



RESEARCH

Exploring Onset Cluster Preferability

Jonathan Bauman¹

¹ Dallas International University

Keywords: Sonority, Ease of Articulation, Phonetic Difference, Phonotactics, Consonant Cluster

<https://doi.org/10.64830/001c.144564>

Journal of Language, Culture, and Religion

Vol. 6, Issue 1, 2025

This paper critically examines two metrics for measuring phonetic similarity: Net Auditory Distance (NAD) (Dziubalska-Kořaczyk 2014) and the Phonetic Similarity Metric (PSM) (Mielke 2012). Taking insights from these, a new metric for measuring articulatory ease is proposed: the Articulatory Gradience Scale (AGS). The AGS provides a metric to consistently measure the articulatory ease of consonant clusters. Taking into account a description of vocal tract anatomy, changes in airflow/voicing, etc., and real-time sound production overlap, numerical values representing the articulatory ease of consonant clusters are calculated. The results and predictions of the NAD, PSM, and AGS are compared; a need is demonstrated to consider not only the presence of phonetic difference, but the type and degree of that difference when determining underlying motivations for phonotactics in a given language. It is further shown that the resulting AGS values for American English onset clusters show a degree of overlap with predictions made by the Sonority Sequencing Principle.

To view the tables referenced in this article, please download [this file](#).

1. Introduction

The purpose of this paper is to discuss the relationship between phonetic differences and anatomy in American English consonant clusters. More specifically, can we measure these considerations in such a way that can be used to predict cluster preferability? Three different approaches to measuring differences in speech sounds are examined: NAD, phonetic similarity, and the anatomy of the vocal apparatus. Based on the results of reviewing these metrics, I argue for the necessity of a more complex composite metric for measuring speech sound combination preferability: the AGS. These three approaches, while different in their specific focus, all provide useful insights when determining preferability, discussed here in relation to a description of American English consonant onset clusters.

Section 2 critically reviews the work of Dziubalska-Kořaczyk (2014). Her approach to phonotactics relies heavily upon an algorithm for calculating NAD, and is based on both place and manner of articulation of consonants in a cluster. Section 3 discusses phonetic similarity, which is approached by

looking at the similarity metric of Mielke (2012). His approach measures the articulatory and acoustic natures of 58 cross-linguistically common phones. Forty-one of these are consonants, 25 of which are phonemes in English. In Section 4, anatomical realities are considered by describing the basic articulatory movements required to produce a series of consonants. Here, possible differences in airflow, voicing, and place of articulation are considered, as well as possible ways to confirm differences in articulation across consonant clusters. I also introduce a new metric-the Articulatory Gradience Scale- for measuring articulatory ease, based on the principles of gestural overlap and phonetic similarity. This is then discussed in regard to its predictions about or correlations with cluster preferability. Section 5 summarizes the three approaches discussed. Finally, Section 6 contains concluding remarks and possible directions for future research.

For the purposes of this paper, only data on English onset consonant clusters are considered. Vowels are not considered, as they have no direct bearing upon the articulation of the consonant clusters. Data on other languages considered in the studies discussed will be ignored entirely, as they are beyond the scope of this paper.

2. Net Auditory Distance

Dziubalska-Kołaczyk's (2014) work on Net Auditory Distance considers the phonemic inventories of both English and Polish. Phonemes are ranked on two dimensions: place of articulation and manner of articulation. Phonemes are assigned a value from 0.0 to 5.0 on both dimensions. The table of English phonemes and their values taken from page 5 of Dziubalska-Kołaczyk's paper¹ is given in Table 1 (in the supplementary file).

As is shown by the table, consonant phones are categorized by manner of articulation in such a way that lines up very closely with many commonly posited versions of the sonority hierarchy. For example, it is very similar to the scale posited by Parker (2008), as additional divisions are added in the form of half-point values for laterals vs. rhotics and affricates vs. fricatives and stops (though Dziubalska-Kołaczyk does assign a flat value of 0 to all vowels).

Her categorization of phones based on place of articulation is quite a bit more ad hoc, however, and there are three primary arguments to be levied against it. First of all, it could be argued that round vowels such as /o/, /u/, and the like do, in fact, have a labial node, like the labial consonants. Thus, there exists little to no basis for separating them entirely from labial consonants such as /b/, /m/, etc. Another problem arises in the ranking of glottal consonants. The patterning of this natural class is dubious, with them patterning in some ways more with sonorants and in others more with obstruents (Miller 2012,

¹ Dziubalska-Kołaczyk includes non-phonemic [ʔ]. Her reasons for this are not explained, and the phone is not considered here.

275). In the former case, one may expect these to be closer to vowels in terms of production, yet they are assigned a value of 5.0. Another argument arises when considering the fact that /w/ is categorized as both a velar and a labial. While this is phonetically accurate, it complicates the process by which NAD values are calculated for this sound. In the algorithm's current form, it is necessary to calculate NAD between other phones in clusters based on both the labial and velar places of articulation. Unfortunately, this interferes with the intended standardized nature of the NAD. Dziubalska-Kołaczyk does not present a solution to this problem.

If we look at all possible consonant clusters in American English, we can calculate NAD between the segments in those clusters. This should, in theory, yield results that reflect the preferability of the clusters in question. Dziubalska-Kołaczyk provides the following algorithm for calculating a bi-segmental cluster's NAD value:

$$\text{NAD}_{CC} = |(\text{MOA1} - \text{MOA2})| + |(\text{POA1} - \text{POA2})|$$

(Dziubalska-Kołaczyk 2014, 4)

This paper takes into account the NAD values of both bi-segmental and tri-segmental clusters; the former was calculated using the above NAD algorithm, while the latter were calculated by taking the resulting values of all adjacent segments in the cluster and averaging them (Dziubalska-Kołaczyk does not provide a simple means of calculating tri-segmental cluster NAD values without taking the following vowel into account). All values were rounded to two decimal places. The results are collated in Table 2, from lowest NAD values (least different, and thus less marked) to highest (most different, and thus more marked). Several unattested clusters are also included for comparison.

- (1): These values were calculated based on /w/ being a labial.
- (2): These values were calculated based on /w/ being a velar.

Several of the clusters included here occur only as a result of variation that facilitates ease of articulation, such as the variation between [st₁] and [st̪₁]. It is interesting to note that, in this case, the NAD values of both the underlying form and the surface representation are exactly the same. This contrasts with the pairs [s₁] vs. [ʃ₁], [t₁] vs. [t̪₁], and [d₁] vs. [d̪₁], where the non-affricated forms display much greater NAD values (more auditory difference) than the occurring ones (possibly reflecting articulatory ease).

How, then, does the NAD approach relate to sonority? Dziubalska-Kołaczyk's claim is that NAD is actually a more specific predictor of phonotactic preferability than sonority sequencing (2014, 1). One claim she makes is that syllable-initial /p₁/ and /t₁/ are preferred over /ps/ or /ɪt/. This seems to hold true for /ps/, which comes out to 2.30. However, /ɪt/ has the NAD value of 3.30 when considered as a stand-alone consonant cluster

(not taking vowels into account). She also states that initial /p₁/ is better than /t₁/. However, if we take into account the fact that /t₁/ is often realized phonetically as [t̪₁], we see that [t̪₁] is actually far more preferable than /p₁/ (based on the NAD). While the manner of articulation scale follows the sonority hierarchy quite nicely, the place of articulation is a factor that is generally not considered for this purpose. It is also clear that voicing differences should at least be considered, as it has been claimed that voicing plays a role in sonority (see the sonority hierarchy in Parker [2008, 60]). Thus, it is at least the case that before we can confidently say NAD is a better predictor of cluster preferability than sonority, more factors need to be considered.

3. Phonetic Similarity

Mielke (2012) posits a measure of phonetic similarity, in part to justify the groupings of natural classes on the world's languages. His approach is based on laboratory measurements of oral airflow (milliliters/second), nasal airflow (milliliters/second), EGG signal intensity (decibels), and larynx height (relative, based on the position at the beginning and end of articulation). The study takes into account 58 cross-linguistically common phones. Of these, 25 are found in American English, and these segments' phonetic measurements are the focus of the current section.

In Mielke's procedure, measurements were taken for each segment from four individuals, all of whom were native speakers of American English. Thus, while the sample is obviously small, it still represents a good source of data for the current paper. Individual speaker measurements were taken for each segment and averaged by Mielke. The averages for American English consonants are presented in Table 3.

Using these values, we can calculate the differences between consonants in clusters. In Table 4, differences in bi-segmental clusters are calculated by finding the difference between each individual measurement and averaging those values together. For tri-segmental clusters, the difference averages of C1,C2 and C2,C3 are averaged together. This allows only adjacent segments to be taken into account. If similarity alone is equal to preferability, clusters displaying lower values in the Difference Average column are more similar and thus should theoretically be more preferred².

Several patterns are observed in these results. First, note that according to this application of Mielke's raw phonetic measurements, s+stop clusters are actually the most preferred types among all the bi-segmental clusters, with s+labial being most preferred among these. This preference declines the further back in the oral cavity the stop is produced. Among the C+glide

2 The Difference AVG values were calculated by the author of the present paper to compare the phonetic similarity values across clusters.

combinations, [dw] is most preferred, followed by [kw], [tw], and finally [sw]. If we also calculate the phonetic difference for the cluster [gw] (a marginal cluster in American English), we get the following results:

Figure 1. Phonetic preference patterns in American English consonant clusters.

Cluster	Nasal Airflow	Oral Airflow	EGG Intensity	Larynx Height	Difference AVG
g w	0.01	0.02	4	0.08	1.0275

This places [gw] between [sk] and [dw]. This is especially telling when one takes the following points into consideration:

1. Both [g] and [w] have a dorsal node and share voicing.
2. Both [d] and [w] share voicing, but differ in POA, as [d] lacks a velar node.
3. Both [k] and [w] share a velar node, but differ in voicing.
4. [t] and [w] differ in place and voicing.
5. [s] and [w] also differ in both POA and voicing, and [s] is a sibilant.

Thus, among the C+glide combinations, the following preferences seem to emerge³.

1. It is better for stop+glide clusters to have the same POA than the same voicing.
2. It is better for stop+glide clusters to have the same voicing than differ in voicing.
3. Stop+glide clusters are better than sibilant+glide clusters.

Another pattern for consideration is that of C+liquid clusters, where the pattern [gl]>[bl]>[kl]>[pl]>[sl]>[fl] is suggested. Unfortunately, EGG and larynx height measurements for [ɹ] seem to be absent from Mielke's study, making it impossible to compare with C+l clusters. This should be followed up in future research. Also note that the current analysis seems to least prefer s+nasal clusters, although, somewhat strangely given their sonority profiles, [fl]>[sm] is suggested. Finally, as interesting as these patterns are on paper, it must be determined whether the typological facts and frequency data back these preferences up before any conclusions are drawn. In the next section,

³ While it is beyond the scope of this paper to provide an optimality theory explanation to these preferences, they may lend themselves to such an analysis.

we will explore possible anatomical factors affecting articulation, then tie all three approaches together, comparing these results to traditional notions of sonority.

4. Articulatory Ease and Vocal Tract Anatomy: An Issue of Gradience

In addition to the NAD and phonetic similarity approaches, one must consider that certain sequences of sounds are less common simply because they are more difficult to produce in succession. For instance, take some examples of sonority reversals in American English. The sequence /plop/ is possible, while */lpop/ is not. While the bearing sonority has on constraining syllable structure and possible onset/coda clusters has been discussed exhaustively, these discussions do little to explain why sonority works in the way it does. In this section, possible anatomical realities are taken into account to partly explain this.

In the /pl/ vs. /lp/ example above, we can begin by laying out what happens in the vocal tract in each sequence:

Articulatory Sequence of /pl/:

1. The lips come together, blocking the oral cavity.
2. Air pressure builds up in the oral cavity.
3. Air pressure is released at the lips, resulting in an explosion of air. In American English, this is generally accompanied by aspiration. While aspiration occurs, the tongue moves quickly into position at the alveolar ridge.
4. Airflow continues uninterrupted around the tongue, producing a voiceless lateral.
5. Voicing begins, fully producing the /l/.

This contrasts with the articulatory sequence of /lp/:

1. The tongue moves into position at the alveolar ridge.
2. Airflow and voicing begin.
3. The lips come together, blocking the current flow of air producing the lateral and voicing; air pressure immediately begins to build in the oral cavity.
4. Air pressure is released at the lips, producing the aspirated plosive.

In the /lp/ sequence, if the /l/ is held too long, it will be in danger of being perceived as syllabic. Yet if the /l/ is not held long enough, it will not be perceived at all. Additionally, the closure of the lips forces both the lateral and voicing to stop abruptly without possibility of transition. In contrast, /pl/ allows the necessary articulatory changes to occur in gradient stages: *Lips Close & Air Pressure Builds & Air is Released & Airflow Begins & Tongue Moves & Voicing Begins*

To perceptibly produce /lp/, on the other hand, the articulators are forced to make many changes within a short period of time or even simultaneously, turning some features such as voicing on and then off at the same time as oral airflow:

Airflow Begins & Voicing Begins & Lips Close & Airflow Stops & Voicing Stops & Air Pressure Builds & Air is Released

Here, the concept of gradience must be introduced. For our purposes, ‘gradience’ refers to the propensity of a given sequence of phones to allow more gradual or overlapping articulatory transition from one to the next in that sequence. This can be applied to other SSP-obeying and -defying clusters in American English.

Recall that the NAD and phonetic similarity approaches produce different groupings of certain cluster types. For example, Mielke’s approach lumps all s+stop clusters together, giving them a very high preferability value. The NAD, on the other hand, scatters s+stop sequences throughout the table. If the difference between constituent segments alone is indicative of cluster preferability, we have two very different preferability scales: one based on the NAD, and the other based on phonetic similarity. For this reason, we need to design a scale that can take into account the type of difference that exists between adjacent cluster consonants, using the gradience approach above. To accomplish this, I propose the Articulator Gradience Scale (AGS). The AGS evaluates two adjacent segments in a cluster primarily in terms of the following:

1. Number of Articulatory Steps: The more steps required, the less preferable.
2. Gradience: The easier the sounds can overlap in actual articulatory time, the more preferable.

Here, I define the term ‘articulatory step’ in the following way: Any single change or set of changes in the vocal tract, such as moving of articulators (rounding of the lips, moving or curling of the tongue, etc.), turning voicing on/off (vibration of the vocal folds or lack thereof), opening or closing of the velum (nasalization), etc. Much like Optimality Theory constraints, each pair of adjacent consonants receives a mark against it per the number of

articulatory steps required to produce the cluster.⁴ Additionally, clustered sounds receive a mark against them if the articulation of the two sounds does not overlap in real-time (i.e., the transition is not gradient). For example, articulatory overlap occurs in American English /pl/ (phonetically [p^hl]), where the lateral airflow can begin at the same time as the aspiration of the plosive release (i.e., it overlaps).⁵ In contrast, /lp/ does not allow overlap, as the plosive causes a complete stop of airflow. A rough equation for calculating this then looks as follows:

$$\text{AGS} = n \text{ Steps} + 1 \text{ (if No Overlap)}$$

Thus, for the /pl/ vs. /lp/ onset examples, we end up with the following scores:

Figure 2. AGS scores for /pl/ and /lp/ clusters, illustrating effects of articulatory sequencing and overlap

CLUSTER	ARTICULATORY SEQUENCE	OVERLAP?	SCORE
pl	Lips Close & Air Pressure Builds & Air is Released Continuously & Tongue Moves & Voicing Begins	Yes	5
*lp	Tongue Moves & Airflow Begins & Voicing Begins & Lips Close & Airflow Stops & Voicing Stops & Air Pressure Builds & Air is Released	No	9

The /pl/ cluster receives a score of 5 (5 steps), while /lp/ receives a score of 9 (8 steps + no overlap). A more detailed account of the variables for /pl/ is shown below:

/pl/:
 $n \text{ Steps} = 5$:
 Lips Close & Air Pressure Builds & Air is Released
 Continuously & Tongue Moves & Voicing Begins
 Overlap: Yes (0)
 Total: $5 + 0 = 5$

Compare this with the AGS value for the cluster /lp/:

/lp/:
 $n \text{ Steps} = 8$:
 Tongue Moves & Airflow Begins & Voicing Begins & Lips Close & Airflow
 Stops & Voicing Stops & Air Pressure Builds & Air is Released
 Overlap: No (1)
 Total: $8 + 1 = 9$

⁴ It would in theory actually be quite easy to convert these into OT constraints.

⁵ As an anonymous reviewer points out, this sequence doesn't phonetically overlap in this way in many other languages (i.e. Spanish) where cluster-initial stops aren't aspirated. This is, admittedly, a complication requiring further investigation. However, the dispreferred nature of the reversed onset cluster /lp/ remains for the same reasons it is dispreferred in English.

Nasal consonants are ambiguous in this area, as they may overlap in continued airflow, but the airflow continues in a different cavity (i.e., airflow stops in the oral cavity, and continues in the nasal cavity in the transition between consonants in the cluster /sn/). Cavity airflow changes are considered here to only partially overlap, with this represented by a half-point mark against the cluster. For tri-segmental clusters, values for each constituent bi-segmental pair is added together. This supports the idea that cluster complexity is correlated with dispreferability. With this in mind, we can compare the scores of allowed and some disallowed American English onset clusters as shown in Table 5.

It is acknowledged here that a number of issues exist with this early version of the AGS. First, it does not yet account for the actual distance the tongue must move between different articulations; only that it does move. This is especially important to deal with when analyzing s+coronal clusters, as often the only tongue movement required is simply pressing the tongue against the same POA. Second, the preferability of affricate+rhotic combinations is not captured; clearly, similarity in continuancy needs to be accounted for.

Further, it could be debated that simply adding the scores for constituent bi-segmental pairs within tri-segmental clusters gives too much weight to onset complexity when calculating preferability. This also unnecessarily causes several tri-segmental clusters to be judged as less preferable than unattested ones (*/lb/ > /stʃ/). This may be best handled by averaging the scores of the constituent bi-segmental clusters together, in a manner similar to that used in the NAD approach. This allows us to take into account only the transitions between adjacent segments in the cluster (i.e., in /spl/, there is no need to factor in the articulatory ease of transitioning from /s/ to /l/). Averaging the values results in a score of 5.5 for /stʃ/, /st/, /spl/, and /sp/, and a score of 6 for /sk/ and /skw/. This seems, at first glance, to be more in keeping with phonotactic realities.

Moreover, there seems to be a preference in American English cluster patterns to go from the front of the mouth to the back, rather than vice-versa (labialvelar vs. velarlabial), all else being equal. This needs to be accounted for in future models. Finally, it seems to be the case that the more changes which must occur simultaneously in real-time, the less preferable the sound is. However, a reliable way to give these additional weight in the algorithm has not yet been arrived at.

Nonetheless, several interesting patterns begin to emerge. First, in all cases where an ungrammatical sonority reversal was calculated using this scale, the reversal scored a higher (less preferable) score than the SSP-following cluster: /fl/ > /lf/, /sl/ > /ls/, /pl/ > /lp/, /dw/ > /wd/, /pj/ > /jp/, /sm/ > /ms/, etc. This seems to at least minimally suggest a connection between sonority and articulatory ease. Second, there is a significant tendency under this model for continuant+continuant clusters to be more preferable. Of course, further

study needs to be conducted to confirm whether these predictions hold up in general. Regardless, the AGS allows us to measure articulatory ease and test such predictions in a consistent way.

5. Comparison of Approaches

All three approaches are potentially useful for predicting cluster preferability in some sense. Mielke's phonetic difference metric groups consonants based on oral/nasal airflow, larynx height, and intensity. However, this says nothing directly about manner of articulation, which has been shown to be phonologically important in terms of natural class groupings since the early days of linguistics (Jakobson 1949, 208–9). Further, although place of articulation is rarely invoked in discussions on sonority, anatomical realities must be considered when discussing allowable cluster types regardless. In the case of Dziubalska-Kołaczyk's NAD, the place of articulation scale is far too ad hoc to be useful for determining cluster preferability. Unfortunately, this means a potentially useful area of consideration is overlooked; POA concerns have been noted to influence markedness (de Lacy 2006, 35). Further, while both the NAD and the Phonetic Similarity Metric define sound differences, neither considers the order of these sounds, a vital component of the SSP. It is clear that difference alone in either place of articulation or acoustics is not enough to provide a full understanding of the clusters discussed in this paper; this is where the considerations in Section 4 come into play.

6. Conclusion and Further Study

This paper has detailed the pros and cons of two previous approaches to describing phonetic differences, and has introduced the possible merits of a third layer of factors in terms of consonant cluster preferability. It has been suggested here that a more robust metric needs to be synthesized from at least these three main areas of consideration.

In order to do this, several lines of investigation should be further explored. First, a larger data sample, both in terms of the number of speakers and languages represented, needs to be considered. Second, the daunting task of analyzing all the consonant combinations that would be theoretically possible based on inventory, but do not occur in a language, must be accomplished (what do the calculations for /db/ say about its non-occurrence in American English?). Due to space, it is not possible to include all such combinations here. Additionally, future experiments measuring actual articulatory gestures and positions are needed to confirm the purely descriptive sequences in Section 4. Of course, a number of adjustments will need to be made to the AGS scale itself, with both the MOA measurements of Dziubalska-Kołaczyk's NAD and the four phonetic measurements of Mielke's PSM being integrated with the articulatory measurements. The potential value of collecting and compiling such measurements thoroughly is worth this work; while a number of phonetic and phonological databases exist, many of these focus solely on

inventory distribution such as PHOIBLE (Moran and McCloy 1990) or phonotactics, such as the Onset Cluster Project (Bare et al. 2023). Providing a consistent measure of phones relative to phonetic and articulatory differences within a given language could thus contribute much to the understanding of why sounds pattern the way they do.

Published: October 17, 2025 CDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

- Bare, Katherine, Jonathan Bauman, Jack Campbell, Katya Jablonski, Stephen Jay, David Leonardi, Timothy Palmer, et al. 2023. "The Onset Cluster Project." Dallas International University. 2023. <https://www.diu.edu/academics/special-electronic-publications>.
- Dziubalska-Kořaczyk, Katarzyna. 2014. "Explaining Phonotactics Using NAD." *Language Sciences*, 1–12.
- Jakobson, Roman. 1949. "On the Identification of Phonemic Entities." In *Travaux Du Cercle Liguistique Copenhague*, 205–13. The Hague: Mouton.
- Lacy, Paul de, ed. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge Studies in Linguistics 112. New York: Cambridge University Press.
- Mielke, Jeff. 2012. "A Phonetically Based Metric of Sound Similarity." *Lingua* 122:145–63. <https://doi.org/10.1016/j.lingua.2011.04.006>.
- Miller, Brent. 2012. "Sonority and the Larynx." In *The Sonority Controversy*, edited by Steve Parker, 257–88. Boston: DeGruyter.
- Moran, Steven, and Daniel McCloy. 1990. "PHOIBLE." 1990. <https://phoible.org/contributors/UPSID>.
- Parker, Steve. 2008. "Sound Level Protrusions as Physical Correlates of Sonority." *Journal of Phonetics* 36:55–90. <https://doi.org/10.1016/j.wocn.2007.09.003>.

SUPPLEMENTARY MATERIALS

Supplemental File: Exploring Onset Cluster Preferability

Download: <https://jlcrl.diu.edu/article/144564-exploring-onset-cluster-preferability/attachment/302660.pdf>
