Gradient Metrically Conditioned Lenition in Hidatsa*

Levi Driscoll

MIT

1 Introduction

Literature concerned with synchronic fortition and lenition has broadly concluded that strengthening and weakening processes can be gradient or categorical and interact with a wide variety of prosodic domains above and below the word. Curiously, the way in which these phenomena are discussed is quite segregated; studies that treat fortition and lenition as gradient investigate the variation of continuous measures like degrees of constriction and VOT in relation to phrasal prosodic boundaries (Silva 1992, Fougeron & Keating 1997, Cho & Keating 2001). Work concerned with structure below the word—particularly foot-governed lenition and fortition—treat these processes as categorical, often expressed only through feature changes (Prince 1980, Kager 1993, Das 2001, Davis & Cho 2003, Harris 2004). This leaves us with a consensus that featurally identical segments can be realized differently depending on their position in a phrasal domain and that word-internal structure can manipulate the featural identity of a segment. What is largely absent in this literature is discussion of how the same matrix of features is realized in different metrical positions.

One might argue that this gap is filled by the abundance of studies that measure the phonetic realization of segments in relation to stress. It is indeed the case that many processes analyzed as foot-based can be recast in terms of stress; examples of (pre-/post-)tonic fortition and lenition have been discussed at length and the competing units of analysis, stress or foot, are often empirically indistinguishable. I present here a case of gradient lenition that not only can, but *must* be analyzed as conditioned by metrical structure rather than metrical prominence.

The present study[1] examines the interaction of voicing duration and foot structure in Hidatsa, an understudied Siouan language spoken in North Dakota. The objectives of this paper are (i) to document a *bona fide* instance of a gradient lenition process that is unambiguously governed by

* I owe thanks to Brian Agbayani, Adam Albright, Ido Benbaji, John Boyle, Megan Crowhurst, Edward Flemming, Sean Fulop, Chris Golston, Martin Hackl, Jonah Katz, Björn Köhnlein, Anton Kukhto, Katie Martin, Zach Metzler, Norvin Richards, Giovanni Roversi, Donca Steriade, and Jad Wehbe for their many insightful suggestions and comments. I am additionally grateful to the audiences of mfm 26, NAPhCX, NELS 52, and MIT Phonology Circle for their feedback on this and earlier iterations of this work. Any flaws in analysis, reasoning, or argumentation are my fault.

[1] Unless otherwise noted all data in this paper are original, taken from recordings from my fieldwork on Hidatsa in 2017.

metrical structure (ii) to describe the foot structure of Hidatsa and a grammar capable of generating its voicing pattern.

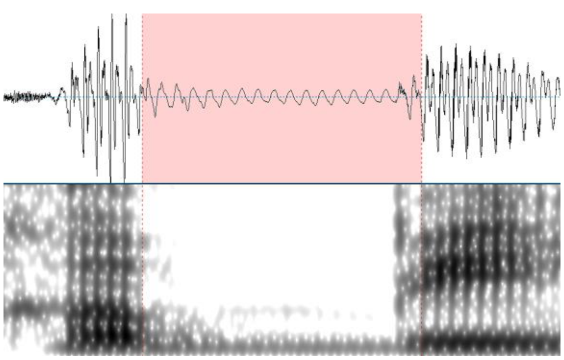
2 Hidatsa

The phonology of Hidatsa is characterized by widespread lenition. Hidatsa contrasts plain and voiced stop series, both of which undergo lenition intervocalically. Plain labial stops lenite to their voiced counterpart (1a) while voiced labial stops lenite to an approximant (1b).

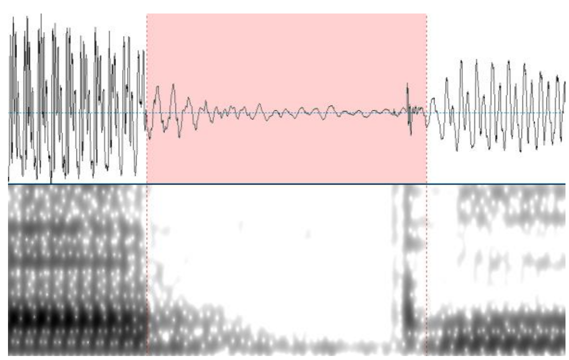
(1)	a.	**p**aːhi-ts	naː-**b**aːhi-ts	
		sing-DECL	2-sing-DECL	
		‘s/he sings’	‘you sing’	(Boyle 2007: 149)
	b.	naː**b**	naː**w**i	
		come.in.a.direction.IMP	come.in.a.direction	
		‘come in!’	‘come in a direction’	(Park 2012: 58)

Although all labial stops lenite between vowels, intervocalic voicing as in (1a) is a gradient process where the duration of voicing varies significantly. In some cases voicing persists for the duration of the stop closure and in others, voicing dies out fairly early.

(2) ˈnaːdagiːha**b**i ‘have a heart attack’ ˈiːdabaː**b**uʃi ‘bobcat’



Voicing (ms): 106 ms



Voicing (ms): 71 ms

The waveforms and spectrograms above show words of identical shapes (HLHLL) with stop closures of approximately equal duration. However, when [b] is in the final syllable (HLHL**L**), voicing duration is greater than when [b] is in the penultimate syllable (HLH**L**L). Such differences are not found only in this pair, but are characteristic of the full set of data (novel to the literature and described in greater detail in §3). The table below illustrates how voicing duration in intervocalically voiced stops systematically fluctuates in different positions and word shapes. Light and heavy syllables are abbreviated as L and H respectively, and bolded letters identify the syllable whose onset is being measured (e.g. a CVCVCV word where the onset of the third syllable is voiced for 75 ms is written as LˈL**L**, 75 ms in the table).

(3)

Word Shape	Voicing Duration	Word Shape	Voicing Duration
(Lˈ**L**)	99 ms	(ˈH)**L**	59 ms
(ˈH)(L**L**)	85 ms	(ˈH)(**L**L)	72 ms
(ˈH)(H)(L**L**)L	93 ms	(ˈH)(**H**)(LL)L	41 ms
(Lˈ**L**)(LL)L	100 ms	(LˈL)(**L**L)L	75 ms
(LˈL)(L**L**)L	98 ms	(LˈL)(LL)**L**	77 ms
(ˈH)(L**L**)(H)L	85 ms	(ˈH)(LL)(**H**)L	40 ms
(ˈH)(H)(L**L**)(LL)L	94 ms	(ˈH)(H)(**L**L)(LL)L	75 ms

Voicing duration in the words in the left two columns is much greater than the words in the right two columns. At first glance it is not obvious why onsets of different syllables in words with the same word shape should be more voiced than others. For instance in a series of five light syllables LLLLL, there is no expectation that the second and fourth syllables should have greater voicing duration than the third and fifth syllables. I argue that this pattern is attributable to foot structure; what the words on the left have in common is that the prominently voiced onsets are foot-medial while the highlighted syllables on the right fall in stray or foot-initial positions. Given the parse above (which I construct an argument for in §3), foot-medial stops have greater voicing duration than stops in other prosodic environments.

Discerning what the foot structure of Hidatsa looks like is quite challenging. The most direct indicator of foot structure, stress, is an area of Hidatsa's phonology where the literature is extremely inconsistent in its findings. Stress has been reported as phonemic/contrastive (Harris & Voegelin 1939, Matthews 1965, Boyle 2007), syntactically predictable (Bowers 1996), phonologically predictable (Voegelin & Robinett 1954, Boyle et al. 2016, Metzler 2021), and nonexistent (Park 2012). The stress analysis I adopt for this paper is that of Metzler (2021). Metzler reports one stress per word on an initial quantity-sensitive (QS) iamb, even if the syllable due to receive stress is final.

(4)	Initial LL	(maˈxu)ba	'bag'	Initial HL	(ˈaː)gugaː	'above'	
		(iˈxu)	'be tired'		(ˈnaː)ga	'calf'	
	Initial LH	(tsaˈgaː)ga	'bird'	Initial HH	(ˈkoː)xaːdi	'corn'	
		(aˈbiː)	'cornsilk'		(ˈhiː)taː	'absent'	(Metzler 2021)

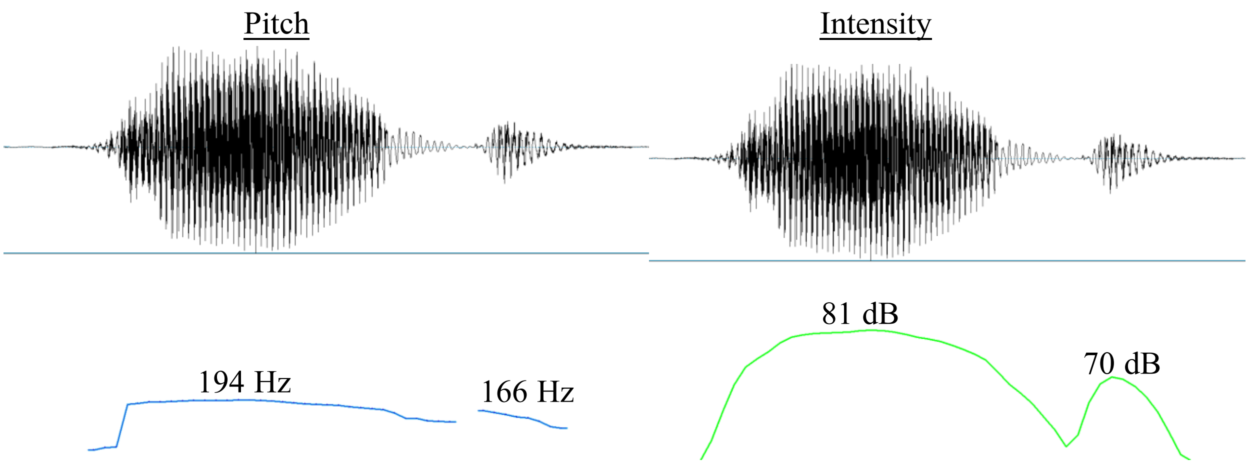
He provides acoustic measurements of monomorphemic words showing that stress is marked by increased intensity. I favor this account because it is the most comprehensive, makes use of the most concrete data, and paints an explicit, detailed, and coherent picture of Hidatsa's stress system.

Harris & Voegelin (1939), Voegelin & Robinett (1954), Matthews (1996), Boyle (2007), and Park (2012) base their claims on impressionistic transcriptions and do not identify an acoustic correlate of stress. Although Bowers (1996) explicitly says that stress is realized as increased intensity, acoustic measurements were either withheld or never performed. Moreover, the analysis does not describe the full range of data.

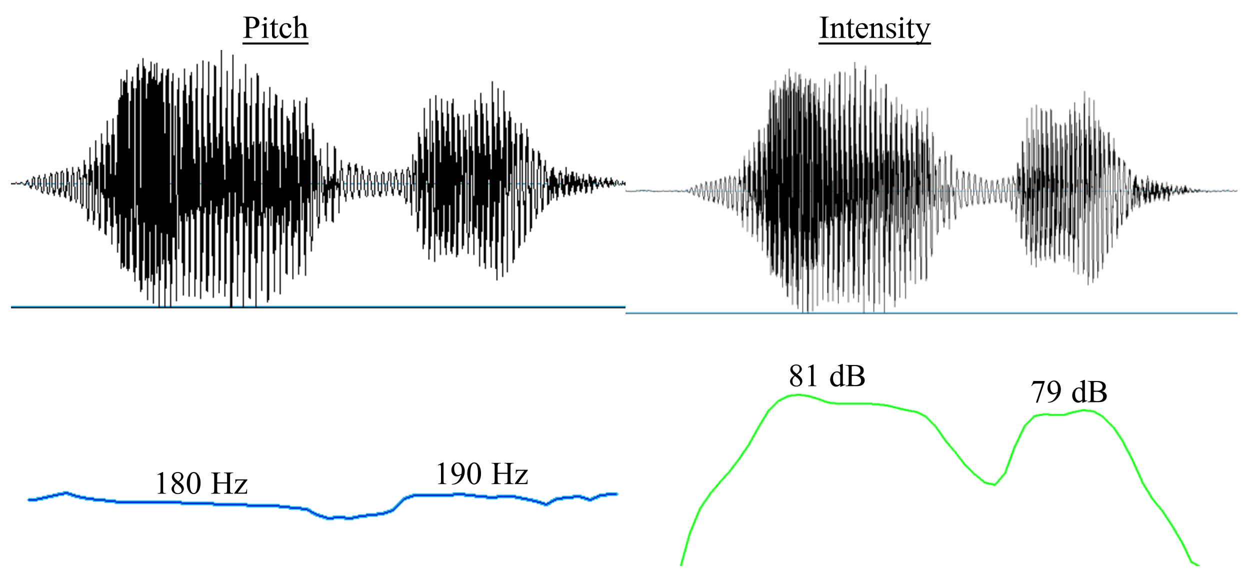
Boyle et al. (2016) provide a complete description of stress assignment based on acoustic analysis, but report that the markers of stress are different depending on syllable weight. It is difficult to ascertain how exactly they propose stress surfaces because the results and conclusions are not communicated clearly. They state only that "stress is marked in long vowels via lengthening" and "stress is marked in short vowels via quality" (Boyle et al. 2016: 30). These statements are supported by graphs of acoustic measurements, though they do not provide any mean duration or formant values and they report statistics on unreliably small samples. Even if we take all of their claims at face value, Hidatsa would be the first case of a stress system with distinct, nonoverlapping acoustic cues to stress for syllables of different weights (as far as I am aware). Boyle et al.'s analysis is vague and the limited data presented do not convincingly support the peculiar stress system they describe.

In contrast, Metzler reports consistent and well-attested manifestations of stress and a regular assignment pattern, with ample data and explicit interpretations of his results. Furthermore, the analysis is not inherently incompatible with the body of literature claiming that Hidatsa has contrastive stress. Recall that the authors who assert that stress is contrastive (Harris & Voegelin 1939, Matthews 1965, Boyle 2007) do not specify how stress is realized and Metzler (2021) identifies stress as increased intensity aligned to initial QS iambs. According to Metzler, stress falls on peninitial syllables if the initial syllable is light and initial syllables if the initial syllable is heavy. There is, however, some kind of contrastive prominence in Hidatsa as evidenced by minimal pairs such as *áaba* 'neck' and *aabá* 'plant'. Acoustic measurement of this pair reveals that intensity, the correlate of stress identified by Metzler, consistently appears on the initial heavy syllable while pitch peaks appear on the initial syllable of *áaba* ‘neck’ and the final syllable of *aabá* ‘plant’ (intensity is marked by the IPA symbol from primary stress (ˈ) and pitch by an acute accent (´)).

(5) [ˈá:ba] ‘neck’



(6) [ˈa:bá] ‘plant’



Given that Harris & Voegelin (1939), Matthews (1965), and Boyle (2007) do not specify what phonetic cues they identify as contrastive ‘stress’, it is conceivable that both a contrastive prominence system (realized as a pitch excursion) and phonologically predictable stress (realized as an intensity peak) coexist in Hidatsa. I will use the term ‘stress’ in this paper as Metzler (2021) does, to refer to syllables with increased intensity.

Because I adopt Metzler's foot-based analysis of stress in Hidatsa, my starting assumptions about feet in Hidatsa in this paper are primarily what can be derived from Metzler's stress data. Specifically, I assume that (i) there is a single QS iamb at the left edge of the word which bears stress (ii) final stress is tolerated (iii) final feet are tolerated. A limited number of additional assumptions beyond what Metzler discusses are adopted. As (3) implicitly indicates, I begin assuming—and later corroborate with data—that feet are strictly binary (i.e., no degenerate or ternary feet). This conclusion is supported not only as an unmarked analytical default, but also because Hidatsa's minimal word is bimoraic. This can be interpreted to mean that each prosodic word must minimally contain a foot, which is never smaller than two moras. Further, there is no evidence suggestive of ternarity, so this is not a parse I consider seriously.

While few (if any) of the words in Metzler's data are long enough to allow secondary stresses to appear, I will still assume that there is only one stress in the longer words in this study. Despite the many conflicting findings in the literature regarding stress in Hidatsa, one thing they have in common is that there are no claims that stress is recurring. I take this as a preliminary indication that there are no secondary stresses (but the absence of secondary stress is not crucial to the analysis, as I spell out in §3.3).

Although there are no secondary stresses offering cues to foot structure beyond the initial iamb, it appears that it is necessary for predictable phenomena to refer to foot structure. Iterative foot construction independent of stress is not unheard of (Poser 1990, Crowhurst 1996), but is there evidence for such a thing in Hidatsa and if so, what does that structure look like? These are the questions this study aims to address.

3 The Study[2]

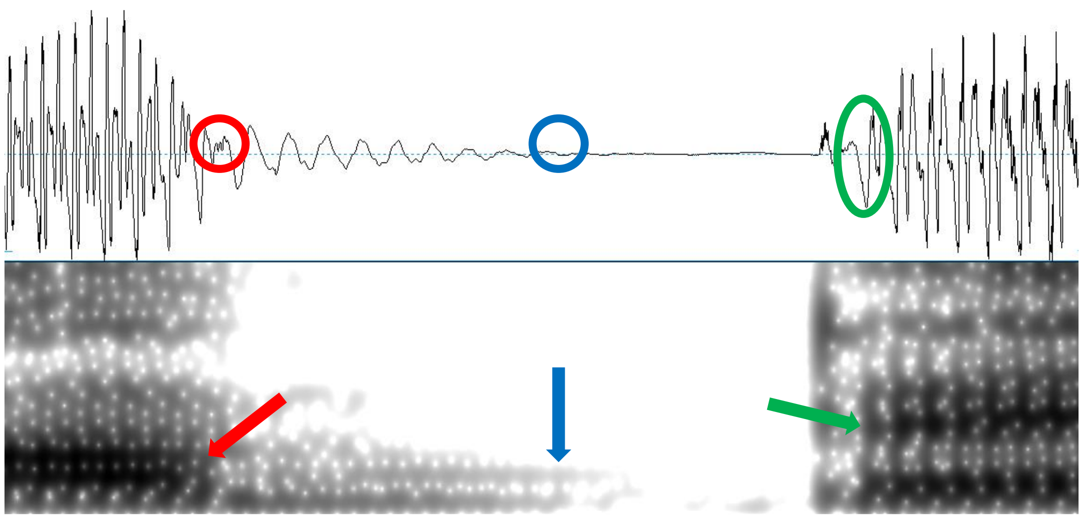
The data presented here are drawn from a pool of 399 intervocalic [b]s (i.e., underlying plain labial stops /p/ that surface as [b] as a result of lenition). Measurements are taken from recordings of words produced in isolation for a dictionary project. In addition to closure duration (Closure (ms) in the examples that follow), three crucial metrics of voicing duration were noted: voicing duration in milliseconds (Voicing (ms)), voicing duration as a percentage of closure duration (Voicing (%)), and the frequency of fully voiced stops where voicing duration is equal to closure duration (FV/N).

---

[2] LH syllable sequences that could plausibly be parsed as (LH) are excluded from the data until §3.2.3 for two reasons: first, whether LH is an acceptable QS iamb has been a topic of much debate (Hayes 1985, 1995; Kager 1993) and I aim to minimize the number of assumed permissible structures in Hidatsa. Second, a primary objective of this paper is to argue against the grammaticality of (LH) iambs and I wish to avoid skewing the data against or in favor of my proposed analysis.

Recordings were analyzed in Praat (Boersma & Weenink 2018) according to the procedure illustrated in (7): the starting point of the closure is marked where disruption in the waveform is observed (red circle) and vowel formants disappear (red arrow), indicating that the upper and lower lips have met to form a labial closure. Closure release is marked where vowel formants reappear (green arrow) and there is a sharp downward spike in the waveform (green circle), marking the resumption of outward airflow. The time between these two points is recorded as closure duration in milliseconds. The cessation of voicing is marked as the point where there the voicing bar disappears or significantly lightens (blue arrow) and there is no longer regular periodic frequency in the waveform (blue circle). The time between cessation of voicing and closure formation is recorded as Voicing (ms).

(7)

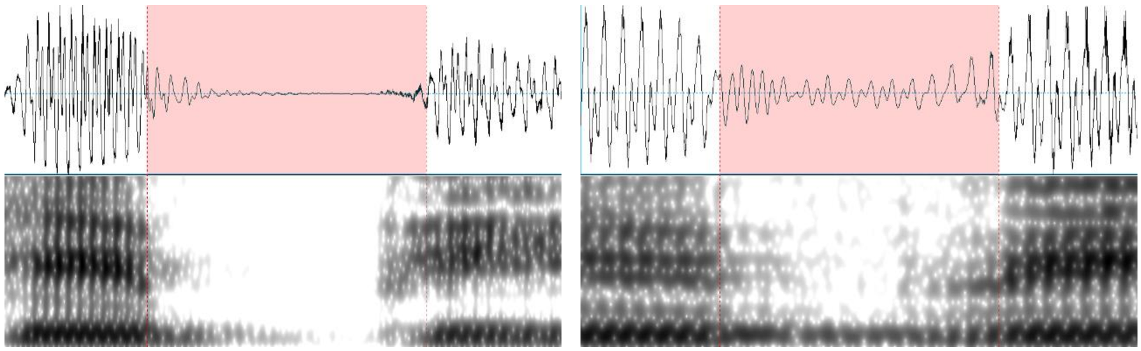


Voicing duration is divided by closure duration to derive Voicing (%). If regular periodic frequency and a dark voicing bar are present from closure formation to release, the stop is recorded as fully voiced. FV/N is a fraction (unreduced) representing the number of such fully voiced stops out of the total number of recordings.

3.1 Distribution of Voicing Duration

Recall from (3) that the duration of voicing of intervocalically voiced stops is widely variable. I argue that this variation is regulated by prosodic structure and that foot-medial [b]s are more voiced than [b]s in other environments.

(8) (aˈgi)b̲i ‘team them up’ (tsiˈb̲i) ‘to drown’



Voicing (ms): 62 ms Voicing (ms): 94 ms

The [b]s above fall in different positions with respect to foot structure. The [b] in *tsibi* is foot-medial while the [b] in *agibi* is in a stray syllable, and there is a substantially greater degree of voicing in the foot-medial [b].

Statistical tests comparing stops in the second syllable of a word-initial LL foot and stray L syllables illustrate that the effect of foot structure on voicing holds generally. ‘Stray’ in this context is used fairly conservatively; strays included in this test consist only of final trapped light syllables (Mester 1994). This pair of environments was chosen because they have the most concrete evidence to indicate their structural position. It further ensures that no matter what assumptions we make about the iterativity or directionality of syllable parsing, the syllables in question are unambiguously stray. Word shapes that meet this criterion include (LˈL)**L** and (LˈL)H**L**, where the final syllable will never belong to any foot given a QS iambic parse.

(9)

	Voicing (ms)	Closure(ms)	Voicing (%)	FV/N
Initial LˈL	88 ms	100 ms	90%	46/60
Stray	72 ms	106 ms	69%	18/53
p-value	$<.001$	.08	$<.001$	$<.001$

Significance for continuous metrics (Voicing (ms), Closure (ms), and Voicing (%)) is calculated with two-tailed t-tests and the categorical metric (FV/N) with a chi-square test. All three metrics return significant results, providing evidence that foot-medial [b]s are more voiced than [b]s in unparsed syllables.

It is prudent to pause here to flag two potential confounds in this test, which I will address later in the paper. The environments being compared differ in ways that this preliminary test cannot control for: finality and stress. All strays included in this test belong to word-final syllables, which could decrease voicing duration. Although some of the initial LˈL data includes measurements of final syllables (disyllabic words #(LˈL)#), we cannot be certain that there is no effect of finality. Similarly, there is categorically no overlap between strays and initial LˈLs with respect to stress. The onsets of initial LˈLs all belong to stressed syllables and there are no stray syllables bearing stress. It is important to eliminate the possibility that this effect is conditioned by stress rather than foot structure, but it is crucial to first establish that there are feet beyond the initial iamb to do so properly. I leave these issues here for now.

3.2 Voicing as a Diagnostic

What little research has been done on the metrical structure of Hidatsa is not very informative about feet beyond the initial iamb as there is only one stress per word. I show here that voicing can be used to learn more about feet in such cases. Specifically, voicing patterns can reveal whether there are multiple feet in languages that have one stress per word and whether those feet are left- or right-aligning.

3.2.1 Iterativity

There are two tests that can function as evidence for whether feet are parsed iteratively. One is comparing stray syllables to syllables that *would be* foot-medial if we assume iterative parsing (#(LˈL)**L**# vs. #(LˈL)(L**L**)#). If the results show that the stray syllables do not pattern with the non-word-initial feet, this indicates that there are multiple feet per word. The other test compares voicing in foot-medial [b]s in word-initial feet and non-word-initial feet. Here if the test does not find a significant difference in voicing duration, this suggests that feet are built across the word and [b]s in both contexts are foot-medial.

Just as stray environments are restricted to avoid ambiguity, so are the noninitial feet. In words that could contain stray syllables like LLLLL, two parses are available: (LˈL)(L**L**)L and (LˈL)L(L**L**). Under the first parse, the onset of the fourth syllable is foot-medial and predicted to be realized with greater voicing duration. Under the second, the onset of the fifth syllable is foot-medial and expected to be prominently voiced. Word shapes are therefore constrained to those where exhaustive parsing is possible, e.g. (LˈL)(L**L**)(L**L**), (LˈL)(H)(L**L**), (ˈH)(L**L**), etc. This guarantees that if syllables are indeed parsed iteratively into feet, data categorized as belonging to noninitial feet will certainly be foot-medial.

Tests comparing voicing duration in stray syllables and foot-medial syllables in noninitial feet show that [b]s in noninitial feet contrast with strays.

(10)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
Stray	72 ms	106 ms	69%	18/53
Noninitial LL	83 ms	102 ms	83%	26/39
p-value	.037	.35	.01	.002

The difference in voicing duration is significant, suggesting that footing is iterative. This also provides addition evidence in support of a foot-based analysis rather than a stress-based analysis, as the contrast in voicing duration is still observed even though neither environment is stressed.

Further evidence for the existence of noninitial feet comes from testing the difference in voicing duration of [b]s that are medial in initial feet versus those that would be medial in noninitial feet.

(11)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
Initial LˈL	88 ms	100 ms	90%	46/60
Noninitial LL	83 ms	102 ms	83%	26/39
p-value	.29	.44	.09	.27

The results are not significant. Foot-medial voicing found in initial (LˈL) is also present in noninitial LL, which indicates that feet are built across the word. Note also that the voicing pattern persists in noninitial LLs despite the absence of stress. This again supports the argument that prosodic structure is responsible for the effect, not stress.

From the results of the tests above, we can safely conclude that syllables are parsed iteratively.

3.2.2 Directionality

Another property of metrical structure that voicing can diagnose is the direction of alignment. Although there is a stressed foot at the left edge of the word, this does not rule out the possibility of a bidirectional parse in which one foot appears at one word edge and the other feet align to the opposite edge (LˈL)L(LL)(LL)(LL).

Checking the direction of alignment requires two test: one comparing initial feet to *left*-aligned iambs (LˈL)(L**L**)(L**L**)L and one comparing initial feet to *right*-aligned iambs (LˈL)L(L**L**)(L**L**). In both cases, words must contain stray syllables to avoid the ambiguity of cases like (LˈL)(L**L**)(L**L**) where the stop in question is foot-medial whether noninitial feet are left- or right-aligning.

(12)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
Initial LˈL	88 ms	100 ms	90%	46/60
Left	94 ms	106 ms	88%	16/22
p-value	.34	.10	.89	.71

[b]s that appear in foot-medial position given a left-aligning parse pattern with foot-medial initial LˈLs, suggesting that this is the appropriate footing.

(13)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
Initial LˈL	88 ms	100 ms	90%	46/60
Right	83 ms	126 ms	68%	4/19
p-value	.34	$< .001$	$< .001$	$< .001$

Two of the three metrics of voicing return significant results indicating that the right-aligning parse does not pick out foot-medial stops. However there is some conflicting evidence; the difference in Voicing (ms) is not significant which suggests initial LˈLs and right-aligning iambs pattern together. One notable insight from these results is that Voicing (ms) does not change between these environments but initial LˈL closures are significantly shorter than the stops in the right-aligning parse. Shorter closures is a known property of constituent-medial lenition (Lavoie 2001). While these results are not entirely conclusive, right alignment in iambic systems is extremely rare, if it exists at all (Alber (2005) for instance claims that such systems are unattested). Left alignment should be our default expectation in a QS iambic system, and the evidence in favor of it is stronger than the evidence for a bidirectional parse. Therefore, I will tentatively assume that feet in Hidatsa are left-aligning.

3.2.3 Asymmetrical Iambs

The prior sections have created a nearly complete sketch of the foot structure of Hidatsa. The only remaining matter to consider is what foot shapes are permitted. This study adopts a conservative theoretical approach in which data that are ambiguous or rely on controversial assumptions are excluded without undergoing some form of test. Foot typology is one area where this has been particularly relevant. (LL) and (H) feet are universally accepted as QS iambs. While (LH) iambs are regarded as well-formed in most research in metrical phonology, there have been notable challenges to its inclusion in foot typology. The most prominent of these objections is made by Kager (1993), who argues that all QS feet are strictly bimoraic[3]. Hayes' (1995) proposal that (LH) is not only acceptable, but the optimal iamb, has become a standard assumption, however the issue is far from settled. Foot typologies with and without asymmetrical iambs make identical predictions for iambic stress.

(14) Hayes: (ˈH)(LˈH)(LˈL)(LˈH)
Kager: (ˈH)L(ˈH)(LˈL)L(ˈH)

As stress is the primary cue to foot structure, it has been difficult to rule out one of these hypotheses on empirical grounds.

Foot-medial voicing in Hidatsa is able to distinguish between the two. Hayes' foot typology predicts that there should be no contrast in voicing between LL and LH because both are acceptable iambs, thus the second syllable of both pairs is foot-medial. Kager's typology, however, predicts that only LL should be prominently voiced since a foot boundary intervenes between L and H syllables.

(15) Hayes: (H)(L**H**)(L**L**)(L**H**)
Kager: (H)L(H)(L**L**)L(H)

By comparing voicing duration in LH syllable pairs, whose status as a foot has been disputed, to voicing duration in LL syllable pairs, whose status as a foot is unchallenged, we are afforded new insight on foot typology. If LL and LH pattern together, this indicates that both are grammatical iambs. If they contrast, this is evidence that LH is not a foot.

(16)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
LL	86 ms	101 ms	87%	104/151
LH	78 ms	110 ms	73%	32/81
p-value	.008	$<.001$	$<.001$	$<.001$

[3] Kager prohibits (LH) as a 'parsing iamb' (always) and as a 'surface iamb' when (LH) has foot-internal moraic lapse. Because most Hs bear prominence on their first mora, surface (LH) usually does not have foot-internal moraic lapse (μ.μ́μ). When H has final prominence, (LH) contains moraic lapse *(μ.μμ́) and is unacceptable. Hidatsa has rising diphthongs [ia] and [ua] but not falling diphthongs *[ai] and *[au]. If permissible diphthongs are indicative of the prominence structure of H syllables, relegating the more sonorous vowel to the right edge of a diphthong suggest that Hs in Hidatsa have final prominence. According to Kager, (LH) would be ungrammatical as both a parsing and surface iamb for the relevant data.

The robust voicing found in foot-medial stops is absent in LH pairs, suggesting that there is in fact a foot boundary between L and H syllables.

This can be further tested by examining the predictions for voicing if L and H syllables are not parsed together. The ungrammaticality of LH implies that H syllables always form their own feet in a QS system like that found in Hidatsa. If asymmetrical iambs are illicit, LH should behave the same as foot-initial stops.

(17)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
Foot-initial	75 ms	115 ms	68%	24/87
LH	78 ms	110 ms	73%	32/81
p-value	.40	.16	.23	.10

This prediction is borne out which further indicates that LH is not an acceptable QS iamb. Voicing data in Hidatsa therefore support the strictly bimoriac foot typology proposed by Kager (1993) in which only (LL) and (H) are well-formed QS iambs.

3.3 Summary

There is now evidence that informs how all possible strings of syllables are parsed in Hidatsa, which enables us to consider the full set of data. Descriptive statistics of voicing duration by foot position for all 399 [b]s is shown below.

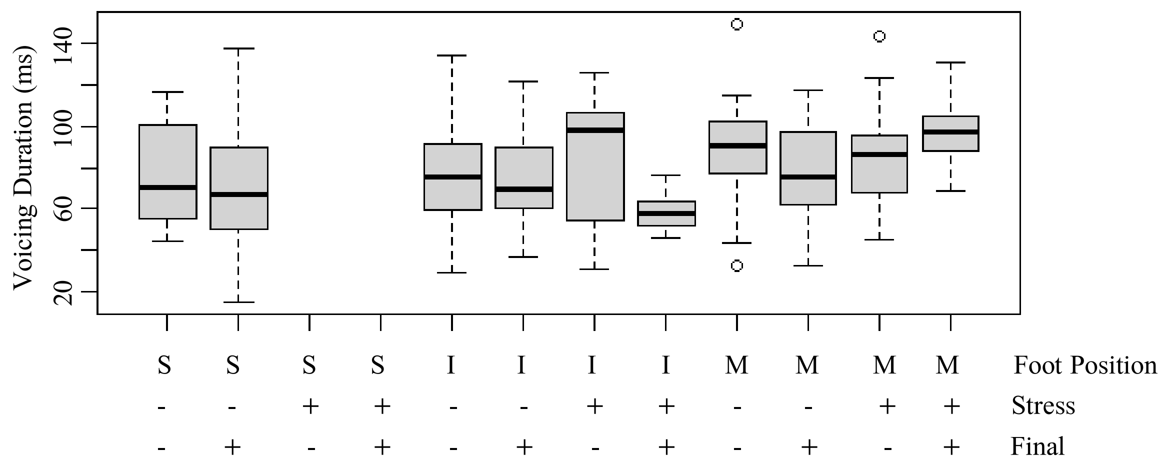
(18)

	Voicing (ms)		**Closure (ms)**		**Voicing (%)**		**FV/N**
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	
Foot-initial	76 ms	2.39	112 ms	2.34	70%	24.4	56/168
Foot-medial	86 ms	2.15	101 ms	1.65	87%	21.1	104/151
Stray	73 ms	2.74	110 ms	2.41	68%	25.3	18/80

I now return to the potential confounds mentioned in §3.1. The initial test establishing the interaction of voicing duration and foot structure in (9) was unable to control for the influence of stress and finality. This test compared sets of data that differed both in foot structure and stress, so it was impossible to disentangle the effects of stress from those of foot structure. Similarly, the fact that one set of data consisted entirely of final syllables may have undermined the integrity of the test. For instance, it may have been the case that the culmination of differences in stress, foot position, and finality led to an illusory significant result even though no single factor had significant effects on its own. While I would argue that the battery of subsequent tests throughout §3.2 which consistently upheld an effect of foot position should assuage these suspicions, they still deserve to be addressed directly.

The full dataset was coded for foot position, stress, and finality and fit in a linear model in R investigating these factors' impact on voicing duration and their interactions. The boxplot below shows voicing duration averages for each combination of the three factors (S = stray, I = initial, M = medial, - = unstressed/nonfinal, + = stressed/final).

(19) Voicing duration by foot position, stress, and finality



The overall model returned significant results ($r^2 = 0.094$, $F(9, 389) = 4.469$, $p < .001$). Foot position was found to be the only reliable predictor of voicing duration with significant main effects between medial and initial positions ($\beta = -14.905$, $p < .001$) and medial and strays ($\beta = -9.328$, $p = .032$). Stress ($\beta = -4.296$, $p = .073$) and finality ($\beta = -0.684$, $p = .775$) were not significant predictors.

Additional significant interactions were found. There is a two-way interaction which reduces the main effect of foot position between foot-medial and foot-initial [b]s in nonfinal syllables ($\beta = 7.374$, $p = .032$). In final stressed syllables, the difference in voicing duration between foot-initial [b]s and foot-medial [b]s is much greater ($\beta = -12.414$, $p < .001$). The table below details the results of interactions among foot structure, stress, and finality.

(20)

	β	Std. Error	t value	p value
M vs. I	-14.905	3.433	-4.342	<.001 *
M vs. S	-9.328	4.344	-2.148	.032 *
Finality	-0.684	2.387	-0.286	.775
Stress	-4.296	2.387	-1.800	.072
M vs. I: Final	7.374	3.433	2.148	.032 *
M vs. S: Final	-3.260	4.344	-0.751	.453
M vs. I: Stressed	6.343	3.433	1.848	.065
M vs. I: Final, Stressed	-12.414	3.433	-3.616	<.001 *

The linear model shows that voicing duration patterns in Hidatsa are best captured by a foot-based analysis. Stress and syllable finality do not have a significant impact on voicing duration, and even when the effects of these factors are accounted for, foot structure is still an accurate predictor of voicing duration. We can therefore conclude that intervocalically voiced [b]s have greater voicing duration in foot-medial position than other prosodic environments.

The statistical model in (19) and (20) assumes that there are no secondary stresses, however this is not essential to rule out the possibility that stress is responsible for the voicing duration effect. Suppose there is secondary stress in Hidatsa (whether it has acoustic correlates that have gone unnoticed in earlier work or its character is more like an abstract notion of headedness); heavy syllables attract stress as Metzler (2021) shows, so the secondary stresses would be assigned as in (21) (secondary stresses are indicated with the IPA character for secondary stress (ˌ)).

(21) (LˈL)(ˌH) (LˈL)(LˌL)(ˌH)
(ˈH)(LˌL) (ˈH)(ˌH)(LˌL)L
L(ˈH)(LˌL) (ˈH)(LˌL)(ˌH)L

To get increased voicing duration on the second syllable of each (LL) foot without referring to foot structure, this pattern would need to be described as having increased voicing duration in the onset of stressed syllables. This means that each of the (noninitial) Hs are predicted to be likewise prominently voiced, which I have already shown does not happen for L(ˈH)(LˌL) in example (16) and (ˈH)(ˌH)(LˌL)L and (ˈH)(LˌL)(ˌH)L in (3). Comparison between mean voicing duration in Hs and Ls shows that Hs are generally less voiced than Ls.

(22)

	Voicing (ms)	Closure (ms)	Voicing (%)	FV/N
H	75 ms	112 ms	70%	37/108
LL	86 ms	101 ms	87%	104/151

Characterizing the voicing data in terms of stress requires stipulations that stress compels these effects but heavier syllables that are more prone to receiving stress somehow resist a phonetic enhancement associated with prominence, that the magnitude of the effect is identical in syllables with primary and secondary stress, and that there is secondary stress for which no one has previously found evidence. It is further unclear how a grammar would derive this pattern on stressed light syllables to the exclusion of stressed heavy syllables if what crucially conditions the effect is stress.

The foot-based analysis need not make any claims of this sort. All that it requires is that Hidatsa have iteratively parsed left-aligning quantity-sensitive foot structure and that a single environment, foot-medial intervocalic stops, be marked by increased voicing duration. This analysis is descriptively concise and easy for a grammar to refer to the necessary structure, as I show below.

4 Motivating and Modelling Lenition

While the above adequately describes voicing duration patterns in Hidatsa and accounts for the distribution of data, the question of why this occurs and how to represent the pattern remains unaddressed. Due to the gradient nature of the pattern, treating voicing lenition in Hidatsa as a simple feature change is insufficient. It seems implausible that foot-medial stops are specified as voiced while stops that undergo a lesser degree of lenition are specified as voiceless, especially considering that there is clear voicing through more than half of the closure in all intervocalic environments and stops in word-initial position lack prevoicing. While this does not necessarily preclude the possibility of voicing lenition involving a feature change or contextual constraints on phonological features, an adequate analysis must be more complex. I consult phonetic literature to derive this additional complexity.

Two approaches to lenition heavily grounded in phonetics attribute synchronic lenition processes to the minimization of effort in production (Kirchner 2004) or facilitating perception (Kingston 2008, Katz 2016). On a production-based account, lenition results from a desire to reduce articulatory effort. For intervocalic obstruents, ceasing voicing at the end of the first vowel and restarting voicing at the start of the second vowel requires more effort than maintaining voicing through the constriction. Perceptual motivations for lenition hold that fortition and lenition act as cues to otherwise inaudible prosodic structure. Low intensity sounds often produced by fortition processes interrupt the continuity of a stream of speech sounds, affording listeners the opportunity to parse a new constituent. Conversely lenition processes eliminate low intensity sounds to preserve the continuity of the speech stream as a sign that the current prosodic constituent is ongoing. These explanations yield domain-initial fortition and medial lenition, respectively.

Neither fits the voicing duration patterns of Hidatsa perfectly. If the purpose of lenition is to reduce articulatory effort, there is motivation for a general intervocalic voicing process but Hidatsa's sensitivity to prosody remains unexplained. If consonants undergo lenition as an indicator that a constituent is ongoing, foot-medial intervocalic voicing is motivated but foot-initial intervocalic voicing is not.

Because this study fills a gap in the literature regarding how relatively small prosodic constituents impact the phonetic detail of featurally identical segments, extant analyses generally provide solutions that are too coarse or do not explain in detail how the grammar determines phonetic values. The nearest analyses to achieving these goals are the works of Kaplan (2010) and Katz (2021), though neither meets the specific needs of a grammar of Hidatsa. Kaplan focuses on accounting for systems of contrast neutralizing lenition, which includes some discussion of gradient intervocalic voicing duration. The crucial mechanism utilized by Kaplan is an effort avoidance constraint that penalizes extreme articulations: VOT values that are too close to 0 and values that are too close to some stipulated upper limit. This approach is inappropriate for Hidatsa for two reasons. First, the present study is not concerned with how contrasting sounds differ along a single phonetic dimension. Second, foot-medial [b]s in Hidatsa are regularly fully voiced, suggesting that articulations at the higher extreme are not penalized, but encouraged by the grammar.

Katz (2021) proposes a grammar in which consonant duration is sensitive to phrasal prosodic boundaries, but it is not explained how the grammar determines the magnitude of prosodic duration adjustments. The numbers used in the analysis are derived from observed durational variation in a corpus and are stipulated as being part of the grammar. A grammar of Hidatsa must provide a more concrete account of what determines variation in voicing duration.

I entertain two analyses of the Hidatsa voicing pattern, both of which incorporate Kingston (2008) and Katz's (2016) proposals that lenition is perceptually motivated. My goal in presenting two analyses is not to debate the merits of each and decide between them; rather my intent is to show that it is possible to write a grammar that generates Hidatsa's voicing pattern within the existing architecture of phonological/phonetic grammars, and in fact there are multiple ways to approach this problem. The most natural point of divergence between analyses of a gradient realization phenomenon such as Hidatsa's voicing pattern seems to be whether the computation of gradient values takes place in a phonological grammar or in a distinct post-phonological phonetic grammar. I pursue both possibilities for a functional grammar of Hidatsa.

The first analysis assumes that categorical phonology and gradient phonetic realization are evaluated in the same module of the grammar. This approach to lenition is largely familiar in the literature in that intervocalic voicing involves changing the featural specification of a stop from voiceless to voiced to avoid the markedness of a voiceless stop between two vowels. Each potential output is featurally specified as voiced and varies by the degree of voicing. When a stop is intervocalic and foot-medial, additional pressure to prolong voicing is exerted to maintain continuous intensity within the foot domain.

The alternative view is that gradient lenition is strictly phonetic and variation in voicing duration is entirely motivated by the grammar's demand for uninterrupted constituent-internal intensity. On such an account, reference to phonological features is unnecessary; all intervocalic stops are targeted for lenition to promote continuous word-medial intensity (23). Foot-medial stops (23b) are privileged because they are medial at multiple levels of the prosodic hierarchy (word and foot) and undergo a more significant degree of lenition as a result.

(23)		(a) Foot-initial	(b) Foot-medial	(c) Stray
	Word	[ ]	[ ]	[ ]
	Foot	[ara][**b**agi]	[kiː][ra**b**i]	[huː]**b**a

This idea builds on Fougeron & Keating's (1997) study of American English where they found that fortition effects at initial prosodic boundaries are amplified when several initial boundaries coincide.

In what follows, I present detailed accounts of how phonological and phonetic grammars of Hidatsa determine voicing duration in different prosodic environments. While I discuss the predictions and consequences of each analysis, I do not advocate for one approach over the other.

4.1 A Phonological Grammar of Hidatsa

To achieve the gradient effects of structure on voicing, I depart from the ranked constraints of standard OT (Prince & Smolensky 1993/2004) in favor of a weighted constraint grammar (cf. Flemming 2001). In a grammar where phonetic detail is integrated with categorical phonology, the general process of intervocalic voicing can be captured by standard interactions of contextual markedness and faithfulness constraints.

(24) *T/V_V *Cost of violation* = $w(n)$
No voiceless stops between vowels

(25) IDENT[VOI] *Cost of violation* = $w(n)$
Do not change the voicing specification of a segment

The cost of violating each of these constraints is calculated by multiplying the weight of the constraint (w) by the number of times the constraint is violated (n).

While the grammar compels feature changes with weighted analogs of familiar categorical constraints, variation in voicing duration is modeled as a series of conflicting targets.

(26) *VOICED *Cost of violation* = $w(\text{Voicing}(\%) - 0)^2$
Stops must not be voiced

(27) CONTINUOUSINTENSITY-φ (CONTINTφ) *Cost of violation* = $w(\text{Voicing}(\%) - 1)^2$
Intensity must be continuous within a foot

(28) [VOI] ≥ 70% *Cost of violation* = 0 IF Voicing (%) ≥ 70
Segments specified as voiced must be voiced for = $w(\text{Voicing}(\%) - .7)^2$ IF
at least 70% of their duration Voicing (%) < 70

The cost of violating target constraints is calculated by squaring the difference between the target and the voicing duration of a candidate, then multiplying the result by the weight of the constraint (see Flemming 2001 for discussion of this method of assessing violations). *VOICED is a 0% voicing target which seeks to eliminate stop voicing from the acoustic signal. CONTINTφ requires foot-medial segments to maintain a certain level of intensity; for our purposes this constraint is satisfied by foot-medial stops that are 100% voiced[4]. Stops that fall at an initial foot boundary or in a stray syllable are not subject to this constraint since they occupy a peripheral position in the foot or are foot-external. The final target constraint, [VOI] ≥ 70%, is perceptually motivated. To ensure that segments that are specified as voiced are perceived as voiced segments, [VOI] ≥ 70% states that at least 70% of the duration of that segment must consist of voicing. Unlike other targets where deviating from a specified value in either direction (overshoot or undershoot) is penalized, [VOI] ≥ 70% is not violated by overshooting the 70% target as this does not impede

[4] Although this constraint was developed for a fairly narrow purpose, it could be parameterized like Katz's (2016) boundary disruption constraints to apply generally to various types of lenition at different prosodic levels. An example of such parameterization might look like the following: CONTINT(Z, ω)—this constraint would require all word-medial segments to be at least as intense as a voiced continuant.

the perception of a voiced segment as voiced. Only undershooting the 70% target results in accrual of violations.

In the tableaux that follow, I derive the mean Voicing (%) values for foot-medial [b] (87%) and foot-initial [b] (70%). Each candidate is a potential output for the mean Voicing (%) of a given structural configuration. The subscript number in each candidate indicates what percentage of the preceding segment's duration is voiced (e.g. in $ab_{50}a$, the [b] is 50% voiced). The total cost of violations for each candidate is displayed in the far right column labeled *H.* I exclude from the tableaux constraints responsible for the construction and placement of feet and candidates that deviate from the foot structure described in §3. I take for granted that constraints producing the observed metrical system are active in the grammar; while the Hidatsa data may bear on interesting theoretical issues regarding the typology of foot structure, these problems exceed the scope of this paper.

When an underlyingly plain stop falls between two vowels in foot-medial position, it is compelled to become voiced by *T/V_V, and the duration of voicing is determined by the weights of target constraints.

(29) Intervocalic voicing, foot-medial

apa	[VOI] ≥ 70% $w = 5.5$	*T/V_V $w = 4$	CONTINTφ $w = 3.18$	*VOICED $w = 0.42$	IDENT[VOI] $w = 0.1$	*H*
(ap_0a)		1	1	0		7.18
($ab_{10}a$)	0.36		0.81	0.01	1	4.66
($ab_{20}a$)	0.25		0.64	0.04	1	3.527
($ab_{30}a$)	0.16		0.49	0.09	1	2.576
($ab_{40}a$)	0.09		0.36	0.16	1	1.807
($ab_{50}a$)	0.04		0.25	0.25	1	1.22
($ab_{60}a$)	0.01		0.16	0.36	1	0.815
($ab_{70}a$)	✓		0.09	0.49	1	0.592
($ab_{80}a$)	✓		0.04	0.64	1	0.496
☞ ($ab_{87}a$)	✓		0.017	0.757	1	0.471
($ab_{90}a$)	✓		0.01	0.81	1	0.472
($ab_{100}a$)	✓		0	1	1	0.52

Because the markedness of intervocalic voiceless stops outweighs faithfulness, the underlyingly plain stop becomes specified as voiced, which subjects the [b] to the demands of the perceptual target [VOI] ≥ 70%. CONTINTφ pushes foot-medial [b] beyond the 70% voicing duration required by [VOI] ≥ 70%, but the winning candidate falls short of CONTINTφ's 100% voicing target because the cost of violating *VOICED exceeds the reduction in markedness achieved by prolonging voicing further.

Outside of foot-medial contexts CONTINTφ no longer applies and is unable to cause a boost in voicing duration. This results in shorter voicing duration for foot-initial and stray syllables alike[5].

(30) Intervocalic voicing, foot-initial

apaː	[VOI] ≥ 70% $w = 5.5$	*T/V_V $w = 4$	CONTINTφ $w = 3.18$	*VOICED $w = 0.42$	IDENT[VOI] $w = 0.1$	H
a(p_0aː)		1		0		4
a(b_{10}aː)	0.36			0.01	1	2.084
a(b_{20}aː)	0.25			0.04	1	1.492
a(b_{30}aː)	0.16			0.09	1	1.018
a(b_{40}aː)	0.09			0.16	1	0.662
a(b_{50}aː)	0.04			0.25	1	0.425
a(b_{60}aː)	0.01			0.36	1	0.3062
☞ a(b_{70}aː)	✓			0.49	1	0.3058
a(b_{80}aː)	✓			0.64	1	0.369
a(b_{87}aː)	✓			0.757	1	0.418
a(b_{90}aː)	✓			0.81	1	0.44
a(b_{100}aː)	✓			1	1	0.52

As before the intervocalic [b] is specified as voiced to satisfy markedness and so must minimally reach the perceptual voicing target of 70%. Absent the influence of CONTINTφ, *VOICED guarantees that the [b] will be as voiced as perceptual constraints require and no more. This places the winning candidate's voicing duration at 70%.

---

[5] I do not provide separate treatments for intervocalic voicing duration in foot-initial and stray syllables. These environments are viewed as identical by the grammar(s) I have proposed and do not have significantly different voicing duration means (70% vs. 68%). I assume that what statistically negligible variation is found can be accounted for by some combination of the absence of stress in stray syllables, the greater proportion of final stray syllables compared to the proportion of final foot-initial syllables, and an inherent margin of error in the production of each individual utterance.

The choice to represent voicing duration candidates as a percentage is not crucial to the function of the grammar, though I argue it addresses the motivations of lenition most directly. In §3, I showed that Voicing (ms) and closure duration systematically differ according to foot position. In foot-medial [b], Voicing (ms) is greater and closure duration is shorter relative to other environments. As a result, the magnitude of the difference in Voicing (%) between foot-medial [b] and other [b]s is larger than the difference in either voicing duration or closure duration. This allows a speaker to fill a greater portion of the stop closure with voicing while simultaneously minimizing deviation from any potential articulatory VOT or closure duration targets. The perceived difference in voicing duration is enhanced without significantly compromising faithfulness, which is a sensible way to satisfy the objectives of lenition if prosody-sensitive lenition is perceptually motivated. Thus representing candidates as percentages appears to coincide with the goals of the grammar and efficiently capture the increase in Voicing (ms) and the decrease in closure duration in foot-medial position.

Placing the calculation of phonetic details such as these in the same module of grammar responsible for structure building and stress assignment might be a point of concern. Prosodic structure can impact the phonetic realization of segments to enhance the prominence of marked features. A reversal in the weighting of constraints that causes structure to augment marked articulations then predicts that languages could refuse to build structure that conditions the production of a marked segment rather than produce the marked segment.

(31) Hypothetical grammar: segmental markedness outweighs structure building

apa	CONTINTφ $w=5$	*VOICED $w=5$	PARSEσ $w=1$	H
(ap$_0$a)	1	0	0	5
(ab$_{50}$a)	0.25	0.25	0	2.5
(ab$_{100}$a)	0	1	0	5
☞ ap$_0$a		0	2	2

This possibility predicts the existence of a language much like Hidatsa: foot-medial stops are subject to increased voicing prominence, but every foot is marked by a stress and prohibitions on voicing outweigh constraints that parse syllables into feet. The resulting system manifests stress only on feet that do not contain obstruents in medial position (e.g. (maˈri)(waˈ**ra**)(ˈbaː)(giˈ**wi**) but (maˈri)wa**ba**(ˈbaː)gi**bi**). While there are abundant examples of languages in which structure influences segmental identity, there are no languages in which segmental properties determine where or whether prosodic structure appears.

The observation of unidirectional prosody-segmental interactions is neither novel nor is it unique to gradient processes like voicing in Hidatsa; this is simply another instance of a too-many-solutions problem (Steriade 2001, de Lacy 2003, Blumenfeld 2006). Several paths are available to limit the number of repairs a grammar may deploy. One could stipulate a fixed weighting (a move independently motivated by implicational hierarchies) between constraints on structure and segments such that constraints responsible for building prosodic structure (PARSEσ) always outweigh those that refer to this structure (CONTINTφ). This guarantees that a grammar like that in (31) is impossible.

Another option is to appeal to modularity, either that the phonetic details of a segment's realization are calculated based on the output of categorical phonology (as in §4.2 of this paper), or that structure building is segregated from the rest of the grammar. This is similar to and compatible with The Modularity Hypothesis (Rasin 2018):

(32) *The Modularity Hypothesis*

Stress is computed in an informationally encapsulated module with the following properties:

a. The input to the stress module excludes representations of segmental features
b. Outside of the stress module, stress representations cannot be changed

Given The Modularity Hypothesis, foot-based accounts of stress assignment already take for granted that foot construction occurs separately from the rest of the grammar because the computation of stress is dependent on foot structure. This alone rules out grammars like (31), though it does not likewise prevent other bidirectional prosody-segmental interactions. To achieve this end, The Modularity Hypothesis could be augmented to generally include all prosody-building operations in the portion of the grammar where stress is computed.

For our purposes, the particular choice between these options is irrelevant; what is crucial to observe is that the constraints and architecture of the analysis do not uniquely predict unattested phenomena. The analysis laid out here derives Hidatsa's predictable variation in voicing duration and its undesirable typological consequences are readily mitigated by existing proposals.

4.2 A Phonetic Grammar of Hidatsa

Rather than adopting a weighted constraint grammar for both phonology and gradient phonetics, voicing duration could be determined separately from the portion of the grammar that builds structure, assigns stress, and manipulates segmental features. This enables us to maintain a ranked constraint grammar for categorical phonological processes and account for noncontrastive fine-grained details with a gradient weighted constraint grammar at a later derivational stage. Grammatical architecture of this nature allows phonetic constraints to refer to existing prosodic structure to calculate phonetic values while rendering such structure immutable by the phonetic grammar.

In a post-phonological grammar of Hidatsa, many aspects of the analysis remain the same; I still accept perceptual motivations for lenition (*à la* Kingston (2008) and Katz (2016)), represent voicing duration with percentages, and treat constraints on voicing duration as targets. The analyses differ in that a phonetic grammar need not refer to phonological features to trigger intervocalic voicing or specify minimal voicing duration targets. Instead, it is the structural environment where a segment appears that determines whether and to what degree voicing occurs. I remain agnostic as to whether underlyingly plain /p/ exits the phonological module as featurally voiced. This is not necessary to account for the observed duration of voicing in any prosodic context. Intervocalic voicing occurs regardless of metrical position due to pressure from a constraint analogous to CONTINTφ that applies to the word domain: CONTINTω.

(33) CONTINUOUSINTENSITY-ω (CONTINTω) *Cost of violation* = $w(\text{Voicing}(\%) - 1)^2$
Intensity must be continuous within a word

CONTINTω requires word-medial stops to be voiced. Intervocalic stops are inherently word-medial, so they meet the conditions for CONTINTω to influence their voicing duration. Word-initial and word-final stops are not medial at the word or any lower domain, CONTINTω fails to apply, and stops in these environments surfaces as voiceless.

Foot-medially, stops are pushed towards a 100% voicing target by CONTINTφ and CONTINTω while *VOICED repels stops from this 100% target.

(34) Intervocalic voicing, foot-medial

(apa)	CONTINTφ $w = 2.7$	CONTINTω $w = 1.4$	*VOICED $w = 0.6$	H
(ap$_0$a)	1	1	0	4.1
(ab$_{10}$a)	0.81	0.81	0.01	3.33
(ab$_{20}$a)	0.64	0.64	0.04	2.65
(ab$_{30}$a)	0.49	0.49	0.09	2.06
(ab$_{40}$a)	0.36	0.36	0.16	1.57
(ab$_{50}$a)	0.25	0.25	0.25	1.18
(ab$_{60}$a)	0.16	0.16	0.36	0.87
(ab$_{70}$a)	0.09	0.09	0.49	0.66
(ab$_{80}$a)	0.04	0.04	0.64	0.55
☞ (ab$_{87}$a)	0.017	0.017	0.757	0.52
(ab$_{90}$a)	0.01	0.01	0.81	0.53
(ab$_{100}$a)	0	0	1	0.6

The weights of these constraints determine where the optimal compromise between the three targets lies and produce the mean voicing duration of 87% for foot-medial stops.

(35) Intervocalic voicing, foot-initial

a(paː)	CONTINTφ $w = 2.7$	CONTINTω $w = 1.4$	*VOICED $w = 0.6$	H
a(p$_0$aː)		1	0	1.4
a(b$_{10}$aː)		0.81	0.01	1.14
a(b$_{20}$aː)		0.64	0.04	0.92
a(b$_{30}$aː)		0.49	0.09	0.74
a(b$_{40}$aː)		0.36	0.16	0.6
a(b$_{50}$aː)		0.25	0.25	0.5
a(b$_{60}$aː)		0.16	0.36	0.44
☞ a(b$_{70}$aː)		0.09	0.49	0.42
a(b$_{80}$aː)		0.04	0.64	0.44
a(b$_{87}$aː)		0.017	0.757	0.48
a(b$_{90}$aː)		0.01	0.81	0.5
a(b$_{100}$aː)		0	1	0.6

In foot-initial and stray syllables, only the weaker CONTINTω target compels stops to become voiced and the result is a lower mean voicing duration.

Fougeron & Keating (1997), whose work inspired this analysis, might lead us to expect that similar weaker effects may be found in Hidatsa phrase-medially. Fougeron & Keating found that there is more fortition at higher initial prosodic boundaries and where more initial boundaries coincide. The parallel prediction regarding lenition is that being medial in lower constituents and a greater number of prosodic domains yields a greater degree of lenition (i.e., more lenition inside of a foot than in a word, more lenition in words than phrases, etc.). Thus an underlying word-initial /p/ should undergo some amount of voicing when it is phrase-medial but not when it is phrase-initial, and the degree of voicing should be lesser than what is observed word-medially.

Though I lack data bearing on the issue[6], this is not a necessary consequence of the analysis. An additional constraint referring to phrasal structure as in (36) could be weighted either above or below *VOICED.

(36) CONTINUOUSINTENSITY-Φ (CONTINTΦ) *Cost of violation* = $w(\text{Voicing}(\%) - 1)^2$
Intensity must be continuous within a phonological phrase

If it is the case that there is some amount of active voicing across word boundaries phrase-medially, this is an indication that CONTINTΦ is weighted between CONTINTω and *VOICED (*w*=X in (37)). However, if voicing is deliberately avoided or there is only incidental carryover voicing, this suggests that CONTINTΦ is weighted below *VOICED (*w*=Y in (37)).

(37) Voicing across word boundaries

$[[\text{mawa}]^{\omega}\ [\text{paː}]^{\omega}]^{\Phi}$	CONTINTω *w*=1.4	CONTINTΦ *w*=X	*VOICED *w*=0.6	CONTINTΦ *w*=Y	*H*
$[[\text{mawa}]^{\omega}\ [\text{p}_0\text{aː}]^{\omega}]^{\Phi}$		1	0	1	
$[[\text{mawa}]^{\omega}\ [\text{b}_{10}\text{aː}]^{\omega}]^{\Phi}$		0.81	0.01	0.81	
$[[\text{mawa}]^{\omega}\ [\text{b}_{20}\text{aː}]^{\omega}]^{\Phi}$		0.64	0.04	0.64	
$[[\text{mawa}]^{\omega}\ [\text{b}_{30}\text{aː}]^{\omega}]^{\Phi}$		0.49	0.09	0.49	
…					

[6] The data I have access to are dictionary recordings of words produced in isolation. Every word is also a phonological and intonational phrase, so differentiating between phrasal and word-level phonology is impossible. However, to the extent that this prediction is currently testable, the data are consistent with expected outcomes.

Whatever the nature of the phrasal data, the weights of constraints as outlined above can be adjusted to accommodate further CONTINT constraints. Critically, the core aspects of the analysis need not be altered to account for phrasal data, as the weights of constraints penalizing voicing are free to interact with those that require it within specific prosodic domains.

In the interest of preserving parallelism between Fougeron & Keating (1997)'s observations and the effects of prosodic structure on lenition, it should be noted that a sensible restriction to place on the weights of CONTINT constraints is that the weight of constraints referring to lower prosodic domains is always greater than the weight of constraints referring to higher prosodic domains. This generates a hierarchy of the magnitude of lenition effects ($\varphi > \omega > \Phi > \iota$), and similarly predicts that any lenition process present within prosodic constituent X also occurs within every lower prosodic domain.

5 Conclusion

I have argued here that voicing duration in Hidatsa varies by foot position and that voicing duration can be utilized as a diagnostic of numerous aspects of foot structure. I show that there is phonetic evidence in favor of a strictly bimoraic QS iambic, iterative, left-aligning metrical system in Hidatsa.

Although I have stated that the specific analytical tools required to account for Hidatsa's voicing pattern have not been previously proposed in fortition and lenition literature, this should not be interpreted as a suggestion that gradient foot-governed lenition itself is unique to Hidatsa. To the contrary, I suspect that such phenomena are widespread cross-linguistically and believe that this topic merits more extensive study to ascertain the variety and pervasiveness of the interaction of metrical structure and gradient lenition. It is exactly this kind of intense scrutiny of gradient prosody-segmental interactions that enables us to make informed decisions between the analytical options offered in §4.1 and §4.2.

While I have left to future work the question of what implications the Hidatsa data have for the typology of foot structure, the constraints on permissible foot shapes within and across languages is certainly a theoretically interesting avenue of research. It is important not to overlook the potential significance of Hidatsa with respect to foot typology; the voicing pattern detailed in this paper bears on longstanding tensions between symmetrical and asymmetrical foot typologies and illustrates the value of revisiting this area with new types of evidence.

References

Alber, Birgit. 2005. Clash, lapse and directionality. *Natural Language and Linguistic Theory* 23, 485-542.

Blumenfeld, Lev. 2006. *Constraints on phonological interactions*. Ph.D. dissertation, Stanford University. Stanford, CA.

Bowers, Norman. 1996. *Hidatsa suprasegmentals: A phonological analysis of a Siouan native North American language*. Ph.D. dissertation, The University of Idaho. Moscow, ID.

Boyle, John, Rayan Kasak, Sarah Lundquist, Armik Mirzayan, Jonnie Torres and Brittany Williams. 2016. A Preliminary Study on Accentuation in Hidatsa. Paper presented at the annual meeting of the Society for the Studies of Indigenous Languages of the Americas at the Linguistic Society of America annual meeting, Washington D.C.

Boyle, John. 2007. *Hidatsa Morpho-Syntax and Clause Structure*. Ph.D. dissertation, University of Chicago. Chicago. IL.

Cho, Taehong and Patricia Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29, 155-90.

Crowhurst, Megan. 1996. An optimal alternative to conflation. *Phonology* 13, 409-24.

Das, Shyamal. 2001. *Some aspects of the prosodic phonology of Tripura Bangla and Tripura Bangla English*. PhD dissertation, Central Institute of English and Foreign Languages, Hyderbad. Hyderbad, India.

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, 607-52.

de Lacy, Paul. 2003. Fixed ranking and the 'Too Many Solutions' problem. CASTL Kick-off conference handout.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18, 7-44.

Fougeron, Cécilia and Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101, 3728-40.

Harris, John. 2004. Release the captive coda: The foot as a domain of phonetic interpretation. *Phonetic interpretation: Papers in laboratory phonology* 6, 103-29.

Harris, Zellig and Charles Voegelin. 1939. Hidatsa texts collected by Robert H. Lowie with grammatical notes and phonograph transcriptions. *Indiana Historical Society, Prehistory Research Series* 1, 173-239.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, 429-46.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. University of Chicago Press: Chicago.

Kager, René. 1993. Alternatives to the Iambic-Trochaic Law. *Natural Language and Linguistic Theory* 11, 381-432.

Kaplan, Abby. 2010. *Phonology shaped by phonetics: The case of intervocalic lenition*. Ph.D. dissertation, UC Santa Cruz. Santa Cruz, CA.

Katz, Jonah. 2016. Lenition, perception, and neutralisation. *Phonology* 33, 43–85.

Katz, Jonah. 2021. Intervocalic lenition is not phonological: Evidence from Campidanese Sardinian. *Phonology* 38, 651-92.

Kingston, John. 2008. Lenition. In Laura Colantoni & Jeffrey Steele (eds.), *Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, 1–31. Somerville: Cascadilla.

Kirchner, Robert. 2004. Consonant lenition. In Bruce Hayes, Robert Kirchner, & Donca Steriade (eds.), *Phonetically based phonology*. Cambridge: Cambridge University Press.

Lavoie, Lisa. 2001. *Consonant strength: phonological patterns and phonetic manifestations*. Garland: New York.

Matthews, Hubert. 1965. *Hidatsa Syntax*. The Hague: Mouton.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12, 1-61.

Metzler, Zach. 2021. *Iambic Intensity in Hidatsa*. Fresno, CA: California State University, Fresno thesis.

Park, Indrek. 2012. *A Grammar of Hidatsa*. Ph.D. dissertation, Indiana University. Bloomington, ID.

Poser, William. 1990. Evidence for foot structure in Japanese. *Language* 66, 78-105.

Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11, 511-62.

Rasin, Ezer. 2018. *Modular interactions in phonology*. Ph.D. dissertation, MIT. Cambridge, MA.

Silva, David. 1992. *The phonetics and phonology of stop lenition in Korean*. Ph.D. dissertation, Cornell University. Ithaca, NY.

Steriade, Donca. 2001. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Ms., UCLA/MIT.

Voegelin, Charles and Florence Robinett. 1954. 'Mother Language' in Hidatsa. *International Journal of American Linguistics* 20, 65-70.