In Search of Phonetic Evidence for Prosodically-Motivated Aspiration

McKinley Sprinkle and Anya Hogoboom

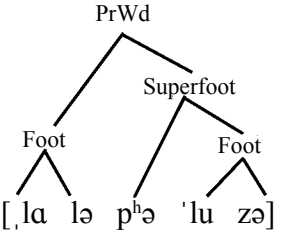
1. Introduction

Aspiration is an acoustic phenomenon primarily associated with the distinction of voiceless stops from voiced stops in English, manifesting as a burst of air before the onset of voicing in voiceless stops. While a single rule-based analysis of aspiration based on stress environment and word position fails to account for where aspiration surfaces in English, tying aspiration to prosodic structure in the form of a foot-based analysis largely resolves this problem. Aspiration has been proposed by Kiparsky (1979), Withgott (1982), and Jensen (2000) to be linked to the left edge of prosodic feet, condensing the disparate environments of aspiration in English (e.g. onset of word-initial and stressed syllables) into a single description based on prosodic structure. Withgott also identifies a further aspirated environment of the onset of the second unstressed syllable before a stressed syllable which is accounted for by this generalization; e.g. the [p] in *lolla[p]alooza* as shown in (1). This forms the basis of the accepted analysis of aspiration in English (e.g. Davis and Cho 2003, Kager and Martínez-Paricio 2018).

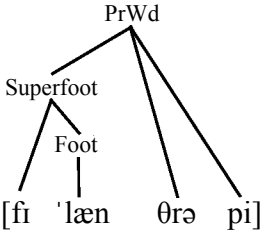
Withgott (1982)'s assertion that the environment of voiceless stops in (1) is aspirated is somewhat surprising given the disjunction in acoustic strength between aspiration and that environment. Aspiration is a phonologically strong feature, and in all other posited environments appears either in a stressed syllable or word-initially, both perceptually salient positions. By contrast, the environment of the [p] in words like *lolla[p]alooza* (1) is quite weak; it is not word-initial, nor stressed, and it is in a stress lapse position; the second consecutive stressless syllable. The contrast between this perceptually weak position and the proposed presence of a phonologically strong feature like aspiration is surprising.

To our knowledge, no mention has been made in the literature to the environment shown in (2): the onset of a word-final syllable in stress lapse. While this environment would not be expected to aspirate given its right-edge position, its status of having a voiceless stop in the onset of a stress lapse syllable makes it a somewhat close parallel to the environment posited as hosting aspiration in (1), and the level of aspiration that surfaces in that position seems perceptually similar, suggesting that an acoustic comparison between these two environments might be fruitful.

(1) Lollapalooza



(2) Philanthropy



The purpose of the current series of studies is twofold. First, it aims to examine the realization of voiceless stops in onsets in the lapse positions and determine whether aspiration does indeed manifest

*McKinley Sprinkle, William & Mary, mtsprinkle@wm.edu. Anya Hogoboom, William & Mary, ahogoboom@wm.edu. Thank you to Kate Harrigan for help with experiment design and Dan Parker, Erin Webster, the W&M Linguistics research group, the audience of VALing 2022, and the WCCFL 40 reviewers and participants for helpful comments and suggestions. We are grateful to Kim Love for statistical consultation and to the Roy Charles Center for their generous funding of this study. We also thank our study participants. A full thesis with statistical analysis on this topic is available here: https://scholarworks.wm.edu/honorstheses/1749/

 Proceedings of the 40th West Coast Conference on Formal Linguistics, ed. Jiayi Lu et al., 295-305. Somerville, MA: Cascadilla Proceedings Project.

in the acoustically weak stress lapse position in (1) but not in (2). Secondly, it aims to record and perceptually compare the phonetic levels of voice onset time (VOT) that appear in all different aspiration environments. Outside of phonological and prosodic theory, little phonetic work has been done on acoustically measuring VOT in different word positions, especially beyond relatively short words (e.g. Kim et al. 2018, which focused on two-syllable words). The presence of aspiration in the third syllable of a five-syllable word (1) is well beyond the environments that have been phonetically examined, especially in comparison to the final lapse environment (2).

To accomplish these aims we carried out one production study and two linked perception studies. The production study aims to record the amount of voice onset time present in every possible aspiration position, as well as every attested environment of voiceless stops in English that do not aspirate, to compare their acoustic manifestations. The two perception studies use recordings produced by participants in the production study as stimuli for a categorical and a gradient listening study, probing the manifestation of a specific aspiration threshold and gauging the phonetic sensitivity of English speakers to changes in VOT at different word positions.

2. Experiment 1: Production

The production experiment is intended to measure the VOT for voiceless stops [p] and [k] in the recorded speech of participants at all possible stress levels and word positions, including positions where aspiration is not expected to surface. Through these measurements, we compare the relative strengths of various factors in influencing the presence and strength of aspiration, and create recordings that are used as stimuli for later perception studies.

While the production study also measures the VOT levels of [t] in these positions, /t/ is expected to surface as allophonic variants (glottal stop and flap) in certain environments, preventing the full analysis of [t] in relation to [p] and [k]. This inability to analyze all possible environments means that it is not as rigorously examined as [p] or [k], and test words with [t] in target positions are not as frequent in the word list used for this study.

2.1. Stimuli

Words were selected for the production study to fit into 14 different conditions; seven where aspiration was hypothesized to occur, and seven where it was not, as shown in (3). These conditions were based on the environments described by Davis and Cho (2003) but were expanded to incorporate variation based on stress level and word position, as well as the environment of word-final lapse not considered in that or, to our knowledge, in previous descriptions of aspiration environments.

Aspiration Expected	Example	Aspiration Not Expected	Example
1. σ́σσ	*[k]onduit*	8. in coda	*a[k]ne*
2. σσ́σ	*a[p]earance*	9. ...(σ́σ)...	*ra[p]id*
3. σ̀σσ	*[k]angaroo*	10. #sT́	*s[k]ydiver*
4. σ̀σσσσ	*[k]omplimentary*	11. #sT̆	*s[p]oradic*
5. σσσ̀	*motor[k]ade*	12. ...sT́...	*es[k]ape*
6. σσ́σ	*[k]abana*	13. ...sT̆...	*ex[p]osition*
7. σσσσ́σ	*lolla[p]alooza*	14. (σ)σσ́σσ	*harmoni[k]a*

2.2. Procedure

There were 25 participants for the production study (14 female, 10 male, 1 non-binary; aged 18-25, mean age=19). Participants were compensated with course research credit or with $5. The purpose of the study was explained and participants were allowed to withdraw at any time. Participants read 23 question and answer pairs.

This study was a self-paced reading task where participants read the 23 question and answer pairs (one fifth of the full set) from slides on a computer. Participants were seated in a sound attenuated booth and wore a Shure WH30 condenser headset microphone recorded into a TASCAM DR-100 portable audio recorder. Participants were instructed to speak naturally, as if talking to a friend, to speak the entire

phrase again if they mispronounced a word, and to read through all slides before beginning recording to ensure that all words were familiar.

Incorrectly pronounced target words were excluded in cases where the incorrect pronunciation occurred in a syllable adjacent to the target sound or if it resulted in an added or dropped syllable, possibly affecting VOT measurement. Cases where a pronunciation was not corrected but the mispronunciation was more than a syllable after the target syllable (such as pronouncing a different vowel) were included in the data. If participants corrected a pronunciation, the corrected pronunciation was always the one measured. VOT durations were measured using the program Praat (Boersma & Weenick 2019).

2.3. Results

The z-scored duration of VOT across all conditions is shown in Figure 1, separated by consonant. The z-scores show the change above or below each consonant's average VOT duration in terms of standard deviation from the mean, normalizing the variation between each consonant and speaker and thereby accounting for the speaker VOT variation. There are several clear groupings across all three consonants; in Conditions 1-5, there is consistently a robust amount of VOT, aligning with the expectation of categorical aspiration. In Conditions 8-13 there is a clearly much lower VOT, aligning with the expectation that these conditions are not categorically aspirated. Three conditions behave somewhat in-between these two larger groupings; Conditions 6, 7, and 14. The VOT levels of the target consonants [p] and [k] for all three ambiguous conditions are markedly below the length of Conditions 1-5 and above the length of Conditions 8-13.

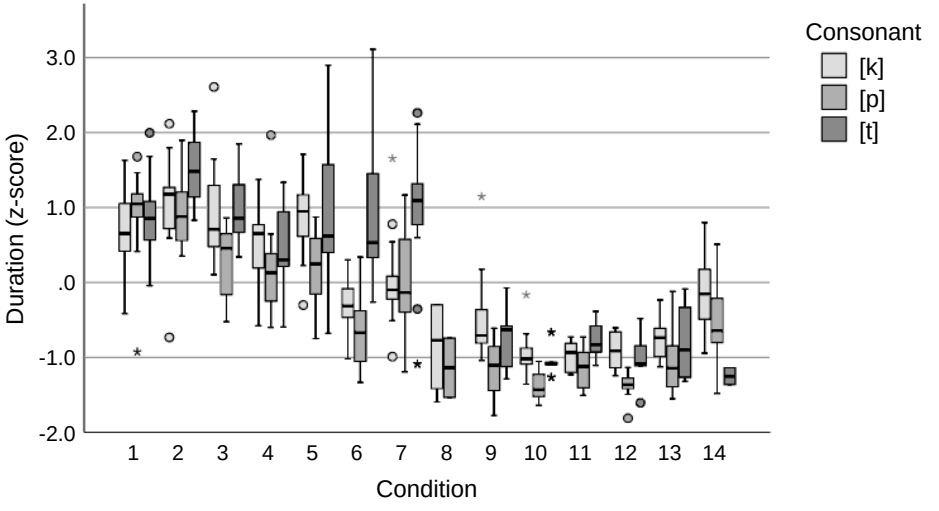


Figure 1: Z-score of voice onset time (VOT) duration by condition for consonants [k], [p], and [t]

The distribution of VOT for [t] in Figure 1 is distinct from that of [p] or [k]. All of the expected aspiration conditions, including 7, behave similarly, are distinct from the unaspirated conditions, and none of the test conditions behave ambiguously. However, [t] test words were intentionally underrepresented in the word list and, as expected, /t/ frequently surfaced as one of its allophonic variants, a voiced alveolar flap [ɾ] or a glottal stop [ʔ]. Even though it behaves differently in some conditions from [k] and [p], because its behavior cannot be analyzed relative to [p] and [k] in Conditions 8, 9, and (crucially) 14, we set aside data from [t] going forward.

Collapsed z-score measurements of VOTs in the recordings of both [p] and [k] pronunciations from the production study are shown in Figure 2. There is a clear group of conditions where aspiration is expected (Conditions 1-5) and a clear group where it is not (8-13), largely aligning with the left-edge aspiration hypothesis. There is, however, a group of outliers from this trend, where the raw duration of VOT for both [p] and [k] falls between the categorically aspirated and unaspirated groups. These outliers consist of the initial stressless condition (*[k]abana,* Condition 6), the medial lapse condition

(*lolla[p]alooza,* 7), and the final lapse condition (*harmoni[k]a,* 14). While Condition 14 is not predicted to be aspirated as a rightward foot adjunct, its comparability with supposed aspirated Conditions 6 (both stressless syllables) and 7 (both stress lapse conditions) suggest it might behave in a similar manner.

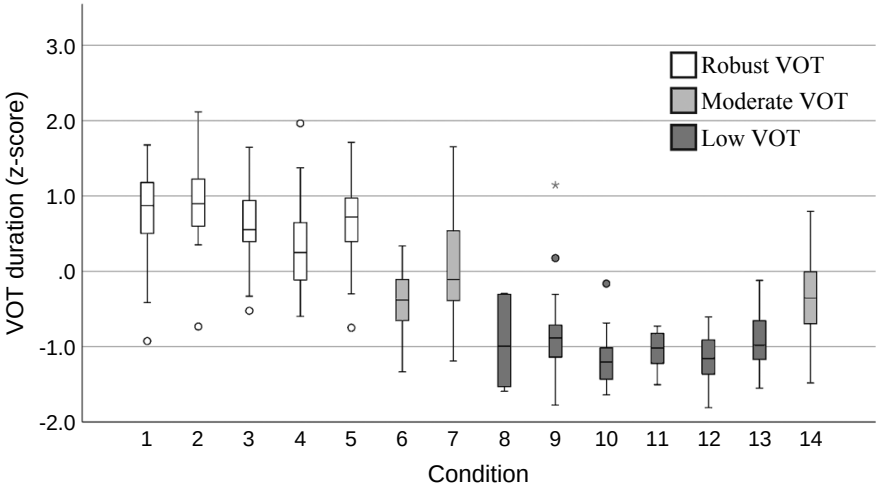


Figure 2: Production study VOT duration by condition, by aspiration group of robust, moderate, or low

The existence of a visual group of conditions for the consonants [p] and [k] in which the presence of aspiration is unclear, in contravention of our understanding of aspiration as fundamentally categorical in English, motivates further exploration into how aspiration surfaces and is perceived in these conditions. While the phonetic realization of aspiration in these conditions is somewhat unclear, approaching them from a perceptual standpoint, i.e. focusing on how they are heard and interpreted, should allow us to disambiguate whether they are functioning more as aspirated or unaspirated syllables.

3. Experiments 2 and 3: Perception studies

In order to disambiguate the aspiration status of the three marginal conditions from the production experiment, Conditions 6, 7, and 14, we deployed two linked perception studies. These studies seek to gauge whether the presence of aspiration in these positions is important for the listener in identifying the nature of the test words. If changes in VOT in these positions are perceptually salient to speakers of English, then aspiration is likely categorically encoded for these positions even if the acoustic realization is somewhat weaker than expected relative to what is recorded in other conditions. However, if changes in VOT are not salient to listeners, then the presence of aspiration in these positions does not significantly impact the identification of words, and it is likely not encoded in these conditions. We approach the question of perceptual salience from two different perspectives in the two perception studies.

3.1. Stimuli components

Stimuli for both perception studies were created from production study recordings. Stimuli consist of a word with the target sound's level of VOT altered to four different levels: 15ms, 35ms, 55ms, and 75ms. These levels allow comparison of participant response at different intervals, both above and below the perceptually important level of 35ms as the threshold between aspirated and unaspirated (selected based on Reetz and Jongman 2009 and Johnson 2012). These four levels allow us to observe a large (40ms) interval both across the 35ms threshold (by comparing 15ms and 55ms) and without crossing it (by comparing 35ms and 75ms), as well as examine small changes both below (15ms and 35ms) and above (55ms and 75ms) the threshold.

For consistency, only one speaker per word list was used as the stimuli for the perception studies. We selected the first subject for each word list, unless they had no data for a study condition (such as a mispronounced word) or if the level of VOT on a study condition was a statistical outlier. Based on the

production results, we selected 9 conditions to be tested in the perception studies: the seven expected aspiration conditions, one comparison unaspirated condition (*ra[p]id*, 9), and the ambiguous final lapse condition (*Ameri[k]a*, 14). Condition 9 is a syllable onset that is not part of a cluster, making the VOT measurable, as well as being the second member of a foot, thus not surfacing as aspirated, and therefore it provides the best comparison case for a categorically unaspirated voiceless stop.

3.2. Categorical Perception

The categorical perception study examines listeners' ability to recognize phonological aspiration in the absence of any external frame of reference or comparison. In this study, participants were presented with a VOT-modified pronunciation of a word and responded with a judgment based on whether they perceived a target syllable as voiced or not.

3.2.1. Procedure

There were 40 participants (31 female, aged 18-23, mean age=19) for the categorical perception study. Participants received course research credit for participating.

Stimuli consisted of the modified VOT tokens described in §3.1. There were two versions of the study, each consisting of 120 stimuli played individually, including two pronunciations of each of the 60 test words. Each version included two VOT levels out of the four possible levels for each word, with VOT differences of 40ms between the two levels present in the same version (eg. 15ms/55ms or 35ms/75ms). Each study version included an equal balance of all four VOT levels within each condition.

Participants heard stimuli and were presented with two buttons representing the standard spelling of the word and a spelling corresponding to a voiced stop at the target position (e.g. *cabana* or *gabana*). The version with unconventional spelling was presented on the left so that participants reading left to right encountered the altered spelling first to maximize their awareness of the alternative voicing. Participants were told that the stimuli included some pronunciations that had been altered to match the unconventional spelling on the screen. Participants were asked to click on the spelling that matched what they heard.

3.2.2. Results

The identification of the altered stimuli versions aligned almost entirely with the expected distribution of categorical aspiration. In conditions where strong aspiration was present in the original recordings (Conditions 1-5), listeners were much more sensitive to the level of VOT present, and consistently identified the 15ms level as the word's voiced equivalent (e.g. identifying the 15ms level of *consonant* as *gonsonant*). This also occurred for the initial stressless condition (*[p]otato,* 6), which was ambiguous in the production study, indicating that aspiration is categorically present in this condition despite its less forceful acoustic realization. The other three environments in the study behaved as a group: medial stress lapse (*lolla[p]alooza,* 7), onset of weak member of a foot (*ra[k]quet,* 9), and final tstress lapse (*Ameri[k]a,* 14). All three were identified as being the altered spelling at a much lower rate.

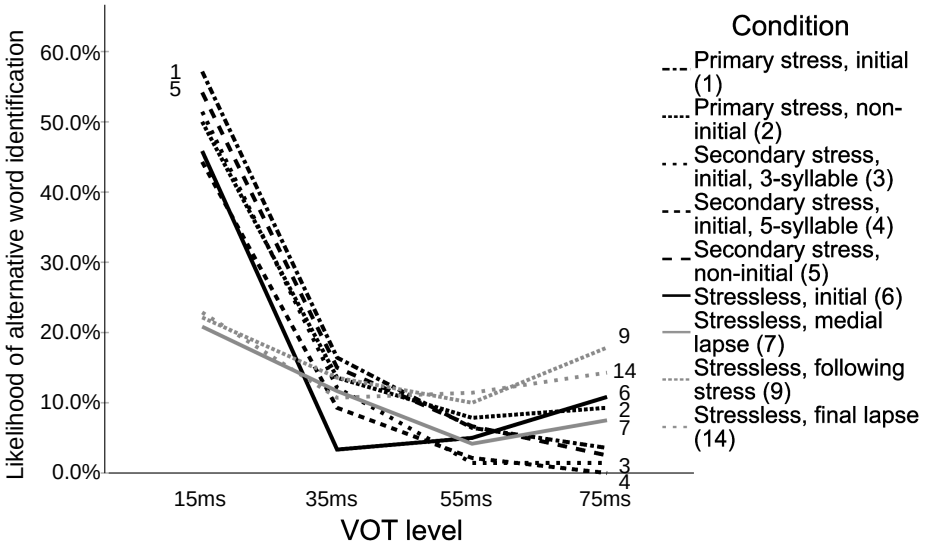


Figure 3: Categorical perception study likelihood of alternative (voiced) word identification by stimuli VOT level

Figure 3 shows the likelihood of a word in each condition being identified as its voiced equivalent over the four VOT lengths. A higher percentage of alternative word identification indicates a more frequent response of the voiced spelling for a given stimulus. Above the categorical aspiration threshold at 35ms VOT and above there are no groupings of condition behavior, indicating that in all conditions words are being perceived as voiceless. There is a clear split in the 15ms group between a categorically aspirated set of conditions (1-6) and a categorically unaspirated set of conditions (7, 9, 14).

This data shows a split in behavior from the three environments that showed moderate levels of VOT for [p] and [k] words in the production study. Condition 6 behaves like other aspirated environments (1-5) while Conditions 7 and 14 behave like Condition 9, the non-aspiration environment included as a control. We see a clear grouping of categorically aspirated syllables (Conditions 1-6) and categorically unaspirated syllables (Conditions 7, 9, and 14), with both stress lapse environments unambiguously perceived as unaspirated and patterning together despite their supposed different positions in relation to foot boundaries.

3.3. Experiment 3: Gradient Perception

The gradient perception study is intended to probe the perception of aspiration through a same/different task, and is focused on participants' sensitivity to VOT change at target locations in stimuli words. The study presents pairs of stimuli with various relationships to the threshold, and seeks to determine the importance of a threshold-based perception of VOT in different environments. This should reveal whether categorical aspiration is a perceptually important cue to listeners in these conditions, regardless of whether it is produced by speakers.

3.3.1. Procedure

There were 60 participants for the gradient perception experiment (39 female, 1 non-binary; aged 18-26, mean age=19). Each participant heard 180 stimuli (half of the full set). Participants received course research credit for participating.

Stimuli consisted of the modified VOT tokens described in §3.1. Pairs consisted of two modified pronunciations of the same word, each word being presented three times: once in a pair with identical VOT lengths (either 15ms/15ms or 55ms/55ms), once with VOT lengths contrasting across the 35ms threshold (either 15ms/35ms or 15ms/55ms), and once with VOT lengths contrasting without crossing the 35ms threshold (either 35ms/55ms or 35ms/75ms). The ordering of all three pairs were reversed in equal numbers, so that the longer VOT stimuli appeared first and second in equal numbers. Each study

version included a balance of identical pair VOT levels and of contrastive VOT orderings. Stimuli were separated by a silent 100ms inter-stimulus interval.

Stimuli pairings were broken down by type, where Type 1 and 2 pairings were identical VOT pairs, Type 1 at 15ms and Type 2 at 55ms, Type 3 and 4 pairings were VOT distinctions crossing the aspiration threshold, Type 3 with a 20ms difference and Type 4 with a 40ms difference, and Type 5 and 6 pairings were VOT distinctions not crossing the threshold, Type 5 with a 20ms difference and Type 6 with a 40ms difference. These Types are described in Table 1.

Type	VOT level distinction	Description	Type	VOT level distinction	Description
1	15ms/15ms	Same VOT level below the typical aspiration threshold	4	15ms/55ms 55ms/15ms	40ms VOT difference crossing the aspiration threshold
2	55ms/55ms	Same VOT level above the typical aspiration threshold	5	35ms/55ms 55ms/35ms	20ms VOT difference not crossing the aspiration threshold
3	15ms/35ms 35ms/15ms	20ms VOT difference crossing the aspiration threshold	6	35ms/75ms 75ms/35ms	40ms VOT difference not crossing the aspiration threshold

Table 1: Gradient perception study stimuli pairing types

Participants should consistently identify Types 4 and 6, which have 40ms acoustic differences, as different more frequently than they identify Types 3 and 5, which have 20ms acoustic differences, as different. In conditions where aspiration is an important cue to word identity, differences in VOT that cross the threshold (Types 3 and 4) should be identified as different more frequently than their equivalent length distinction (Types 5 and 6) are identified as different. In conditions where aspiration is not an important cue, we do not expect to see significant differences in behavior between the threshold-crossing and non-threshold-crossing stimuli, and only see differences based on raw VOT length differences.

Participants were primed by hearing two sample pairs of stimuli: *[k]onduit* (Condition 1) with a Type 4 distinction (15ms/55ms) and *pro[p]ulsion* (Condition 2) with a Type 6 distinction (35ms/75ms). Participants were played each priming stimuli twice, and were specifically directed to the difference. They were told that differences between the two pronunciations, if present, would manifest as a difference in the strength of a 'p' or 'k' sound somewhere in the word.

The criteria for inclusion was that subjects had to respond "different" more to Type 4 (15ms/55ms) pairings than to Type 2 (55ms/55ms) pairings by a margin of 3 judgements (10% of the 30 stimuli pairs of each type). This excluded 30 additional participants who failed to identify an identical stimulus pairing (Type 2) as being the same reasonably more frequently than the maximally different stimuli pairing (Type 4), which had both a 40ms difference in VOT between stimuli and crossed the aspiration threshold.

3.3.2. Results

As a baseline metric for the effectiveness of the gradient perception task, we first compare the responses to 40ms VOT difference types (Types 4 and 6) with 20ms difference types (Types 3 and 5). If participants are reacting to the VOT level changes between the two stimuli, then the 20ms types should both have a higher "same" response rate than the 40ms types. In both pairings relative to the threshold, participants identified stimulus pairs with a 40ms difference as being the same less frequently than pairs with a 20ms difference; 48% same for Type 4 relative to 65% same for Type 3, and 56% same for Type 6 relative to 69% same for Type 5.

A binary logistic Generalized Linear Mixed Model (GLMM) was run for dependent variable (DV) *response* and factors *condition*, *type*, and their interaction, with individual intercepts fitted for the random effects of *subject* and *word* (Table 2). We see that the interaction of condition and type is significant.

Source	F	*df1*	*df2*	Sig.
Corrected model	6.493	47	326	p<0.001
Condition	.919	7	37	p=0.503
Type	45.276	5	460	p<0.001
Condition * Type	2.722	35	810	p<0.001

Table 2: Fixed effects chart for gradient perception GLMM

Condition	Type 3 and 5 (20ms) significantly different?	Type 4 and 6 (40ms) significantly different?
1	No (p=0.396)	Yes (p=0.002)
2	No (p=1.000)	Yes (p=0.008)
3	No (p=0.326)	Yes (p<0.001)
4	No (p=0.170)	Yes (p=0.032)
5	Yes (p=0.018)	Yes (p=0.014)
6	No (p=0.208)	No (p=0.742)
7	No (p=1.000)	Yes (p=0.012)
9	No (p=0.292)	No (p=0.072)
14	No (p=0.154)	No (p=1.000)

Table 3: Comparison of significance between Type 3/5 and Type 4/6 by condition. All p-values are multiplied by 2 for the Bonferroni correction.

Based on this analysis, there is a significant difference ($p<0.001$) between the behavior of both 40ms VOT difference types and both 20ms VOT difference types, indicating that participants are significantly more likely to respond "different" to a larger VOT change than a smaller change. With evidence that participants are able to consistently identify 40ms changes in VOT as "different" more than 20ms changes in VOT, we now compare how participants respond to in cases where the raw VOT is the same, within and across the attested threshold for aspiration.

We are particularly interested in the comparison of Type 3 to Type 5, as both have 20ms differences within the pair, but Type 3 is across the aspiration threshold (of ~30ms) while Type 5 is not. Looking at the pairwise comparisons for *condition*type*, only Condition 5 showed a statistically significant difference between Types 3 and 5. The second comparison of interest is between that of Type 4 to Type 6, which both have a 40ms difference, with the former being across the aspiration threshold and the latter not. Again looking within each condition, we see that all the known aspiration environments show a statistically significant difference. Condition 9, however, does not. Our test cases are split, where Condition 7 behaves as the aspirated ones do, and Condition 14 does not. Both Type 3 to 5 and Type 4 to 6 significance comparisons are shown in Table 3.

The identification of stimuli pairs as "different" when crossing the aspiration threshold largely aligned with the expected aspiration status of each condition. When asked to differentiate between a pair of stimuli with a difference of 40ms whose VOT levels crossed the threshold, participants consistently responded that pairs were different in Conditions 1, 2, 3, 4, 5, and 7, indicating that categorical aspiration is an important cue in these conditions. Participants were not better at noticing a distinction when stimuli pairs crossed the categorical aspiration threshold in Condition 9, the control, providing an example of what a categorical lack of aspiration looks like perceptually. Participants were also unable to mark a difference between threshold-crossing and non-threshold-crossing stimuli pairs in Conditions 6 and 14, suggesting that neither are categorically aspirated. For Condition 6, this is in opposition to the behavior observed in the categorical perception study.

A comparison of participants' likelihood of returning a "same" response for each of the conditions in the two 40ms-difference stimuli pairings (Type 4, 15ms/55ms; and Type 6, 35ms/75ms) is shown in Figure 4 below. Conditions that are categorically aspirated should see the 15ms/55ms stimuli pairing being identified as "same" markedly less than the 35ms/75ms stimuli pairing. Conditions that are not categorically aspirated should see no clear distinction between the two stimuli pairing types, as aspiration is not an important acoustic cue in these environments. Distinctions are seen in the difference between the two types rather than in the numerical proportion of "same" responses.

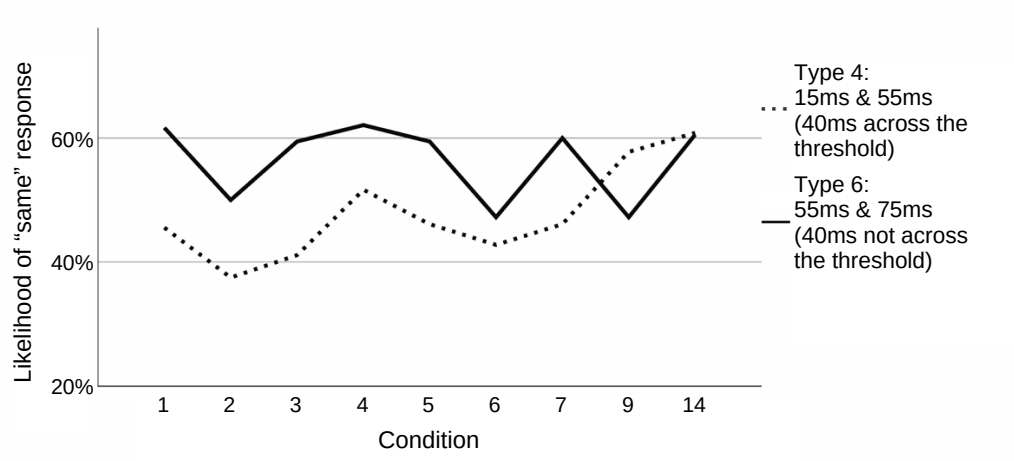


Figure 4: Gradient perception study likelihood of a "same" response by condition for 40ms-difference stimuli pairs

In Figure 4, responses for Types 4 and 6 in Conditions 1-5 and 7 are all distinctly different, with Type 4 (40ms across threshold) consistently identified as "different" more than Type 6 (40ms not across threshold), showing that a threshold effect is important to participants in these conditions. This relationship between the behaviors of Type 4 and Type 6 stimuli pairs is not present for Conditions 6, 9, and 14, suggesting that there is no important threshold effect in these conditions. For Condition 6, responses to Type 4 and Type 6 pairings have the same relation to each other as the categorically aspirated conditions, with Type 4 being identified as "same" more frequently than Type 6, but their difference from each other does not reach statistical significance.

4. Discussion

We deployed three experiments in this study on aspiration. In the production task, we examined how robustly aspiration is produced in all possible environments, and how the VOT duration in categorically aspirated conditions compares to that of categorically unaspirated conditions. These phonetic measurements resulted in three ambiguous conditions, with levels of VOT intermediate to the otherwise categorical distribution: the initial stressless environment (6, *[p]otato*), word-medial lapse environment (7, *lolla[p]alooza*), and the word-final lapse environment (14, *harmoni[k]a*).

In the categorical perception task, we examined the perception of raw VOT duration change in voiceless stops. Categorically aspirated conditions were overwhelmingly identified as voiced when they occurred with a VOT of 15ms, while categorically unaspirated conditions were generally identified as voiceless regardless of the level of VOT. Condition 6 patterned with other aspirated conditions, and Conditions 7 and 14 both patterned with the unaspirated comparison condition (9, *ra[p]id*).

In the gradient perception task, we examined the sensitivity of listeners to VOT change in different conditions, determining whether the same absolute VOT change would be perceived differently depending on whether it crossed the aspiration threshold or not. Categorically aspirated Conditions 1-5 exhibited a threshold effect, showing a clear perceptual difference between threshold-crossing and non-threshold-crossing stimuli pairs. While Condition 6 exhibited the same general relationship between Types as the other aspirated conditions, this relationship did not reach statistical significance. Condition 7 clearly patterned with aspirated conditions and Condition 14 patterned with unaspirated Condition 9.

Based on the results of these three experiments, there are a number of conclusions that can be drawn about the behavior of the different conditions. For the most part, conditions that were expected to be aspirated behaved as such; Conditions 1-5 were robustly produced and perceived as aspirated in all three experiments. Likewise, most conditions that were not expected to exhibit aspiration behaved distinctly differently from the aspirated conditions; Conditions 8-13 behaved uniformly in the production task, and, through the representative behavior of Condition 9, were always distinct from aspirated conditions in the perception tasks.

Condition 6 behaved ambiguously in the production study, with a lower level of VOT than expected given its presumed aspiration status, but its relationship to other aspirated conditions in the categorical

perception study clarify its status as aspirated. In the gradient perception study Condition 6 again behaved strangely, not exhibiting a threshold effect, however the relationship between the two 40ms VOT difference pairs was the same as the other aspirated conditions, just less robust.

Since the categorical perception study demonstrated that a threshold exists for Condition 6, we must assume that it is aspirated. In the gradient perception study, even in the most phonetically distinct and acoustically prominent aspiration conditions (1 and 2) the mean response hovers at around 40% "same." The distinctions between stimuli are clearly perceptually subtle, and Condition 6 still exhibits the same general behavior as Conditions 1-5 and 7, with Type 4 identified as "different" more than Type 6, though not to a statistically significant level. Given the difficulty of the task, it is possible that more participants are needed in order for Condition 6 to resolve to be the same as other categorically aspirated conditions.

Both Conditions 7 and 14 also behaved ambiguously in the production study, with VOT levels falling between the aspirated and unaspirated groupings. For both conditions this is surprising; Condition 7 is expected to behave as aspirated, and 14 as unaspirated. To clarify their status, we examine their behavior in both perception studies relative to Condition 9, the most structurally related unaspirated condition, and Conditions 1-5, which are clearly aspirated. In the categorical study, neither condition behaves as aspirated; listeners consistently identify both as voiceless at the 15ms level even without robust VOT. Low levels of VOT are tolerated without the consonant being heard as voiced. In contrast, the gradient study shows a significant aspiration threshold for Condition 7, and no such threshold for Condition 14. For Condition 7, participants were much better at distinguishing between two stimuli when the VOT difference crossed the threshold than when it did not, showing a threshold effect akin to that of Conditions 1-5. Condition 14 showed no such threshold effect, in parallel to Condition 9.

Condition 7 can thus be described as categorically aspirated. While in the production study it lacked a robust amount VOT, and in the categorical study listeners tolerated a low level of VOT while still identifying it as voiceless (like Conditions 9 and 14), in the gradient perception study it robustly behaved as a voiceless stop, and exhibited the same significant threshold effect as other aspirated conditions.

This behavior could be explained as being an effect of the weak position in which the aspirated stop in Condition 7 surfaces. Unlike other aspiration locations, which are word-initial or in stressed syllables, aspiration in Condition 7 surfaces in a stress lapse in the middle of a word, a perceptually and phonetically weak position, and so it would be unsurprising if speakers pronounced aspiration less robustly here. Perceptually, this structural weakness could mean that listeners do not expect to hear strong VOT in this position despite its categorical status, and therefore interpret stops here as voiceless regardless of the VOT level. However, even if not produced or expected, VOT can certainly be perceived in this position, with listeners clearly able to differentiate stops above and below the aspiration threshold.

The empirical evidence from this study thus shows that Conditions 1-7 all behave as categorically aspirated, and Conditions 8-14 all behave as categorically unaspirated. While aspiration in Condition 7 is produced somewhat weaker than in other aspirated conditions, it is perceived as aspirated, with a clear threshold effect, and is therefore a member of the class of categorically aspirated conditions. Though it had not been previously considered in literature on aspiration, Condition 14 behaves as categorically unaspirated, in line with expectations. This provides clear support for the left-edge prosodic description of aspiration posited by Kiparsky (1979), Withgott (1982), and Jensen (2000), and affirms that all examined environments for aspiration can be accurately described under the prosodic definition.

5. Conclusion

The purpose of this study was to acoustically examine the existing theoretical understanding of aspiration as appearing on the left edge of prosodic feet. We used a production study to measure the amount of voice onset time (VOT) in each possible environment for aspiration as well as in environments for unaspirated voiceless stops in English. From this study, we concluded that aspiration largely aligned with this left-edge description, although the initial stressless and the word-medial lapse conditions did not have as robust a level of VOT as the other aspirated conditions, and the word-final stress lapse condition had a higher level of VOT than would be expected from an unaspirated environment.

We next deployed two perception studies, one categorical and one gradient, to examine how listeners perceive VOT in relation to the threshold for aspiration. The categorical study revealed that the initial stressless environment was patterning with other aspirated conditions, while both lapse conditions were behaving like the unaspirated control condition. The gradient study revealed that the medial lapse

condition showed a strong threshold effect like other aspirated conditions, while the final lapse condition behaved like the unaspirated control. Based on these results, we conclude that the medial lapse condition is categorically aspirated and the final lapse condition is categorically unaspirated.

These phonetic and perceptual observations provide empirical, quantitative support for the description of aspiration as occurring at the left edge of prosodic feet, including recursive feet, as described in Kiparsky (1979), Withgott (1982), and Jensen (2000).

References

Boersma, Paul & David Weenink. 2019. Praat: Doing phonetics by computer [Computer program]. Version 6.1.47. Retrieved from http://www.praat.org/

Davis, Stuart and Mi-Hui Cho. 2003. The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics* 41:607-652. https://doi.org/10.1515/ling.2003.020

Jensen, John. 2000. Against ambisyllabicity. *Phonology* 17:187–235. http://www.jstor.org/stable/4420169

Johnson, Keith. 2012. *Acoustic and Auditory Phonetics*. 3rd ed. Wiley-Blackwell.

Kager, René and Violeta Martínez-Paricio. 2018. The internally layered foot in Dutch. *Linguistics* 56:69-114. https://doi.org/10.1515/ling-2017-0037

Kim, Sahynag, Jiseung Kim and Taehong Cho. 2018. Stop voicing contrast in American English: Data of individual speakers in trochaic and iambic words in different prosodic structural contexts. *Data in Brief* 21:980-988, https://doi.org/10.1016/j.dib.2018.10.053.

Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10: 421–441. https://www.jstor.org/stable/4178120

Reetz, Henning and Allard Jongman. 2009. *Phonetics: Transcription, production, acoustics, and perception.* Wiley-Blackwell.

Withgott, Mary Margaret. 1982. Segmental Evidence for Phonological Constituents. PhD thesis, University of Texas at Austin.

Proceedings of the 40th West Coast Conference on Formal Linguistics

edited by Jiayi Lu, Erika Petersen, Anissa Zaitsu, and Boris Harizanov

Cascadilla Proceedings Project Somerville, MA 2024

ISBN 978-1-57473-482-9 hardback

Ordering information

Orders for the printed edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Sprinkle, McKinley and Anya Hogoboom. 2024. In Search of Phonetic Evidence for Prosodically-Motivated Aspiration. In *Proceedings of the 40th West Coast Conference on Formal Linguistics*, ed. Jiayi Lu et al., 295-305. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3721.