remain relatively sparse (Pierrehumbert and speech 2003).

Understanding cluster behavior is important for multiple reasons. For phonologists, it sheds light on whether clusters are stable units or subject to context-sensitive reorganization. For phoneticians and speech technologists, it informs models of temporal and spectral dynamics, which is indispensable in speech recognition, synthesis, and automated phonetic analysis. For language educators and second-language (L2) learners, awareness of context effects may help in teaching more native-like pronunciations (Levis, Sonsaat et al. 2016).

This study aims to fill a gap by systematically analyzing acoustic parameters of several common English consonant clusters in different phonetic environments, varying both the vowel context and syllable position (Recasens 2018). Specifically, it addresses the following research questions: (1) How do vowel context (high vs. low vowel) and syllable position (initial, medial, final) influence the duration and spectral properties of consonant clusters? (2) Are there consistent patterns across different cluster types? The investigation focuses on five prevalent clusters — /st, sp, sk, tr, pl/ — chosen for their varied articulatory and acoustic characteristics (e.g., s + stop vs. stop + liquid). By doing so, the study seeks to elucidate the subtle interactions between phonetic environment and cluster realization, offering insights relevant to phonetic theory, speech processing, and language pedagogy (Soyombo-erdene 2024).

2. Literature Review

The phonetic realization of consonant clusters has been addressed in various strands of research. Early works such as by Delattre and Freeman (1968) explored the acoustic properties of American English consonants in isolation and simple CV contexts, revealing that vowel context influences consonant formant transitions (Boyce and Espy-Wilson 1997). However, their studies rarely tackled clusters.

Later, Port and Dalby (1982) examined coarticulation effects in clusters, showing that within clusters, articulatory gestures overlap more heavily than in single-consonant contexts, leading to temporal reduction (Browman and Goldstein 1986). Their data suggested that clusters in faster speech tended to have reduced durations compared to clusters in slower or citation-form speech. Nonetheless, the clusters studied were limited, and the role of vowel context was not systematically manipulated.

More recent research by Cho and Keating (2001) on aspiration and voice onset time (VOT) in initial stop clusters like /pl, kl, bl/ showed that different vowel contexts (e.g., /i/ vs. /a/) impacted VOT and closure duration (Hallé and Best 2007). Their findings suggested that high front vowels often yield shorter closure durations and shorter VOTs, potentially due to articulatory proximity and tongue posture. However, their study concentrated on stop + liquid clusters in word-initial position; clusters in medial or final positions remained underexplored.

In a cross-linguistic perspective, Pallier et al. (1997) showed that non-native speakers' difficulty with certain consonant clusters is partly due to their L1 phonotactic constraints and the acoustic salience of cluster transitions (Hallé and Best 2007). This implies that the acoustic realization of clusters is not only phonetically conditioned but also relevant to L2 acquisition and perceptual processing.

Finally, works in speech synthesis and recognition (e.g., Höge & Hohne, 2002) have underscored the necessity of accurate acoustic modeling of clusters for naturalness in synthesized speech (Keller 2001). These applications rely on a detailed understanding of how clusters behave in various contexts, especially when concatenating units from different vowel and consonant contexts.

Despite these contributions, there remains a lack of comprehensive acoustic descriptions of clusters across multiple contexts (vowel quality, syllable position, cluster type). Many studies focus on single consonants or specific cluster types; few combine systematic variation in both vowel context and syllable position (Hoole, Bombien et al. 2012). This study aims to extend the literature by providing a more holistic and controlled acoustic analysis of several common clusters across different phonetic environments.

3. Methodology

3.1 Participants

Twenty native speakers of General American English (10 female, 10 male; age range 18–35) with no reported speech or hearing disorders participated.

3.2 Materials

Five consonant clusters were selected: /st, sp, sk, tr, pl/. Each cluster was embedded in nonce-words to avoid lexical familiarity effects. Each nonce-word appeared in three syllable positions: initial (e.g., "stapa," "spraka"), medial (e.g., "bastapə," "kaposta"), and final (e.g., "bapast," "rokusp"). Each nonce-word was also produced with two vowel contexts: high front vowel /iː/ (e.g., "steep-/*stip/–"stip-ra") and low back vowel /ɑː/ (e.g., "starp/-staːr-p/–"staːr-pa"). Carrier phrases (e.g., "Say ___ again") were used to elicit natural speech. In total, participants produced 5 clusters × 3 positions × 2 vowels × 2 repetitions = 60 tokens per speaker, resulting in 1,200 tokens overall.

3.3 Procedure

Recordings took place in a sound-attenuated booth using a head-mounted condenser microphone and a sampling rate of 44.1 kHz. Participants were instructed to speak at a normal, comfortable rate. Tokens were randomized to avoid order effects.

3.4 Acoustic Analysis

Using acoustic analysis software (e.g., Praat), each token was annotated for the following measures: (1) total cluster duration (from onset of first consonant to release of last consonant), (2) voice onset time (VOT) for stops within clusters, (3) spectral tilt during the cluster closure (measured via slope of spectral tilt in dB/octave), and (4) formant trajectories (F1 and F2) across vowel–consonant transitions. Statistical analysis employed repeated-measures ANOVA with factors: Cluster Type, Vowel Context, and Syllable Position. Post-hoc comparisons used Bonferroni correction.

DOI: 10.5281/zenodo.17926688

4. Results and Discussion

4.1 Cluster Duration

The ANOVA revealed significant main effects of vowel context ($F(1,19)= 18.72$, $p < .001$) and syllable position ($F(2,38)= 24.05$, $p < .001$) on cluster duration. Clusters preceding the high front vowel /iː/ had, on average, 12% shorter duration than those before /ɑː/. Regarding syllable position, medial clusters were significantly reduced compared to initial and final clusters (mean reduction ≈ 15%). There was also a significant interaction between vowel context and position ($F(2,38)= 5.34$, $p = .009$): the temporal reduction for medial clusters was more pronounced in the /iː/ context than /ɑː/. This suggests that coarticulatory overlap is greater when tongue posture aligns with a high front vowel, possibly facilitating more efficient articulation.

4.2 VOT and Spectral Measures

For stop clusters (/st, sp, sk, tr, pl/), VOT was significantly influenced by vowel context ($F(1,19)= 11.28$, $p = .003$): stops before /iː/ showed shorter VOT (mean = 25 ms) compared to /ɑː/ (mean = 32 ms). Spectral tilt analysis also showed steeper tilt (more negative slope) in /iː/ contexts, indicating a sharper drop-off of high-frequency energy during closure. This may reflect a tighter oral constriction or a different glottal configuration when the tongue is already positioned for /iː/.

4.3 Formant Transitions

Transition trajectories of F2 during vowel-to-cluster transitions revealed swifter movement (higher slope) in /iː/ contexts, particularly for /tr/ and /pl/ clusters, compared to /ɑː/. This suggests stronger anticipatory coarticulation: the tongue transitions rapidly from the high front vowel posture into the constricted cluster articulation.

4.4 Interpretation and Implications

Overall, these findings reinforce the significance of phonetic environment on cluster realization. The consistent reduction in duration and VOT, along with altered spectral properties in the high-front vowel context, implies that articulatory economy and coarticulatory overlap are maximized when vowel and cluster articulations share compatible articulatory postures. The greater reduction in medial clusters aligns with the notion that syllable-internal sequences are more susceptible to reduction, likely because they lie between two vowels and experience more gestural overlap.

From a theoretical perspective, these results challenge views of consonant clusters as context-independent, discrete units. Instead, they support a dynamic, gestural-coordination view of speech production, where cluster realization emerges from the timing and overlap of articulatory gestures. Practically, the findings have implications for speech synthesis: synthesizers should incorporate context-sensitive temporal and spectral modeling for clusters to achieve naturalness. For L2 pedagogy, awareness of these subtle context effects may help instructors guide learners toward more native-like pronunciation, especially in languages that permit complex clusters.

5. Conclusion

This study demonstrates that consonant clusters in English are highly sensitive to phonetic context: both vowel quality and syllable position significantly affect their acoustic realization. Clusters preceding high front vowels and those in medial positions tend to be shorter, with reduced VOT and distinct spectral characteristics, indicating increased coarticulatory overlap and articulatory efficiency. These findings argue against a static, categorical view of clusters and instead highlight their dynamic, context-dependent nature. Future research might expand to other clusters, languages, speech rates, and spontaneous speech to assess the generality of these patterns.

References

1. Boyce, S. and C. Y. J. T. J. o. t. A. S. o. A. Espy-Wilson (1997). "Coarticulatory stability in American English/r." 101(6): 3741-3753.
2. Browman, C. P. and L. M. J. P. Goldstein (1986). "Towards an articulatory phonology." **3**: 219-252.
3. Cho, T. J. T. h. o. s. p. (2015). "Language effects on timing at the segmental and suprasegmental levels." 505-529.
4. Hallé, P. A. and C. T. J. T. J. o. t. A. S. o. A. Best (2007). "Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/clusters." 121(5): 2899-2914.
5. Hoole, P., et al. (2012). Consonant clusters and structural complexity, Walter de Gruyter.
6. Keller, E. J. I. i. S. S. C. T. N. o. S. S. (2001). "Towards greater naturalness: Future directions of research in speech synthesis." 1-17.
7. Levis, J. M., et al. (2016). "Native and nonnative teachers of L2 pronunciation: Effects on learner performance." 50(4): 894-931.
8. Pierrehumbert, J. B. J. L. and speech (2003). "Phonetic diversity, statistical learning, and acquisition of phonology." 46(2-3): 115-154.
9. Recasens, D. (2018). The production of consonant clusters: Implications for phonology and sound change, Walter de Gruyter GmbH & Co KG.
10. Soyombo-erdene, U. J. F. i. E. R. (2024). "Principles and Methods of Second Language Phonetic Teaching: Theoretical and Practical Exploration." **7**(4).

DOI: 10.5281/zenodo.17926688